MULTIDISIPLIN

ANTARA KECERDASAN BUATAN (AI) DAN ILMU HUKUM

Dr. Agus Wibowo, M.Kom, M.Si, MM.

MULTIDISIPLIN

ANTARA KECERDASAN BUATAN (AI) DAN ILMU HUKUM

Dr. Agus Wibowo, M.Kom, M.Si, MM.

YAYASAN PRIMA AGUS TEKNIK

PENERBIT :
YAYASAN PRIMA AGUS TEKNIK
Jl. Majapahit No. 605 Semarang
Telp. (024) 6723456. Fax. 024-6710144
Email : penerbit_ypat@stekom.ac.id

ISBN 978-634-7227-67-6 (PDF)
9 786347 227676

Multidisiplin antara Kecerdasan Buatan (AI) dan Ilmu Hukum

Penulis :
Dr. Agus Wibowo, M.Kom, M.Si, MM.

ISBN : 978-634-7227-67-6 (PDF)

Editor :
Dr. Joseph Teguh Santoso, S.Kom., M.Kom.

Penyunting :
Dr. Mars Caroline Wibowo. S.T., M.Mm.Tech

Desain Sampul dan Tata Letak :
Irdha Yunianto, S.Ds., M.Kom

Penebit :
Yayasan Prima Agus Teknik Bekerja sama dengan
Universitas Sains & Teknologi Komputer (Universitas STEKOM)

Anggota IKAPI No: 279 / ALB / JTE / 2023

Redaksi :
Jl. Majapahit no 605 Semarang
Telp. 08122925000
Fax. 024-6710144
Email : penerbit_ypat@stekom.ac.id

Distributor Tunggal :
Universitas STEKOM
Jl. Majapahit no 605 Semarang
Telp. 08122925000
Fax. 024-6710144
Email : info@stekom.ac.id

Hak cipta dilindungi undang-undang
Dilarang memperbanyak karya tulis ini dalam bentuk dan dengan cara apapun tanpa ijin dari penulis

KATA PENGANTAR

Puji syukur kami panjatkan ke hadirat Tuhan Yang Maha Esa atas karunia-Nya yang tak terputuskan, sehingga buku "*Multidisiplin Antara Kecerdasan Buatan (AI) dan Ilmu Hukum"* ini dapat diselesaikan dengan baik. Karya ini muncul sebagai respons mendesak terhadap percepatan revolusi digital yang dipicu oleh kecerdasan buatan (AI), yang kini meresap ke seluruh lapisan kehidupan manusia, termasuk ilmu hukum, etika sosial, pengambilan keputusan peradilan, dan tata kelola global. Di era di mana AI tidak hanya menjadi alat, melainkan aktor potensial dalam dinamika hukum seperti prediksi residivisme, deepfake di pengadilan, atau dokter otonom. Buku ini hadir untuk menyediakan kerangka multidisiplin yang komprehensif, mengintegrasikan perspektif ilmiah, teknologi, etika, dan regulasi bagi akademisi, hakim, pengacara, regulator, inovator teknologi, serta pembuat kebijakan.

Latar belakang penulisan buku ini didasari oleh kesenjangan pengetahuan yang semakin lebar antara kemajuan AI eksponensial seperti pembelajaran mendalam, pemrosesan bahasa alami (NLP), dan sistem rekomendasi dengan kerangka hukum yang masih adaptif secara lambat. Saat ini, regulasi seperti AI Act Uni Eropa, GDPR, dan isu antimonopoli algoritmik menuntut pemahaman holistik yang tidak hanya teknis, tetapi juga filosofis dan yuridis. Buku ini terdiri dari 23 bab yang terstruktur rapi dalam tiga bagian utama, yang dirancang untuk membimbing pembaca dari fondasi pencapaian AI hingga solusi regulasi masa depan, sambil menyoroti risiko seperti bias algoritma, keburaman metakognisi, dan kolusi harga otomatis.

Bagian Pertama: Pencapaian AI Ilmiah, Teknologi, Sosial membentuk pondasi fundamental melalui enam bab. Bab 1 menyajikan konteks historis AI, mulai dari perkembangan awal, kemampuan mesin berpikir, keberatan filosofis, manipulasi simbol, pembelajaran mesin, revolusi deep learning, aplikasi analisis-otomasi, hingga peran AI sebagai penggerak Revolusi Industri Keempat. Bab 2 mendalami dampak teknologi bahasa dalam hukum, mencakup Legal AI via NLP untuk data tekstual dan lisan, serta tantangan masa depan. Bab 3 menganalisis implikasi sosial sistem rekomendasi machine learning, risiko pengambilan keputusan, kondisi kegagalan, langkah perbaikan, dan dampak sosial. Bab 4 membahas data kesehatan dengan paradigma perawatan pasien berbasis nilai (P4), layanan data-driven, tantangan etika-hukum AI, serta tren investasi. Bab 5 mengeksplorasi keamanan siber versus privasi melalui CIA Triad dan GDPR, definisi istilah, masalah, pencapaian teknologi, integrasi machine learning, resistensi sensor, serta tren digital.

Bagian Kedua: Tantangan Etika dan Hukum dalam Kecerdasan Buatan mengupas dilema mendalam melalui tujuh bab. Bab 6 merefleksikan pra-anggapan AI sebelum-sesudah, imitasi manusia-mesin, etika hukum relasi tersebut, serta evaluasi kritis risiko, pendidikan, dan hukum. Bab 7 membahas robot otonom dan cerdas dari perkembangan industri-layanan, kebangkitan sosial, dampak etika-sosial-hukum, serta Hukum Robotika Asimov. Bab 8 menyoroti AI sehari-hari dalam sistem rekomendasi, tantangan etika-hukum-regulasi, strategi

solusi, dan GDPR. Bab 9 menganalisis metakognisi, akuntabilitas, kepribadian hukum AI melalui status agen, persamaan agensi, tindakan sukarela, dan kriteria tanggung jawab. Bab 10 membahas AI keputusan kesehatan dengan bioetika kompleks (CBM), risiko, dan batas kecerdasan. Bab 11 fokus pada dokter AI otonom: regulasi, patologi, parameter, implikasi etis-hukum, dan pengaturan akuntabel. Bab 12 mengeksplorasi keunggulan AI medis, keburaman algoritmik, etika data dalam hukum perdata, serta masa depan pendidikan digital.

Bagian Ketiga: Hukum, Tata Kelola, dan Regulasi Kecerdasan Buatan memberikan blueprint kebijakan melalui 10 bab terakhir. Bab 13 membongkar mitos regulasi UE: empat rumus AI, kedaulatan digital, konstitusionalisme, Efek Brussel, HAI manusia-sentris, dan pelajaran. Bab 14 menguraikan penalaran kontrafaktual untuk kausalitas hukum, latar sosial-sejarah, konflik populasi, perburuan rusa, permainan evolusioner, dan koordinasi stag-hunt. Bab 15 membahas AI pengambilan keputusan peradilan, algoritma otomatis, proses hukuman, kasus S v Loomis, efek teknologi, bias otomatisasi-penjangkaran. Bab 16 mengeksplorasi tanggung jawab AI: presentasi masalah, subjektif kausalitas alternatif, mutlak, pembebasan kerusakan. Bab 17 analisis risiko bahasa alami Swiss: teknis NLG, aspek hukum, tantangan perdata. Bab 18 membahas AI pidana prediksi residivisme, prediktabilitas perilaku, bias teknologi. Bab 19 mengupas deepfake: definisi, AI pembuatan-deteksi, relevansi peradilan pidana, solusi penegakan. Bab 20 membahas hukum antimonopoli AI pricing: koordinasi algoritmik, kolusi AS-UE, analisis komparatif, prospek. Bab 21 proposal AI Act e-Justice Eropa: digitalisasi pro-kontra, sistem peradilan, risiko tinggi manusia-sentris, etika. Bab 22 kritik AI Act UE terhadap risiko sistemik fintech. Bab 23 regulasi sandbox: penjinakan AI, inovasi terintegrasi, evolusi disruptif.

Kami menyampaikan ucapan terima kasih yang sebesar-besarnya kepada seluruh penulis bab, reviewer akademik dari berbagai universitas, institusi penelitian AI dan hukum di Indonesia serta internasional, sponsor publikasi, keluarga yang sabar mendukung, dan komunitas open-source yang berkontribusi data. Tanpa dukungan mereka, karya ini takkan terwujud. Buku ini tentu saja bukan akhir, melainkan undangan diskursus; kritik, saran, dan masukan konstruktif sangat diharapkan untuk menyempurnakan edisi mendatang. Semoga buku ini menjadi cahaya pencerahan, katalisator regulasi bijaksana, dan fondasi ekosistem AI yang etis, inklusif, akuntabel, serta berkeadilan bagi generasi mendatang di Indonesia dan dunia.

Semarang, Desember 2025

Penulis

Dr. Agus Wibowo, M.Kom, M.Si, MM.

DAFTAR ISI

BAGIAN 1
PENCAPAIAN AI: ILMIAH, TEKNOLOGI, SOSIAL

Teknologi AI telah berkembang selama lebih dari setengah abad, tetapi dua dekade terakhir telah menyaksikan perkembangan yang mengubah sifat masyarakat. Perkembangan ini meliputi pembelajaran mesin, robotika, visi komputer, pemrosesan bahasa alami, dan banyak aplikasi dalam analisis data, keuangan, dan kesehatan. Tujuan bab-bab dalam bagian ini bukanlah untuk memberikan deskripsi teknis yang terperinci tentang teknologi yang terlibat, tetapi untuk memberikan latar belakang teknis minimal yang akan memungkinkan pembaca memahami apa yang dapat dan tidak dapat dilakukan oleh teknologi ini. Banyak analisis tentang konsekuensi teknologi kecerdasan buatan sangat cacat karena penulis mengabaikan beberapa fakta sederhana tentang cara kerja teknologi tersebut.

Bab-bab dalam bagian ini membahas beberapa teknologi yang telah memberikan dampak signifikan terhadap masyarakat dalam beberapa dekade terakhir. Namun, teknologi ini pasti akan memiliki dampak yang jauh lebih besar dalam beberapa dekade mendatang, seiring dengan semakin matangnya teknologi dan semakin banyaknya aplikasi dalam analisis data dan otomatisasi. Bab karya *Arlindo L. Oliveira dan Mário Figueiredo* memberikan gambaran umum tentang teknologi kecerdasan buatan, dengan tinjauan historis dan deskripsi perkembangan terkini, yang menjadi latar belakang untuk bab-bab yang lebih spesifik berikutnya. Bab ini membahas secara mendalam salah satu area yang lebih sentral dalam kecerdasan buatan: pembelajaran mesin, teknologi yang mendorong perkembangan terkini dalam kecerdasan buatan, dengan banyak aplikasi dalam analitik dan otomatisasi pekerjaan.

Kemajuan terbaru dalam model bahasa berskala besar telah menarik banyak perhatian, dengan dirilisnya ChatGPT, tetapi ini bukan satu-satunya teknologi bahasa yang akan mengubah cara kita bekerja dan berinteraksi dengan manusia dan komputer. Bab karya *Isabel Trancoso, Nuno Mamede, Bruno Martins, H. Sofia Pinto, dan Ricardo Ribeiro* menjelaskan perkembangan terkini dalam teknologi bahasa alami dan dampaknya di beberapa bidang, dengan fokus khusus pada bidang hukum. Bab ini mencakup beberapa aplikasi teknologi bahasa alami seperti peringkasan, pengambilan dokumen, prediksi hasil, dan ekstraksi informasi. Teknologi bahasa lisan, yang semakin relevan, juga dibahas dalam bab ini.

Meskipun teknologi ini bermanfaat, dengan aplikasinya di berbagai domain yang akan meningkatkan kualitas hidup kita secara keseluruhan, teknologi ini juga menimbulkan beberapa pertanyaan penting. Salah satu tantangan paling utama terkait dengan sistem rekomendasi dan privasi. Penerapan analitik yang ekstensif terhadap data yang diperoleh dari berbagai sumber, seperti jejaring sosial, mesin pencari, dan aplikasi ponsel, memiliki dampak signifikan terhadap privasi individu, sebuah isu yang menimbulkan banyak isu hukum, etika, dan moral. Beberapa isu ini dibahas dalam bab karya *Joana Gonçalves-Sá dan Flávio L. Pinheiro*, yang berfokus pada risiko yang melekat dalam sistem rekomendasi. Bab ini berfokus, khususnya, pada risiko yang diakibatkan oleh sistem rekomendasi yang gagal berfungsi

sebagaimana mestinya dan pada risiko sistem yang berfungsi tetapi menimbulkan ancaman bagi individu dan bahkan bagi masyarakat demokratis.

Domain aplikasi penting lainnya adalah kesehatan. Penerapan teknik kecerdasan buatan dalam domain kesehatan menimbulkan banyak pertanyaan berbeda, banyak di antaranya terkait dengan tanggung jawab hukum, privasi, dan keamanan. Bab karya *Ana Teresa Freitas* membahas beberapa tantangan ini, khususnya pentingnya persetujuan berdasarkan informasi, kebutuhan akan keamanan dan transparansi dalam semua sistem yang menangani data kesehatan, serta kebutuhan terkait privasi data.

Tantangan penting terakhir datang dari isu keamanan. Seiring dengan semakin pentingnya teknologi digital bagi masyarakat modern, risiko yang ditimbulkan oleh aktivitas peretas, perusahaan, dan pemerintah semakin meningkat dan membutuhkan pemahaman yang semakin mendalam tentang isu-isu yang terlibat. Beberapa risiko ini dianalisis dalam bab karya *Luís Rodrigues dan Miguel Correia* yang secara formal mendefinisikan tiga karakteristik yang membentuk dasar keamanan: kerahasiaan, tidak adanya pengungkapan data yang tidak sah; integritas, tidak adanya data atau modifikasi sistem yang tidak sah; dan ketersediaan, atau kesiapan sistem untuk menyediakan layanannya.

Tentu saja, tidak semua teknologi yang digunakan dalam bidang seluas kecerdasan buatan dibahas dalam lima bab ini. Namun, yang paling signifikan yang dibahas dalam bab-bab ini, pemrosesan bahasa alami, pembelajaran mesin, analitik, dan keamanan siber, memberikan titik awal yang baik bagi pembaca non-teknis yang tertarik untuk lebih memahami teknologi dan aplikasi kecerdasan buatan yang mengubah dunia kita.

BAB 1
KECERDASAN BUATAN: KONTEKS HISTORIS DAN KEADAAN TERKINI

Abstrak Gagasan bahwa kecerdasan adalah hasil dari proses komputasi dan, oleh karena itu, dapat diotomatisasi, telah ada sejak berabad-abad lalu. Kami meninjau asal-usul historis gagasan bahwa mesin dapat menjadi cerdas, dan kontribusi paling signifikan yang diberikan oleh Thomas Hobbes, Charles Babbage, Ada Lovelace, Alan Turing, Norbert Wiener, dan lainnya. Keberatan terhadap gagasan bahwa mesin dapat menjadi cerdas telah diajukan dan ditanggapi berkali-kali, dan kami memberikan tinjauan singkat tentang argumen dan kontra-argumen yang diajukan dari waktu ke waktu.

Kecerdasan awalnya dipandang sebagai manipulasi simbol, yang mengarah pada pendekatan yang memiliki beberapa keberhasilan dalam masalah spesifik, tetapi tidak dapat digeneralisasi dengan baik ke masalah dunia nyata. Untuk mengatasi kesulitan yang dihadapi oleh sistem awal, yang rapuh dan tidak mampu menangani kompleksitas yang tak terduga, teknik pembelajaran mesin semakin banyak diadopsi.

Baru-baru ini, sub-bidang pembelajaran mesin yang dikenal sebagai pembelajaran mendalam telah mengarah pada perancangan sistem yang dapat berhasil belajar untuk mengatasi masalah-masalah sulit dalam pemrosesan bahasa alami, penglihatan, dan (namun pada tingkat yang lebih rendah) interaksi dengan dunia nyata. Sistem-sistem ini telah menemukan aplikasi dalam domain yang tak terhitung jumlahnya dan merupakan salah satu teknologi utama di balik revolusi industri keempat, juga dikenal sebagai Industri 4.0.

Aplikasi dalam analitik memungkinkan sistem kecerdasan buatan untuk mengeksploitasi dan mengekstrak nilai ekonomi dari data dan merupakan sumber pendapatan utama bagi banyak perusahaan terbesar saat ini. Kecerdasan buatan juga dapat digunakan dalam otomatisasi, memungkinkan robot dan komputer untuk menggantikan manusia dalam banyak tugas. Kami menyimpulkan dengan memberikan beberapa petunjuk untuk kemungkinan perkembangan di masa depan, termasuk kemungkinan pengembangan kecerdasan umum buatan, dan memberikan petunjuk tentang implikasi potensial dari teknologi ini di masa depan umat manusia.

1.1 PERKEMBANGAN KECERDASAN BUATAN

Gagasan bahwa kecerdasan dapat diotomatisasi memiliki akar sejarah yang panjang. Referensi tentang mesin berpikir non-manusia terdapat dalam Iliad karya Homer, dan Thomas Hobbes dengan jelas menyatakan, dalam Leviathan, bahwa pemikiran manusia tidak lebih dari sekadar komputasi aritmatika. Baik Pascal maupun Leibnitz, di antara banyak lainnya, merancang mesin untuk mengotomatisasi komputasi aritmatika, yang dapat dianggap sebagai cikal bakal kalkulator modern. Namun, baru pada pertengahan abad ke-19 muncul proposal pertama komputer yang benar-benar umum, yang diciptakan oleh Charles Babbage.

Tujuan awal Babbage adalah membangun perangkat tabulasi canggih, yang ia sebut Mesin Perbedaan. Sebagaimana dibayangkan oleh Babbage, ini adalah perangkat mekanis

yang dapat diprogram untuk melakukan serangkaian komputasi yang telah ditentukan sebelumnya, melalui susunan roda gigi, roda penyok, dan tuas yang rumit. Meskipun ia hanya berhasil membangun beberapa bagian dari Mesin Perbedaan, Babbage membayangkan mesin yang bahkan lebih canggih, Mesin Analitik. Seandainya mesin itu benar-benar diciptakan, mesin itu akan mampu melakukan komputasi umum dengan cara yang hampir sama seperti komputer modern, meskipun dengan kecepatan yang jauh lebih lambat yang ditentukan oleh komponen mekanisnya.

Meskipun Babbage yang merancang mesin tersebut, Ada Lovelace, seorang matematikawan yang bersahabat dengannya, lah yang menulis analisis paling mendalam tentang kekuatan mesin tersebut, dengan argumen bahwa mesin itu dapat melakukan lebih dari sekadar komputasi numerik. Secara khusus, ia mengamati bahwa mesin itu dapat melakukan hal-hal selain angka jika hal-hal tersebut memenuhi aturan matematika yang terdefinisi dengan baik. Ia berpendapat bahwa mesin itu dapat menulis lagu atau melakukan aljabar abstrak, selama tugas-tugas tersebut dapat diungkapkan menggunakan bahasa simbolik. Namun, Lovelace juga berpendapat bahwa mesin itu tidak dapat menciptakan sesuatu yang baru, melainkan hanya melakukan tugas-tugas yang telah diprogramkan untuknya, sehingga mengesampingkan kemungkinan bahwa perilaku cerdas, entah bagaimana, dapat diprogram ke dalam mesin tersebut. Argumen ini dianalisis jauh di kemudian hari oleh seorang matematikawan yang bahkan lebih berpengaruh, Alan Turing.

1.2 MESIN MEMILIKI KEMAMPUAN UNTUK BERPIKIR

Sekitar satu abad kemudian, Alan Turing, salah satu matematikawan paling mendalam dan kreatif sepanjang masa, mengembangkan beberapa gagasan sentral komputasi modern dan sampai pada kesimpulan yang berbeda dari yang dicapai oleh Lovelace. Turing, yang dikenal karena memainkan peran penting dalam upaya Sekutu pada Perang Dunia II untuk memecahkan kode pesan musuh yang dikodekan oleh mesin sandi Enigma Jerman, mencapai beberapa hasil paling signifikan dalam matematika, terutama dalam fondasi matematika ilmu komputer, hasil yang sama pentingnya saat ini seperti pada saat diperoleh.

Dalam sebuah makalah yang sangat penting, Turing menunjukkan bahwa komputer digital apa pun dengan memori yang cukup besar, yang menangani simbol dan memenuhi beberapa kondisi sederhana, dapat melakukan kalkulasi dan menghitung rangkaian fungsi yang sama seperti komputer digital lainnya. Konsep ini kemudian dikenal sebagai universalitas Turing. Ia mendeskripsikan jenis komputer tertentu, yang kini dikenal sebagai mesin Turing, yang menggunakan pita untuk menulis dan membaca simbol sebagai memori, dan mendemonstrasikan bahwa jenis komputer ini (pada prinsipnya, dengan asumsi pita tak terbatas) dapat melakukan operasi dan kalkulasi yang sama, seperti komputer lain yang memanipulasi simbol.

Pada tahun yang sama, Alonzo Church menerbitkan deskripsi tentang apa yang disebut kalkulus lambda, sebuah sistem formal untuk mengekspresikan komputasi berdasarkan abstraksi dan aplikasi fungsi, yang juga merupakan model komputasi universal dengan daya ekspresif yang sama dengan mesin Turing. Kombinasi kedua hasil ini menghasilkan apa yang

kemudian dikenal sebagai tesis Church-Turing, yang dapat dinyatakan secara informal sebagai berikut: setiap hasil yang dapat dihitung secara aktual dapat dihitung oleh mesin Turing, atau oleh komputer lain yang memanipulasi simbol dan memiliki memori yang cukup. Hasil teoretis dan murni matematis ini memiliki konsekuensi filosofis yang penting.

Perhatikan bahwa terdapat definisi yang agak melingkar dalam rumusan ini: apa sebenarnya arti kalimat "hasil yang dapat dihitung secara aktual"? Adakah hasil numerik yang tidak termasuk dalam kategori ini? Karya Alonzo Church, Alan Turing, dan Kurt Gödel menunjukkan bahwa terdapat hasil yang, meskipun terdefinisi dengan sempurna, tidak dapat dihitung. Pada tahun 1931, Gödel membuktikan bahwa tidak ada sistem aksioma yang konsisten yang cukup untuk membuktikan semua kebenaran tentang aritmatika bilangan asli dan bahwa, untuk sistem formal yang konsisten tersebut, akan selalu ada pernyataan tentang bilangan asli yang benar, tetapi tidak dapat dibuktikan dalam sistem tersebut.

Terdapat hubungan erat antara hasil Gödel dan masalah yang dibahas Turing dalam makalahnya tahun 1937, dan yang dapat dinyatakan secara sederhana (yang kemudian dikenal sebagai masalah penghentian): mungkinkah menentukan apakah eksekusi program komputer dengan masukan tertentu akan berakhir? Turing menunjukkan bahwa mustahil menjawab pertanyaan ini dalam kasus umum. Pertanyaan ini mungkin dapat dijawab dalam kasus-kasus khusus, tetapi ada program yang tidak memungkinkan untuk menentukan apakah eksekusinya akan berakhir atau tidak. Berbekal wawasan penting tentang hakikat dan kekuatan komputer digital ini, Turing melanjutkan analisisnya dengan pertanyaan penting lainnya, yang merupakan inti filosofis bidang kecerdasan buatan: dapatkah komputer berperilaku cerdas?

Sebelum menjelaskan karya Turing tahun 1950, di mana ia mengusulkan jawaban atas pertanyaan ini, penting untuk memahami konsekuensi dari mempertimbangkan gagasan mekanistik Thomas Hobbes dan tesis Church-Turing secara bersamaan. Hobbes berpendapat bahwa penalaran yang dilakukan oleh otak manusia tidak lebih dari manipulasi simbol matematika. Church dan Turing menunjukkan bahwa semua mesin yang memanipulasi simbol setara satu sama lain, selama mereka memenuhi persyaratan minimum tertentu dan tidak dibatasi waktu yang diizinkan untuk melakukan tugas tertentu, maupun memori yang tersedia.

Hasil dari kedua gagasan ini dapat dikatakan mengarah pada kesimpulan bahwa komputer, dalam arti luas, seharusnya mampu melakukan manipulasi simbol yang sama seperti otak manusia dan, oleh karena itu, sama cerdasnya dengan manusia. Namun, terdapat beberapa perbedaan pendapat di komunitas ilmiah mengenai kesimpulan ini, sebagaimana telah disebutkan. Beberapa orang percaya bahwa jenis substrat tempat komputasi dilakukan (biologis atau digital) mungkin penting, sementara yang lain berpendapat bahwa kesimpulan tersebut mungkin benar secara prinsip tetapi tidak relevan dalam praktik karena beberapa jenis kesulitan.

Dalam makalahnya yang terkenal, Turing mengajukan pertanyaan persis seperti ini: dapatkah mesin berpikir? Untuk menghindari kesulitan yang melekat dalam mendefinisikan arti "berpikir", Turing mengusulkan untuk merumuskan kembali pertanyaan tersebut menjadi masalah yang berbeda dan lebih terdefinisi. Secara khusus, ia mengusulkan untuk menganalisis sebuah permainan imitasi hipotetis, sebuah eksperimen pikiran yang mengarah

pada tes Turing yang kini terkenal. Dalam permainan yang diusulkan Turing, seorang interogator, di ruangan terpisah, berkomunikasi dengan seorang pria dan seorang wanita, melalui teks yang diketik. Tujuan interogator adalah membedakan pria dari wanita tersebut dengan mengajukan pertanyaan kepada mereka.

Turing bertanya-tanya apakah suatu hari nanti di masa depan, sebuah komputer yang berada di posisi pria tersebut dapat membuat interogator melakukan kesalahan sesering yang akan dilakukannya jika ada pria dan wanita. Variasi tes ini diusulkan oleh Turing sendiri dalam teks-teks selanjutnya, tetapi esensi tesnya tetap sama: mungkinkah seorang interogator membedakan antara jawaban yang diberikan oleh komputer dan jawaban yang diberikan oleh manusia? Turing berpendapat bahwa pertanyaan ini, pada dasarnya, setara dengan pertanyaan awal "Bisakah mesin berpikir?" dan memiliki keuntungan menghindari prasangka antropomorfik yang dapat mengkondisikan respons.

Bahkan, mengingat pengalaman individu dan sejarah manusia kita, wajar saja jika berasumsi bahwa hanya manusia yang dapat berpikir. Prasangka ini dapat menghalangi kita untuk mendapatkan jawaban objektif atas pertanyaan awal. Bisa saja interogator memutuskan bahwa komputer tidak dapat berpikir karena alasan sederhana bahwa komputer tidak seperti kita karena tidak memiliki kepala, lengan, dan kaki, seperti manusia. Penggunaan permainan imitasi mengurangi kemungkinan prasangka yang berakar pada pengalaman kita sebelumnya menghalangi kita mengenali mesin sebagai makhluk berpikir, bahkan dalam kasus-kasus yang memungkinkan hal ini.

Turing tidak hanya mengusulkan jawaban positif untuk pertanyaan "dapatkah mesin berpikir?", tetapi juga menunjukkan perkiraan waktu di masa depan ketika hal ini mungkin terjadi. Ia berpendapat bahwa dalam setengah abad akan ada mesin dengan memori satu Gigabita yang tidak dapat dibedakan dari manusia dalam uji Turing lima menit. Kita sekarang tahu bahwa Turing agak optimis. Pada akhir abad ke-20 (50 tahun setelah makalah Turing), memang ada mesin dengan memori 1GB, tetapi tidak satu pun dari mereka yang mungkin lulus uji Turing lima menit. Bahkan hingga saat ini, lebih dari 70 tahun setelah artikel Turing, kita masih belum memiliki mesin seperti itu, meskipun model bahasa besar terbaru (yang akan dijelaskan nanti), seperti ChatGPT, bisa dibilang tidak jauh dari lulus uji Turing dengan interogator yang bukan ahli dalam kelemahan mereka.

1.3 KEBERATAN TERHADAP KECERDASAN BUATAN

Bagian menarik dari artikel Turing tahun 1950 adalah ketika ia mengklasifikasikan, menganalisis, dan menanggapi serangkaian keberatan terhadap usulannya bahwa, suatu saat nanti, komputer mungkin dapat berpikir. Daftar keberatan ini bersifat instruktif dan relevan saat ini seperti ketika artikel tersebut ditulis.

Keberatan pertama bersifat teologis, yang menyatakan bahwa kecerdasan manusia adalah hasil dari jiwa abadi yang diberikan Tuhan kepada setiap manusia, tetapi tidak kepada hewan atau mesin. Turing menyadari bahwa ia tidak dapat menjawab keberatan ini secara ilmiah, tetapi tetap mencoba memberikan semacam jawaban, menggunakan apa yang ia anggap sebagai penalaran teologis.

Turing berpendapat bahwa klaim bahwa Tuhan tidak dapat menganugerahkan jiwa kepada hewan atau mesin merupakan pembatasan yang tidak dapat diterima terhadap kekuasaan Yang Mahakuasa. Mengapa Tuhan tidak dapat, tanya Turing, menganugerahkan jiwa kepada hewan, jika hewan tersebut dikaruniai kemampuan berpikir yang serupa dengan manusia? Argumen serupa berlaku untuk mesin: bukankah Tuhan memiliki kapasitas untuk menganugerahi mesin dengan jiwa, jika Dia menghendakinya dan mesin tersebut dapat bernalar?

Keberatan kedua didasarkan pada gagasan bahwa konsekuensi dari kemampuan berpikir mesin akan begitu mengerikan sehingga lebih baik berharap hal ini tidak akan pernah terjadi. Turing merasa argumen ini tidak cukup kuat untuk dibantah secara eksplisit. Namun, tujuh dekade setelah artikelnya, muncul usulan untuk menghentikan pengembangan beberapa teknologi kecerdasan buatan, karena khawatir akan kemungkinan konsekuensi negatifnya. Oleh karena itu, keberatan yang dianalisis Turing tidak sepenuhnya tidak relevan dan kebijakan yang ditentukan manusia dapat menjadi hambatan bagi pengembangan mesin cerdas.

Keberatan ketiga bersifat matematis dan kemudian dikaji ulang oleh John Lucas dan Roger Penrose. Keberatan ini didasarkan pada teorema Gödel (yang telah disebutkan di atas), yang menyatakan bahwa terdapat hasil matematika yang tidak dapat diperoleh oleh mesin atau prosedur apa pun. Keberatan ini didasarkan pada argumen bahwa keterbatasan ini tidak berlaku untuk otak manusia. Namun, seperti yang dikemukakan Turing, tidak ada bukti yang diberikan bahwa otak manusia tidak tunduk pada keterbatasan ini. Turing kurang memberikan kredibilitas pada keberatan ini, terlepas dari prestise beberapa pendukungnya.

Keberatan keempat didasarkan pada gagasan bahwa hanya kesadaran yang dapat mengarah pada kecerdasan dan bahwa tidak akan pernah mungkin untuk membuktikan bahwa sebuah mesin memiliki kesadaran. Sekalipun sebuah mesin dapat menulis puisi, hanya ketika mesin tersebut menyadari makna puisi tersebut, mesin tersebut dapat dianggap cerdas, demikian argumen tersebut. Turing mencatat bahwa, dalam versi paling radikal dari keberatan ini, hanya dengan menjadi mesin tersebut kita dapat yakin bahwa mesin tersebut memiliki kesadaran. Namun, begitu kita menjadi mesin, akan sia-sia untuk menggambarkan perasaan atau sensasi kesadaran, karena kita akan diabaikan oleh seluruh dunia, yang tidak akan mengalami sensasi ini secara langsung. Jika dibawa ke ekstrem, keberatan ini mencerminkan posisi solipsistik, yang menyangkal perilaku sadar tidak hanya pada mesin tetapi juga pada semua manusia lainnya karena keberadaan kesadaran pada manusia lain tidak dapat dibuktikan tanpa keraguan. Meskipun mengakui bahwa fenomena kesadaran masih belum dapat dijelaskan (sesuatu yang masih berlaku hingga saat ini), Turing tidak menganggap keberatan ini sebagai sesuatu yang menentukan.

Keberatan kelima muncul dari beragam argumen yang tidak berdasar tentang perilaku yang tidak dapat dimiliki oleh mesin mana pun. Kategori ini berisi argumen seperti "tidak ada mesin yang akan memiliki selera humor", "tidak ada mesin yang akan jatuh cinta", "tidak ada mesin yang akan menyukai stroberi dan krim", dan "tidak ada mesin yang akan menjadi objek pemikirannya sendiri". Salah satu argumen yang aneh (dan membingungkan) dalam kategori

ini adalah bahwa "mesin tidak membuat kesalahan", sedangkan manusia melakukannya. Sebagaimana ditunjukkan Turing, tidak ada pembenaran yang diberikan secara eksplisit untuk batasan-batasan ini, yang seharusnya umum bagi semua mesin.

Menurut Turing, keberatan-keberatan tersebut mungkin muncul dari penerapan prinsip induksi yang salah: sejauh ini, tidak ada mesin yang pernah jatuh cinta, jadi tidak akan pernah ada mesin yang jatuh cinta. Keberatan yang umum dalam kategori ini adalah bahwa tidak ada mesin yang akan pernah memiliki perasaan yang tulus. Seperti yang lainnya, keberatan ini tidak memiliki dasar ilmiah, hanya mencerminkan pandangan terbatas yang kita miliki, berdasarkan pengetahuan kita tentang mesin yang ada saat ini.

Keberatan keenam diajukan oleh Ada Lovelace dan telah disebutkan di bagian sebelumnya. Meskipun ia menyadari bahwa mesin analitik dapat memproses banyak jenis informasi selain angka, Lovelace berpendapat bahwa mesin tersebut tidak akan pernah dapat menciptakan sesuatu yang baru, karena hanya menjalankan operasi yang telah diprogram sebelumnya. Turing tidak membantah klaim Lovelace, tetapi berpendapat bahwa Lovelace tidak pernah berpikir bahwa instruksi tersebut bisa begitu rumit sehingga akan mengarahkan mesin untuk benar-benar menciptakan sesuatu yang baru.

Keberatan ketujuh, yang mungkin bahkan lebih populer saat ini daripada di zaman Turing, didasarkan pada gagasan bahwa otak tidak setara dengan mesin yang memanipulasi simbol. Kita sekarang tahu bahwa otak tidak bekerja seperti komputer tradisional, karena prinsip kerja otak dan komputer digital sangat berbeda. Lebih lanjut, otak tidak secara langsung memanipulasi simbol-simbol diskrit, melainkan variabel fisik dengan nilai kontinu dan hasil teoretis dari matematika memberi tahu kita bahwa mesin yang memanipulasi nilai kontinu (riil) tentu lebih kuat daripada mesin yang hanya memanipulasi simbol-simbol diskrit.

Jawaban Turing adalah bahwa setiap mesin yang memanipulasi simbol, jika diprogram dengan benar, akan mampu memberikan jawaban yang cukup mendekati jawaban yang diberikan oleh mesin yang memanipulasi nilai kontinu. Terlepas dari tanggapan ini, argumen ini tetap berbobot selama beberapa dekade. Banyak filsuf dan ilmuwan masih percaya bahwa tidak ada mesin yang memanipulasi simbol yang dapat secara akurat meniru perilaku otak manusia dan lulus uji Turing.

Keberatan kedelapan didasarkan pada argumen bahwa tidak ada seperangkat aturan yang cukup untuk menggambarkan kekayaan perilaku manusia. Karena mesin selalu mengikuti seperangkat aturan, tidak ada mesin yang dapat mereproduksi perilaku manusia, yang akan selalu tidak dapat diprediksi. Turing berpendapat bahwa tidaklah sulit bagi mesin untuk berperilaku tidak terduga dan tidak mungkin untuk membuktikan bahwa, jauh di lubuk hati, otak kita tidak berfungsi sesuai dengan seperangkat aturan.

Keberatan kesembilan, dan terakhir, yang anehnya merupakan keberatan yang tampaknya lebih dikuatkan oleh Turing, didasarkan pada (dugaan) keberadaan persepsi ekstrasensori dan kekuatan telepati. Meskipun ESP dan telepati semakin tidak dihargai di kalangan ilmiah, Turing tampaknya mempercayai bukti yang diketahui pada saat itu, yang tampaknya menunjukkan keberadaan fenomena ini. Sebagaimana Turing argumenkan dengan sangat baik, jika ada persepsi ekstrasensori, tes Turing harus dimodifikasi untuk menghindari

kemungkinan komunikasi telepati. Namun, kini kita tahu bahwa ESP tidak ada, yang membuat keberatan kesembilan tidak relevan dengan pembahasan saat ini.

Lebih dari setengah abad setelah karya Turing, keberatan-keberatan terhadap kemungkinan adanya mesin berpikir pada dasarnya tetap sama dan jawaban-jawaban yang diberikan Turing tetap valid seperti dulu. Tak satu pun keberatan yang diajukan tampaknya cukup kuat untuk meyakinkan kita bahwa mesin tidak dapat berpikir, meskipun tentu saja hal ini sama sekali tidak membuktikan bahwa mesin dapat berpikir.

Dalam beberapa dekade terakhir, beberapa keberatan lain diajukan terhadap tes Turing sebagai mekanisme untuk mengidentifikasi kecerdasan. Keberatan pertama, dan terpenting, adalah bahwa tes tersebut tidak benar-benar menilai kecerdasan, melainkan apakah subjek yang diuji memiliki kecerdasan yang analog dengan kecerdasan manusia. Sebuah komputer cerdas (atau bahkan individu hipotetis dari spesies cerdas non-manusia) tidak akan lulus tes Turing kecuali ia dapat meyakinkan penguji bahwa ia berperilaku seperti manusia. Sebuah sistem bahkan bisa jauh lebih pintar daripada manusia namun tetap gagal dalam tes, misalnya, karena ia gagal menyembunyikan kemampuan matematika yang luar biasa.

Keberatan kedua adalah bahwa tes tersebut tidak membahas semua kemampuan yang dapat diekspresikan oleh kecerdasan manusia, melainkan hanya kemampuan yang dapat diekspresikan melalui bahasa tulis. Meskipun tes tersebut dapat digeneralisasi untuk mencakup bentuk komunikasi lain (misalnya, pertanyaan dapat diajukan menggunakan bahasa lisan), tetap akan ada kesulitan dalam menguji keterampilan manusia yang tidak dapat diekspresikan melalui antarmuka yang dipilih. Di sisi lain, Turing secara eksplisit mengusulkan tes berdurasi terbatas, beberapa menit, yang sangat berbeda dari tes yang interaksinya berlangsung berjam-jam, berhari-hari, atau bahkan bertahun-tahun.

Keberatan ketiga berkaitan dengan hubungan antara kecerdasan dan kesadaran. Meskipun Turing membahas pertanyaan tentang kesadaran ketika menganalisis keberatan keempat dalam daftarnya, ia tidak secara eksplisit menyatakan bahwa mesin yang lulus tes tersebut pasti memiliki kesadaran. Turing menghindari pembahasan eksplisit tentang masalah ini, suatu sikap yang dapat dianggap bijaksana, mengingat hubungan antara kecerdasan dan kesadaran saat ini masih hampir sama misteriusnya seperti pada tahun 1950. Namun, proposal-proposal terbaru membahas hubungan antara model komputasional dan kesadaran, dan bidang ini sedang dipelajari secara aktif saat ini.

Terlepas dari keberatan-keberatan ini, dan keberatan-keberatan lain yang belum kami analisis, uji Turing tetap penting, lebih sebagai instrumen filosofis yang memungkinkan kita untuk meneliti argumen-argumen terkait kemungkinan keberadaan kecerdasan buatan, daripada sebagai mekanisme nyata untuk menganalisis kapabilitasnya.

Selain mengusulkan uji Turing, artikel yang ditulis Turing pada tahun 1950 memberikan satu saran terakhir yang, secara profetik, menunjukkan arah yang akhirnya akan diambil oleh kecerdasan buatan, beberapa dekade kemudian. Turing mengusulkan bahwa alih-alih mencoba menulis program yang memungkinkan mesin melewati permainan imitasi, akan lebih mudah untuk menulis program yang memungkinkan mesin belajar dari pengalaman, seperti yang dilakukan bayi. Program semacam itu, menurut Turing, akan jauh lebih mudah ditulis dan

akan memungkinkan mesin mempelajari apa yang dibutuhkannya untuk akhirnya lulus uji Turing.

Saran ini, yang begitu penting dan berwawasan luas, mendahului pendekatan paling sukses dalam masalah penciptaan sistem cerdas: pembelajaran mesin, beberapa dekade sebelumnya. Sebaliknya, pendekatan pertama yang diadopsi untuk mencoba menciptakan sistem cerdas didasarkan pada gagasan bahwa kecerdasan manusia, dalam bentuknya yang paling rumit dan berkembang, terdiri dari manipulasi representasi simbolis pengetahuan tentang dunia dan deduksi pengetahuan baru melalui manipulasi simbol-simbol ini.

1.4 KECERDASAN SEBAGAI MANIPULASI SIMBOL

Gagasan bahwa mesin cerdas dapat eksis, yang dipelopori oleh Turing dan banyak lainnya, dengan cepat mendorong proyek pembangunannya. Dimulai pada tahun 1950an, komputer digital menjadi lebih canggih dan semakin mudah diakses. Komputer pertama didedikasikan untuk kalkulasi ilmiah dan militer, tetapi secara bertahap penerapannya menyebar ke area aktivitas manusia lainnya. Dengan berakhirnya Perang Dunia Kedua, kemungkinan penggunaan komputer dalam aktivitas yang tidak terkait dengan aplikasi militer menjadi kenyataan. Salah satu bidang yang patut mendapat perhatian signifikan adalah ranah kecerdasan buatan yang sedang berkembang.

Pada tahun 1956, sebuah lokakarya ilmiah diadakan di Dartmouth, New Hampshire, yang mempertemukan beberapa pelopor di bidang kecerdasan buatan. Bahkan, dalam proposal penyelenggaraan konferensi ini, yang ditulis oleh John McCarthy, Marvin Minsky, Nathaniel Rochester, dan Claude Shannon (bapak teori informasi dan komunikasi yang terkenal), istilah kecerdasan buatan dicetuskan. Banyak dari mereka yang hadir pada pertemuan ini kemudian membentuk kelompok-kelompok riset di bidang kecerdasan buatan di universitas-universitas terpenting di Amerika Serikat. Pendekatan-pendekatan awal tersebut mencoba mereproduksi bagian-bagian penalaran manusia yang pada saat itu dianggap paling maju, seperti pembuktian teorema, perencanaan urutan tindakan, dan permainan papan, seperti dam dan catur.

Tidak mengherankan, upaya-upaya pertama untuk mereproduksi kecerdasan manusia justru berfokus pada masalah-masalah yang membutuhkan manipulasi simbol dan pencarian solusi. Pada tahun yang sama, sebuah program yang ditulis oleh Allen Newell dan Herbert Simon (yang juga menghadiri lokakarya Dartmouth), yang disebut Ahli Teori Logika, mampu mendemonstrasikan teorema matematika, termasuk beberapa di antaranya dalam Principia Mathematica karya Whitehead dan Russell yang berpengaruh.

Pada tahun 1959, Arthur Samuel (yang juga menghadiri lokakarya Dartmouth) menulis sebuah program yang mampu bermain catur dengan cukup baik hingga mampu mengalahkan penciptanya. Program ini menggabungkan beberapa konsep yang dikembangkan di bidang kecerdasan buatan, termasuk kemampuan untuk mencari solusi dalam ruang pencarian yang sangat besar dan kompleks.

Untuk bermain catur dengan baik, perlu untuk memilih, di antara semua langkah yang memungkinkan, langkah-langkah yang menghasilkan hasil terbaik. Karena, untuk setiap

langkah, lawan dapat merespons dengan salah satu dari beberapa langkah, yang juga harus dijawab oleh program, proses ini menyebabkan peningkatan yang sangat pesat dalam jumlah posisi yang perlu dianalisis. Percabangan proses pencarian ini mengambil bentuk pohon, yang kemudian disebut pohon pencarian. Mengembangkan metode untuk mengeksplorasi pohon pencarian ini secara efisien menjadi salah satu tujuan instrumental terpenting dalam bidang kecerdasan buatan.

Metode pencarian yang efisien sama pentingnya saat ini seperti ketika pertama kali dipelajari dan dikembangkan. Metode-metode ini juga diterapkan di banyak bidang lain, terutama untuk masalah perencanaan. Robot perlu melakukan pencarian untuk mengetahui cara menumpuk balok, dalam dunia balok yang disederhanakan, atau untuk menemukan jalan dari satu ruangan ke ruangan lain. Banyak hasil kecerdasan buatan dihasilkan dari studi yang dilakukan dengan lingkungan yang disederhanakan, di mana robot diajarkan untuk memanipulasi balok guna mencapai tujuan tertentu atau bergerak dalam lingkungan yang terkendali. Salah satu proyek pertama yang membuat robot melakukan tugas-tugas tertentu, dalam dunia balok yang disederhanakan, mengarah pada pengembangan sistem yang dapat memanipulasi dan menyusun balok dalam konfigurasi tertentu menggunakan penglihatan dan pemrosesan bahasa alami.

Pemrosesan bahasa alami, yang bertujuan agar komputer memproses (misalnya, menerjemahkan) dan bahkan memahami kalimat tertulis, merupakan salah satu masalah yang dipelajari dalam fase pertama kecerdasan buatan ini. Terlepas dari kesulitan yang melekat pada pemrosesan ini, yang terutama disebabkan oleh adanya ambiguitas yang besar dalam cara manusia menggunakan bahasa, sistem yang melakukan percakapan sederhana, dalam bahasa Inggris sehari-hari, dirancang. Sistem paling terkenal dari sistem-sistem awal ini, ELIZA, dirancang oleh Joseph Weizenbaum dan mampu berkomunikasi dengan pengguna, dalam bahasa Inggris tertulis yang sederhana. ELIZA menggunakan serangkaian mekanisme yang sangat sederhana untuk menjawab pertanyaan, menggunakan kalimat-kalimat yang telah ditulis sebelumnya atau sekadar merumuskan ulang pertanyaan dengan istilah yang sedikit berbeda. Meskipun sistem tersebut tidak benar-benar memahami percakapan tersebut, banyak pengguna tertipu dan mengira mereka sedang berbicara dengan manusia. ELIZA bisa dibilang merupakan salah satu sistem pertama yang lulus uji Turing, meskipun uji tersebut dilakukan dalam kondisi yang sangat spesifik dan relatif ringan.

Proyek-proyek lain bertujuan untuk menciptakan cara merepresentasikan pengetahuan manusia, sehingga dapat dimanipulasi dan digunakan untuk menghasilkan pengetahuan baru. Melalui penerapan aturan penalaran deduktif pada basis pengetahuan, misalnya, dimungkinkan untuk membangun sistem yang mampu membuat diagnosis medis dalam kondisi-kondisi tertentu yang terkontrol secara khusus, di mana pengetahuan dapat diungkapkan secara simbolis, dan dikombinasikan menggunakan aturan-aturan untuk manipulasi simbol. Beberapa sistem pakar dikembangkan berdasarkan teknik-teknik ini dan memainkan peran relevan di berbagai bidang, terutama pada tahun 1970an dan 1980an.

Proyek-proyek ini dan proyek-proyek lainnya menunjukkan bahwa beberapa kemampuan otak manusia yang tampak lebih kompleks dan canggih, seperti

mendemonstrasikan teorema matematika atau bermain permainan papan, dapat diprogram ke dalam komputer. Hasil-hasil ini menghasilkan beberapa prediksi yang terlalu optimistis tentang evolusi kecerdasan buatan di masa depan. Pada tahun 1960an, beberapa peneliti ternama, termasuk Marvin Minsky dan Herbert Simon (yang juga pernah menghadiri lokakarya Dartmouth), meramalkan bahwa dalam tiga dekade, kecerdasan serupa manusia dapat dikembangkan dalam komputer dan sistem yang dapat menjalankan fungsi apa pun yang dijalankan oleh manusia dapat diciptakan. Namun, prediksi-prediksi tersebut ternyata terlalu optimistis.

Penelitian yang dilakukan pada dekade-dekade tersebut akhirnya menunjukkan bahwa banyak tugas yang mudah dilakukan manusia ternyata sangat sulit direplikasi di komputer. Khususnya, terbukti sangat sulit untuk menerjemahkan hasil yang diperoleh dalam lingkungan yang disederhanakan, seperti dunia balok, ke lingkungan yang lebih kompleks dan tidak pasti, seperti kamar tidur, dapur, atau pabrik. Tugas sederhana seperti mengenali wajah atau memahami bahasa lisan terbukti sangat rumit dan tidak pernah dipecahkan dengan pendekatan yang hanya didasarkan pada manipulasi simbol.

Faktanya, hampir semua kemampuan otak manusia yang berkaitan dengan persepsi dan interaksi dunia nyata terbukti sangat sulit untuk direplikasi. Misalnya, menganalisis pemandangan yang ditangkap kamera dan mengidentifikasi objek-objek relevan di dalamnya merupakan tugas yang sangat sulit bagi program komputer, dan baru sekarang hal tersebut akhirnya mulai dapat dicapai oleh sistem kecerdasan buatan paling modern. Meskipun demikian, kita melakukannya tanpa upaya nyata atau pelatihan khusus. Tugas-tugas lain yang kita lakukan dengan mudah, seperti mengenali wajah yang familiar atau memahami kalimat di lingkungan yang bising, sama sulitnya untuk direplikasi.

Kesulitan ini kontras dengan relatif mudahnya menulis program komputer yang mereproduksi manipulasi simbol secara cerdas, yang dijelaskan dalam beberapa pendekatan yang telah disebutkan. Kesulitan yang agak tak terduga dalam mereproduksi perilaku yang sepele bagi manusia dan banyak hewan di komputer ini disebut paradoks Moravec: lebih mudah mereproduksi perilaku di komputer yang, bagi manusia, memerlukan penalaran matematika kompleks yang eksplisit daripada mengenali wajah atau memahami bahasa alami, sesuatu yang dilakukan seorang anak dengan sangat mudah dan tanpa instruksi khusus.

Kesulitan dalam memecahkan sebagian besar masalah yang melibatkan persepsi dan karakteristik lain dari kecerdasan manusia menyebabkan beberapa kekecewaan dengan bidang kecerdasan buatan, yang disebut musim dingin AI. Terlepas dari fase-fase negatif ini, yang ditandai oleh keputusasaan dan kurangnya dana untuk proyek-proyek di bidang tersebut, pengembangan sistem kecerdasan buatan berdasarkan manipulasi simbol berkontribusi, dengan berbagai cara, pada penciptaan banyak algoritma yang dieksekusi oleh komputer saat ini, dalam aplikasi yang paling bervariasi. Bidang ini telah mengembangkan banyak metode pencarian dan representasi pengetahuan yang memungkinkan untuk membuat banyak program yang melakukan tugas-tugas yang sering tidak kita kaitkan dengan sistem cerdas.

Misalnya, optimalisasi jadwal kereta api, pesawat terbang, dan sistem transportasi lainnya sering dilakukan oleh sistem berdasarkan algoritma pencarian dan perencanaan yang

dikembangkan oleh komunitas kecerdasan buatan. Demikian pula, sistem yang dibuat pada dekade terakhir abad kedua puluh untuk bermain catur menggunakan teknik pencarian yang pada dasarnya diusulkan oleh komunitas yang sama ini.

Metode dan algoritma yang memungkinkan terciptanya mesin pencari yang kini menjadi salah satu pilar utama Internet, dan yang memungkinkan kita menemukan dokumen relevan tentang topik tertentu dalam sepersekian detik, juga berkat sub-area spesifik kecerdasan buatan: pengambilan informasi. Sistem ini mengidentifikasi dokumen relevan berdasarkan istilah yang digunakan dalam pencarian, dan menggunakan berbagai metode untuk menentukan dokumen mana yang paling penting. Versi terbaru mesin pencari ini menggunakan model bahasa yang luas (yang akan kami jelaskan nanti) untuk lebih memahami maksud pengguna dan memberikan jawaban yang paling bermakna.

Namun demikian, dengan beberapa pengecualian, sistem yang berbasis manipulasi simbol bersifat kaku, tidak adaptif, dan rapuh. Hanya sedikit, jika ada, yang mampu berkomunikasi dalam bahasa alami, lisan maupun tulisan, dan memahami esensi pertanyaan yang kompleks. Sistem ini juga tidak mampu melakukan tugas-tugas yang membutuhkan pemrosesan gambar yang kompleks dan tidak terstruktur, atau memecahkan tantangan yang membutuhkan adaptasi terhadap lingkungan dunia nyata yang tak terkendali. Meskipun para peneliti kecerdasan buatan telah mengembangkan banyak sistem robotik, hanya sedikit yang berinteraksi dengan manusia di lingkungan yang tidak terkendali. Sistem robotik digunakan secara luas di pabrik dan lingkungan industri lainnya, tetapi secara umum, mereka melakukannya di lingkungan yang terkendali, tunduk pada aturan yang ketat dan tidak fleksibel, yang memungkinkan mereka melakukan tugas-tugas berulang, berdasarkan manipulasi komponen dan instrumen yang selalu muncul di posisi dan pengaturan yang sama.

Baru-baru ini, setelah puluhan tahun penelitian, kita mulai memiliki sistem dan robot yang berinteraksi dengan dunia nyata, dengan segala kompleksitas dan ketidakpastiannya. Meskipun mereka juga memanipulasi simbol, mereka didasarkan pada gagasan lain, gagasan bahwa komputer dapat belajar dari pengalaman, dan mengadaptasi perilaku mereka secara cerdas, seperti yang dilakukan anak-anak.

1.5 PEMBELAJARAN MESIN

Konsep Dasar

Dalam artikel yang ditulis Alan Turing pada tahun 1950, ia menunjukkan kesadaran yang jelas akan kesulitan yang melekat dalam memprogram sistem untuk berperilaku cerdas. Turing mengusulkan bahwa, sebaliknya, mungkin lebih mudah untuk membangun program yang mensimulasikan otak anak. Dengan proses pendidikan yang semestinya, otak orang dewasa akan terbentuk, yang mampu bernalar dan memiliki kecerdasan yang lebih tinggi. Turing membandingkan otak anak dengan buku kosong, yang di dalamnya tersimpan pengalaman-pengalaman seumur hidup. Turing berpendapat bahwa mungkin akan lebih mudah membangun sistem adaptif yang menggunakan pembelajaran mesin untuk memperoleh kemampuan bernalar dan memecahkan masalah-masalah kompleks yang kita kaitkan dengan kecerdasan manusia.

Apa sebenarnya gagasan pembelajaran mesin ini, gagasan bahwa komputer dapat belajar dari pengalaman? Sekilas, gagasan ini bertentangan dengan intuisi kita tentang komputer, yang kita anggap sebagai mesin yang secara membabi buta mematuhi serangkaian instruksi tertentu. Ini juga merupakan gagasan yang dimiliki Ada Lovelace, yang sangat dipengaruhi oleh komputer mekanis yang ia miliki, yang membawanya pada kesimpulan bahwa komputer tidak fleksibel dan tidak dapat menciptakan sesuatu yang baru.

Konsep kunci pembelajaran mesin adalah bahwa suatu sistem, jika dikonfigurasi dengan benar, dapat menyesuaikan perilakunya untuk memperkirakan hasil yang diinginkan untuk serangkaian masukan tertentu. Intinya, konsep ini mudah dijelaskan. Bayangkan sebuah sistem yang sangat sederhana yang menerima masukan berupa satu angka dan menghasilkan keluaran berupa satu angka, yang bergantung pada angka pertama. Jika sistem ini diperlihatkan beberapa contoh korespondensi yang diinginkan antara angka masukan dan angka keluaran, sistem dapat belajar menebak korespondensi ini.

Misalkan angka pada masukan adalah lintang suatu kota dan angka pada keluaran adalah suhu rata-rata di kota tersebut selama musim dingin. Jika sistem diberikan beberapa contoh pasangan lintang/suhu, sistem dapat mempelajari korespondensi perkiraan antara lintang dan suhu. Tentu saja, kecocokannya, secara umum, tidak akan tepat, karena variasi lokal dalam lingkungan dan lokasi kota.

Namun, dengan menggunakan teknik matematika yang familiar bagi banyak pembaca, seperti regresi, dimungkinkan untuk memperkirakan suhu rata-rata musim dingin dari lintang kota. Sistem semacam ini mungkin mewakili sistem pembelajaran mesin paling dasar yang dapat dibayangkan. Korespondensi antara garis lintang dan suhu ini tidak diprogram secara eksplisit oleh programmer mana pun, melainkan disimpulkan dari data menggunakan rumus atau algoritma matematika. Namun, program pembelajaran mesin ini sangat umum. Program ini dapat menyimpulkan korespondensi ini atau korespondensi antara pendapatan rata-rata suatu negara dan penggunaan energi per kapitanya. Setelah ditulis, program yang sama ini dapat belajar menentukan hubungan jenis tertentu (misalnya, linear) antara pasangan angka, terlepas dari masalah konkret yang dianalisis.

Dalam pembelajaran mesin, kumpulan contoh yang digunakan untuk melatih sistem disebut set pelatihan, dan kumpulan contoh yang kemudian disajikan kepada sistem untuk menguji kinerjanya disebut set pengujian. Setelah dilatih, sistem dapat digunakan berkali-kali untuk memprediksi keluaran dari nilai masukan baru, tanpa perlu pelatihan tambahan. Dalam banyak kasus (yaitu jika set pelatihan sangat besar dan hubungan yang dipelajari sangat kompleks), proses pelatihan dapat berlangsung relatif lambat, tetapi menggunakan sistem untuk menentukan kecocokan yang diinginkan untuk contoh-contoh baru dapat berlangsung cepat dan efisien.

Kembali ke contoh lintang dan suhu rata-rata musim dingin, sekarang bayangkan sistem pembelajaran diberikan, sebagai masukan, tidak hanya lintang, tetapi juga konsumsi energi rata-rata musim dingin per rumah tangga, jarak dari laut, dan variabel relevan lainnya. Mudah untuk melihat bahwa program sekarang dapat belajar, dengan presisi yang jauh lebih tinggi, untuk menghitung hubungan antara rangkaian variabel ini dan suhu rata-rata musim dingin

suatu kota. Secara matematis, permasalahannya lebih kompleks, tetapi formulasinya sama: dengan serangkaian data masukan, tujuannya adalah untuk mendapatkan program yang menghasilkan estimasi keluaran. Dengan menggunakan algoritma yang tepat, permasalahan ini tidak jauh lebih sulit daripada permasalahan sebelumnya.

Jika keluaran (yaitu, variabel yang diprediksi) adalah variabel kuantitatif (atau kumpulan variabel kuantitatif), sistem yang diperoleh melalui eksekusi algoritma pembelajaran disebut regresor, dan permasalahannya dikenal sebagai regresi. Jika tujuannya adalah menetapkan kelas tertentu pada objek yang dikarakterisasi oleh input, sistem tersebut disebut pengklasifikasi. Mari kita pertimbangkan masalah yang jauh lebih rumit: bayangkan Anda diberi gambar dengan, katakanlah, satu juta piksel. Tujuannya adalah mempelajari cara mengklasifikasikan gambar berdasarkan isinya; misalnya, apakah setiap gambar berisi kucing atau anjing atau tidak. Sekali lagi, kita memiliki beberapa variabel sebagai input, dalam hal ini, tiga juta variabel untuk gambar berwarna satu megapiksel, dan tujuannya adalah menghasilkan variabel dalam output yang menunjukkan isi foto, misalnya anjing, kucing, atau mobil. Memang, terdapat kumpulan data yang sangat besar, seperti ImageNet yang memiliki lebih dari 14 juta gambar dalam lebih dari 20.000 kategori, yang digunakan untuk melatih sistem pembelajaran mesin.

Pembaca yang cermat akan menyadari bahwa masalah ini pada dasarnya tidak berbeda dengan masalah pertama yang dibahas di atas. Dalam kasus ini, perlu dihitung korespondensi antara tiga juta angka, masukan, dan kelas yang diinginkan, keluaran. Meskipun memiliki formulasi yang sama, masalah ini jauh lebih sulit. Saat ini tidak ada korespondensi langsung, melalui rumus matematika yang kurang lebih sederhana, antara masukan dan keluaran yang diinginkan.

Untuk mendapatkan kelas yang tepat, kita perlu mampu mengidentifikasi beragam karakteristik gambar, seperti mata, roda, atau kumis, dan bagaimana fitur-fitur ini saling terkait secara spasial. Di sini, ide perintis Alan Turing juga berhasil. Meskipun sangat sulit (bahkan mungkin mustahil bagi manusia) untuk menulis program yang memetakan masukan ke keluaran, gambar ke kategori, adalah mungkin untuk mempelajarinya dari label dan deskripsi yang dibuat manusia untuk gambar-gambar ini.

Gagasan bahwa komputer dapat belajar dari contoh dikembangkan sejak tahun 1960an dan seterusnya oleh banyak peneliti dan ilmuwan. Pendekatan pertama, yang menggunakan representasi simbolis untuk mempelajari korespondensi antara masukan dan keluaran, akhirnya digantikan oleh metode statistik dan/atau numerik. Ada banyak cara untuk mempelajari korespondensi antara nilai-nilai dalam masukan dan keluaran yang diinginkan, dan di sini mustahil untuk menjelaskan secara minimal dan lengkap, bahkan sebagian kecil dari algoritma yang digunakan. Namun, dimungkinkan untuk menyajikan, secara singkat, filosofi yang mendasari sebagian besar pendekatan ini.

Pendekatan Statistik

Sekelompok pendekatan yang berasal dari statistika didasarkan pada estimasi, dari data latih, hubungan statistik antara masukan dan keluaran. Beberapa pendekatan statistik ini (dalam subkelas yang biasanya disebut generatif) didasarkan pada hukum Bayes yang terkenal.

Hukum ini, yang sebenarnya merupakan sebuah teorema, menghitung probabilitas kemunculan suatu peristiwa (misalnya, kelas suatu citra tertentu) dari pengetahuan sebelumnya tentang peristiwa tersebut (misalnya, probabilitas bahwa suatu citra acak berisi seekor anjing, seekor kucing, seseorang, ...) dan tentang bagaimana masukan terkait dengan observasi (misalnya, apa statistik dari citra yang berisi anjing).

Kelompok pendekatan statistik lainnya (biasanya disebut diskriminatif) mengabaikan penerapan hukum Bayes dan, dari data latih, secara langsung mengestimasi model tentang bagaimana probabilitas nilai keluaran bergantung pada masukan (misalnya, probabilitas bahwa suatu citra berisi seekor anjing, dengan mempertimbangkan semua nilai piksel citra tersebut). Dalam permasalahan kompleks (seperti klasifikasi citra), pendekatan diskriminatif jauh lebih lazim, karena memiliki dua keuntungan penting: pendekatan ini memanfaatkan data yang tersedia secara lebih efisien dan tidak terlalu bergantung pada asumsi tentang bentuk hubungan statistik yang mendasarinya, sehingga lebih robust dan bersifat umum.

Pendekatan Berbasis Kesamaan

Pendekatan lain berfokus pada penilaian kesamaan antara berbagai contoh dalam set pelatihan. Bayangkan kita ingin menentukan berat badan seseorang berdasarkan tinggi badan, jenis kelamin, usia, dan lingkar pinggangnya. Dan misalkan kita ingin menentukan berat badan individu baru yang belum pernah terlihat sebelumnya. Pendekatan yang sederhana dan pragmatis, jika terdapat banyak contoh dalam set pelatihan, adalah dengan mencari seseorang (atau sekumpulan orang) dalam set tersebut dengan karakteristik yang sangat mirip dengan individu yang dimaksud, dan menduga bahwa berat badan individu tersebut sama dengan berat badan orang yang paling mirip dalam set pelatihan.

Pendekatan ini, pembelajaran melalui analogi, efektif jika terdapat set pelatihan yang luas, dan dapat dilakukan dengan berbagai cara, dengan algoritma yang penunjukannya mencakup metode tetangga terdekat (atau k tetangga kearest). Perluasan kelas metode ini, yang didasarkan pada penilaian kesamaan antara objek yang akan diklasifikasikan dan objek dalam set pelatihan, menghasilkan kelas metode yang dikenal sebagai mesin kernel (yang anggota paling terkenalnya adalah mesin vektor pendukung), yang sangat berpengaruh dan memiliki dampak signifikan pada dekade terakhir abad ke-20 dan awal abad ini.

Pohon Keputusan

Kelas metode lain, yang dikenal sebagai pohon keputusan, bekerja dengan membagi data menjadi serangkaian keputusan biner. Mengklasifikasikan objek baru sama dengan menelusuri pohon berdasarkan keputusan-keputusan ini, bergerak melalui keputusan-keputusan tersebut hingga mencapai simpul daun, yang akan mengembalikan prediksi (kelas atau nilai). Pohon keputusan dibangun dengan memanfaatkan heuristik yang dikenal sebagai partisi rekursif, yang memanfaatkan rasional "bagi dan kuasai". Pohon keputusan memiliki beberapa keunggulan penting, seperti dapat diterapkan secara mulus pada data heterogen (dengan variabel kuantitatif dan kategoris, yang tidak berlaku untuk sebagian besar metode lain), agak analog dengan proses berpikir pengambilan keputusan manusia, dan, yang sangat penting, transparan, karena rantai keputusan yang mengarah pada prediksi akhir memberikan penjelasan untuk prediksi tersebut.

Pohon keputusan juga dapat digabungkan menjadi apa yang disebut hutan acak, suatu jenis model yang menggunakan beberapa pohon keputusan, yang masing-masing dipelajari dari subset data yang berbeda. Prediksi hutan diperoleh dengan merata-ratakan pohon-pohon, yang diketahui dapat meningkatkan akurasinya, dengan mengorbankan beberapa hilangnya transparansi dan keterjelasan. Hutan acak, hingga saat ini, masih menjadi salah satu metode pilihan dalam permasalahan yang melibatkan data heterogen dan yang jumlah data pelatihannya tidak banyak.

Jaringan Syaraf Tiruan

Jaringan saraf tiruan merupakan salah satu pendekatan paling fleksibel dan canggih yang digunakan saat ini. Pendekatan ini dirintis pada tahun 1980an dan sangat terinspirasi oleh karya-karya terdahulu tentang model matematika neuron biologis, yaitu oleh McCulloch dan Pitts, dalam sebuah makalah terkenal tahun 1943 berjudul "Kalkulus Logika Ide-Ide yang Imanen dalam Aktivitas Saraf", yang mengusulkan model matematika pertama neuron biologis. Pada tahun 1940an, Donald Hebb mengusulkan proses pembelajaran pertama yang masuk akal secara biologis untuk jaringan saraf (aturan Hebbian), yang kemudian dirangkum secara terkenal dalam kalimat "sel yang bekerja bersama akan terhubung bersama".

Meskipun fungsi neuron biologis kompleks dan sulit dimodelkan, dimungkinkan untuk membangun model matematika abstrak dari pemrosesan informasi yang dilakukan oleh masing-masing sel ini, di mana model karya McCulloch dan Pitts adalah yang pertama. Dalam model yang disederhanakan ini, setiap sel mengakumulasi eksitasi yang diterima pada inputnya dan, ketika nilai eksitasi ini melebihi ambang batas tertentu, sel tersebut akan aktif, menstimulasi neuron yang terhubung ke outputnya.

Dengan mengimplementasikan atau mensimulasikan interkoneksi beberapa unit ini (neuron buatan) di komputer, dimungkinkan untuk mereproduksi mekanisme pemrosesan informasi dasar yang digunakan oleh otak biologis. Dalam otak nyata, neuron saling terhubung melalui sinapsis dengan kekuatan yang bervariasi. Dalam jaringan saraf yang digunakan dalam pembelajaran mesin, kekuatan koneksi antar neuron ditentukan oleh angka yang mengontrol bobot yang memengaruhi status neuron yang terhubung dengannya.

Metode matematika, yang disebut algoritma pelatihan, digunakan untuk menentukan nilai bobot interkoneksi ini guna memaksimalkan akurasi korespondensi antara keluaran yang dihitung dan keluaran yang diinginkan, pada suatu set pelatihan. Faktanya, algoritma ini pada dasarnya merupakan bentuk umpan balik, di mana setiap kesalahan yang dilakukan pada sampel pelatihan diumpan balikkan ke jaringan untuk menyesuaikan bobot sedemikian rupa sehingga kemungkinan kesalahan ini menjadi lebih kecil. Algoritma pelatihan yang lazim dalam pembelajaran mesin modern berakar kuat pada karya Norbert Wiener, salah satu matematikawan dan ilmuwan terhebat abad ke-20.

Karya penting Wiener dalam sibernetika (bidang yang ia ciptakan dan namakan) mencakup formalisasi gagasan umpan balik, yang merupakan salah satu landasan sebagian besar teknologi modern. Wiener juga memengaruhi karya awal pada jaringan saraf dengan membawa McCulloch dan Pitts ke MIT dan menciptakan kelompok penelitian pertama di mana

para ahli saraf, matematikawan, dan biofisikawan bergabung dalam upaya untuk mencoba memahami cara kerja otak biologis.

Contoh pertama algoritma pembelajaran yang berhasil untuk jaringan saraf tiruan, mengikuti rasional umpan balik kesalahan untuk memperbarui bobot jaringan, diusulkan oleh Frank Rosenblatt pada tahun 1958 untuk jenis jaringan tertentu, yang disebut perceptron. Perceptron, bersama dengan algoritma Rosenblatt, merupakan cikal bakal jaringan saraf modern dan algoritma pembelajaran. Keberhasilan awal perceptron memicu gelombang antusiasme dan optimisme yang besar, yang kemudian berubah menjadi kekecewaan ketika menjadi jelas bahwa optimisme ini terlalu dilebih-lebihkan. Sebuah momen simbolis dalam runtuhnya ekspektasi adalah penerbitan buku terkenal "Perceptrons", karya Minsky dan Papert, pada tahun 1969.

Buku ini memberikan bukti matematis tentang keterbatasan perceptron serta pernyataan yang tidak didukung bukti mengenai tantangan pelatihan perceptron multi-lapis (yang mereka sadari akan mengatasi keterbatasan tersebut). Kekecewaan ini menyebabkan penurunan drastis minat dan pendanaan untuk penelitian jaringan saraf, yang berlangsung selama lebih dari tiga dekade. Jaringan saraf tiruan modern muncul dari kesadaran bahwa memang memungkinkan untuk mempelajari/melatih jaringan dengan beberapa lapisan (saat ini dikenal sebagai jaringan saraf tiruan dalam, yang akan dibahas di bagian selanjutnya), yang secara matematis dan eksperimental dapat dibuktikan sebagai struktur yang sangat fleksibel, yang mampu memecahkan banyak masalah prediksi.

Inti dari kemungkinan ini adalah prosedur, yang dikenal sebagai backpropagation, yang memungkinkan penerapan umpan balik yang telah disebutkan sebelumnya dari kesalahan prediksi dalam contoh pelatihan hingga penyesuaian bobot jaringan, yang bertujuan untuk mengurangi kemungkinan kesalahan ini. Istilah backpropagation dan penerapannya dalam jaringan saraf tiruan berasal dari Rumelhart, Hinton, dan Williams, pada tahun 1986, tetapi teknik ini ditemukan kembali secara independen berkali-kali, dan memiliki banyak pendahulu yang berasal dari tahun 1960an, yaitu dalam teori kendali umpan balik.

Jaringan saraf tiruan modern dapat memiliki hingga jutaan neuron dan miliaran bobot. Jaringan saraf tiruan yang cukup kompleks dapat digunakan untuk memproses gambar, suara, teks, dan bahkan video. Misalnya, ketika digunakan untuk memproses gambar, setiap neuron di salah satu jaringan ini akhirnya belajar mengenali karakteristik tertentu dari gambar tersebut. Satu neuron mungkin mengenali garis pada posisi tertentu, neuron lain (yang lebih dalam, yaitu, lebih jauh dari input) mungkin mengenali garis luar hidung, dan neuron ketiga, yang lebih dalam lagi, mungkin mengenali wajah tertentu. Sekali lagi, beberapa karya modern tentang jaringan saraf untuk menganalisis gambar berakar dari karya-karya dari akhir 1950an dan awal 1960an, yaitu model saraf korteks visual mamalia yang dijelaskan oleh Hubel dan Wiesel, yang karenanya mereka menerima Hadiah Nobel pada tahun 1981.

Meskipun otak biologis telah menginspirasi jaringan saraf tiruan dan mekanisme pembelajaran, penting untuk disadari bahwa, dalam keadaan teknologi saat ini, jaringan-jaringan ini tidak bekerja dengan cara yang sama seperti otak biologis. Meskipun jaringan jenis ini telah dilatih untuk mengemudikan kendaraan, memproses teks, mengenali wajah dalam

video, dan bahkan memainkan permainan tingkat juara seperti catur atau Go (permainan papan yang sangat kompleks dan populer di Asia), keliru jika berpikir bahwa jaringan ini menggunakan mekanisme yang sama dengan yang digunakan otak manusia untuk memproses informasi. Dalam kebanyakan kasus, jaringan ini dilatih untuk memecahkan masalah yang sangat spesifik dan tidak mampu menangani masalah lain, apalagi membuat keputusan secara mandiri tentang masalah mana yang harus ditangani.

Bagaimana otak manusia mengorganisir dirinya sendiri, melalui proses perkembangan dan pembelajaran yang terjadi selama masa kanak-kanak dan remaja, bagaimana ingatan baru disimpan sepanjang hidup, bagaimana berbagai tujuan dikejar dari waktu ke waktu, oleh siapa pun di antara kita, bergantung pada mekanisme yang pada dasarnya tidak diketahui dan tidak ada dalam jaringan saraf tiruan. Namun, dekade terakhir telah menyaksikan munculnya teknologi yang memungkinkan kita menciptakan sistem yang sangat kompleks yang, setidaknya di permukaan, menunjukkan perilaku yang agak lebih cerdas.

1.6 REVOLUSI PEMBELAJARAN MENDALAM

Dalam dekade terakhir, pembelajaran mesin telah menghasilkan perkembangan luar biasa dalam aplikasi di mana metode simbolik kurang berkinerja baik, seperti visi komputer, pengenalan suara, dan pemrosesan bahasa alami. Perkembangan ini secara kolektif dikenal sebagai pembelajaran mendalam. Kata sifat "dalam" mengacu pada penggunaan beberapa lapisan dalam jaringan saraf tiruan, meskipun pendekatan lain yang tidak menggunakan jaringan saraf tiruan juga termasuk dalam lingkup pembelajaran mendalam.

Pembelajaran mendalam umumnya menggunakan arsitektur jaringan saraf tiruan dengan banyak lapisan (pada dasarnya penggabungan banyak perseptron dalam struktur multilapis), yang menghasilkan aplikasi baru dan implementasi yang dioptimalkan, terutama karena tersedianya kumpulan data yang sangat besar untuk melatih struktur besar ini dengan banyak parameter (bobot yang disebutkan di atas) dan prosesor komputer tujuan khusus yang sangat efisien, seperti GPU.

Dalam pembelajaran mendalam, setiap lapisan jaringan saraf tiruan belajar untuk mengubah masukannya menjadi representasi yang lebih abstrak dan komposit. Misalnya, dalam pengenalan atau klasifikasi gambar, masukan mentahnya dapat berupa matriks piksel, lapisan pertama dapat mengabstraksi piksel dan mengodekan tepi atau sudut, lapisan kedua dapat menyusun dan mengodekan susunan tepi dan sudut menjadi garis dan bentuk lain, dan seterusnya hingga konsep yang bermakna secara semantik, seperti wajah atau objek, atau tumor dalam citra medis. Inti dari algoritma yang mempelajari jaringan saraf dalam ini adalah algoritma backpropagation yang telah disebutkan di bagian sebelumnya.

Pembelajaran mendalam telah meraih banyak keberhasilan luar biasa dalam beberapa tahun terakhir. Permainan papan telah populer dalam penelitian kecerdasan buatan sejak awal kemunculannya di tahun 1950an. Permainan yang lebih sederhana, seperti dam dan backgammon, telah dikuasai oleh mesin beberapa dekade yang lalu, tetapi permainan lain yang lebih kompleks, seperti catur dan Go, membutuhkan waktu lebih lama untuk

diselesaikan. Deep Blue milik IBM adalah program catur pertama yang mengalahkan juara dunia, ketika mengalahkan Garry Kasparov pada tahun 1997 dalam pertandingan ulang.

Namun, Deep Blue, seperti banyak program catur lainnya, sangat bergantung pada strategi bermain yang dirancang manusia dan sangat bergantung pada pencarian dengan kekuatan kasar. Dimulai pada tahun 2016, serangkaian pengembangan oleh DeepMind, sebuah perusahaan milik Google, menghasilkan peluncuran AlphaGo, sebuah sistem yang mengalahkan pemain Go terbaik di dunia, setelah belajar dari permainan para ahli dan bermain sendiri. Perkembangan selanjutnya menghasilkan AlphaGo Zero dan AlphaZero, sistem yang unggul dalam Go, catur, dan permainan lainnya, tetapi tidak perlu belajar dari para ahli manusia dan belajar secara unik dari bermain sendiri, menggunakan pembelajaran mendalam dan teknik untuk pengambilan keputusan berurutan yang dikenal sebagai pembelajaran penguatan.

Hasil yang dicapai oleh program-program ini luar biasa, karena mereka mampu mempelajari teknik dan strategi dalam hitungan hari yang telah luput dari manusia selama ribuan tahun sejak dimulainya permainan ini. Permainan yang dimainkan mesin-mesin ini saat ini sedang dipelajari untuk memahami strategi dan teknik baru yang mereka kembangkan dan belum pernah dilakukan manusia.

Bidang lain di mana pembelajaran mendalam telah menghasilkan kemajuan signifikan adalah visi komputer, yang tujuannya adalah memungkinkan komputer untuk memproses, menganalisis, dan bahkan memahami gambar dan video. Salah satu fitur penting dari arsitektur jaringan saraf yang dikembangkan untuk visi komputer adalah bahwa lapisan pertama bersifat konvolusional, terinspirasi oleh arsitektur lapisan pertama sistem saraf visual mamalia, seperti yang ditemukan oleh Hubel dan Wiesel pada tahun 1950-1960an. Lapisan konvolusional memiliki jenis invariansi tertentu (lebih tepatnya, ekivariansi, tetapi itu di luar cakupan pengantar ini), yang berarti, secara sederhana, bahwa pemrosesan yang dilakukan di setiap lokasi gambar adalah sama di seluruh gambar.

Lapisan konvolusional merupakan perwujudan dari apa yang disebut bias induktif: properti jaringan yang dirancang alih-alih dipelajari, berdasarkan pengetahuan tentang data yang akan diproses dan tujuan jaringan. Lebih khusus lagi, bias induktif dalam hal ini adalah bahwa untuk mengenali keberadaan suatu objek dalam suatu gambar, lokasinya tidak relevan. Implikasi krusial lain dari sifat konvolusional jaringan ini adalah, karena invariansi ini, jumlah parameter yang perlu dipelajari jauh lebih kecil dibandingkan jaringan arbitrer dengan ukuran yang sama.

Dengan menggabungkan fitur arsitektur baru, seperti lapisan konvolusional (dan berbagai trik lainnya) dengan kumpulan data besar dan mesin pemrosesan canggih berbasis GPU dan TPU (unit pemrosesan tensor), pembelajaran mendalam telah menghasilkan ledakan dalam berbagai aplikasi visi komputer. Aplikasi ini mencakup pengenalan wajah (banyak digunakan dalam ponsel pintar modern dan pengawasan otomatis), pengenalan dan klasifikasi gambar, pengawasan, inspeksi fasilitas otomatis, rekonstruksi dan analisis citra medis, serta pengemudian otonom, dan masih banyak lagi.

Beberapa arsitektur pembelajaran mendalam modern, seperti transformer, telah dikembangkan untuk pemrosesan bahasa alami dan digunakan untuk membangun model bahasa berskala besar. Model bahasa besar ini, yang bersifat statistik, seperti GPT-3, telah dilatih dalam korpus dengan triliunan kata, dan dapat secara akurat menjawab pertanyaan, melengkapi kalimat, dan menulis artikel dan catatan tentang berbagai topik. Mereka menjadi semakin berguna dalam pengembangan alat interaksi pelanggan, seperti yang ditunjukkan oleh rilis terbaru ChatGPT dan GPT-4, yang meniru dengan cara yang sangat mengesankan perilaku agen manusia dalam menganalisis dan menjawab permintaan yang dibuat dalam bahasa alami.

Dalam beberapa hal, model bahasa besar ini mendekati visi Turing tentang mesin yang berinteraksi dengan cara yang, jika hanya berdasarkan teks, tidak dapat dibedakan dari interaksi dengan manusia. Minat yang sangat besar yang ditimbulkan oleh rilis ChatGPT menunjukkan potensi pendekatan ini, meskipun sistem ini mewakili, sebenarnya, hanya satu langkah maju dalam evolusi teknologi yang digunakan dalam model bahasa besar, yang merupakan hasil dari revolusi pembelajaran mendalam.

1.7 APLIKASI DALAM ANALISIS DAN OTOMASI

Kecerdasan buatan modern memiliki banyak aplikasi, di berbagai bidang yang paling beragam, dan tidak dapat dengan mudah diklasifikasikan menggunakan taksonomi sederhana. Namun, aplikasi-aplikasi tersebut secara garis besar dapat dikelompokkan menjadi dua area besar yang tidak eksklusif: analitik dan otomasi.

Analisis memiliki hubungan yang kuat dengan bidang-bidang lain dengan sebutan seperti ilmu data, data besar, penambangan data, atau intelijen bisnis. Tujuan mendasar analitik adalah untuk mengorganisasikan data yang ada tentang orang, organisasi, bisnis, atau proses, untuk mengekstraksi nilai ekonomi dari data tersebut, dengan mengidentifikasi keteraturan, kecenderungan, atau sensitivitas yang rentan terhadap eksploitasi. Banyak perusahaan terbesar di dunia, termasuk yang umumnya dikenal sebagai GAFAM (Google, Apple, Facebook, Amazon, dan Microsoft) berutang banyak nilai mereka pada kemampuan untuk mengorganisasikan data yang disumbangkan oleh pengguna mereka dan menjualnya kembali untuk mendukung iklan bertarget atau kampanye penjualan lainnya.

Namun, aplikasi analitik jauh melampaui periklanan dan pemasaran. Jika dieksplorasi dengan tepat, data yang diperoleh dengan beragam cara juga dapat digunakan untuk menemukan pengetahuan ilmiah baru dan mengoptimalkan proses desain, manufaktur, distribusi, atau penjualan, yang merupakan komponen penting dari area lain yang dikenal sebagai Industri 4.0. Setiap perusahaan yang ingin bersaing secara internasional saat ini atau setiap institusi yang ingin memberikan layanan yang baik kepada penggunanya harus menggunakan perangkat analitik untuk mengeksplorasi dan menggunakan data yang mereka miliki.

Otomatisasi, di sisi lain, berkaitan dengan penggantian sebagian atau seluruh manusia dalam tugas-tugas yang biasanya membutuhkan kecerdasan. Area ini, yang dampak ekonominya mungkin masih lebih kecil daripada analitik, akan tumbuh pesat di tahun-tahun

mendatang, karena perusahaan dan institusi terus menghadapi tekanan untuk menjadi lebih efisien dan mengurangi biaya. Area yang beragam seperti dukungan pelanggan, layanan hukum, sumber daya manusia, logistik, distribusi, perbankan, layanan, dan transportasi akan secara bertahap bertransformasi karena fungsi-fungsi yang sebelumnya dilakukan oleh karyawan manusia secara bertahap diotomatisasi oleh mesin. Proses penggantian ini akan berlangsung progresif dan bertahap, memberikan waktu bagi perusahaan untuk beradaptasi. Namun, mau tidak mau, tugas-tugas seperti layanan pelanggan, pengawasan fasilitas, analisis dokumen hukum, diagnosis medis, pengemudian kendaraan, dan banyak lagi lainnya akan secara bertahap dilakukan oleh sistem otomatis berbasis kecerdasan buatan. Kondisi teknologi saat ini belum memungkinkan penggantian sepenuhnya tenaga profesional di sebagian besar tugas ini, tetapi kemajuan teknologi tampaknya tak terelakkan dan konsekuensinya, dalam jangka panjang, tak terbantahkan.

Baru-baru ini, Komisi Eropa mengusulkan dua dokumen untuk kemungkinan diadopsi oleh Parlemen Eropa yang bertujuan untuk mengatur berbagai aspek penerapan teknologi kecerdasan buatan. Dokumen-dokumen ini sebagian selaras dengan taksonomi ini. Undang-Undang Pasar Digital, yang diusulkan pada Desember 2020, berfokus pada kebutuhan untuk mengatur akses ke data, untuk tujuan analitik, dan mencegah kontrol berlebihan atas data ini oleh platform besar (dokumen tersebut menyebutnya Gatekeeper), yang akan mengarah pada situasi dominasi pasar yang luar biasa.

Undang-Undang Kecerdasan Buatan, yang diusulkan pada April 2021, lebih berfokus pada masalah yang disebabkan oleh potensi aplikasi berisiko tinggi yang terutama berada di bidang otomatisasi. Sistem yang dianggap berisiko tinggi oleh dokumen tersebut mencakup, antara lain, sistem yang mengidentifikasi orang, mengoperasikan infrastruktur penting, merekrut atau memilih kandidat untuk posisi atau tunjangan, mengendalikan akses ke fasilitas dan negara, atau berperan dalam pendidikan atau administrasi peradilan.

Ambisi Uni Eropa adalah untuk mengatur teknologi kecerdasan buatan, menjaga keamanan dan privasi warga negara, menjamin persaingan, dan menjaga keterbukaan pasar sekaligus mendorong pengembangan aplikasi baru yang aman dan non-invasif. Namun, kesenjangan antara regulasi esensial dan regulasi yang berlebihan sangatlah kecil, dan kompromi seringkali sulit dicapai. Misalnya, Peraturan Perlindungan Data Umum (GDPR) tentu saja merupakan bagian penting dalam perlindungan hak-hak warga negara dan telah menempatkan Eropa di garis depan dalam bidang ini. Namun, peraturan ini juga merupakan peraturan yang kepatuhannya menimbulkan banyak tantangan, tuntutan, dan kesulitan bagi perusahaan dan lembaga. Mengenai regulasi kecerdasan buatan, harapannya adalah Eropa akan berhasil menemukan keseimbangan yang tepat, menjaga hak-hak individu tetapi juga menciptakan kondisi untuk penciptaan inovasi, yang akan terus menjadi mesin pendorong peningkatan produktivitas dan pembangunan ekonomi.

1.8 AI: PENGGERAK REVOLUSI INDUSTRI KEEMPAT

Kecerdasan buatan, bidang yang dikembangkan untuk mewujudkan gagasan bahwa mesin suatu hari nanti juga akan mampu berpikir, kini merupakan teknologi penting yang

berada di balik perubahan besar yang akan memengaruhi masyarakat dalam beberapa dekade mendatang, yang secara global dikenal sebagai revolusi industri keempat.

Aplikasi dalam analitik dan otomatisasi akan berkembang pesat dalam beberapa dekade mendatang, yang akan membawa perubahan dalam cara kita hidup, bekerja, dan berinteraksi. Meskipun dunia tampak sangat berubah berkat teknologi yang sudah ada, kita harus siap menghadapi perubahan yang lebih mendalam di tahun-tahun mendatang, yang dibawa oleh konvergensi beragam teknologi, di mana kecerdasan buatan merupakan salah satu yang paling penting.

Meskipun kita masih belum tahu bagaimana mereproduksi perilaku cerdas setingkat manusia pada mesin, sebuah tujuan yang dikenal sebagai kecerdasan umum buatan, upaya skala besar untuk mengembangkan teknologi yang dapat menghasilkan hasil tersebut sedang dilakukan oleh semua blok ekonomi utama, perusahaan, lembaga penelitian, dan universitas. Sangat mungkin bahwa upaya gabungan jutaan peneliti pada akhirnya akan menjelaskan salah satu pertanyaan terpenting yang dihadapi umat manusia: apa itu kecerdasan dan dapatkah ia direproduksi pada mesin? Jika kecerdasan umum buatan memang memungkinkan dan menjadi kenyataan di masa depan, hal ini akan menimbulkan pertanyaan-pertanyaan praktis, etis, dan sosial yang signifikan, yang harus dibahas dan ditangani dari berbagai sudut pandang.

BAB 2
DAMPAK TEKNOLOGI BAHASA DALAM RANAH HUKUM

Abstrak Di era digital saat ini, teknologi bahasa memainkan peran yang semakin vital dalam ranah hukum, membantu pengguna, pengacara, hakim, dan profesional hukum untuk memecahkan berbagai masalah dunia nyata. Meskipun kumpulan data terbuka dan metodologi pembelajaran mendalam yang inovatif telah menghasilkan terobosan terbaru di bidang ini, upaya signifikan masih dilakukan untuk mentransfer perkembangan teoretis/algoritmik, yang terkait dengan pemrosesan teks dan ucapan umum, ke dalam aplikasi nyata dalam ranah hukum. Bab ini menyajikan tinjauan singkat tentang teknologi bahasa untuk menangani tugas-tugas hukum, yang mencakup studi dan aplikasi yang terkait dengan pemrosesan teks dan ucapan.

2.1 LEGAL AI: TRANSFORMASI HUKUM MELALUI NLP

Hukum merupakan salah satu bidang yang dapat memperoleh manfaat besar dari kemajuan pesat Kecerdasan Buatan (AI), terutama yang berkaitan dengan teknologi bahasa. Bahkan, dapat dikatakan bahwa AI sedang mengubah bidang ini. Perubahan-perubahan ini tercermin dalam istilah-istilah yang baru-baru ini muncul seperti "Legal AI", yang mencakup ratusan metode yang diusulkan untuk pengambilan informasi, penambangan teks/pengetahuan, dan Pemrosesan Bahasa Alami (NLP). Dalam literatur, NLP seringkali terbatas pada pemrosesan teks, tetapi kami mengambil pandangan menyeluruh yang mencakup pemrosesan bahasa tulis dan lisan, karena pemrosesan teks dan ucapan memainkan peran penting dalam membentuk masa depan AI hukum.

Oleh karena itu, kami menyusun tinjauan singkat ini menjadi dua bagian utama, yang mencakup analisis teks dan ucapan. Kami menjelaskan bagaimana bidang ini telah berubah dalam dekade terakhir, dan bagaimana berbagai teknologi bahasa dapat berkontribusi dalam penyusunan, pendiktean, analisis, dan anonimisasi dokumen hukum, penyederhanaan penelitian hukum, prediksi putusan, transkripsi proses pengadilan, dll. Lebih lanjut, bab ini juga berupaya menyoroti potensi penyalahgunaan teknologi bahasa, dan dampaknya dalam ranah hukum.

2.2 TEKNOLOGI PEMROSESAN BAHASA UNTUK MEMPROSES DATA TEKSTUAL

Pemrosesan Bahasa Alami (NLP) dalam ranah hukum telah menangani tugas-tugas analisis teks seperti prediksi putusan hukum, klasifikasi topik hukum, pengambilan dokumen hukum dan tanya jawab, atau pemahaman kontrak, dan masih banyak lagi. Seperti di area aplikasi NLP lainnya, kemajuan sering kali telah dibuat sehubungan dengan kumpulan data yang tersedia untuk umum, yang dapat digunakan para peneliti untuk mengevaluasi kinerja sistem dengan cara yang terstandarisasi.

Inisiatif evaluasi bersama (yaitu, tugas bersama) juga populer di area ini. Dalam kompetisi ini, tim peneliti mengirimkan sistem yang mengatasi tantangan spesifik yang telah

ditentukan sebelumnya untuk tugas bersama, dan hasilnya kemudian dievaluasi terhadap "standar emas" yang sebelumnya disiapkan oleh penyelenggara tugas bersama. Contoh untuk tugas bersama yang terkait dengan NLP hukum meliputi Kompetisi Ekstraksi dan Entailmen Informasi Hukum, tantangan AI dan Hukum Tiongkok, yang berlangsung setiap tahun sejak 2018, atau seri tugas bersama Kecerdasan Buatan untuk Bantuan Hukum, yang dimulai pada 2019. Bidang ini juga memiliki sejarah panjang, yang mencerminkan perubahan yang juga terjadi dalam bidang NLP secara umum selama bertahun-tahun.

Hingga tahun 1980an, sebagian besar sistem NLP didasarkan pada pendekatan simbolis yang memanfaatkan aturan tertulis. Dimulai pada akhir 1980an, terjadi pergeseran dengan diperkenalkannya algoritma pembelajaran mesin untuk NLP, yang menggunakan inferensi statistik untuk mempelajari aturan secara otomatis melalui analisis korpus besar. Pada tahun 2010an, pembelajaran representasi dan metode pembelajaran mesin bergaya jaringan saraf dalam (JNN) menjadi luas dalam NLP. Teknik-teknik populer meliputi penggunaan penyisipan kata untuk menangkap sifat semantik kata, dan peningkatan pembelajaran menyeluruh (end-to-end learning) untuk tugas-tugas tingkat tinggi (misalnya, menjawab pertanyaan), alih-alih mengandalkan serangkaian tugas perantara yang terpisah (misalnya, penandaan jenis kata dan penguraian ketergantungan sintaksis).

Perbedaan-perbedaan ini dapat memengaruhi kinerja model NLP generik, memotivasi penelitian di bidang spesifik ini. Bahkan dalam kasus metode modern berdasarkan pembelajaran ujung ke ujung, model pra-pelatihan dengan teks hukum dapat membantu menangkap karakteristik yang disebutkan di atas dengan lebih baik, menyediakan pengetahuan dalam domain yang hilang dari korpora generik lainnya.

Faktanya, beberapa model bahasa hukum pra-pelatihan, berdasarkan jaringan saraf yang sangat besar, telah diperkenalkan baru-baru ini. Pendekatan NLP mutakhir didasarkan pada jenis model ini, mengikuti desain yang didasarkan pada model bahasa saraf pra-pelatihan pada sejumlah besar teks (idealnya dalam domain), misalnya. dengan mempertimbangkan tujuan tanpa pengawasan seperti memprediksi kata-kata yang tersamar dari kalimat nyata, diikuti dengan penyempurnaan model-model ini secara terawasi untuk tugas-tugas hilir tertentu.

Anonimisasi Teks

Anonimisasi data adalah proses penyembunyian atau penghapusan data sensitif dari suatu dokumen dengan tetap mempertahankan format aslinya. Proses ini penting untuk berbagi dokumen hukum dan putusan pengadilan tanpa mengekspos informasi sensitif apa pun. Teks bentuk bebas adalah jenis dokumen khusus di mana data tersimpan secara tidak terstruktur, sebagaimana direpresentasikan dalam bahasa alami.

Putusan pengadilan adalah contoh dari jenis dokumen ini. Dari isi dokumen-dokumen ini, perlu untuk mengidentifikasi struktur teks yang mewakili nama atau pengenal unik, yang dikenal sebagai entitas. Tugas ini umumnya disebut sebagai NER (Pengenalan Entitas Bernama). Tiga kelas utama NER adalah: orang, lokasi, dan organisasi. Kelas penting lainnya meliputi tanggal, nomor telepon, plat nomor mobil, referensi rekening bank (misalnya IBAN), dan situs web.

Kegunaan utama sistem anonimisasi teks otomatis adalah untuk mendeidentifikasi rekam medis dan putusan pengadilan. Sistem anonimisasi generik biasanya terdiri dari hingga empat modul: (1) modul yang menormalkan teks dan melakukan ekstraksi fitur; (2) seperangkat pengklasifikasi NE; (3) jajak pendapat untuk memilih kelas NE yang paling mungkin; dan (4) modul yang menerapkan metode anonimisasi pada NE dan mengganti kemunculan entitas-entitas ini dalam teks.

Salah satu sistem anonimisasi otomatis pertama adalah Scrub. Sistem ini diperkenalkan oleh Sweeney dan menggunakan pencocokan pola dan kamus. Sistem ini menjalankan beberapa algoritma secara paralel untuk mendeteksi berbagai kelas entitas. Pada tahun 2006, bagian dari Tantangan i2b2 (Informatika untuk Mengintegrasikan Biologi ke Tempat Tidur) didedikasikan untuk de-identifikasi data klinis. Tujuh sistem berpartisipasi dalam tantangan ini. Sistem MITRE, yang dikembangkan oleh Wellner dkk., mencapai kinerja tertinggi. Sistem MITRE menggunakan dua alat NER berbasis model, satu berdasarkan Medan Acak Bersyarat (CRF) dan yang lainnya berdasarkan Model Markov Tersembunyi.

Gardner dan Xiong mengembangkan kerangka kerja *Health Information DE-Identification* (HIDE) untuk de-identifikasi informasi kesehatan pribadi (PHI), yang menggunakan alat NER berbasis CRF. Neamatullah dkk. mengembangkan paket MIT De-id. Paket ini merupakan sistem berbasis kamus dan aturan dan tersedia gratis di internet oleh PhysioNet. Uzuner dkk. mengembangkan Stat De-id yang menjalankan serangkaian pengklasifikasi secara paralel.

Setiap pengklasifikasi memiliki spesialisasi dalam mendeteksi kategori entitas yang berbeda. *Best-of-Breed System* (BoB) oleh Ferrández dkk., sebuah sistem desain hibrida, menggunakan aturan dan kamus untuk mendapatkan perolehan yang lebih tinggi, dan juga menggunakan pengklasifikasi berbasis model untuk mendapatkan presisi yang lebih tinggi.

Michaël Benesty menyoroti pentingnya kecepatan pemrosesan sistem anonimisasi. Studi kasus ini dilakukan bekerja sama dengan pemerintah Prancis dan Mahkamah Agung Prancis (*Cour de cassation*). Baru-baru ini, Glaser dkk. menyajikan pendekatan pembelajaran mesin untuk identifikasi otomatis elemen teks sensitif dalam putusan pengadilan Jerman. Strategi yang diadopsi mencakup beberapa jaringan saraf tiruan dalam (DNN) berdasarkan embedding kontekstual generik yang telah dilatih sebelumnya.

Metode anonimisasi yang paling umum meliputi: penekanan, penandaan, substitusi acak, dan generalisasi. Metode penekanan adalah cara sederhana untuk menganonimkan teks yang terdiri dari penekanan NE menggunakan indikator netral yang menggantikan teks asli, misalnya 'XXXXXX'. Metode penandaan terdiri dari penggantian NE dengan label yang dapat menunjukkan kelasnya dan pengenal unik. Metode ini dapat diimplementasikan dengan menggabungkan kelas yang diberikan oleh alat NER dan pengenal numerik unik, misalnya `[**Organization123**]`.

Metode substitusi acak menggantikan NE dengan entitas acak lain dari kelas dan fitur morfosintaksis yang sama. Metode ini dapat diimplementasikan menggunakan daftar default yang berisi entitas acak dari setiap kelas. Dalam bahasa yang sangat terinfleksi, penting untuk mengganti entitas dari setiap kelas dengan entitas lain dengan jenis kelamin dan nomor yang

sama. Generalisasi adalah metode penggantian entitas dari teks dengan entitas lain yang menyebutkan objek dari kelas yang sama tetapi dengan cara yang lebih umum, misalnya: Universitas Lisbon dapat digeneralisasi menjadi Universitas, atau bahkan menjadi Institusi.

Beberapa masalah utama dalam pengembangan sistem anonimisasi untuk dokumen hukum dan putusan pengadilan adalah: (1) kurangnya kumpulan data non-anonim, sehingga mustahil untuk membandingkan pendekatan dan mempersulit evolusi sistem ini; (2) setiap yurisdiksi memiliki distribusi tipe entitas bernama yang berbeda dan menimbulkan tantangan anonimisasi khusus pengadilan; (3) semua entitas yang merujuk pada objek yang sama dalam dokumen harus diganti dengan label yang sama, yang menyiratkan keberadaan modul resolusi ko-referensi yang juga merupakan tantangan besar; (4) Substitusi acak menyiratkan ekstraksi jenis kelamin gramatikal dan nomor NE yang diberikan oleh kata utamanya. Kata utama NE dan fitur-fiturnya harus ditentukan pada tahap pra-pemrosesan. Menentukan jenis kelamin kata utama penting untuk mengganti NE yang merujuk orang dengan NE lain dengan jenis kelamin yang sama, misalnya mengganti John dengan Peter, atau mengganti Mary dengan Anna.

Klasifikasi Dokumen

Klasifikasi teks (yaitu, pengelompokan dokumen berdasarkan taksonomi yang telah ditentukan sebelumnya) memiliki banyak potensi aplikasi dalam ranah hukum, khususnya untuk mengkategorikan dokumen dan kasus legislatif. Hal ini dapat membantu proses penelitian hukum dan pengembangan sistem manajemen pengetahuan. Beberapa studi telah berfokus pada konten legislatif atau kasus pengadilan, dengan beberapa penulis menyoroti bahwa klasifikasi dokumen hukum dapat jauh lebih sulit daripada masalah klasifikasi teks yang lebih umum.

Khusus untuk konten legislatif, banyak penelitian tentang klasifikasi topik berfokus pada dokumen legislasi Uni Eropa, baik dalam konteks monolingual yang berfokus pada bahasa Inggris, maupun dalam konteks multilingual. Upaya-upaya sebelumnya ini membahas tugas mengklasifikasikan hukum Uni Eropa ke dalam konsep EuroVoc, melihat masalah ini sebagai contoh yang menantang dari Klasifikasi Teks Multilabel Skala Besar (LMTC), mengingat kebutuhan untuk menetapkan, untuk setiap dokumen yang diberikan, subset label dari set besar yang telah ditentukan sebelumnya (yaitu, ribuan kelas yang terorganisir secara hierarkis), dan juga mengingat kebutuhan untuk menangani skenario sedikit dan tanpa percobaan (yaitu, distribusi penetapan label sangat miring, dan beberapa label memiliki sedikit atau tidak ada contoh pelatihan).

Serangkaian metode LMTC mutakhir telah dievaluasi secara empiris, dengan hasil yang sangat baik saat ini diperoleh dengan pendekatan yang didasarkan pada penggabungan model bahasa neural besar yang telah dilatih sebelumnya (yaitu, model yang didasarkan pada pendekatan berbasis Transformer yang telah dilatih sebelumnya seperti BERT) dengan jaringan atensi label-wise (yaitu, menggunakan parameter yang berbeda untuk memberi bobot pada representasi dokumen, sesuai dengan setiap label yang memungkinkan). Misalnya, dalam eksperimen dengan 57 ribu dokumen legislatif Inggris dari EURLEX, studi telah melaporkan nilai 80.3 dalam hal metrik evaluasi R-Precision@K.

Teknologi klasifikasi dokumen saat ini juga diterapkan dalam banyak pengaturan praktis. Salah satu contoh menarik adalah JRC EuroVoc Indexer (JEX), yaitu alat sumber terbuka yang saat ini digunakan di berbagai pengaturan, yang dikembangkan oleh Pusat Penelitian Gabungan (JRC) Komisi Eropa untuk mengklasifikasikan dokumen secara otomatis menurut deskriptor EuroVoc, yang mencakup 22 bahasa resmi Uni Eropa. JEX dapat digunakan sebagai alat untuk penugasan deskriptor EuroVoc multi-label interaktif, yang khususnya berguna untuk meningkatkan kecepatan dan konsistensi proses kategorisasi manusia, atau dapat digunakan sepenuhnya otomatis.

Pengambilan Informasi

Kebutuhan untuk menangani dokumen digital dalam jumlah besar telah membuat sektor hukum tertarik untuk mengembangkan metodologi khusus dalam pengelolaan, penyimpanan, pengindeksan, dan pengambilan informasi hukum. Semua tugas ini termasuk dalam ranah pengambilan informasi, yang berfokus pada masalah pencarian informasi di mana deskripsi situasi terkini (yaitu, kebutuhan informasi) digunakan untuk meminta sistem otomatis mengambil informasi yang paling sesuai, dalam repositori besar, untuk permintaan masukan.

Penelitian tentang pengambilan informasi hukum telah dilakukan sejak tahun 1960an, tetapi perkembangan ilmiah terkini berkaitan erat dengan Kompetisi Ekstraksi/Entailmen Informasi Hukum (COLIEE), dan aplikasi spesifik terkait pengambilan yurisprudensi. Dari tahun 2015 hingga 2017, tugas COLIEE adalah mengambil artikel-artikel KUH Perdata Jepang berdasarkan sebuah pertanyaan, dan sejak saat itu, tugas utama pengambilan COLIEE adalah mengambil kasus-kasus pendukung berdasarkan deskripsi singkat dari kasus yang belum pernah dilihat sebelumnya. Sebagian besar sistem yang diajukan memanfaatkan representasi dokumen dan kueri yang jarang, berdasarkan kemunculan kata, bersama dengan statistik numerik sederhana yang mencerminkan seberapa penting sebuah kata bagi sebuah dokumen dalam suatu koleksi (misalnya, statistik seperti TF-IDF atau BM25).

Studi yang lebih baru, baik dalam konteks COLIEE maupun dalam publikasi terpisah lainnya, telah mulai mengeksplorasi kemajuan terbaru yang terkait dengan model pemeringkatan neural (misalnya, menggunakan model bahasa besar yang dilatih dalam data pencocokan teks, untuk merangking ulang hasil metode yang lebih sederhana berdasarkan statistik tingkat kata). Sebuah studi menarik baru-baru ini, misalnya, berfokus pada tugas pengambilan informasi regulasi, yang berkaitan dengan pengambilan semua undang-undang relevan yang harus dipatuhi oleh suatu organisasi, atau sebaliknya (misalnya, jika ada undang-undang baru, ambil semua kontrol kepatuhan regulasi, dalam suatu organisasi, yang terpengaruh oleh undang-undang ini).

Aplikasi seperti ini jauh lebih menantang daripada tugas pengambilan informasi tradisional, di mana kueri biasanya berisi beberapa kata informatif dan dokumennya relatif pendek. Dalam kasus pengambilan informasi regulasi (dan juga tugas hukum lainnya, seperti pencocokan kasus serupa, kuerinya juga berupa dokumen panjang (misalnya, sebuah peraturan) yang berisi ribuan kata, yang sebagian besar tidak informatif. Akibatnya, mencocokkan kueri dengan dokumen panjang lainnya, di mana kata-kata informatifnya juga

jarang, menjadi sangat sulit untuk pendekatan tradisional berdasarkan pencocokan tingkat kata.

Dengan memanfaatkan kumpulan data yang terdiri dari arahan Uni Eropa dan peraturan Inggris, yang dapat berfungsi baik sebagai kueri maupun dokumen (misalnya, undang-undang Inggris relevan dengan arahan Uni Eropa yang ditransposisikannya, dan sebaliknya), penulis melaporkan hasil yang sangat baik (misalnya, dengan merata-ratakan berbagai kueri, sekitar 86,5% dokumen yang diambil pada 100 posisi teratas adalah relevan) dengan sistem yang menggabungkan pengambilan BM25 standar dengan pemeringkatan ulang hasil melalui model bahasa saraf yang disesuaikan dengan dokumen dan tugas tertentu dari domain hukum.

Ekstraksi Informasi

Ekstraksi informasi berkaitan dengan pengumpulan dan penataan fakta-fakta penting secara otomatis dari dokumen tekstual (misalnya, tentang jenis peristiwa tertentu, atau tentang entitas dan hubungan antar entitas yang telah didefinisikan sebelumnya), yang memfasilitasi pengembangan aplikasi tingkat tinggi. Hal ini berarti bahwa aplikasi tersebut kini dapat berfokus pada analisis informasi terstruktur, alih-alih teks hukum tradisional yang tidak terstruktur atau semi-terstruktur.

Beberapa studi telah mengeksplorasi ekstraksi informasi dari kontrak, misalnya untuk mengekstrak elemen informasi seperti para pihak dalam kontrak, jumlah pembayaran yang disepakati, tanggal mulai dan berakhir, atau hukum yang berlaku. Studi lain berfokus pada ekstraksi informasi dari undang-undang atau kasus pengadilan.

Ringkasan

Ringkasan teks menyampaikan kepada pembaca konten yang paling relevan dari satu atau lebih sumber informasi tekstual, secara ringkas dan mudah dipahami. Tujuan sistem peringkasan teks adalah untuk membuat dokumen semacam itu secara otomatis. Dokumen baru ini, ringkasan, dicirikan oleh beberapa aspek, seperti asal konten, jumlah unit masukan, atau cakupan ringkasan.

Mengenai kontennya, ringkasan dapat disusun dari ekstrak peringkasan ekstraktif, yang diambil langsung dari masukan, atau parafrase peringkasan abstrak, yang menyampaikan konten suatu bagian masukan dengan kata-kata yang berbeda. Terkait jumlah unit masukan, jika masukan hanya terdiri dari satu dokumen, tugas ini disebut peringkasan dokumen tunggal; ketika menangani beberapa dokumen masukan, tugas ini disebut peringkasan multi-dokumen. Terakhir, terkait cakupan sumber masukan, ringkasan dapat bersifat komprehensif, ketika membuat ringkasan generik, atau selektif, jika didorong oleh kueri masukan.

Beberapa kesulitan muncul ketika menangani tugas ini, tetapi salah satu yang terpenting adalah bagaimana menilai konten yang relevan. Berbagai metode telah dieksplorasi sejak percobaan pertama yang dilaporkan oleh Luhn dan Edmundson. Mereka menetapkan arah penelitian penting, yang dapat disebut sebagai penilaian bagian berbasis fitur, yang mengeksplorasi fitur berdasarkan pembobotan istilah, posisi kalimat, panjang kalimat, atau informasi linguistik. Pada tahun 2000an, metode berbasis sentralitas, terlepas dari representasi yang mendasarinya, menarik banyak perhatian. Sentroid geometris, metode

pemeringkatan berbasis grafik, atau, secara umum, representasi di mana bagian yang "direkomendasikan" menjadi fokus dari bidang penelitian ini. Semua ini merupakan pendekatan tanpa pengawasan, yang juga mencakup metode-metode penting seperti Maximal Marginal Relevance atau metode berbasis Latent Semantic Analysis.

Selain pendekatan-pendekatan tersebut, beberapa metode terbimbing juga telah dieksplorasi. Akhir-akhir ini, sebagian besar penelitian tentang topik ini berbasis jaringan saraf tiruan, dengan model sekuens-ke-sekuens yang menarik banyak perhatian, serta penelitian yang menggunakan model bahasa pra-latih sebagai encoder. Dalam hal peringkasan dokumen hukum, beberapa karya tertua kembali ke tahun 2004. Farzindar dan Lapalme menyajikan pendekatan untuk menghasilkan ringkasan singkat dari catatan proses pengadilan federal di Kanada.

Pekerjaan ini berfokus pada mengekstraksi unit tekstual yang paling penting, mengikuti pendekatan penilaian bagian berbasis fitur. Alur kerja mencakup segmentasi tematik yang secara khusus diarahkan pada dokumen hukum, yang bertujuan untuk menemukan struktur catatan putusan. Segmentasi ini diikuti oleh tahap penyaringan dan pemilihan. Yang pertama menyaring kutipan dan konten berisik lainnya dan yang terakhir mengekstraksi unit tekstual berdasarkan skor mereka, dihitung menggunakan fitur-fitur seperti posisi paragraf dalam dokumen, posisi paragraf dalam segmen tematik, posisi kalimat dalam paragraf, distribusi kata-kata dalam dokumen, dan TF-IDF.

Terkait erat dengan pekerjaan ini, adalah sistem yang diusulkan oleh Hachey dan Grover, dalam arti bahwa penulis ini juga mengembangkan pengklasifikasi untuk informasi retoris. Ekstraksi kalimat-kalimat yang relevan juga dianggap sebagai masalah klasifikasi, mengikuti pendekatan berbasis pembelajaran mesin terawasi. Fitur-fitur seperti lokasi, kata tematik, panjang kalimat, tingkat kutipan, entitas, dan fase isyarat dieksplorasi baik dalam pengklasifikasi Naïve Bayes maupun pengklasifikasi Entropi Maksimum.

Penelitian yang lebih baru dalam peringkasan dokumen hukum mengikuti tren yang sama yang kita amati dalam peringkasan teks otomatis generik dan pemrosesan bahasa alami secara umum, yang berarti bahwa hal itu didasarkan pada jaringan saraf dan model bahasa yang telah dilatih sebelumnya. Penelitian yang dilaporkan oleh Glaser dkk. berfokus pada putusan pengadilan Jerman. Sistem ini mencakup langkah pra-pemrosesan khusus di mana norma, token anonimisasi, dan referensi ke dokumen hukum lainnya, misalnya, ditangani. Mengenai metode peringkasan, penulis mengeksplorasi pendekatan ekstraktif dan abstraktif.

Kata-kata dikodekan menggunakan penyematan GloVe dan untuk representasi kalimat, tiga pendekatan dieksplorasi: CNN, GRU, atau atensi. Representasi kalimat akhir diberikan oleh CNN atau RNN lintas kalimat yang menangkap informasi dari kalimat-kalimat tetangga. Skor seleksi diberikan oleh fungsi sigmoid. Pendekatan abstraktif serupa, tetapi alih-alih menggunakan CNN atau RNN lintas kalimat, penyisipan kalimat diagregasi menggunakan RNN, sehingga menghasilkan penyisipan untuk dokumen. Pendekatan ini mengikuti struktur encoder-decoder.

Seperti yang diperkirakan, pendekatan abstraktif memiliki kinerja terburuk jika dibandingkan dengan pendekatan ekstraktif dan baseline, bahkan dengan

mempertimbangkan pendekatan lama berbasis sentralitas seperti LexRank. Pendekatan ekstraktif yang diusulkan mencapai hasil terbaik yang menunjukkan bahwa pendekatan berbasis jaringan saraf tiruan juga memadai untuk domain tertentu.

Sebuah karya menarik berdasarkan model bahasa pra-latihan adalah sistem yang diusulkan oleh Savelka dan Ashley. Karya mereka berkaitan erat dengan peringkasan, tetapi mereka berfokus pada gagasan tentang seberapa baik sebuah kalimat menjelaskan suatu konsep hukum, berdasarkan data dari proyek akses Caselaw (kasus hukum AS dari berbagai jenis pengadilan). Data yang dipilih diklasifikasikan secara manusiawi ke dalam empat kategori: bernilai tinggi, bernilai pasti, bernilai potensial, dan tidak bernilai.

Mereka menyempurnakan RoBERTa, sebuah model yang diturunkan dari BERT yang terkenal, sebagai dasar eksperimen mereka. Mereka mengeksplorasi tiga pendekatan: pendekatan pertama yang memprediksi kelas hanya menggunakan kalimat masukan; pendekatan kedua menggunakan pasangan konsep hukum-kalimat; dan, terakhir, variasi terakhir didasarkan pada pasangan yang terdiri dari "keseluruhan ketentuan hukum tertulis" dan kalimat.

Satu kesimpulan penting adalah, mengingat keberhasilan eksperimen pertama, kalimat-kalimat tersebut membawa informasi tentang kegunaannya, yang sangat berkaitan dengan pemahaman apakah suatu kalimat merupakan kandidat yang baik untuk dimasukkan dalam ringkasan.

Sistem Tanya Jawab dan Percakapan

Sistem tanya jawab hukum berkaitan dengan pengambilan dan analisis informasi dalam repositori pengetahuan (misalnya, koleksi dokumen besar), untuk memberikan jawaban yang akurat atas pertanyaan hukum. Pengguna umum sistem tanya jawab hukum dapat mencakup litigator yang mencari jawaban atas pertanyaan hukum spesifik kasus, atau orang awam yang ingin lebih memahami hak-hak hukum mereka.

Tugas ini memerlukan identifikasi undang-undang, yurisprudensi, atau dokumen hukum lain yang relevan, dan mengekstrak elemen-elemen di dalam dokumen tersebut yang menjawab pertanyaan tertentu. Dalam beberapa kasus, elemen informasi yang diekstraksi juga perlu diringkas lebih lanjut menjadi jawaban yang ringkas. Sebagaimana terlihat dari definisi masalah sebelumnya, sistem tanya jawab hukum biasanya melibatkan penggabungan teknik-teknik dari pengambilan dan ekstraksi informasi, dan Kompetisi Ekstraksi dan Entailmen Informasi Hukum yang telah disebutkan sebelumnya juga telah menjadi wadah penting untuk melaporkan kemajuan dalam bidang ini.

Terkait aplikasi industri, menarik untuk dicatat bahwa perusahaan seperti IBM Watson Legal, LegalMation, atau Ross Intelligence telah mengembangkan produk komersial tanya jawab khusus berbasis Watson, yaitu sistem tanya jawab yang dikembangkan oleh IBM yang, pada tahun 2011, memenangkan tantangan Jeopardy melawan dua juara kuis TV terbesar sepanjang masa.

Arsitektur Watson menampilkan beragam teknologi NLP, termasuk penguraian, klasifikasi pertanyaan, dekomposisi pertanyaan, akuisisi dan evaluasi sumber otomatis, deteksi entitas dan relasi, pembangkitan bentuk logis, serta representasi dan penalaran pengetahuan.

Setelah kesuksesan Jeopardy, sistem ini direorganisasi dan dikomersialkan, dengan menggabungkan dan menyesuaikan modul-modul spesifiknya dengan domain dan tugas tertentu (yaitu, tidak hanya menjawab pertanyaan, tetapi juga tugas-tugas lainnya).

Prediksi yang Didukung Bukti Tekstual

Gagasan bahwa komputer dapat memprediksi hasil kasus hukum berawal dari awal tahun 60-an, ketika Lawlor menganggap "analisis dan prediksi putusan pengadilan" sebagai salah satu tugas terpenting yang dapat dibantu oleh teknologi komputer. Namun, sebagian besar penelitian tentang tugas ini belum terlalu lama. Bahkan, kita dapat memahami kesulitannya karena tinjauan pustaka terbaru yang berfokus pada pemahaman bahasa alami dalam ranah hukum membahas topik-topik seperti konstruksi sumber daya atau tugas ekstraksi informasi sederhana.

Pada tahun 2004, Ruger dkk. membandingkan kinerja pohon klasifikasi dengan sekelompok pakar hukum untuk memprediksi hasil Sidang Mahkamah Agung Amerika Serikat tahun 2002. Fitur-fitur yang digunakan adalah sebagai berikut: asal usul kasus hingga mencapai Mahkamah Agung, area permasalahan kasus, jenis pemohon, jenis termohon, orientasi politik putusan pengadilan yang lebih rendah (liberal atau konservatif), dan klaim inkonstitusionalitas suatu undang-undang atau praktik oleh pemohon.

Hasilnya mengejutkan: metode otomatis dengan tepat memprediksi 75% hasil persetujuan/pembatalan Mahkamah Agung, sementara akurasi para ahli hanya 59,1%. Hasil terbaik dicapai untuk kasus-kasus kegiatan ekonomi dan yang terburuk untuk federalisme, di mana akurasi untuk para ahli dan pendekatan otomatis serupa.

Pendekatan dua langkah yang berbeda diikuti oleh Brüninghaus dan Ashley, dengan fokus pada informasi tekstual. Langkah pertama mengekstrak representasi kasus hukum, berdasarkan kata-katanya (representasi bag-of-words yang diperoleh dengan menghilangkan tanda baca, angka, dan kata henti); pengenalan entitas bernama (di mana nama dan contoh kasus spesifik diganti dengan jenisnya); dan hubungan sintaksis.

Representasi ini diserahkan ke serangkaian pengklasifikasi yang menangkap beberapa aspek yang digunakan untuk merepresentasikan kasus-kasus tersebut, yang ditetapkan sebagai Faktor (misalnya, Disetujui-Tidak-Untuk-Mengungkapkan, Langkah-Langkah Keamanan, atau Perjanjian-Tidak-Spesifik). Data yang digunakan terdiri dari basis pengetahuan Hukum Rahasia Dagang, yang mencakup 146 kasus dari sistem CATO, sebuah lingkungan belajar cerdas bagi mahasiswa baru yang mulai belajar hukum, yang sudah direpresentasikan dalam bentuk Faktor. Langkah kedua memprediksi hasil hukum berdasarkan representasi Faktor, menggunakan penalaran berbasis kasus.

Berfokus pada Mahkamah Hak Asasi Manusia Eropa dan hanya mengeksplorasi konten tekstual, Aletras dkk. bereksperimen dengan pengklasifikasi Support Vector Machines (kernel linear) berdasarkan representasi n-gram dan berbasis topik. Tujuannya adalah untuk memprediksi apakah suatu kasus tertentu melanggar pasal tertentu dari Konvensi, mencapai tingkat akurasi mendekati 80% dalam masalah klasifikasi biner ini. Sulea dkk. juga menggunakan pengklasifikasi Support Vector Machines (kernel linear), tetapi berfokus pada putusan Mahkamah Agung Prancis (*Cour de Cassation*).

Para penulis menangani tiga tugas: memprediksi bidang hukum suatu kasus, memprediksi putusan pengadilan, dan memperkirakan kapan deskripsi kasus dan putusan dikeluarkan. Pra-pemrosesan meliputi penghapusan diakritik dan tanda baca, serta mengubah semua kata menjadi huruf kecil. Fitur terdiri dari unigram dan bigram. Dalam hal memprediksi putusan pengadilan, dua variasi tugas dengan mempertimbangkan enam kelas (putusan kata pertama) dan delapan kelas (putusan lengkap) ditangani. Hasilnya menjanjikan karena dalam kedua percobaan, pendekatan yang diusulkan mencapai skor f1 dan akurasi lebih dari 90%.

Sebagaimana disebutkan sebelumnya, sebagian besar penelitian tentang topik ini masih baru dan dengan demikian mengeksplorasi model jaringan saraf tiruan dalam. Chalkidis dkk. mengeksplorasi arsitektur jaringan yang berbeda Gated Recurrent Units dwiarah dengan arsitektur berbasis self-attention, jaringan atensi hierarkis, jaringan atensi label-biwise, dan dua arsitektur berbasis BERT (versi reguler dan hierarkis) untuk memprediksi pelanggaran pasal-pasal Mahkamah Hak Asasi Manusia Eropa. Berbeda dengan Aletras dkk., mereka tidak membatasi diri pada pasal-pasal tertentu, melainkan mengeksplorasi tampilan klasifikasi multi-label dari tugas tersebut.

Pendekatan berbasis BERT yang sederhana merupakan arsitektur dengan kinerja terburuk, dengan hasil terbaik dicapai oleh arsitektur hierarkis. Mengonfirmasi pentingnya pendekatan hierarkis, Zhu dkk. juga mengeksplorasi arsitektur berbasis jaringan atensi hierarkis dalam prediksi putusan hukum dalam konteks kasus pidana yang diterbitkan oleh pemerintah Tiongkok dari China Judgment Online. Alghazzawi dkk. menggabungkan memori jangka pendek dengan jaringan saraf konvolusional untuk mengatasi masalah yang sama seperti Ruger dkk., yaitu untuk memprediksi hasil positif/negatif/berbeda dari putusan Mahkamah Agung AS.

Terakhir, Medvedeva dkk. memberikan tinjauan menarik tentang topik ini, sekaligus membahas klarifikasi konsep-konsep yang relevan. Para penulis berpendapat bahwa terminologi yang jelas dan terdefinisi dengan baik penting untuk kemajuan penelitian dalam topik ini, yaitu membedakan tiga tugas yang berbeda: identifikasi luaran, kategorisasi penilaian berbasis luaran, dan peramalan luaran.

Ringkasan

Bagian ini menjelaskan status perkembangan teknologi bahasa yang menargetkan berbagai tugas pemrosesan teks hukum. Meskipun survei kami sebagian besar berfokus pada perkembangan akademis terkini, sejumlah besar perusahaan, termasuk ratusan perusahaan rintisan, juga saat ini beroperasi di industri "AI Hukum" yang sedang berkembang, menyediakan layanan analisis teks yang menargetkan beragam kasus penggunaan yang saat ini kurang ditangani, misalnya karena jumlah data yang berlebihan (yaitu, dokumen) yang menimbulkan tantangan bagi analisis manusia.

Dalam semua tugas yang disurvei, perkembangan terkini terkait metode pembelajaran mendalam (misalnya, model bahasa saraf pra-latih yang secara khusus menargetkan teks hukum) telah menghasilkan peningkatan yang signifikan. Tantangan terkini di bidang ini berkaitan, misalnya, dengan kombinasi jaringan saraf dalam dengan metode berbasis pengetahuan (misalnya, untuk meningkatkan interpretabilitas dan untuk memperhitungkan

pengetahuan ahli dan penalaran hukum dengan lebih baik), atau dengan teknik yang memungkinkan pengendalian potensi bias yang lebih baik dalam hasil model (misalnya, bias gender atau diskriminasi rasial).

2.3 TEKNOLOGI BAHASA LISAN

Penggunaan teknologi bahasa lisan dalam ranah hukum juga semakin meluas, terutama karena peningkatan signifikan dalam kinerja yang dicapai oleh teknik pembelajaran mendalam. Hal ini membenarkan tinjauan atas kemajuan terbaru ini dan dampaknya dalam ranah hukum.

Pengenalan Ucapan Otomatis

Kondisi terkini dalam pengenalan ucapan otomatis (ASR) sebelum munculnya pembelajaran mendalam sebagian besar didasarkan pada paradigma GMM-HMM (Model Campuran Gaussian Model Markov Tersembunyi). Dengan memasukkan fitur-fitur yang bermakna secara perseptual ke dalam model akustik ini, dan menggabungkannya dengan sumber pengetahuan tambahan yang disediakan oleh model bahasa n-gram dan model leksikal (atau pelafalan), seseorang mencapai tingkat kesalahan kata (WER) yang membuat sistem ASR dapat digunakan untuk tugas-tugas tertentu.

Dikte adalah salah satu tugas ini, terutama untuk ranah hukum dan layanan kesehatan (misalnya laporan radiologi), yang dicirikan oleh kondisi perekaman yang bersih dan dokumen yang relatif formal. Model akustik dapat diadaptasi untuk pembicara, dan model leksikal dan bahasa dapat diadaptasi untuk domain, memungkinkan penggunaan ASR oleh pengacara untuk mendiktekan dokumen, catatan kasus, ringkasan, kontrak dan korespondensi. Namun, penerimaan aplikasi tersebut tidak signifikan dan sangat bergantung pada ketersediaan sumber daya untuk melatih model untuk bahasa/aksen yang berbeda. Selama hampir tiga dekade, kemajuan relatif basi untuk ASR, sampai munculnya apa yang disebut 'paradigma hibrida', yang memasangkan jaringan saraf dalam dengan HMM. Saat ini, model yang dilatih dengan hampir 1000 jam buku audio yang dibaca mencapai WER sebesar 3,8%, hasil yang tidak terpikirkan satu dekade lalu.

Seperti di banyak domain AI lainnya, arsitektur ujung ke ujung sepenuhnya juga telah diusulkan untuk melakukan seluruh jalur ASR, dengan pengecualian ekstraksi fitur, tetapi kinerjanya secara signifikan lebih buruk ketika data pelatihan pendek. Faktanya, "Tidak ada data yang lebih baik daripada data lainnya" adalah kutipan dari peneliti ASR di tahun 80-an, yang masih berlaku hingga saat ini, yang merupakan tantangan besar saat melakukan porting ke domain lain.

Transkripsi buku audio atau mendiktekan dokumen hukum merupakan tugas yang relatif mudah bagi sistem ASR. Penerapannya pada percakapan jauh lebih menantang, dengan tingkat kesalahan hampir tiga kali lipat dari hasil di atas. Kehadiran faktor-faktor lain seperti aksen non-penutur asli, atau mikrofon yang jauh di ruang rapat, juga dapat berdampak sangat negatif.

Semua tantangan ini memotivasi penggunaan beragam pendekatan pembelajaran mesin: augmentasi audio, pembelajaran transfer, pembelajaran multitugas, dll. Pendekatan

tanpa pengawasan terbaru yang memanfaatkan representasi ucapan untuk melakukan segmentasi audio tanpa label dan mempelajari pemetaan dari representasi tersebut ke fonem melalui pelatihan adversarial juga patut disebutkan.

Peningkatan kinerja ASR yang sangat besar ini berdampak besar dalam ranah hukum. Untuk dikte di ranah ini, di mana data pelatihan semakin terkumpul, beberapa perusahaan kini mengklaim WER di bawah 1%, mengumumkan cara yang lebih efisien untuk mendiktekan dokumen, 3–5 kali lebih cepat daripada mengetik, sehingga secara keseluruhan mengurangi kewajiban dan kepatuhan. Kemajuan ini khususnya relevan selama pandemi baru-baru ini yang telah memaksa lebih banyak pengacara untuk bekerja jarak jauh, tanpa akses mudah ke kumpulan juru ketik.

Namun, penggunaan ASR dalam ranah hukum tidak terbatas pada tugas dikte saja. Transkripsi bukti video dan audio menjadi transkrip hukum merupakan tugas yang semakin penting untuk presentasi di pengadilan, pengajuan banding, dan sebagainya. Proses manual dapat memakan waktu lebih dari 5 kali waktu nyata, yang menjadi motivasi signifikan untuk mempercepatnya dengan mengoreksi transkrip otomatis, alih-alih mentranskripsi dari awal. Transkripsi proses pengadilan audio juga merupakan penggunaan ASR yang semakin umum dalam ranah hukum. Menemukan kata kunci dalam percakapan yang disadap mungkin juga relevan bagi badan intelijen.

Selain tantangan yang telah disebutkan sebelumnya dalam mengenali ucapan spontan, semua jenis transkrip ini memerlukan tugas diarisasi pembicara sebelumnya mengenali siapa yang berbicara dan kapan. Tugas ini berkaitan erat dengan pengenalan pembicara otomatis dan mungkin sangat rumit dalam skenario di mana tumpang tindih pembicara sering terjadi.

Karena kompleksitasnya yang tinggi, sistem ASR biasanya berjalan di cloud. Pada asisten suara pribadi, tugas mendeteksi kata kunci bangun saat berada dalam "mode selalu mendengarkan" dilakukan di perangkat menggunakan pendekatan yang jauh lebih sederhana. Kesadaran publik akan kekhawatiran privasi atas "mode selalu mendengarkan" ini semakin meningkat. Permintaan rekaman ini oleh tersangka yang ingin menunjukkannya sebagai bukti di pengadilan sejauh ini telah ditolak oleh perusahaan seperti Amazon.

Pengenalan Pembicara dan Profil Pembicara

Pentingnya mengidentifikasi pembicara dalam rekaman telah disadari oleh lembaga penegak hukum dan badan intelijen yang dulu mengandalkan para ahli untuk menganalisis secara manual apa yang disebut 'sidik jari suara', jauh sebelum sistem pengenalan pembicara otomatis mencapai tingkat kinerja yang memungkinkan penggunaannya dalam domain tersebut. Sebagian besar kemajuan terbaru ini dapat dikaitkan dengan pembelajaran representasi, yang disebut 'penyisipan pembicara', yang mengodekan karakteristik pembicara dari suatu ucapan dengan durasi variabel ke dalam vektor dengan panjang tetap.

Teknik paling populer untuk mencapai representasi ringkas ini saat ini adalah pendekatan vektor-x. Embedding ini diekstraksi dari lapisan tersembunyi jaringan saraf dalam (*deep neural network*), ketika dilatih untuk membedakan lebih dari ribuan penutur. Faktanya, pendekatan ini telah diterapkan pada Voxceleb, sebuah korpus multimoda klip YouTube yang mencakup lebih dari 7000 penutur dari berbagai etnis, aksen, pekerjaan, dan kelompok usia,

mencapai tingkat kesalahan yang sama (EER) yang mengesankan, mendekati 3%. Metrik ini mendapatkan namanya dari kesesuaiannya dengan ambang batas di mana tingkat kesalahan positif palsu dan negatif palsu sama. Ukuran kepercayaan mungkin sangat penting dalam konteks hukum.

Bidang pembuatan profil penutur adalah salah satu yang terbaru dalam pemrosesan ucapan. Ucapan adalah sinyal biometrik yang mengungkapkan, selain arti kata, banyak informasi tentang pengguna, termasuk preferensi, ciri kepribadian, suasana hati, kesehatan, dan opini politiknya, di antara data lain seperti jenis kelamin, rentang usia, tinggi badan, aksen, dll. Selain itu, sinyal input juga dapat digunakan untuk mengekstrak informasi relevan tentang lingkungan pengguna, yaitu suara latar belakang. Pengklasifikasi pembelajaran mesin yang canggih dapat dilatih untuk secara otomatis mendeteksi ciri-ciri penutur yang mungkin sangat penting bagi lembaga penegak hukum dan badan intelijen. Kemungkinan tak terbatas untuk memprofil penutur berdasarkan suaranya ini menimbulkan banyak kekhawatiran privasi tentang penyalahgunaan teknologi tersebut.

Sintesis Ucapan dan Konversi Suara

Satu dekade yang lalu, perkembangan terkini dalam *text-to-speech* (TTS) didominasi oleh teknik konkatenatif yang memilih segmen-segmen terbaik untuk digabungkan dari korpus kalimat yang sangat besar yang dibaca oleh seorang pembicara. Modul sintesis konkatenatif biasanya didahului oleh rangkaian modul pemrosesan linguistik yang kompleks yang menerima teks sebagai masukan dan menghasilkan serangkaian fonem beserta informasi prosodi yang menentukan intonasi turunan. Meskipun terdapat peningkatan yang signifikan, yaitu melalui penggunaan pendekatan hibrida yang menggabungkan teknik parametrik statistik dan konkatenatif, kualitas ucapan sintetis masih sangat berbeda dalam hal kealamian dibandingkan ucapan manusia, ekspresinya sangat terbatas, dan biaya untuk membangun suara sintetis baru seringkali mahal.

Sebuah terobosan besar dicapai pada pertengahan 2010an dengan mengganti modul sintesis konkatenatif tradisional dengan modul jaringan saraf tiruan dalam yang menerima representasi spektrogram waktu-frekuensi sebagai masukan. Kemudian, seluruh paradigma berubah menjadi arsitektur encoder-decoder, dengan mekanisme atensi yang memetakan skala waktu linguistik ke skala waktu akustik.

Kemajuan ini menghasilkan sistem TTS multi-speaker yang memanfaatkan penempatan pembicara, dan membuka kemungkinan membangun suara sintetis hanya dengan beberapa detik suara baru, misalnya menggunakan model berbasis alur. Kualitas suara sintetis menjadi sangat mendekati suara manusia, mencapai nilai di atas 4 pada skala 1–5, yang disebut skala *'Mean Opinion Score'* (MOS).

Kemungkinan untuk memisahkan konten linguistik dan penempatan pembicara memang krusial, tidak hanya untuk sistem teks-ke-suara tetapi juga untuk sistem konversi suara (VC). Dalam VC, masukannya adalah ucapan, bukan teks, dan tujuannya mungkin mengubah identitas suara, emosi, aksen, dll.

Penguraian ini dapat dicapai dengan menggunakan, misalnya, skema auto-encoder variasional, di mana encoder konten linguistik mempelajari kode laten dari ucapan pembicara

sumber, dan encoder pembicara mempelajari penyisipan pembicara dari ucapan pembicara target. Saat dijalankan, kode laten dan penyisipan pembicara digabungkan untuk menghasilkan ucapan dalam suara pembicara target. Saat ini, pendekatan baru mencoba mempertimbangkan penyisipan prosodi atau penyisipan gaya juga. Lebih lanjut, suara buatan mungkin tidak sesuai dengan pembicara target, tetapi, misalnya, dengan rata-rata dari sekumpulan penyisipan pembicara yang dipilih di antara yang terjauh dari penyisipan pembicara asli. Ini sebenarnya salah satu dari banyak pendekatan yang diusulkan untuk anonimisasi pembicara.

Dampak dari kemajuan TTS/VC ini dalam kerangka hukum berpotensi besar karena dapat digunakan untuk tujuan inkriminasi, pencemaran nama baik, atau misinformasi. Mendeteksi suara palsu yang mendalam dari suara asli akan semakin sulit, dan orang mungkin bertanya-tanya kapan bukti audio tidak lagi dapat diterima di pengadilan. Di saat yang sama, peniruan identitas pembicara dapat menyebabkan kejahatan yang tidak mungkin dilakukan dengan teknologi yang kita miliki satu dekade lalu. Bahkan, pencurian besar telah dilaporkan.

Di sisi lain, suara sintetis dapat digunakan untuk menyerang (atau memalsukan) sistem verifikasi pembicara otomatis. Faktanya, seiring meningkatnya kualitas TTS/VC dengan sedikit materi lisan dari pembicara target, kebutuhan akan anti-spoofing yang lebih canggih juga meningkat. Terakhir, perlu disebutkan juga kemungkinan adanya perintah suara tersembunyi yang disuntikkan ke dalam sinyal masukan. Ancaman ini selalu ada, yaitu dengan memanfaatkan fakta bahwa pendengaran manusia tidak dapat mendeteksi sinyal tertentu. Namun, saat ini serangan adversarial meningkatkan ancaman ini ke tingkat yang baru, menyadarkan kita betapa rentannya teknik pembelajaran mendalam terhadap gangguan masukan pengklasifikasi pada waktu pengujian, sehingga pengklasifikasi menghasilkan prediksi yang salah.

Di masa lalu, gangguan semacam itu umumnya dapat dirasakan, tetapi dengan serangan adversarial, kini gangguan yang sangat tak terlihat dapat dihasilkan, yang sangat efektif untuk menyesatkan pembicara maupun sistem pengenalan suara. Teknik-teknik semacam itu hanyalah salah satu contoh dari kemungkinan tak terbatas penyalahgunaan teknologi suara berbasis AI. Kita baru menyentuh permukaannya, dalam hal serangan yang mungkin menargetkan aplikasi berbasis suara.

2.4 TANTANGAN MASA DEPAN TEKNOLOGI BAHASA HUKUM AI

Bab ini mencoba memberikan gambaran singkat tentang teknologi bahasa untuk ranah hukum, yang berisiko segera menjadi usang, mengingat kemajuan pesat di bidang ini saat ini. Gambaran singkat tentang kemajuan terbaru ini mungkin menyesatkan, memberikan kesan bahwa metode berbasis embedding akan menyelesaikan semua jenis masalah, baik untuk pemrosesan bahasa tulis maupun lisan, dalam ranah hukum, asalkan terdapat set data pelatihan yang cukup besar.

Faktanya, banyak peneliti sedang mengembangkan alternatif pembelajaran mesin (misalnya pembelajaran zero-shot atau few-shot) untuk menangani tugas-tugas yang tidak memiliki set data tersebut. Namun, menggabungkan pendekatan berbasis embedding dengan

metode berbasis simbol tetap menjadi tantangan yang dapat berkontribusi signifikan pada interpretabilitas yang lebih baik.

Tantangan lain yang harus diatasi oleh perangkat hukum AI generasi mendatang adalah melacak perubahan regulasi dengan menyebarkan konsekuensi perubahan ini pada isu-isu yang bergantung padanya, karena hal ini membutuhkan integrasi yang lancar antara pendekatan berbasis embedding dan berbasis simbol. Seiring dengan kemajuan, kesadaran yang lebih besar akan isu-isu etika teknologi bahasa dalam ranah hukum juga meningkat. Khususnya, terkait bias gender dan diskriminasi rasial, yang mungkin sangat penting untuk tugas-tugas seperti prediksi penilaian. Kami juga telah memperingatkan potensi penyalahgunaan teknologi ucapan untuk peniruan identitas atau spoofing, dan yang tak kalah pentingnya, masalah privasi yang terkait dengan pemrosesan jarak jauh sinyal seperti ucapan yang secara hukum harus dianggap sebagai PII (Informasi Identifikasi Pribadi).

BAB 3
IMPLIKASI SOSIAL SISTEM REKOMENDASI: PERSPEKTIF TEKNIS

Abstrak Salah satu aplikasi algoritma kecerdasan buatan yang paling populer adalah sistem rekomendasi (RS). Sistem ini memanfaatkan data pengguna dalam jumlah besar untuk belajar dari masa lalu guna membantu kita mengidentifikasi pola, mengelompokkan profil pengguna, dan memprediksi perilaku serta preferensi pengguna. Arsitektur algoritmik RS telah begitu sukses sehingga telah diadopsi dalam berbagai konteks, mulai dari tim sumber daya manusia yang mencoba memilih kandidat terbaik, hingga peneliti medis yang ingin mengidentifikasi target obat. Meskipun peningkatan penggunaan AI dapat memberikan manfaat besar, hal ini merepresentasikan pergeseran dalam interaksi kita dengan data dan mesin yang juga menimbulkan ancaman sosial yang fundamental. Ancaman ini dapat berasal dari kesalahan teknologi atau implementasi, tetapi juga dari perubahan besar dalam pengambilan keputusan.

Di sini, kami mengulas beberapa risiko tersebut, termasuk tantangan etika dan privasi, dari perspektif teknis. Kami membahas dua kasus yang sangat relevan: (1) RS yang gagal berfungsi sebagaimana mestinya dan kemungkinan konsekuensi yang tidak diinginkan; (2) RS yang efektif, tetapi dengan kemungkinan mengorbankan ancaman bagi individu dan bahkan masyarakat demokratis. Terakhir, kami mengusulkan langkah ke depan melalui daftar periksa sederhana yang dapat digunakan untuk meningkatkan transparansi dan akuntabilitas algoritma AI.

3.1 RISIKO SISTEM REKOMENDASI ML PADA PENGAMBILAN KEPUTUSAN

Layaknya Revolusi Industri sebelumnya, Revolusi Digital pasti akan berdampak besar pada masyarakat, di berbagai tingkatan. Dengan belajar dari data tingkat individu yang tak tertandingi jumlahnya yang saat ini dibagikan dan dikumpulkan, mesin akan semakin mampu mengidentifikasi pola, membuat profil, memprediksi perilaku, dan membuat keputusan. Oleh karena itu, penting untuk memahami keterbatasan alat-alat ini guna mengantisipasi dan meminimalkan konsekuensi negatif.

Dalam bab ini, kami berfokus pada model pembelajaran mesin (ML), khususnya sistem rekomendasi (atau sistem rekomendasi), dan bagaimana penggunaannya dalam proses pengambilan keputusan dapat menawarkan layanan yang lebih baik tetapi juga menciptakan risiko penting.

Biasanya, RS merujuk pada algoritma yang merekomendasikan beberapa item X ke pengguna A, sangat sering barang konsumsi (seperti merekomendasikan buku yang diidentifikasi algoritma sebagai sesuai dengan minat kita) atau dalam konteks jejaring sosial (teman atau postingan baru); namun, di sini kami menggunakan istilah ini dalam arti yang lebih luas, untuk merujuk pada algoritma apa pun yang menggunakan kumpulan data besar pada orang untuk mengidentifikasi kecocokan kesamaan dan merekomendasikan keputusan, dalam banyak konteks berbeda. Pertama, kami menjelaskan cara kerja RS dan keterbatasannya yang

tidak dapat dihindari. Kedua, kami fokus pada RS yang bekerja sebagaimana mestinya dan membahas bagaimana pembuatan profil individual dapat menyebabkan iklan bertarget yang kasar dan bahkan ancaman terhadap demokrasi, dari disinformasi hingga pengawasan negara. Ketiga, kami menjelaskan apa yang terjadi ketika sistem ini salah, tetapi masih digunakan untuk membuat generalisasi probabilistik dan membantu dalam pengambilan keputusan berbasis AI. Kami akan menawarkan contoh spesifik tentang bagaimana kesalahan dalam pemilihan data atau pengkodean dapat menyebabkan diskriminasi dan ketidakadilan.

Pada bagian terakhir, kami merangkum beberapa ide tentang cara membuat AI lebih bertanggung jawab dan transparan serta berpendapat bahwa keputusan penting ke depan tidak boleh dibuat oleh sekelompok terbatas pemimpin AI yang tidak dipilih, tetapi seharusnya menjadi peran para ahli AI untuk meningkatkan kesadaran terhadap ancaman tersebut, yang membuka jalan bagi keputusan regulasi yang penting.

3.2 SISTEM REKOMENDASI

Tujuan umum sistem rekomendasi adalah memprediksi, seakurat mungkin, item baru bagi pengguna sambil mengoptimalkan tingkat penerimaan. Sistem ini memanfaatkan informasi tentang pengguna (demografi, pilihan sebelumnya) dan/atau item (misalnya, film) untuk menemukan kecocokan yang akurat di antara keduanya (misalnya: jika Anda menyukai film X, Anda mungkin menyukai film Y, atau orang-orang "seperti Anda" telah menikmati buku Z). Inti dari RS adalah asumsi bahwa item dan/atau pengguna suatu layanan dapat dipetakan berdasarkan kesamaannya dan bahwa orang A (atau item X) dapat berfungsi sebagai proksi untuk orang B (item Y).

Dalam hal ini, sistem rekomendasi menyarankan item yang dalam ruang kesamaan tersebut lebih dekat dengan pilihan atau preferensi pengguna sebelumnya dan hanya sebaik yang dapat memberikan rekomendasi paling akurat kepada pengguna (orang A akan benar-benar menikmati film Y dan buku Z).

Selama beberapa dekade terakhir, kita telah menyaksikan peningkatan penggunaan teknik ML/AI untuk mendukung pengembangan dan implementasi RS yang lebih cepat, lebih andal, dan lebih mumpuni. Hal ini dimungkinkan karena (a) akumulasi data yang besar tentang pilihan pengguna sebelumnya; (b) kumpulan data yang besar tentang detail item; dan (c) algoritma yang semakin canggih yang memanfaatkan data tersebut dan mengambil nilai dari semakin banyaknya fitur dan instans.

Secara umum, sistem rekomendasi dapat dibagi menjadi tiga kelompok besar: penyaringan berbasis kolaboratif (yang mencoba memprediksi apakah orang A akan menyukai produk X berdasarkan preferensi "orang yang mirip" B); penyaringan berbasis konten (yang mencoba memprediksi apakah orang A akan menyukai barang X berdasarkan preferensi orang A yang telah terungkap sebelumnya untuk barang serupa); dan sistem hibrida, yang menggabungkan keduanya. Dalam hal algoritma yang digunakan, sistem ini dibagi lagi berdasarkan apakah didukung oleh pendekatan heuristik atau berbasis model.

Sistem rekomendasi penyaringan kolaboratif bergantung pada kesamaan antar pengguna untuk melakukan rekomendasi. Artinya, jika pengguna A dan B serupa, maka pilihan

B di masa lalu dapat menjelaskan apa yang akan direkomendasikan kepada pengguna A. Oleh karena itu, tantangan teknis inti adalah memperkirakan kesamaan antar pengguna atau barang dari data preferensi pengguna di masa lalu yang terungkap (misalnya, film favorit di masa lalu).

Pendekatan ini telah populer di portal berbasis web seperti Netflix dan Reddit di mana karakteristik pengguna dan suara positif atau negatif digunakan untuk memperkirakan kesamaan. Algoritma populer berkisar dari model Grafik jejaring sosial yang menunjukkan kesamaan antara pengguna dan Tetangga Terdekat hingga penggunaan regresi Linier, teknik Pengelompokan, Jaringan Syaraf Tiruan, dan jaringan Bayesian.

Seringkali, data tambahan digunakan untuk meningkatkan sistem penyaringan kolaboratif, atau untuk mengatasi beberapa keterbatasannya. Informasi konteks (misalnya, lokasi atau waktu) dapat membantu sistem mencapai tingkat keberhasilan yang lebih tinggi dan, dalam skenario yang memiliki lebih banyak pengguna daripada item, rekomendasi sering kali dilakukan melalui kesamaan item-item. Selain itu, ketika informasi masa lalu tentang aktivitas pengguna terbatas (misalnya, dalam kasus pengguna baru suatu layanan baru), informasi pengguna tentang hubungan sosial dan karakteristik mereka (misalnya, jenis kelamin, usia, pendapatan, lokasi, pekerjaan, dll.) dapat membantu sistem ini membangun kesamaan bahkan tanpa aktivitas historis yang spesifik.

Penyaringan berbasis konten tidak memerlukan informasi tentang pengguna, melainkan memetakan kesamaan antar item untuk menghasilkan rekomendasi. Dengan kata lain, pengguna direkomendasikan item yang serupa dengan pilihan mereka sebelumnya (orang A menyukai milkshake vanila, sehingga mungkin juga menyukai es krim vanila). Secara algoritmik, permasalahan ini didekati menggunakan berbagai teknik, mulai dari TF-IDF dan pengelompokan untuk pemodelan dan inferensi topik, serta menggunakan model klasifikasi berdasarkan pengklasifikasi Bayesian, pohon keputusan, dan Jaringan Syaraf Tiruan. Aplikasi tradisional dari teknik-teknik ini adalah pada mesin rekomendasi buku yang mengukur kesamaan konten antar buku.

Solusi hibrida menggabungkan aspek penyaringan konten dan kolaboratif. Solusi ini muncul dalam situasi yang praktis dan bermanfaat untuk mengembangkan meta-algoritma guna menyeimbangkan rekomendasi yang berasal dari sistem berbasis kolaboratif dan konten. Hal ini semakin banyak menggunakan algoritma pembelajaran mendalam yang kompleks dan umum digunakan dalam rekomendasi media sosial, termasuk konten umpan berita, iklan, dan teman.

Ada beberapa keterbatasan dan masalah yang terkait dengan pengembangan RS. Dari perspektif teknis, masalah dapat muncul di dua ujung spektrum. Pertama, kurangnya data awal dapat menyebabkan "cold start", yang mencegah penyiapan seluruh sistem rekomendasi (misalnya, film baru yang belum dinilai oleh siapa pun atau berasal dari studio atau sutradara yang kurang dikenal) atau membatasi rekomendasi yang dapat diberikan kepada pengguna baru. Kedua, ketika ada "masalah kelangkaan" dan jumlah item yang akan direkomendasikan sangat besar, algoritma mungkin kurang skalabel dan pengguna terus melihat beberapa rekomendasi yang sama, baik karena mereka adalah beberapa yang paling banyak dinilai atau karena masing-masing pengguna hanya menilai beberapa.

Ketiga, dan lebih konseptual, asumsi implisit dalam solusi ML/AI bahwa masa depan dapat diprediksi dari tindakan masa lalu, membuat solusi tersebut sulit untuk dijalankan dalam kondisi baru (misalnya, memperluas layanan ke lingkungan budaya baru). Keempat, beberapa model yang dijelaskan di atas dapat belajar dari kesalahan masa lalu (pengguna A membenci buku Z) dan, oleh karena itu, terus berkembang, tetapi hal ini tidak berlaku pada beberapa contoh RS lainnya, terkadang dengan konsekuensi yang mengerikan. Contoh-contoh penting tentang dampak keterbatasan yang tercantum akan dibahas lebih rinci di bagian-bagian berikut.

3.3 KAPAN SISTEM REKOMENDASI BERFUNGSI

Implikasi terhadap Konsumsi

Meskipun berawal dari masalah yang tampak intuitif dan sederhana, sistem rekomendasi telah berkembang menjadi solusi algoritmik yang sangat kompleks yang mampu memanfaatkan beragam sumber data untuk meningkatkan layanan yang mendasari kesuksesan beberapa perusahaan terbesar di dunia.

Keberhasilan Sistem Rekomendasi, dan keberadaannya di mana-mana, berasal dari kemampuannya untuk meningkatkan retensi pengguna dan tampaknya membantu pengguna menemukan konten yang relevan untuk profil mereka. Selain itu, ekstraksi informasi dan pola dari informasi terstruktur (misalnya, keranjang belanja daring) dan informasi tidak terstruktur (misalnya, teks bebas, gambar, dan video) sudah dimungkinkan. Dengan demikian, Sistem Rekomendasi berkembang lebih cepat dan akan menawarkan lebih banyak keuntungan bagi penyedia konten.

Tentu saja, spesifikasi setiap sistem sangat bergantung pada konteks dan tujuan penerapannya. Ambil contoh, Amazon yang awalnya menggunakan pendekatan per item, kini memanfaatkan informasi dari pesanan, profil, dan aktivitas pengguna sebelumnya untuk menawarkan berbagai jenis rekomendasi personal, mulai dari email tertarget hingga rekomendasi belanja. Pada tahun 2018, perusahaan konsultan McKinsey memperkirakan bahwa RS Amazon menyumbang 35% dari penjualannya. Netflix meraih popularitas internasional di kalangan insinyur dan penggemar dengan merilis, pada tahun 2006, kumpulan data berisi 100 juta rating film pengguna dan menawarkan hadiah Rp 10 miliar kepada tim yang dapat mengembangkan RS terbaik.

Sepuluh tahun kemudian, RS Netflix diperkirakan bernilai hingga 1 miliar dolar AS dan mendorong 75% pilihan tontonan pengguna. Namun, perusahaan-perusahaan besar ini bergantung pada penggunaan data secara bebas dan sering kali dengan sukarela dibagikan oleh penggunanya, yang mungkin menyerahkan kendali atas privasi dan keputusan mereka dengan imbalan kenyamanan dan produktivitas. Faktanya, kita semua tahu bahwa data kita sedang digunakan, tetapi kita mungkin tidak tahu sejauh mana ini terjadi atau masalah yang dapat ditimbulkannya. Pertama, dan meskipun proses ini melibatkan persetujuan, ketentuan layanan seringkali tidak dapat dipahami, berbagi tidak selalu bersifat sukarela, dan mungkin merupakan persyaratan untuk mengakses konten atau layanan.

Kedua, dan bahkan ketika bersifat sukarela, hal itu dapat memiliki implikasi yang tidak terduga. Pada tahun 2012, New York Times melaporkan sebuah kasus di mana jaringan Target yang berbasis di AS menghasilkan prediksi tentang kehamilan pelanggannya, tepatnya dengan menganalisis profil belanja. Salah satu toko tersebut dikunjungi oleh seorang ayah, yang marah karena putrinya yang remaja telah menerima promosi untuk produk bayi; Kemudian, ketika manajer toko menelepon untuk meminta maaf, pria itu dengan malu menjawab bahwa putrinya memang hamil: jaringan supermarket sudah tahu sebelum keluarga. Situasi anekdot semacam itu memperkuat meningkatnya ketergantungan kita pada sistem semacam itu, yang juga membuat kita rentan terhadap manipulasi. Misalnya, tidak ada yang menghalangi toko daring untuk menampilkan produk yang lebih mahal kepada orang-orang yang sebelumnya tidak membandingkan harga secara daring.

Memang, berbagai layanan web umumnya memperdagangkan informasi pengguna untuk tujuan pemasaran, dan merupakan hal yang umum bagi pengguna yang mencari jeans di Amazon untuk langsung ditargetkan dengan iklan jeans di Facebook atau Google. Yang penting, "sistem pengawasan" ini begitu lazim dan semakin canggih sehingga bahkan ketika Anda berhati-hati saat mempublikasikan secara daring, kehati-hatian itu sendiri dapat informatif.

Sebagaimana telah disebutkan, aplikasi tradisional RS adalah untuk mendorong konsumsi konten dan produk dan, dengan demikian, RS mewakili pengembangan algoritma semacam itu yang paling umum (identifikasi kesamaan, ketergantungan pada proksi, prediksi hasil di masa mendatang). Namun, mereka juga menemukan penerapan dalam jenis pengambilan keputusan algoritmik lainnya (misalnya, skor kredit, atau perdagangan keuangan) dan kami akan menggunakannya dalam konteks yang lebih luas untuk membahas lebih lanjut implikasinya saat ini.

Implikasi bagi Demokrasi

Sebagaimana dijelaskan, RS dapat sangat berguna untuk mengarahkan orang ke produk yang mereka minati, baik film maupun popok. Namun, ada garis tipis antara menginformasikan dan memanipulasi, dan ini khususnya relevan ketika "barang" yang dipromosikan adalah berita atau ide. Jejaring sosial, seperti Facebook atau Instagram, telah lama dikenal dapat meningkatkan perhatian yang adiktif, meskipun dengan mengorbankan penyebaran disinformasi, menciptakan ruang gema, meningkatkan polarisasi, dan mengancam kesehatan mental pengguna.

Dari para peneliti hingga advokat perlindungan data, banyak yang telah menyuarakan kekhawatiran tentang data yang dikumpulkan oleh platform besar dan bagaimana sistem rekomendasi mereka dapat memanipulasi informasi yang diakses individu, baik itu memprioritaskan postingan, atau mesin pencari yang menampilkan iklan bersponsor. Faktanya, Facebook menawarkan kepada siapa pun yang berminat untuk menempatkan iklan kemungkinan untuk memilih lebih dari sepersepuluh karakteristik individu, termasuk (perkiraan) usia atau jenis kelamin, tingkat pendidikan dan mata pelajaran, tingkat pendapatan, hobi, orientasi politik, profil perjalanan, dan bahkan apakah target sedang pergi untuk akhir pekan bersama keluarga atau teman.

Kasus Cambridge Analytica, pada awal 2018, menyoroti publik bagaimana, melalui pembuatan profil individu yang lebih baik, kampanye politik dapat memengaruhi pemungutan suara individu atau konstituen target. Ilmuwan politik berpendapat bahwa penggunaan pendekatan ilmu data modern terhadap politik mewakili pergeseran signifikan dari strategi klasik: teknik pemasaran telah digunakan dalam politik setidaknya sejak tahun 1930an, tetapi kecepatan dan meningkatnya presisi alat AI berarti bahwa pesan politik tidak perlu lagi bersifat umum dan menarik bagi konstituen yang luas; Bahasa Indonesia: alih-alih, mereka dapat mengirim pesan yang sangat personal berdasarkan profil individu, mengatakan satu hal kepada satu demografi dan kebalikannya kepada yang lain, dengan pengawasan yang sangat minim.

Bahwa beberapa pesan ini dapat berisi informasi yang tidak benar bahkan lebih mengkhawatirkan. Tentu saja, penggunaan informasi yang menyesatkan dan bahkan sepenuhnya salah untuk tujuan politik bukanlah hal baru, tetapi lonjakan aktivitas daring, ditambah dengan literasi digital yang buruk, dan profil konsumen tingkat individu, memiliki semua unsur dari badai yang sempurna. Penyebaran disinformasi telah menemukan lahan subur di jejaring sosial, seringkali melalui manipulasi emosi, yang pertama kali terbukti terjadi oleh Facebook sendiri, dan tidak hanya bekerja pada individu yang menjadi target tetapi juga untuk mencemari teman-teman mereka.

Ada juga semakin banyak bukti bahwa beberapa individu mungkin lebih rentan terhadap disinformasi politik daripada yang lain, dengan bias kognitif spesifik memainkan peran penting. Faktanya, algoritma yang dipersonalisasi pada mesin pencarian dan umpan jejaring sosial mungkin memperkuat bias yang sudah ada ini setidaknya dalam tiga cara berbeda: (1) karena informasi disaring berdasarkan riwayat masa lalu (yang berpotensi memperbesar bias ketersediaan, di mana individu cenderung menilai hal-hal yang lebih penting yang dapat mereka ingat dengan lebih mudah, dan kecenderungan konfirmatori, di mana individu mencari atau khususnya mempercayai informasi yang memperkuat atau mengonfirmasi keyakinan mereka; (2) karena manusia cenderung bergaul dengan orang lain yang mirip dengan mereka dan lebih menyukai orang-orang dalam kelompoknya sendiri, daripada orang-orang yang diidentifikasi sebagai bagian dari kelompok luar (bias kelompok dalam); dan (3) dengan keyakinan, bias, dan bahkan disinformasi yang diperkuat dan dikuatkan oleh komunitas homofilik yang tertutup ini, yang mengarah ke ruang gema yang telah disebutkan dan peningkatan permusuhan dan polarisasi daring. Namun, (mis)informasi politik bukanlah satu-satunya jenis informasi yang berdampak besar pada masyarakat dan demokrasi. Pada April 2020, Facebook mengakui bahwa jutaan penggunanya melihat informasi palsu terkait COVID-19 di platform ini; Di Twitter, menurut Yang dkk., "volume gabungan twit yang menautkan ke informasi dengan kredibilitas rendah sebanding dengan volume artikel New York Times dan tautan CDC"; hingga Agustus 2021, YouTube telah menghapus 1 juta video yang memuat misinformasi COVID-19 yang berbahaya. Yang penting, terdapat bukti bahwa misinformasi tersebut berdampak pada keraguan vaksinasi dan kepatuhan terhadap langkah-langkah pengendalian, sejalan dengan gagasan bahwa misinformasi seringkali bertujuan untuk menciptakan konten yang memecah belah dan memicu keresahan sosial. Hampir sepanjang tahun 2020 dan 2021, dunia memerangi dua pandemi secara paralel: satu disebabkan oleh

virus, dan satu lagi disebabkan oleh berita palsu, yang didukung oleh bias manusia dan algoritma yang memaksimalkan perhatian.

Risiko lain yang sangat relevan datang dari kontrol masyarakat. Sebagaimana dijelaskan, politisi dapat menggunakan jejaring sosial dan sistem AI untuk menyasar calon pemilih, tetapi beberapa pemimpin juga telah menyadari potensi AI yang jauh lebih luas, mulai dari peningkatan administrasi publik hingga penciptaan robot perang. Menurut Vladimir Putin: "Kecerdasan buatan adalah masa depan, tidak hanya Rusia, tetapi seluruh umat manusia (...) Siapa pun yang menjadi pemimpin di bidang ini akan menjadi penguasa dunia." Tiongkok berharap menjadi pemimpin pada tahun 2030 (Departemen Kerja Sama Internasional Kementerian Sains dan Teknologi (MOST) 2017) dan sedang merancang dan mengimplementasikan eksperimen sosial skala besar, yang melibatkan penggunaan RS untuk mengklasifikasikan warga negara berdasarkan perilaku sosial mereka, yang hanya dimungkinkan berkat teknologi pengenalan wajah berbasis AI. Model-model ini semakin banyak digunakan di seluruh dunia, seringkali dengan tujuan keamanan. Pada tahun 2020, tentara Israel dan AS menggunakan AI untuk melacak dan membunuh seorang fisikawan Iran.

Semua contoh ini menggambarkan situasi di mana RS mengkhawatirkan karena berfungsi sebagaimana mestinya, baik untuk meningkatkan konsumsi maupun untuk menyasar pemilih. Di bab selanjutnya, kita akan berfokus pada situasi di mana alat pengenalan wajah gagal dan bagaimana hal itu dapat berdampak pada individu dan masyarakat.

3.4 KETIKA SISTEM REKOMENDASI GAGAL

Sistem rekomendasi yang dijelaskan menggunakan informasi terperinci untuk melatih model AI agar dapat menargetkan individu tertentu. Biasanya, sistem ini menghasilkan probabilitas suatu peristiwa tertentu dan membantu dalam pengambilan keputusan. Contohnya antara lain algoritma yang menghitung skor risiko depresi, mencoba mengidentifikasi kandidat terbaik untuk posisi tertentu, atau yang merekomendasikan film berdasarkan pilihan sebelumnya. Algoritma ini dilatih pada set data pelatihan yang besar, dengan kualitas yang bervariasi, dan "belajar" melalui uji coba, dengan definisi kesalahan yang subjektif (misalnya, arti "kandidat terbaik" harus direpresentasikan sebagai objek matematika, padahal seringkali "terbaik" tidak dapat dengan mudah dikuantifikasi).

Ini berarti tidak ada perbedaan nyata antara model, data yang digunakan untuk melatihnya, dan asumsi yang dibuat oleh pembuat kode: jika data atau target bias, model akan bias. Bias ini mungkin muncul pada langkah yang berbeda dan memiliki konsekuensi yang berbeda, tetapi penting untuk disadari bahwa: (1) hampir mustahil untuk memiliki kumpulan data yang lengkap dan semua kumpulan data adalah sampel, yang bias oleh proses pengambilan sampel; (2) terdapat keputusan manusia yang terlibat dalam menentukan target; (3) target seringkali bergantung pada proksi, (4) prediksi dapat berubah menjadi ramalan yang terpenuhi dengan sendirinya karena seringkali memengaruhi hasil dan seringkali sangat sulit untuk mendapatkan validasi eksternal. Sekali lagi, ini mungkin tidak terlalu penting dalam kasus pengguna yang tidak pernah mendapatkan film X karena tidak disarankan, tetapi sangat serius dalam kasus seseorang yang permintaan kreditnya ditolak dan, akibatnya, gagal

membayar pembayaran lain: sistem mungkin menemukan konfirmasi bahwa penolakan kredit adalah keputusan terbaik padahal sebenarnya, itulah yang menyebabkan gagal bayar. Akibatnya, seharusnya tidak ada ilusi "netralitas model". Semua model memiliki masalah, dan mengakuinya merupakan langkah fundamental dan esensial dalam merancang strategi mitigasi. Di bagian ini, kami menjelaskan bagaimana data yang bias menyebabkan algoritma yang bias, bagaimana algoritma yang bias dapat menyebabkan kebijakan diskriminatif, dan menawarkan beberapa contoh dari sektor swasta maupun publik.

Belajar dari Data yang Bias: Implikasi bagi Individu

Karena tidak ada dataset yang sempurna, penting untuk memahami keterbatasannya saat melatih algoritma apa pun dengannya. Mari kita bayangkan sebuah model untuk mengidentifikasi kandidat terbaik untuk masuk sekolah teknik. Kita dapat memulainya dengan mengumpulkan sejumlah besar data, termasuk nilai, skor kebahagiaan, waktu penyelesaian gelar, pendidikan sebelumnya, karier masa depan, dll., dari semua mahasiswa yang pernah menempuh pendidikan di universitas tertentu.

Dataset ini tetap tidak memiliki informasi tentang seberapa baik kandidat yang ditolak (bias pengambilan sampel) atau berapa banyak dari mereka yang akhirnya mengalami kelelahan (keterbatasan dan subjektivitas dalam pemilihan fitur): ini berarti, jika sistem secara sistematis menolak kandidat yang menjanjikan di masa lalu, algoritma kemungkinan besar akan terus melakukannya di masa mendatang; dan jika, misalnya, ia lebih mementingkan hadiah daripada menciptakan lingkungan kerja yang aman, hal itu mungkin membuka jalan bagi lebih banyak kecelakaan di masa mendatang.

Mudah juga untuk mengantisipasi bahwa kumpulan data seperti itu akan tidak seimbang dalam hal jenis kelamin dan kemungkinan juga usia, etnis, kebangsaan, dan probabilitas memakai kacamata, dan begitu pula prediksi model. Faktanya, beberapa upaya sebelumnya dalam melatih algoritma tersebut untuk memilih pelamar, untuk sekolah atau pekerjaan, telah menyebabkan praktik diskriminatif, yang memicu diskusi besar. Penting untuk dicatat bahwa, sangat sering, algoritma ini dibuat tidak hanya untuk mempercepat dan mengotomatiskan proses, tetapi juga karena kita tahu bahwa sistem berbasis manusia bias: asumsinya adalah bahwa model akan buta terhadap warna atau jenis kelamin dan, dengan demikian, lebih adil. Namun, RS yang dilatih dengan data yang bias umumnya juga akan bias, dan ini benar bahkan jika model dilatih pada kumpulan data yang sangat besar. Misalnya, aplikasi Chat GPT yang semakin populer dilatih menggunakan data sebesar 570 Gb, tetapi sebagian besar data ini diperoleh melalui internet, yang diketahui memiliki representasi berlebihan dari beberapa negara dan kelompok usia.

Salah satu area di mana diskriminasi semacam itu dapat berdampak buruk adalah kesehatan. Kadambi menemukan sumber bias krusial dalam perangkat medis, termasuk bias komputasional, yang terjadi ketika kumpulan data yang digunakan dalam uji klinis atau ketika algoritma pelatihan untuk memilih kandidat uji klinis tersebut, bias. Secara historis, hal ini terjadi pada kelompok etnis dan perempuan tertentu, yang seringkali kurang terwakili dalam kumpulan data kesehatan (dan bahkan dalam protokol eksperimental).

Kumpulan data yang bias juga telah terbukti memainkan peran penting dalam klasifikasi dan pengenalan wajah. Misalnya, Twitter menghentikan algoritma pemotongan gambarnya setelah adanya kecurigaan bias rasial dan algoritma Flickr maupun Google menandai foto individu kulit hitam sebagai kera. Meskipun contoh-contoh ini sangat buruk, dapat dikatakan bahwa ini adalah harga yang harus dibayar untuk proses pembelajaran: ini adalah kisah peringatan, yang mengingatkan kita bahwa kita masih berada di tahap awal pengambilan keputusan mesin dan banyak kesalahan lain akan terjadi sebelum kita dapat mengandalkan algoritma. Sayangnya, terlepas dari keterbatasannya saat ini, banyak algoritma yang sudah diterapkan, termasuk dalam lingkungan yang menghukum, seperti yang dijelaskan di bagian selanjutnya.

Dari Algoritma yang Buruk hingga Kebijakan Diskriminatif

Konsekuensi individual dari algoritma Netflix yang salah mungkin mudah diminimalkan; Model yang menyeleksi kandidat untuk pekerjaan tertentu dapat memiliki konsekuensi yang jauh lebih buruk, tetapi tidak ada yang sebanding dengan penerapan algoritma tersebut dalam konteks hukuman berskala besar. Kami telah membahas bagaimana pemerintah Tiongkok menggunakan pengenalan wajah dan perangkat AI lainnya untuk mengevaluasi warga berdasarkan perilaku mereka. Jika sistemnya salah, konsekuensinya bagi individu bisa sangat besar.

Contoh lain yang sangat diperdebatkan, yang bergantung pada proksi, adalah COMPAS, sebuah algoritma khusus yang membantu hakim AS menetapkan jaminan berdasarkan skor risiko yang diperkirakan dari pelanggaran di masa mendatang. Pada tahun 2016, COMPAS dianalisis oleh ProPublica dan terbukti mendiskriminasi individu berdasarkan ras mereka: untuk pelanggaran dan kejahatan serupa, terdakwa berkulit hitam lebih mungkin diberi skor risiko yang lebih tinggi. Yang penting, kumpulan data yang digunakan untuk melatih model tersebut tidak mencakup informasi tentang ras: model tersebut mungkin menggunakan kode pos sebagai proksi untuk risiko, sehingga memilihnya sebagai proksi korelasi antara etnis dan status ekonomi dalam masyarakat AS (analisis ini dibantah oleh perusahaan pemiliknya dan tingkat diskriminasinya masih diperdebatkan.

Meskipun begitu banyak kegagalan yang nyata, pemerintah di seluruh dunia telah mensponsori pengembangan algoritma untuk digunakan dalam administrasi publik, di berbagai bidang. Algoritma-algoritma ini seringkali bersifat kepemilikan dan berfungsi sebagai kotak hitam: bahkan pejabat pemerintah pun tidak tahu cara kerjanya dan apa yang membenarkan rekomendasi atau skor risiko mereka. Hal ini menyisakan sangat sedikit ruang bagi masyarakat untuk mengeluh atau bahkan memahami "evaluasi" mereka, sehingga menimbulkan pertanyaan hukum yang mendasar.

Pemerintah Belanda menggunakan algoritma semacam itu, SyRI, dari tahun 2014 hingga 2020, ketika Pengadilan Den Haag menghentikan penggunaannya. Algoritma ini bertujuan untuk mengidentifikasi penipuan kesejahteraan sosial, dilatih dengan kumpulan data pemerintah yang besar, dan mencakup informasi tentang hampir seluruh penduduk Belanda. Algoritma ini akan menghasilkan skor risiko dan, jika skornya tinggi, akan memicu investigasi. Namun, terbukti bahwa algoritma tersebut secara tidak proporsional dan tidak adil

menargetkan komunitas miskin dan minoritas, dengan konsekuensi yang begitu mengerikan hingga menyebabkan pengunduran diri pemerintah Belanda. Algoritma yang cacat seperti itu semakin terungkap, tetapi, tentu saja, dapat dikatakan bahwa algoritma tersebut hanya mengungkap bias masa lalu dan pra-eksis, yang tersembunyi dalam data, dan bahwa pengambilan keputusan manusia sama-sama diskriminatif.

Meskipun argumen pertama kemungkinan besar benar, hal itu tetap menimbulkan pertanyaan penting apakah dapat diterima untuk melestarikan praktik diskriminatif semacam itu di bawah ilusi netralitas matematis. Argumen kedua lebih menarik, karena sulit untuk mengukur apakah manusia atau algoritma saat ini yang lebih diskriminatif, tetapi setidaknya dalam kasus contoh yang dijelaskan di sini, setidaknya ada dua argumen bagus yang mendukung argumen kedua. Pertama, teknis, karena model AI mengidentifikasi pola dominan dan lebih cenderung mengecualikan orang luar yang relevan (misalnya, kandidat brilian dari lingkungan kulit hitam yang sangat miskin). Argumen kedua, skala, karena panel manusia mungkin memiliki biasnya sendiri, tetapi bias ini mungkin berbeda dari satu panel ke panel lainnya dan terdapat batasan manusia terhadap jumlah pelamar yang dapat dilihat oleh panel; Jelas, batasan dan variasi alami ini tidak selalu berlaku untuk pengambilan keputusan oleh mesin.

Kemungkinan ketiga yang kurang diteliti adalah bahwa algoritma, termasuk kuki komersial, dapat digunakan untuk menutupi penargetan individu yang disengaja oleh aktor negara, seperti dalam kasus identifikasi minoritas. Oleh karena itu, terdapat kekhawatiran serius bahwa algoritma yang dilatih pada kumpulan data yang bias tidak hanya akan menghasilkan keputusan yang bias, tetapi juga akan memperkuat diskriminasi dan ketidakadilan sosial yang sudah ada.

3.5 LANGKAH PERBAIKAN DI MASA MENDATANG

Masih banyak ruang untuk perbaikan RS saat ini dan di masa mendatang (dan AI secara umum), dan kami mengusulkan enam langkah, yang dirangkum dalam mnemonik sederhana: (ATI). Langkah pertama adalah mengakui bahwa mereka tidak netral dan sangat rentan terhadap bias. Penerimaan ini seharusnya jelas, tetapi masih diperdebatkan oleh beberapa ahli di bidang ini, yang biasanya menyatakan bahwa mereka (a) tidak bias karena algoritma tidak peka terhadap individu, (b) tidak lebih bias daripada sistem non-algoritmik, atau (c) bahwa ini merupakan masalah bagi ilmuwan sosial dan bahwa insinyur dan pemrogram tidak perlu khawatir dengan isu-isu tersebut.

Faktanya, sebagian besar program pascasarjana Ilmu Data dan Kecerdasan Buatan masih belum mencakup mata kuliah Etika atau bahkan Keadilan Algoritmik, yang secara efektif melatih generasi mahasiswa untuk mengabaikan masalah fundamental dengan kumpulan data, algoritma, dan, akibatnya, sistem rekomendasi dan keputusan yang mereka rancang dan sering implementasikan. Materi semacam itu seharusnya wajib dalam semua pendidikan AI formal, yang membawa kita ke langkah kedua Pelatihan.

Masalah mendasar lainnya adalah kurangnya Inklusi dan Keberagaman. Hal ini tidak hanya diamati pada set data pelatihan, sebagaimana telah dibahas, tetapi juga pada tim

pengodean. Dalam buku "*Racist in the Machine*", Megan Garcia menjelaskan beberapa konsekuensi serius dari titik buta desain dan memberikan contoh empat asisten pribadi ponsel pintar (Siri, Google Now, Cortana, dan S Voice), yang semakin banyak digunakan untuk membantu dalam situasi kesehatan dan darurat, yang tidak dapat mengenali "Saya sedang dilecehkan" atau "Saya dipukuli oleh suami saya".

Tim pembelajaran mesin (ML) harus beragam dan menyatukan orang-orang yang bekerja di berbagai disiplin ilmu dan yang dapat berkontribusi pada komponen teknis dan sosial dari desain algoritma. Lebih lanjut, algoritma yang mencoba menemukan kesamaan dapat menyebabkan polarisasi dan homofili, tetapi juga penyeragaman. Seperti yang dikatakan Thomas Homer-Dixon, "budaya global yang disederhanakan dan seragam pasti akan memiliki lebih sedikit keragaman ide dan kecerdikan yang dapat membantu kita mengatasi tantangan besar yang kita hadapi". Keberagaman harus menjadi nilai di berbagai tingkatan.

Pipeline RS juga harus mencakup Audit Data dan Debiasing yang efisien: menerimanya sebagai bagian integral dari pipeline pemrosesan data, mengakui pentingnya hal tersebut untuk AI yang adil dan efektif, sekaligus mengurangi bias pada dataset. Salah satu upaya pertama tersebut dikembangkan oleh Pedro Saleiro dan Rayid Ghani, melalui algoritma audit data, Aequitas, yang memeriksa dataset untuk berbagai jenis ketidakseimbangan demografis, termasuk usia, jenis kelamin, dan ras.

Dalam ketiga dataset yang dianalisis, mereka menemukan bias penting yang memengaruhi hasil model. Ini merupakan upaya awal yang sangat baik, tetapi penting untuk dicatat bahwa (a) kita hanya dapat mengaudit data dalam jumlah kasus yang sangat terbatas, dan (b) debiasing bahkan lebih menantang. Misalnya, kita dapat memeriksa dataset yang diklasifikasikan berdasarkan gender untuk mendeteksi ketidakseimbangan gender (seperti dalam contoh perekrutan yang dijelaskan di atas), tetapi hal ini mungkin mustahil dilakukan pada dataset yang sepenuhnya anonim atau pada dataset yang tidak menyertakan data yang mungkin relevan seperti yang sering terjadi pada etnisitas atau disabilitas fisik. Yang lebih penting lagi, kita tidak dapat mengidentifikasi bias yang tidak kita sadari keberadaannya, sebagai masyarakat: mungkin saja orang berkacamata dianggap lebih kompeten untuk beberapa pekerjaan; karena kita tidak menyadarinya, kita tidak akan memasukkan "berkacamata" sebagai label dan, bahkan jika kita melakukannya, kemungkinan besar kita tidak akan mengaudit algoritma kita untuk kemungkinan diskriminasi. Namun, mari kita asumsikan bahwa audit lengkap dimungkinkan dan semua kemungkinan ketidakseimbangan diskriminatif dalam dataset kita telah diidentifikasi: kita masih harus membuat keputusan penting terkait cara menghilangkan bias tersebut. Melanjutkan contoh penerimaan perguruan tinggi, seharusnya dapat dipahami bahwa model sedang dilatih pada dataset yang tidak seimbang gender dan bahwa mengoreksinya sekarang akan menghasilkan lebih banyak kandidat perempuan yang terpilih. Tetapi seberapa besar koreksi yang seharusnya?

Haruskah hal ini mencerminkan rasio penerimaan sekolah teknik di masa lalu, yang justru melanggengkan ketidakseimbangan yang ada, atau haruskah tujuannya adalah rasio yang sama seperti yang diamati dalam populasi, yang secara efektif memaksakan kuota gender 50%? Contoh-contoh ini dan lainnya menggambarkan betapa banyak keputusan semacam ini

dapat dan sedang dibuat secara efektif, seringkali secara implisit, dan bagaimana upaya yang kurang tepat untuk mengoreksi bias dapat menghasilkan bentuk-bentuk ketidakadilan baru.

Keputusan-keputusan ini pada dasarnya bermoral, membantu menciptakan masyarakat berdasarkan rancangan. Oleh karena itu, langkah terakhir seharusnya adalah Transparansi. Seperti yang dikatakan Rhema Vaithianathan, "Jika Anda tidak bisa benar, jujurlah".

Algoritma kotak hitam, yang proses dan fiturnya tidak diketahui atau bersifat kepemilikan, sebaiknya dihindari. Namun, algoritma ini semakin banyak digunakan karena dua alasan utama: pertama, dapat dikatakan bahwa jika proses pengambilan keputusan diketahui, individu dan perusahaan dapat menyalahgunakan dan bahkan memanipulasi sistem untuk kepentingan mereka; kedua, semakin kompleks algoritmanya, seperti halnya pembelajaran mendalam, semakin sulit untuk memahami proses pengambilan keputusan. Oleh karena itu, telah dikemukakan bahwa algoritma semacam itu hanya boleh digunakan dalam lingkungan positif dan ketika kinerjanya secara signifikan mengungguli proses tradisional (misalnya untuk diagnosis medis), dan tidak boleh digunakan dalam konteks yang bersifat menghukum (seperti dalam contoh COMPAS dan SyRI).

Bagaimanapun, individu harus selalu memiliki hak untuk mengakses, memverifikasi, mengoreksi kesalahan, dan mengajukan banding atas keputusan algoritmik. Karena proses-proses ini seringkali sangat kompleks, hal ini menimbulkan ketegangan penting, yang secara luas disadari oleh para pemikir yang disebut "Masyarakat Risiko", di mana keahlian teknis merupakan hal mendasar untuk merancang dan mengendalikan sistem tersebut, tetapi kendali ini harus dijalankan oleh masyarakat awam, yang seringkali merupakan masyarakat umum. Oleh karena itu, mereka yang memahami permasalahan ini juga harus menerima tanggung jawab politik dan sosial mereka serta terlibat dalam interaksi aktif dengan masyarakat dan para pengambil keputusan.

3.6 RISIKO SOSIAL SISTEM REKOMENDASI BERBASIS AI

Semakin jelas bahwa penggunaan keputusan berbasis mesin jauh dari netral, dan permasalahannya memiliki implikasi sosial yang penting. Dalam bab ini, kami merangkum beberapa keterbatasan sistem rekomendasi, baik dari perspektif teknis maupun konseptual, serta menawarkan contoh dampak negatifnya di masa lalu, yang sedang berlangsung, dan kemungkinan dampak negatifnya di masa mendatang. Secara keseluruhan, kami berpendapat bahwa risiko-risiko ini perlu dipahami oleh masyarakat umum, dan kami menawarkan panduan khusus untuk meningkatkan RS dan pengawasan sosial.

BAB 4
DATA DALAM KESEHATAN: TANTANGAN & TREN

Abstrak Data mendominasi dan merevolusi industri pelayanan kesehatan dengan cara yang belum pernah terjadi sebelumnya. Berkaitan dengan teknologi baru kecerdasan buatan, teknologi ini menjanjikan akan menciptakan fondasi bagi paradigma baru kedokteran yang berfokus pada individualitas setiap orang. Bab ini dibagi menjadi empat bagian yang bertujuan untuk memperkenalkan pembaca pada topik pendekatan berbasis data di sektor kesehatan. Pada bagian pertama, disajikan tiga ideologi yang, meskipun memiliki beberapa kesamaan, menyajikan pandangan berbeda tentang bagaimana data seharusnya digunakan untuk menjamin layanan kesehatan yang berpusat pada setiap individu. Pada bagian kedua, konsep berbasis data dieksplorasi.

Tantangan yang muncul dalam memproses data dalam volume besar dan dampaknya terhadap individu, institusi, dan masyarakat berkaitan dengan inovasi dalam disiplin ilmu lain seperti kecerdasan buatan dan pengobatan personal. Karena kecerdasan buatan menjadi teknologi yang disruptif di sektor kesehatan, bagian ketiga didedikasikan untuk membahas tantangan etika dan hukum yang ditimbulkan oleh kemajuan teknologi baru ini. Sebagai penutup, bagian keempat menjelaskan bagaimana industri layanan kesehatan telah menjadi ajang pembuktian utama bagi aplikasi kecerdasan buatan, dengan baik perusahaan rintisan maupun investor modal ventura menyadari potensi besar yang ditawarkan oleh teknologi ini.

4.1 PERAWATAN PASIEN, BERBASIS NILAI, DAN PARADIGMA P4

Ideologi seperti perawatan yang berpusat pada pasien, perawatan berbasis nilai, dan kedokteran P4 bukanlah hal baru, berakar pada akhir abad ke-20. Lebih dari 20 tahun yang lalu, proposisi bahwa layanan kesehatan berevolusi dari perawatan penyakit reaktif menjadi perawatan yang berpusat pada pasien atau individu dianggap hipotetis.

Saat ini, elemen inti dari pendekatan ini diterima secara luas dan telah diartikulasikan dalam serangkaian laporan oleh United States Institute of Medicine, atau oleh inisiatif Eropa seperti proyek Innovative Medicines Initiative yang baru-baru ini diluncurkan. Perawatan yang berpusat pada pasien dan perawatan berbasis nilai adalah dua ideologi perawatan yang berbeda namun saling tumpang tindih.

Perawatan yang berpusat pada pasien mencakup berbagai domain yang berpusat pada pasien dan menempatkan pasien dan kerabat sebagai pusat dari semua keputusan dan evaluasi kualitas. Penelitian oleh Picker Institute telah mendefinisikan delapan dimensi perawatan yang berpusat pada pasien, termasuk: (1) penghormatan terhadap nilai-nilai, preferensi, dan kebutuhan yang diungkapkan pasien; (2) informasi dan edukasi; (3) akses ke perawatan; (4) dukungan emosional untuk meredakan rasa takut dan cemas; (5) keterlibatan keluarga dan teman; (6) kontinuitas dan transisi yang aman antar lingkungan perawatan kesehatan; (7) kenyamanan fisik; dan (8) koordinasi perawatan. Meskipun dimensi-dimensi ini

awalnya diterapkan pada perawatan berbasis rumah sakit, dimensi-dimensi ini dapat diterapkan secara setara pada perawatan di fasilitas rawat jalan.

Sebaliknya, perawatan berbasis nilai didefinisikan sebagai kualitas perawatan yang biasanya diukur berdasarkan luaran layanan kesehatan, termasuk biaya. Dalam konsepsi nilai ini, orientasi pada pasien merupakan salah satu ukuran kualitas yang penting, tetapi belum tentu dominan. Dengan menggabungkan model perawatan yang berpusat pada pasien dan berbasis nilai, pasien juga harus dimampukan untuk mengumpulkan data hasil mereka dan memiliki akses serta otonomi untuk menggunakan data kesehatan mereka dengan cara yang bermakna, seperti mengelola kesehatan mereka sendiri dalam kemitraan dengan penyedia layanan kesehatan mereka.

Dalam dunia yang ideal ini, hubungan dokter-pasien ditingkatkan oleh "sistem panduan dan komunikasi berbasis komputer", rekam medis berbasis internet dan tersedia di mana saja, dan pasien secara teratur mengisi survei tentang pengalaman mereka, yang kemudian diumpan balikkan kepada dokter secara "waktu nyata" sehingga mereka dapat meningkatkan perawatan. Perawatan yang berpusat pada pasien merupakan komponen kunci dari sistem kesehatan yang memastikan bahwa semua pasien memiliki akses ke jenis perawatan yang sesuai untuk mereka.

Model ini sangat penting tidak hanya untuk mengobati penyakit, tetapi juga untuk memenuhi kebutuhan sosial dan pribadi pasien guna memastikan hasil terbaik, termasuk kualitas dan kepuasan. Merawat pasien secara menyeluruh dengan tingkat personalisasi ini memerlukan pandangan 360 derajat terhadap data pasien yang dapat diakses oleh pasien dan tim perawatan. Mencapai luaran pasien yang lebih baik untuk layanan kesehatan berbasis nilai membutuhkan kolaborasi yang lebih baik dan lebih cerdas antar tenaga kesehatan profesional.

Visi kedokteran yang prediktif, preventif, personal, dan partisipatif, yang disebut "P4", telah lama diadvokasi oleh Leroy Hood dan pelopor kedokteran sistem lainnya. Pendekatan sistem terhadap biologi dan kedokteran kini mulai menyediakan informasi personal kepada pasien, konsumen, dan dokter tentang pengalaman kesehatan unik setiap individu, baik kesehatan maupun penyakit, pada tingkat molekuler, seluler, dan organ. Informasi ini membuat perawatan penyakit lebih hemat biaya dengan mempersonalisasi perawatan sesuai biologi unik setiap orang dan dengan menangani penyebabnya, alih-alih gejalanya. Informasi ini juga menyediakan dasar bagi tindakan nyata oleh konsumen untuk meningkatkan kesehatan mereka seiring mereka mengamati dampak dari keputusan gaya hidup.

Kedokteran P4 memiliki potensi besar untuk mengurangi beban penyakit kronis dengan memanfaatkan teknologi dan pemahaman yang semakin baik tentang interaksi lingkungan-biologi, intervensi berbasis bukti, dan mekanisme yang mendasari penyakit kronis. Pengobatan P4 menganjurkan bahwa partisipasi individu merupakan kunci untuk menerapkan tiga aspek P4 lainnya pada setiap pasien. Keterlibatan aktif pasien diperlukan untuk menjamin manajemen diri yang efektif, termasuk berbagi keputusan dengan pasien terkait pendekatan klinis atau terapeutik mereka, penggunaan teknologi baru untuk menerapkan partisipasi pasien dalam manajemen penyakit guna memperoleh peningkatan luaran yang signifikan dan relevan.

Perawatan prediktif dan preventif berbasis bukti menggabungkan praktik kedokteran berdasarkan bukti terbaru dengan genomik dan determinan sosial kesehatan untuk menghasilkan rencana perawatan yang benar-benar personal bagi setiap pasien. Hal ini melampaui anggapan bahwa pengujian genomik hanya untuk 5% pasien yang paling sakit, sebuah anggapan yang muncul dari masa ketika pengujian genetik menelan biaya ribuan dolar. Pengujian genetik untuk kesehatan presisi saat ini hanya berharga beberapa ratus dolar dan biayanya terus menurun.

Potensi penghematan biaya di masa mendatang lebih dari sekadar membenarkan biaya awal ini. Penerapan genetika preventif, pada skala populasi, merupakan penentu tidak hanya untuk memahami peningkatan risiko poligenomik seseorang terhadap penyakit atau karakteristik tubuh, tetapi juga untuk mengetahui obat mana yang tidak dapat dimetabolisme, merespons dengan efikasi, dan/atau memiliki efek samping. Bidang penelitian ini, yang disebut farmakogenomik, saat ini menjadi pilar utama penerapan pengobatan preventif dan personalisasi dalam praktik klinis.

Badan Pengawas Obat dan Makanan Amerika Serikat (FDA) baru-baru ini mengakui lebih dari 120 asosiasi farmakogenomik yang data terkininya mendukung perubahan dalam manajemen obat atau potensi dampak pada keamanan, dan daftarnya terus bertambah. *Konsorsium Ubiquitous Pharmacogenomics* (U-PGx), yang didanai oleh program Horizon-2020 Komisi Eropa, bertujuan untuk mengevaluasi utilitas klinis kolektif dari penerapan panel penanda farmakogenomik ke dalam perawatan rutin. Pedoman ilmiah Badan Obat-obatan Eropa tentang farmakogenomik membantu pengembang obat menyiapkan aplikasi otorisasi pemasaran untuk obat-obatan manusia.

Meskipun pasien sangat tertarik pada pengukuran luaran kesehatan yang signifikan, seperti tingkat keparahan gejala atau status fungsional dari waktu ke waktu, sistem kesehatan di seluruh Eropa belum banyak terlibat dalam pengukuran luaran kesehatan jangka panjang tersebut. Akibatnya, keterlibatan pasien dalam pengumpulan dan penggunaan data dan informasi luaran kesehatan yang relevan masih menjadi strategi yang kurang dimanfaatkan untuk mendorong transisi menuju paradigma sistem pelayanan kesehatan baru. Pada saat yang sama, data yang mencakup perspektif pasien perlu disediakan dan dipertimbangkan untuk pengambilan keputusan kebijakan kesehatan. Meskipun menarik lebih banyak minat dari berbagai pemangku kepentingan, model berbasis nilai masih kurang diteliti dan belum diimplementasikan secara luas.

Penyelarasan tujuan dan fokus perawatan yang berpusat pada pasien, perawatan berbasis nilai, dan ideologi pengobatan P4 diperumit oleh beberapa ketegangan, termasuk kurangnya pengalaman pasien, definisi ukuran yang disukai yang tidak jelas, dan konsepsi biaya yang berfokus pada pembayar alih-alih berfokus pada pasien. Namun, implementasi area konvergensi ketiga ideologi ini menawarkan peluang konkret untuk mengubah paradigma sistem perawatan kesehatan yang ada yang menjadi sangat penting di dunia dengan usia rata-rata yang meningkat, begitu pula prevalensi penyakit kronis, termasuk kanker dan penyakit kardiovaskular, semua komorbiditas terkait, dan disabilitas di usia lanjut. Biaya perawatan

kesehatan menyerap proporsi yang signifikan dari produk domestik bruto (PDB) nasional secara global.

Rata-rata, negara-negara anggota Organisasi untuk Kerja Sama dan Pembangunan Ekonomi (OECD) diperkirakan telah menghabiskan 8,8% dari PDB untuk perawatan kesehatan pada tahun 2018, persentase yang kurang lebih tidak berubah sejak tahun 2013. Amerika Serikat adalah negara dengan pengeluaran kesehatan tertinggi, setara dengan 16,9% dari PDB-nya, diikuti oleh Swiss, dengan nilai 12,2%. Setelah Amerika Serikat dan Swiss, sekelompok negara berpenghasilan tinggi, termasuk Jerman, Prancis, Swedia, dan Jepang, semuanya menghabiskan hampir 11% dari PDB mereka untuk layanan kesehatan. Lebih lanjut, diperkirakan bahwa hingga 20% dari pengeluaran kesehatan terbuang sia-sia di negara-negara ini.

Ada semakin banyak bukti dan penerimaan bahwa pembiayaan layanan kesehatan harus difokuskan pada hasil daripada pada penggantian layanan yang diberikan, untuk mencapai alokasi sumber daya yang terbatas secara bijaksana. Pergeseran dari volume ke nilai ini membutuhkan perancangan, pengembangan, dan penerapan produk, layanan, dan solusi terintegrasi yang memberikan nilai dengan meningkatkan hasil pasien secara efisien dan efektif. Untuk menerapkan transformasi ini, akses ke data dalam jumlah besar dan sejumlah besar hasil tentang dampak intervensi klinis diperlukan. Meskipun banyak proyek telah berjalan dengan tujuan memperoleh data ini, investasi dan komitmen yang lebih besar dari pasien dan semua agen yang bekerja di bidang kesehatan masih dibutuhkan.

4.2 LAYANAN KESEHATAN BERBASIS DATA

"Data adalah minyak baru," klaim Clive Humby, seorang matematikawan dan wirausahawan ilmu data asal Inggris, pada tahun 2006. Namun, jika tidak dimurnikan, data tersebut tidak dapat benar-benar digunakan. Untuk menciptakan nilai, data harus dapat dipercaya, akurat, komprehensif, mudah diakses, dapat dibagikan, dan yang terpenting, dapat digunakan.

Seperti di banyak industri lainnya, sektor layanan kesehatan secara rutin menghasilkan data dalam jumlah besar dari berbagai sumber, mulai dari pemeriksaan biokimia, rekam medis elektronik, tanda-tanda vital, hasil yang dilaporkan pasien, survei kesehatan, uji klinis, klaim asuransi, data administratif, dan yang terbaru, omik (genomika, transkriptomik, proteomik, metabolomik, radiomik, mikrobiomik). Secara keseluruhan, kumpulan data ini mewakili koleksi data besar yang merupakan kunci untuk kualitas perawatan pasien yang lebih baik, mengurangi readmisi, mendukung pengambilan keputusan, dan peningkatan hasil secara keseluruhan. Namun, pengumpulan data besar tidak sama dengan layanan kesehatan berbasis data. Meskipun peningkatan dan ketersediaan data dapat mendorong era baru inovasi berbasis fakta dalam layanan kesehatan, otomatisasi diperlukan untuk menyederhanakan proses dan memperjelas pengambilan keputusan dengan cara yang meningkatkan hasil klinis dan kelincahan operasional. Dengan lingkungan regulasi yang semakin bergerak menuju perawatan yang berpusat pada pasien dan berbasis nilai, institusi layanan kesehatan harus

beralih dari pengumpulan data layanan kesehatan menjadi institusi layanan kesehatan berbasis data.

Menjadi institusi layanan kesehatan berbasis data membutuhkan investasi dan sumber daya baru yang memungkinkan anggota tim untuk membuat keputusan yang paling tepat dan organisasi untuk mencapai tujuannya. Manajemen kesehatan berbasis data baru harus digunakan dalam pengambilan keputusan klinis untuk meminimalkan risiko penyakit dan dampak kesehatan yang merugikan di masa mendatang, serta mendorong model perawatan yang berpusat pada pasien dan berbasis nilai. Untuk mencapai status baru ini, diperlukan infrastruktur data, proses berbasis data, budaya yang berpusat pada data, dan kerangka kerja keamanan siber.

Layanan kesehatan berbasis data dapat dipecah menjadi empat pilar berbeda: (1) penggunaan data oleh pasien, tenaga kesehatan profesional, dan organisasi; (2) regulasi data untuk memastikan akuntabilitas, privasi, dan keamanan; (3) teknologi dan metode komputasi yang membantu tenaga kesehatan profesional membuat keputusan berbasis data untuk meningkatkan luaran kesehatan; dan (4) inovasi yang didorong oleh data dan yang mendorong produksi data baru. Untuk meningkatkan penggunaan data oleh semua pemangku kepentingan, pasien, tenaga kesehatan profesional, dan organisasi, strategi layanan kesehatan berbasis data harus mengejar tujuan-tujuan berikut:

(1) promosi edukasi dan literasi untuk meningkatkan kesadaran akan relevansi layanan kesehatan berbasis data di kalangan pasien dan tenaga kesehatan profesional; (2) penciptaan sistem layanan kesehatan terpadu yang berfokus pada pasien dengan menyediakan akses data yang mudah dan luas; (3) promosi inisiatif manajemen dan investasi dalam tata kelola data untuk mengurangi resistensi terhadap perubahan dalam organisasi layanan kesehatan.

Saat ini, layanan kesehatan berbasis data memiliki budaya organisasi sebagai hambatan terbesarnya, yang mengatasi tantangan teknologi atau investasi. Akuntabilitas data, privasi, dan keamanan data dalam layanan kesehatan merupakan topik yang kompleks, yang bertujuan untuk memastikan pertukaran informasi pasien yang aman, melindungi integritas rekam medis dan aplikasi, serta mengendalikan akses ke aplikasi dan sistem layanan kesehatan yang berisi data pribadi.

Pemantauan dan penilaian kepatuhan regulasi untuk pemrosesan data sangat penting, baik di tingkat organisasi maupun personal. Strategi layanan kesehatan berbasis data harus mempertimbangkan definisi kebijakan manajemen data, menyediakan pelatihan bagi mereka yang menangani data kesehatan, dan mendukung penerapan sistem informasi "aman sejak awal". Lebih lanjut, transformasi menuju layanan kesehatan berbasis data tidak terbatas pada bagaimana data dikumpulkan, diproses, dianalisis, dan digunakan, tetapi juga melibatkan pengawasan dan tanggung jawab dari semua pemangku kepentingan, termasuk pemerintah, pasien, tenaga kesehatan profesional, penyedia layanan kesehatan, perusahaan asuransi, dan pemasok layanan kesehatan.

Teknologi yang ada saat ini memungkinkan tenaga kesehatan profesional dan pasien untuk terlibat secara lebih efektif bersama-sama guna meningkatkan hasil kesehatan. Meningkatnya pemantauan medis jarak jauh, dengan penggunaan perangkat seluler medis

untuk memantau kondisi pasien secara aktif, penggunaan penyimpanan dan aplikasi berbasis cloud yang memungkinkan komunikasi yang lebih baik dan meningkatkan pengalaman pasien, serta semakin diterimanya perangkat medis yang dapat dikenakan (wearable) menunjukkan bahwa teknologi kesehatan sedang menjadi komoditas.

Semua teknologi ini didukung oleh metode komputasi inovatif yang dikembangkan untuk melakukan pemantauan cerdas, mengembangkan model prediktif untuk pencegahan, deteksi dini penyakit, diagnosis, pengobatan, dan prognosis, serta menganalisis atau mengusulkan rencana pengobatan dan intervensi. Metode komputasi mampu mengekstraksi nilai dari big data, tetapi seringkali menghadapi tantangan besar seperti ukuran dan relevansi data masukan; aksesibilitas dan keterbacaan data; serta kurangnya interoperabilitas data. Agar data dapat ditindaklanjuti, data yang dikumpulkan harus bersih, lengkap, akurat, dan terstandarisasi untuk digunakan di berbagai sistem.

Data tersebut juga harus mudah diakses dan dibaca oleh berbagai pemangku kepentingan dengan peran yang berbeda, seperti komunitas ilmiah, badan regulasi, dan tenaga kesehatan profesional, untuk menjamin tinjauan sejawat. Tren yang sedang berlangsung ini memberikan landasan penting bagi inovasi generasi berikutnya, termasuk penggunaan kecerdasan buatan untuk pengembangan pengobatan yang tepat, preventif, dan personal, yang melibatkan kombinasi analisis data besar dan metode statistik, yang umumnya dikenal sebagai algoritma pembelajaran mesin.

4.3 TANTANGAN ETIKA DAN HUKUM YANG DIAJUKAN OLEH KECERDASAN BUATAN

Kecerdasan buatan (AI), yang dimulai pada Proyek Penelitian Musim Panas Dartmouth tentang Kecerdasan Buatan pada tahun 1956, belum memenuhi janji utamanya selama bertahun-tahun. Kemajuan terbaru dalam metode pembelajaran mesin, ketersediaan data besar, dan keberadaan infrastruktur superkomputer membantu AI memasuki transisi cepat dari teori ke realitas. Sebagai mesin data besar, kecerdasan buatan mempercepat implementasi layanan aplikasi data dalam.

Pembelajaran mendalam telah memberikan kontribusi signifikan terhadap kemajuan AI terkini. Dibandingkan dengan metode pembelajaran mesin tradisional, metode pembelajaran mendalam telah mencapai peningkatan substansial dalam berbagai tugas prediksi. Namun, metode baru ini relatif lemah dalam menjelaskan proses inferensi dan hasil akhirnya. Dalam banyak aplikasi dunia nyata seperti diagnosis medis, kemampuan menjelaskan dan transparansi menjadi sangat penting bagi pengguna yang terpengaruh oleh keputusan AI.

Selama dekade terakhir, AI, didukung oleh pembelajaran mendalam, secara bertahap telah diintegrasikan ke dalam praktik medis sehari-hari dan telah membuat kemajuan yang cukup besar dalam pemrosesan citra medis, optimasi proses medis, diagnosis medis, pendidikan kedokteran, dan aplikasi lainnya. Namun, implementasi skala besar dalam praktik klinis masih berjuang karena kurangnya proses standar, dan pengawasan etika dan hukum. Masalah utama implementasi AI dalam perawatan kesehatan terkait dengan sifat teknologi itu sendiri, kompleksitas dukungan hukum dalam hal keselamatan dan efisiensi, privasi, masalah

etika dan kewajiban. Penggunaan AI dalam praktik klinis memiliki potensi besar untuk mengubah cara pasien dirawat menjadi lebih baik, tetapi juga menimbulkan tantangan etika, yaitu: (1) persetujuan penggunaan berdasarkan informasi; (2) keamanan dan transparansi; dan (3) privasi data.

Menurut Wikipedia, "persetujuan berdasarkan informasi adalah prinsip dalam etika medis dan hukum medis yang menyatakan bahwa pasien harus memiliki informasi yang cukup sebelum membuat keputusan bebas mereka sendiri tentang perawatan medis mereka. Penyedia layanan kesehatan sering dianggap bertanggung jawab untuk memastikan bahwa persetujuan yang diberikan pasien berdasarkan informasi, dan persetujuan berdasarkan informasi dapat berlaku untuk intervensi perawatan kesehatan pada seseorang, melakukan beberapa bentuk penelitian pada seseorang, atau untuk mengungkapkan informasi seseorang."

Meskipun definisi ini jelas, terdapat situasi, seperti inisiatif pengurutan genom tingkat populasi, di mana persyaratan untuk persetujuan berdasarkan informasi belum didefinisikan dengan baik. Dalam konteks ini, implementasi persetujuan berdasarkan informasi sangat berbeda di antara inisiatif-inisiatif ini, dengan mempertimbangkan formulasi yang disebut sebagai persetujuan luas, persetujuan umum, dan persetujuan berlapis, antara lain. Strategi spesifik yang mengklaim memberikan informasi lengkap dan terus melibatkan partisipan disebut "persetujuan dinamis". Persetujuan dinamis didasarkan pada platform komunikasi personal yang dirancang untuk mendukung komunikasi dua arah yang berkelanjutan antara peneliti dan partisipan.

Meskipun konsensus sedang dibentuk seputar konsep "persetujuan dinamis", penggunaannya dalam skala besar masih menghadirkan beberapa kesulitan teknis yang perlu diatasi. Dalam konteks penggunaan AI yang aman dan transparan untuk membantu intervensi medis, terdapat sejumlah pertanyaan yang masih belum terjawab secara jelas. Pertanyaan-pertanyaan seperti: sejauh mana seorang tenaga kesehatan perlu mengungkapkan bahwa sistem AI akan digunakan untuk diagnosis atau rekomendasi perawatan? Haruskah seorang tenaga kesehatan mengungkapkan bahwa ia tidak dapat menjelaskan hasilnya secara lengkap? Seberapa besar transparansi yang dibutuhkan?

Bagaimana konteks baru ini berinteraksi dengan apa yang disebut "hak atas penjelasan" berdasarkan Peraturan Perlindungan Data Umum (GDPR) Uni Eropa? Pertanyaan-pertanyaan ini khususnya relevan dalam kasus-kasus di mana AI beroperasi menggunakan algoritma “kotak hitam”, yang mungkin dihasilkan dari teknik-teknik pembelajaran mesin yang tidak dapat ditafsirkan sehingga sangat sulit bagi para profesional kesehatan untuk memahaminya secara menyeluruh.

Dalam ranah aplikasi kesehatan AI, baik yang terhubung maupun tidak dengan sensor yang dapat dikenakan, persetujuan berdasarkan informasi (*informed consent*) diubah menjadi perjanjian pengguna. Pertanyaan yang masih perlu dijawab adalah: bagaimana perjanjian pengguna ini berhubungan dengan persetujuan berdasarkan informasi (*informed consent*)? Pertanyaan ini khususnya relevan karena sebagian besar perjanjian pengguna disetujui oleh individu tanpa dialog tatap muka. Selain itu, kebanyakan orang secara rutin mengabaikan

perjanjian pengguna atau tidak memahami ketentuan layanan dan semua kerumitan yang terkait dengan pembaruan perangkat lunak atau perangkat keras.

Menjawab pertanyaan-pertanyaan ini bukanlah hal yang mudah, dan menjadi lebih sulit ketika informasi dari aplikasi kesehatan AI yang berhadapan langsung dengan pasien dimasukkan kembali ke dalam pengambilan keputusan klinis. Keandalan, validitas, dan transparansi kumpulan data yang digunakan untuk melatih algoritma AI menjadikan keselamatan sebagai salah satu tantangan paling relevan dalam layanan kesehatan AI. Sudah menjadi pengetahuan umum bahwa setiap kumpulan data akan bias sampai batas tertentu berdasarkan gender, orientasi seksual, ras, faktor sosiologis, lingkungan, atau ekonomi. Karena model pembelajaran mesin belajar dari data yang dikumpulkan secara historis, populasi yang pernah mengalami bias manusia dan struktural di masa lalu rentan terhadap kerugian akibat prediksi yang salah.

Misalnya, penggunaan data dari Framingham Heart Study untuk memprediksi risiko kejadian kardiovaskular pada populasi non-kulit putih telah menghasilkan hasil yang bias, dengan estimasi risiko yang terlalu tinggi maupun terlalu rendah. Hingga saat ini, sebagian besar penelitian tentang pencegahan primer dan skor risiko penyakit kardiovaskular, seperti Framingham Risk Score dan SCORE Eropa, telah dikembangkan pada populasi yang sebagian besar berkulit putih. Algoritma apa pun yang dirancang untuk memprediksi hasil dari temuan genetik akan bias jika hanya ada sedikit (atau tidak ada) studi genetik pada populasi tertentu.

Transparansi harus menjadi topik utama yang harus dibahas oleh pengembang algoritma AI. Tenaga kesehatan dan pasien harus diinformasikan tentang jenis data yang digunakan dan apakah ada bias dalam sistem. Jika dalam dunia hipotetis dimungkinkan untuk membayangkan model data terbuka yang dapat divalidasi oleh semua orang, pada kenyataannya, terdapat beberapa keterbatasan dalam implementasi jenis ini. Model data terbuka dapat menimbulkan masalah kekayaan intelektual dan investasi, serta perlu menjaga kerahasiaan data pasien yang dipertimbangkan.

Alih-alih hanya melindungi dari kekurangan yang telah diidentifikasi sebelumnya, sistem AI harus digunakan secara proaktif untuk memajukan kesetaraan kesehatan. Agar realitas ini terwujud, prinsip-prinsip keadilan distributif harus diintegrasikan ke dalam perancangan, implementasi, dan evaluasi model. Terlepas dari tantangan yang ada saat ini, AI memiliki potensi luar biasa untuk mentransformasi penyediaan layanan kesehatan di lingkungan miskin sumber daya karena sistem ini dapat bertindak sebagai pengganti pakar manusia jika tidak tersedia, yang sering terjadi di komunitas miskin.

Mengingat informasi medis pribadi merupakan salah satu bentuk data yang paling privat dan dilindungi secara hukum, terdapat kekhawatiran yang signifikan tentang bagaimana akses, kontrol, dan penggunaan oleh pihak-pihak yang mencari keuntungan dapat berubah seiring waktu dengan AI yang dapat meningkatkan dirinya sendiri. Sebagai pemilik data, pasien berhak mengetahui bagaimana dan sejauh mana data kesehatan pribadi mereka dicatat dan digunakan. GDPR di semua Negara Anggota Uni Eropa telah diterapkan sejak 2018 dan memperkenalkan era baru hukum perlindungan data di Uni Eropa.

Peraturan ini khususnya bertujuan untuk melindungi hak perorangan atas perlindungan data pribadi. Peraturan di Kanada mengakui bahwa penyedia layanan kesehatan adalah "penjaga informasi" data kesehatan pribadi pasien, dan kepemilikannya adalah milik pasien. "Perwalian" ini mencerminkan kenyataan bahwa terdapat kepentingan dalam rekam medis pasien, dan kepentingan ini dilindungi oleh hukum. Saat ini, sebagian besar teknologi AI yang ada berada di tangan perusahaan teknologi besar. Google, IBM, Apple, Microsoft, dan perusahaan-perusahaan lain semuanya "mempersiapkan, dengan cara mereka sendiri, tawaran untuk masa depan kesehatan dan berbagai aspek industri layanan kesehatan global."

Konsentrasi inovasi dan pengetahuan teknologi di perusahaan-perusahaan teknologi besar menciptakan ketidakseimbangan kekuatan yang membuat institusi kesehatan publik menjadi semakin bergantung dan kurang mampu mempertahankan kemitraan yang seimbang dalam pengembangan dan implementasi baru. Meskipun beberapa pelanggaran privasi data telah terjadi meskipun terdapat undang-undang, peraturan, dan kebijakan privasi yang berlaku, jelas bahwa perlindungan yang memadai harus diterapkan untuk menjaga privasi dan agensi pasien dalam konteks kemitraan publik-swasta ini.

Perjanjian pembagian informasi, persetujuan berdasarkan informasi dinamis, dan mekanisme lainnya dapat digunakan untuk memberikan akses kepada institusi swasta ini terhadap informasi kesehatan pasien yang akan digunakan untuk melatih algoritma atau mengembangkan produk baru. Selain kemungkinan penyalahgunaan wewenang yang terkait dengan pengendalian data, AI menimbulkan tantangan baru terkait akses terhadap data pasien dalam jumlah besar. Lokasi dan kepemilikan server data yang menyimpan informasi pasien sangat relevan dalam konteks ini. Peraturan saat ini mewajibkan data pasien tetap berada di yurisdiksi asalnya, dengan sedikit pengecualian.

Dalam konteks saat ini, di mana telah terjadi banyak kasus penyalahgunaan oleh perusahaan besar dalam penggunaan data kesehatan pasien tanpa sepengetahuan mereka, tidak mengherankan jika muncul masalah kepercayaan publik. Kurangnya kepercayaan dari pihak pasien, warga negara, dan bahkan institusi kesehatan masyarakat dapat meningkatkan pengawasan publik atau bahkan litigasi terhadap implementasi komersial AI untuk layanan kesehatan.

Beberapa studi menunjukkan bahwa data terkait kesehatan jauh melampaui ekspektasi awal dari undang-undang perlindungan privasi. Misalnya, Aturan Privasi Undang-Undang Portabilitas dan Akuntabilitas Asuransi Kesehatan (HIPAA), yang disetujui oleh Kongres AS pada tahun 1996, memiliki celah yang signifikan dalam lingkungan layanan kesehatan saat ini karena hanya mencakup informasi kesehatan spesifik yang dihasilkan oleh "entitas yang tercakup" atau "rekan bisnis" mereka. (Pusat Pengendalian dan Pencegahan Penyakit 2018). Untungnya, banyak undang-undang baru telah diperkenalkan untuk mengatur perlindungan data AI, penentuan tanggung jawab, dan pengawasan.

Pada bulan Februari 2020, Komisi Eropa menerbitkan buku putih tentang AI. Dalam buku putih tersebut, Komisi menyoroti "Pendekatan Eropa" terhadap AI, menekankan bahwa "sangat penting bahwa AI Eropa didasarkan pada nilai-nilai dan hak-hak fundamental kita, seperti martabat manusia dan perlindungan privasi." (Komisi Eropa 2020). Pada bulan April

2021, sebuah proposal untuk Regulasi AI, "Undang-Undang Kecerdasan Buatan", diajukan. Peraturan ini akan mengatur penggunaan aplikasi AI "berisiko tinggi", yang mencakup sebagian besar aplikasi AI medis.

Pengawasan manusia, kemudahan dijelaskan, privasi sejak awal, dan non-diskriminasi merupakan kriteria utama AI Medis Eropa. Sebagaimana disyaratkan oleh hak-hak fundamental Uni Eropa (lihat juga Pasal 14 Undang-Undang AI yang diusulkan), keputusan AI medis memerlukan penilaian manusia sebelum tindakan apa pun diambil berdasarkan keputusan tersebut. Dengan kata lain, AI Medis Eropa secara hukum mensyaratkan keterlibatan manusia. Yang lebih penting lagi, AI Medis Eropa mensyaratkan persetujuan bebas dan berdasarkan informasi dari pasien. Hal ini menunjukkan adanya pemisahan pengambilan keputusan antara dokter dan pasien, di mana pasien memiliki keputusan akhir.

Poin tentang keputusan pasien yang berdasarkan informasi ini memperjelas bahwa hak-hak fundamental Eropa pada dasarnya mensyaratkan penggunaan AI yang dapat dijelaskan dalam dunia kedokteran. Akibatnya, AI Medis Eropa tidak boleh didasarkan pada "keputusan mesin", melainkan pada "keputusan, temuan diagnostik, atau proposal perawatan yang didukung AI". AI Eropa juga harus dikembangkan dan dioperasikan sesuai dengan persyaratan perlindungan data dan privasi, dan peraturan baru ini mencakup serangkaian ketentuan hak-hak fundamental yang mensyaratkan kesetaraan di hadapan hukum dan non-diskriminasi, termasuk gender, usia, dan tingkat disabilitas.

Meskipun komprehensif, kerangka hak asasi yang ada tidak menjawab semua pertanyaan hukum yang timbul terkait penggunaan AI medis. Sayangnya, terdapat pengecualian besar, yang melibatkan masalah legislasi pertanggungjawaban, yang menjadi perhatian khusus bagi komunitas pengembang perangkat berbasis AI. Siapa yang akan bertanggung jawab secara hukum ketika AI medis menyebabkan kerugian? Pengembang perangkat lunak, produsen, staf pemeliharaan, penyedia TI, rumah sakit, atau tenaga kesehatan profesional?

Terkait hal ini, Komisi mengumumkan bahwa mereka akan segera memberikan kejelasan tambahan, mengusulkan pendekatan pertanggungjawaban ketat yang dipadukan dengan skema asuransi kerusakan AI wajib. Terlepas dari pertanyaan yang masih terbuka, penting untuk dicatat bahwa persyaratan hukum untuk penggunaan AI medis saat ini sudah lebih jelas dibandingkan beberapa tahun yang lalu, jauh lebih jelas daripada yang tampaknya dipertimbangkan oleh komunitas pengembang perangkat berbasis AI.

Di AS, Badan Pengawas Obat dan Makanan (FDA) baru-baru ini merilis Rencana Aksi AI dan Pembelajaran Mesin (ML) pertamanya, sebuah pendekatan multi-langkah yang dirancang untuk memajukan pengelolaan perangkat lunak medis canggih oleh badan tersebut. Rencana aksi regulasi ini bertujuan untuk memaksa produsen perangkat lunak dan perangkat medis agar lebih ketat dalam penilaian keandalan dan keamanan mereka.

"Rencana Aksi Perangkat Lunak Berbasis AI/ML sebagai Alat Kesehatan" menguraikan lima tindakan yang akan diambil FDA, termasuk:

(1) Mengembangkan lebih lanjut kerangka regulasi yang diusulkan, termasuk melalui penerbitan draf panduan tentang rencana pengendalian perubahan yang telah ditentukan sebelumnya (untuk pembelajaran perangkat lunak dari waktu ke waktu);
(2) Mendukung pengembangan praktik pembelajaran mesin yang baik untuk mengevaluasi dan meningkatkan algoritma pembelajaran mesin;
(3) Membina pendekatan yang berpusat pada pasien, termasuk transparansi perangkat kepada pengguna;
(4) Mengembangkan metode untuk mengevaluasi dan meningkatkan algoritma pembelajaran mesin; dan
(5) Memajukan uji coba pemantauan kinerja di dunia nyata. (Badan Pengawas Obat dan Makanan AS 2021).

Banyak isu etika yang dibahas sebelumnya memiliki solusi hukum murni meskipun sulit untuk memisahkan aspek etika dari aspek hukum.

4.4 TREN INVESTASI DALAM KECERDASAN BUATAN DI SEKTOR KESEHATAN

Selama 5 tahun terakhir, industri kesehatan telah menjadi ajang pembuktian utama bagi aplikasi AI. Baik perusahaan rintisan maupun investor modal ventura (VC) menyadari potensi besar yang ditawarkan solusi AI untuk mempercepat rekrutmen uji klinis dan pengembangan vaksin, mendeteksi efek samping obat yang berpotensi mengancam jiwa, meningkatkan perawatan pasien, mengurangi waktu tunggu di rumah sakit dan kebutuhan sumber daya manusia, serta mengurangi biaya kesehatan.

Investor di seluruh dunia menyadari perlunya berinvestasi dalam AI di sektor kesehatan, dan investasi mereka pun beralih ke arah ini. Nilai AI di pasar kesehatan diperkirakan akan tumbuh dari Rp 69 triliun pada tahun 2021 menjadi Rp 674 triliun pada tahun 2027, dengan CAGR sebesar 46,2%. Pendanaan AI untuk kesehatan telah mencatat pertumbuhan rekor seiring dengan lonjakan permintaan. Pada kuartal pertama tahun 2021, setelah 111 transaksi, perusahaan AI di sektor kesehatan telah mengumpulkan modal senilai rekor Rp 25 triliun. Ini merupakan peningkatan sebesar 140% dibandingkan dengan Rp 10 triliun yang terkumpul pada Q1 2020. Investasi dan kasus penggunaan terbaru untuk teknologi ini merupakan bukti bahwa AI akan tetap ada.

Berbagai masalah yang dihadapi kebijakan kesehatan Eropa untuk mengatasi isu-isu seperti penuaan populasi dan pengurangan belanja publik untuk kesehatan, membutuhkan investasi modal di sektor ini. Meskipun investasi harus didorong di sektor kesehatan agar negara-negara dapat mempersiapkan diri dengan baik menghadapi kejadian tak terduga, seperti pandemi Covid-19, ada juga kebutuhan untuk mempromosikan investasi dalam perlindungan sosial dan mendorong gaya hidup sehat.

Ada kebutuhan mendesak untuk mengembangkan model kebijakan insentif yang berfungsi sebagai katalis bagi pertumbuhan sektor kesehatan. Pemerintah Eropa dapat memperoleh manfaat dari pengembalian sosial atas investasi modal VC dalam perawatan kesehatan, serta promosi gaya hidup yang lebih sehat karena pertumbuhan sektor kesehatan berdampak positif pada seluruh perekonomian. Peningkatan layanan kesehatan yang

berkelanjutan dengan fokus utama pada pencegahan penyakit merupakan tujuan yang diinginkan karena investasi dalam kesejahteraan umum suatu populasi berkontribusi pada pertumbuhan ekonomi.

Sebagaimana halnya semua industri, pandemi COVID-19 telah menciptakan krisis yang belum pernah terjadi sebelumnya bagi industri layanan kesehatan. Krisis ini telah mempercepat segmen kesehatan digital, seperti penyediaan layanan kesehatan virtual, dengan fokus yang kuat pada kesehatan mental dan kesejahteraan. Seiring dengan semakin lazimnya vaksin COVID-19, fokus bergeser ke dampak jangka panjang pandemi, yang mencakup tren di bidang kesehatan tertentu, termasuk telehealth, AI dalam layanan kesehatan, perangkat medis, kesehatan mental, kesehatan perempuan, omik, dan keamanan siber.

Selama tahun 2022, investasi dalam AI untuk layanan kesehatan kemungkinan besar akan berfokus terutama pada peningkatan efisiensi backend, seperti penggunaan otomatisasi proses robotik (RPA) untuk mendukung perawatan atau penggunaan AI untuk memproses klaim. Solusi RPA, ketika digunakan bersama AI, menciptakan otomatisasi cerdas, yang bertujuan untuk meniru interaksi manusia secara saksama, seringkali melalui penggunaan bot. Dalam jangka pendek, investasi AI kemungkinan besar akan berfokus pada pengembangan sistem pendukung diagnostik, penyediaan opini kedua, dan pendeteksian kesalahan rutin, alih-alih didedikasikan untuk melakukan diagnosis langsung secara otonom kepada pasien.

Di era di mana AI berkembang pesat di institusi layanan kesehatan dan masyarakat luas, tantangan terbesar dalam adopsi AI bukanlah apakah AI tersebut cukup mumpuni untuk bermanfaat, melainkan memastikan AI tersebut diadaptasi dengan baik ke dalam praktik klinis sehari-hari. Dalam penerapannya ke praktik klinis, terdapat lima topik yang perlu diperhatikan oleh perusahaan dan organisasi ketika berencana untuk bekerja dengan klien di bidang layanan kesehatan atau memulai perjalanan AI layanan kesehatan mereka sendiri. Kelima topik tersebut adalah: (1) teknologi AI mana yang lebih adaptif terhadap tujuan penggunaan; (2) siapa yang menjadi pengguna utama aplikasi AI; (3) perangkat lunak mana yang sebaiknya digunakan untuk membangun solusi AI (sumber terbuka versus komersial); (4) membangun atau membeli perangkat lunak; (5) bagaimana menjamin kepatuhan terhadap regulasi dan standar AI.

Para pemimpin teknis sejati menekankan bahwa organisasi mereka telah menggunakan integrasi data, yaitu pemrosesan bahasa alami (NLP) dan kecerdasan bisnis (BI), anotasi data, serta platform ilmu data. Kini, jelas bahwa organisasi layanan kesehatan yang paling inovatif pun serius dalam mengelola data mereka, memanfaatkan rekam medis elektronik dan teknologi seperti NLP, untuk mengintegrasikan data yang tersimpan di berbagai silo dan mendapatkan gambaran lengkap tentang kondisi proses organisasi dan perjalanan klinis pasien.

Di ranah pengguna, pasien dan dokter menjadi pengguna utama aplikasi AI. Penggunaan aplikasi AI sedang bergeser dari ilmuwan data dan tenaga teknis ke dokter dan pasien. Chatbot dan teknologi interaktif serta otomatisasi lainnya akan terus berkembang seiring dengan semakin matangnya AI. Pertumbuhan ini telah memberikan dampak yang

mendalam pada pemberian layanan pasien, karena menyediakan tingkat akses dan kemudahan yang memfasilitasi berbagai tugas, seperti menjadwalkan janji temu, mengakses rekam medis, dan bahkan mengelola layanan kesehatan jarak jauh.

Dalam hal menentukan perangkat lunak mana yang akan digunakan untuk membangun solusi AI, trennya adalah menggunakan penyedia sumber terbuka dan cloud publik. Tidak mengherankan bahwa solusi sumber terbuka lebih maju daripada penyedia cloud atau solusi komersial lainnya, karena privasi dan keamanan data merupakan beberapa tantangan utama dalam menggunakan layanan ini. Tantangan-tantangan ini khususnya relevan di sektor kesehatan, di mana undang-undang dan peraturan mungkin melarang pembagian data dengan pihak ketiga. Peraturan privasi layanan kesehatan, misalnya, mewajibkan pengguna untuk menghapus rekam medis dari informasi kesehatan yang dilindungi melalui proses yang disebut de-identifikasi.

Proses yang memakan waktu dan rumit ini kini dapat diotomatisasi secara menyeluruh. Perusahaan yang sudah mapan umumnya memilih untuk mengandalkan data dan alat pemantauan mereka sendiri daripada penilaian pihak ketiga atau representasi vendor perangkat lunak. Di sisi lain, perusahaan yang masih mulai mengeksplorasi AI lebih terbuka untuk menggunakan solusi dari vendor perangkat lunak. Menggunakan model yang disesuaikan dengan kebutuhan spesifik masing-masing perusahaan sekaligus menjaga data tetap berada di dalam organisasi merupakan langkah cerdas bagi mereka yang memiliki alat untuk mengelola upaya AI mereka secara internal.

Untuk industri yang sangat teregulasi, seperti layanan kesehatan dan farmasi, teknologi berbasis AI akan sangat penting bagi operasional dan keselamatan. Oleh karena itu, untuk meningkatkan skala penerapan dan penggunaan AI, organisasi harus menetapkan program manajemen kepatuhan untuk memenuhi persyaratan relevan dari aturan resmi AI yang berlaku. Di Eropa, kombinasi kerangka hukum AI dan rencana baru yang dikoordinasikan dengan Negara-negara Anggota bertujuan untuk menjamin keamanan dan hak-hak dasar masyarakat dan bisnis, sekaligus memperkuat penerimaan, investasi, dan inovasi AI. Pada tahap kematangannya saat ini, AI siap untuk mengubah industri layanan kesehatan dan ilmu hayati dengan cara yang tidak dapat kita bayangkan beberapa tahun yang lalu.

BAB 5
KEAMANAN DAN PRIVASI

Abstrak Keamanan komputer atau keamanan siber berkaitan dengan berfungsinya sistem komputer secara wajar meskipun ada tindakan musuh. Privasi berkaitan dengan kemampuan seseorang atau kelompok untuk mengendalikan bagaimana, kapan, dan sejauh mana informasi identitas pribadi mereka dibagikan. Bab ini dimulai dengan mendefinisikan keamanan dan privasi serta menjelaskan mengapa keduanya menjadi masalah. Kemudian, bab ini menyajikan beberapa pencapaian ilmiah dan teknologi di kedua bidang tersebut, dengan menyoroti beberapa tren penelitian. Setelah itu, bab ini menghubungkan keamanan dan privasi dengan topik utama buku ini: pembelajaran mesin sebagai bagian dari kecerdasan buatan. Terakhir, bab ini mengilustrasikan relevansi pembelajaran mesin di bidang ini dengan menggunakan contoh ketahanan sensor.

5.1 KEAMANAN SIBER VS PRIVASI: CIA TRIAD DAN GDPR

Keamanan komputer, juga disebut keamanan siber, berkaitan dengan berfungsinya sistem komputer secara wajar meskipun ada tindakan musuh (peretas, penjahat siber, dll.). Berfungsinya sistem ini dinyatakan dalam hal sifat-sifat seperti kerahasiaan, integritas, dan ketersediaan. Disiplin ini muncul pada akhir 1960an dalam konteks pertahanan AS dengan kekhawatiran tentang kerahasiaan informasi rahasia yang tersimpan di komputer. Sebuah laporan tahun 1970 dari Defense Science Board menyatakan bahwa:

> Dengan munculnya sistem komputer berbagi sumber daya yang mendistribusikan kapabilitas dan komponen konfigurasi mesin di antara beberapa pengguna atau beberapa tugas, dimensi baru telah ditambahkan ke masalah pengamanan informasi rahasia yang tersimpan di komputer.

Privasi adalah istilah lama dengan makna yang terkait tetapi berbeda. Privasi dapat didefinisikan sebagai kemampuan seseorang (atau sekelompok orang) untuk mengendalikan bagaimana, kapan, dan sejauh mana informasi identitas pribadinya dikomunikasikan kepada orang lain.

Definisi aslinya berbicara tentang informasi pribadi, tetapi saat ini istilah yang lebih luas, informasi identitas pribadi (PII), digunakan untuk menunjukkan bahwa beberapa data yang tidak sepenuhnya bersifat pribadi dapat digunakan secara langsung atau tidak langsung untuk mengidentifikasi seseorang (misalnya, alamat IP atau data tentang rekan kerja). Properti yang relevan meliputi anonimitas, ketidakobservabilitasan, dan kerahasiaan PII. Privasi dalam konteks sistem komputer juga muncul pada tahun 1960an atau 1970an. Bidang ini baru-baru ini mendapatkan banyak relevansi dengan Peraturan Perlindungan Data Umum Eropa (GDPR) dan undang-undang serupa di negara lain.

Keamanan dan privasi merupakan disiplin ilmu yang saling terkait erat. Privasi sampai batas tertentu berkaitan dengan keamanan, terutama kerahasiaan, dari pengenal pribadi dan informasi pribadi. Namun, privasi mencakup aspek-aspek yang hampir tidak terkait dengan

keamanan klasik. Dua contohnya adalah pengendalian pengungkapan statistik dan privasi diferensial. Konferensi ilmiah tertua, dan salah satu yang teratas, di bidang ini, menunjukkan hubungan antara kedua topik tersebut, dimulai dengan namanya: Simposium IEEE tentang Keamanan dan Privasi.

Perlu dicatat bagaimana keamanan dan privasi seharusnya disajikan saat ini. Secara tradisional, topik-topik ini disajikan secara negatif: hal-hal buruk dapat terjadi (dan memang terjadi seperti yang terlihat di berita) sehingga kita harus berjuang untuk mencegahnya. Namun, kami berpendapat bahwa hal-hal tersebut seharusnya disajikan secara positif: digitalisasi (transformasi digital) masyarakat kita mengharuskan orang dan organisasi untuk dapat menggunakan sistem berbasis komputer dengan tenang, tanpa kekhawatiran yang berlebihan tentang keamanan dan privasi. Inilah tujuan dari bidang ilmiah dan teknis keamanan dan privasi.

Dalam praktiknya, keamanan dan privasi tidak 100% dapat dicapai dalam lingkungan atau sistem tertentu. Hal ini tidak mengherankan, karena pencurian atau pembunuhan tidak pernah diberantas dalam masyarakat kita. Oleh karena itu, tujuannya bukanlah untuk mencapai keamanan atau privasi 100%, melainkan tingkat risiko yang memadai. Risiko memperhitungkan dua faktor: probabilitas suatu properti dilanggar dan dampak dari pelanggaran tersebut. Probabilitas tersebut bergantung pada tingkat kerentanan dan tingkat ancaman. Upaya untuk meningkatkan keamanan dan privasi sangat substansial, baik dari kalangan akademis maupun industri.

Saat ini, terdapat banyak jurnal dan konferensi akademis yang membahas hal ini, termasuk konferensi-konferensi terkemuka seperti Simposium Keamanan dan Privasi IEEE, Konferensi Keamanan Komputer dan Komunikasi ACM, Simposium Keamanan Jaringan dan Sistem Terdistribusi (NDSS), dan Keamanan Usenix. Industri di bidang ini juga besar. Misalnya, Gartner baru-baru ini memperkirakan pengeluaran sebesar Rp 1,504 kuadriliun untuk keamanan pada tahun 2021, dengan peningkatan sebesar 12,4% dibandingkan tahun 2020. Indikator lainnya adalah banyaknya pameran industri di seluruh dunia. Yang terbesar kemungkinan adalah Konferensi RSA, yang diselenggarakan di beberapa negara setiap tahun, dan menarik lebih dari 40.000 peserta hanya di Amerika Serikat.

Setelah lebih dari 50 tahun, keamanan dan privasi telah menjadi bidang penelitian yang luas, dengan banyak aspek. Oleh karena itu, bab ini memberikan ringkasan yang terbatas. Bab ini memberikan ikhtisar topik-topik penting dan perkembangan terkini, dengan fokus pada teknologi. Sudut pandang lain seperti tata kelola, hukum, manajemen risiko, operasi keamanan, manajemen insiden, dan forensik digital tidak dibahas.

5.2 MENDEFINISIKAN KEAMANAN DAN PRIVASI

Ungkapan seperti "sistem X aman" atau "sistem Y menjamin privasi pengguna" terlalu samar untuk menjadi relevan. Yang relevan adalah menyatakan serangkaian properti keamanan dan privasi mana yang dipenuhi suatu sistem jika diimplementasikan dan dikonfigurasi dengan benar, dengan mempertimbangkan serangkaian asumsi tentang

lingkungan (misalnya, daya komputasi musuh). Keamanan sering diungkapkan dalam istilah yang berasal dari kepercayaan.

Kepercayaan adalah ketergantungan yang diterima dari seseorang atau (sub)sistem pada serangkaian properti (sub)sistem lain. Properti ini dapat terdiri dari beberapa jenis, termasuk keamanan dan privasi. Kepercayaan suatu (sub)sistem adalah ukuran pemenuhan serangkaian properti tersebut.

Properti Keamanan

Tiga properti keamanan inti adalah kerahasiaan, integritas, dan ketersediaan (CIA):

- Kerahasiaan: tidak adanya pengungkapan data yang tidak sah;
- Integritas: tidak adanya modifikasi data atau (sub)sistem yang tidak sah;
- Ketersediaan: kesiapan suatu (sub)sistem untuk menyediakan layanannya.

Perhatikan dua aspek. Pertama, keamanan berkaitan dengan jaminan properti-properti ini di hadapan tindakan jahat dari penyerang. Hal ini diungkapkan dengan istilah tidak sah. Kedua, properti-properti ini terkait dengan dampak tindakan jahat terhadap data (atau informasi) dan (sub)sistem, tetapi tidak harus pada keduanya. Kerahasiaan berkaitan dengan data, ketersediaan berkaitan dengan (sub)sistem, dan integritas berkaitan dengan keduanya.

Dua properti lain yang dapat dianggap terkait dengan integritas:

- Keaslian: tidak adanya modifikasi yang tidak sah atas konten atau informasi tentang sumbernya;
- Non-penyangkalan: tidak adanya penyangkalan atas kepengarangan data atau tindakan.

Properti terakhir yang baru-baru ini semakin terlihat dengan munculnya Bitcoin dan blockchain serta buku besar terdistribusi lainnya:

- Desentralisasi: tidak adanya ketergantungan pada otoritas pusat yang tepercaya.

Desentralisasi tidak menghilangkan kebutuhan akan kepercayaan, tetapi menggantikan kepercayaan pada masing-masing pihak ketiga dengan kepercayaan pada sekelompok pihak.

Properti Privasi

Properti privasi berlaku untuk sistem yang memproses informasi identitas pribadi (PII). Pertimbangkan properti tambahan berikut:

- Ketidaktertautan: mengingat eksekusi sistem tertentu, sekumpulan item PII yang tidak dapat ditautkan tidak boleh lebih atau kurang terkait sebelum dan sesudah eksekusi tersebut.

Berlawanan dengan apa yang terjadi dengan keamanan, terdapat beberapa interseksi antara properti privasi klasik. Pertimbangkan pengidentifikasi seseorang (ID) dan sekumpulan item PII yang ditetapkan sebagai item yang diminati (IOI). Privasi dapat dinyatakan dalam properti berikut:

- Anonimitas: seseorang tidak dapat diidentifikasi dalam sekumpulan orang (kumpulan anonimitas), yaitu, ketidaktertautan sekumpulan IOI dan ID;
- Ketidaktertangkap: IOI tidak dapat dibedakan dari IOI mana pun.

Jika ketidaktertangkap dijamin, maka anonimitas juga dijamin, tetapi kebalikannya tidak berlaku. Istilah terkait lainnya adalah pseudonimitas, yang berarti penggunaan nama samaran

sebagai identitas pribadi. Namun, pseudonimitas bukanlah properti, melainkan mekanisme untuk mendapatkan privasi.

GDPR menetapkan serangkaian hak pengguna di hadapan pengendali dan pemroses data. Hak-hak ini mewujudkan "kemampuan untuk mengendalikan (...) PII" yang tercantum dalam definisi privasi kami. Oleh karena itu, kami menyatakannya sebagai properti (nama-nama tersebut adalah milik kami), merujuk pada pasal GDPR yang mendefinisikannya:

- Aksesibilitas: kemampuan untuk memperoleh informasi tentang PII yang sedang diproses (Pasal 15);
- Dapat diperbaiki: kemampuan untuk mengoreksi PII yang tidak akurat (Pasal 16);
- Dapat dihapus: kemampuan untuk menghapus PII (Pasal 17);
- Dapat dibatasi: kemampuan untuk membatasi cara pemrosesan PII (Pasal 18);
- Portabilitas: kemampuan untuk memperoleh salinan PII yang sedang diproses (Pasal 20);
- Penarikan: kemampuan untuk menarik persetujuan untuk memproses data (Pasal 7) atau untuk menolak pemrosesan tersebut (Pasal 21).

5.3 MASALAH KEAMANAN DAN PRIVASI

Bagian sebelumnya menyatakan bahwa terdapat masalah keamanan dan privasi yang harus dipecahkan. Bagian ini menyajikan masalah-masalah tersebut secara lebih rinci.

Kontrol Akses

Semua masalah keamanan dan privasi berkaitan dengan akses. Misalnya, motivasi awal untuk keamanan adalah akses bersama ke komputer yang berisi informasi rahasia (militer). Contoh lain: masalah keamanan siber saat ini sebagian besar berasal dari konektivitas universal yang disediakan oleh Internet. Pendekatan untuk mengelola akses ada dua. Pertama, melibatkan pemisahan (atau isolasi), yang dapat bersifat logis, kriptografi, fisik, atau temporal. Kedua, akses harus diberikan atau ditolak, mengikuti beberapa kebijakan keamanan.

Di sinilah konteks di mana kontrol akses berperan. Secara abstrak, terdapat objek (atau sumber daya) yang diakses oleh subjek (pengguna, proses). Subjek dan objek dipisahkan secara logis, yaitu, mereka diisolasi satu sama lain menggunakan mekanisme perangkat lunak dan/atau perangkat keras.

Kontrol akses berkaitan dengan validasi izin akses (atau hak) subjek terhadap objek. Kontrol akses dilakukan oleh komponen abstrak yang disebut monitor referensi. Setiap kali subjek ingin mengakses suatu objek (misalnya, pengguna ingin mengakses berkas dalam layanan daring), monitor referensi menggunakan basis data kontrol akses untuk mendapatkan informasi tentang izin dan mengevaluasi apakah pengguna akan diberikan akses atau tidak. Monitor referensi mengambil keputusan dan secara opsional menyimpan data tentang akses tersebut untuk keperluan audit.

Model kontrol akses yang paling umum adalah Daftar Kontrol Akses (ACL). Setiap objek (misalnya, berkas) memiliki ACL yang mencantumkan izin (misalnya, baca, baca-dan-tulis) setiap pengguna (misalnya, pengguna, admin, apa pun) atas objek tersebut. Ada banyak modul lain, misalnya, kapabilitas, Kontrol Akses Berbasis Peran (RBAC), atau standar de facto saat ini,

Kontrol Akses Berbasis Atribut (ABAC). Kontrol akses dapat bersifat diskresioner kebijakan akses ditentukan oleh pemilik objek atau wajib kebijakan akses ditentukan oleh administrator untuk suatu kelas objek.

Kerentanan dan Serangan

Kontrol akses seharusnya efektif jika diimplementasikan dan dikonfigurasi dengan benar yang merupakan tantangan tersendiri tetapi seringkali ada masalah lain yang memungkinkan pengabaian kontrol akses: kerentanan. Sistem komputer itu kompleks, bisa dibilang ciptaan manusia yang paling kompleks. Laptop tempat teks ini ditulis adalah ciptaan ribuan insinyur di seluruh dunia, yang tidak saling mengenal atau sepenuhnya memahami keseluruhan sistem: laptop dalam hal ini. Kompleksitas ini tentu saja menyebabkan kesalahan karena insinyur adalah manusia.

Kerentanan adalah kesalahan yang memungkinkan pelanggaran properti keamanan. Kerentanan dapat muncul selama perancangan, implementasi, atau konfigurasi sistem. Kerentanan dapat dieksploitasi oleh serangan. Jika serangan berhasil, properti keamanan dilanggar. Kerentanan sangat penting sehingga dikatalogkan. Katalog kerentanan yang paling penting adalah Kerentanan dan Paparan Umum (CVE). Misalnya, pada tahun 2014 terdapat kerentanan yang menyebabkan banyak kekacauan yang disebut Heartbleed; menerima pengenal CVE-2014-0160 dalam katalog tersebut (yang ke-160 atau 2014). CVE telah mencatat sekitar 17.000–18.000 kerentanan baru setiap tahunnya dalam beberapa tahun terakhir.

Kelas kerentanan yang sangat berbahaya disebut kerentanan zero-day. Ini adalah kerentanan yang diketahui oleh satu atau lebih kelompok misalnya, badan intelijen atau komunitas peretas tetapi belum diungkapkan kepada publik. Kerentanan ini memungkinkan kelompok-kelompok ini untuk menyerang sistem secara bebas, karena tidak ada perlindungan yang diterapkan terhadap sesuatu yang tidak diketahui.

Kerentanan dapat diungkapkan kepada publik dengan berbagai cara: sebagai bagian dari pembaruan vendor perangkat lunak, dalam katalog CVE, dalam milis seperti Bugtraq, dll. Ketika itu terjadi, ada peluang bagi organisasi dan individu yang menggunakan perangkat lunak yang rentan untuk memperbaikinya atau melindunginya, tetapi pengungkapan ini juga sangat meningkatkan kemungkinan kerentanan tersebut diserang.

Serangan dapat datang melalui berbagai vektor. Vektor yang umum saat ini disebut unduhan drive-by. Korban mengakses halaman web dengan peramban yang memiliki kerentanan. Situs tersebut meluncurkan serangan terhadap peramban, mengeksploitasi kerentanan tersebut. Perlu dicatat bahwa situs tersebut mungkin sah tetapi pernah menjadi korban serangan.

Malware

Banyak serangan melibatkan malware. Istilah ini merangkum banyak serangan lain yang sebelumnya digunakan dengan cara yang seringkali tidak konsisten: virus, worm, Trojan horse, backdoor, dll. Malware berasal dari perangkat lunak berbahaya dan mencakup semua varian ini. Salah satu bentuk malware yang paling aktif saat ini adalah ransomware. Ketika spesimen ransomware memasuki komputer, ia mengenkripsi isi disk dan meminta pembayaran tebusan biaya moneter seringkali dalam mata uang kripto seperti Bitcoin atau Monero, untuk

menyembunyikan identitas penyerang. Tebusan ini bervariasi dari ratusan hingga jutaan euro, misalnya, seperti dalam serangan besar-besaran terhadap Maersk dan Colonial Pipeline. Beberapa serangan ini juga melibatkan pencurian data untuk menekan korban agar membayar.

Bentuk malware penting lainnya adalah *Remote Access Trojan* (RAT) atau bot, yang bersembunyi di komputer dan tetap aktif hingga diperintahkan untuk melakukan suatu tindakan. RAT atau bot ini dikendalikan dari jarak jauh oleh server pusat, membentuk botnet (jaringan bot atau robot). Botnet ini dapat memiliki ribuan komputer dan beroperasi sebagai senjata siber, yang mampu melumpuhkan sistem menggunakan serangan *Distributed Denial of Service* (DDoS) atau mencuri data atau kredensial akses dalam jumlah besar, di antara serangan lainnya.

Faktor Manusia

Pentingnya serangan otomatis yang mengeksploitasi kerentanan dan/atau menggunakan malware tidak dapat disangkal karena terus terjadi. Namun, banyak serangan mengeksploitasi jenis kelemahan lain yang jauh lebih sulit dikelola: manusia yang menggunakan sistem komputer. Serangan ini sering disebut rekayasa sosial.

Phishing adalah serangan umum yang bertujuan mencuri data pribadi. Data ini sering kali berupa kredensial pengguna untuk sistem seperti email perusahaan atau homebanking. Serangan ini hanya berupa pengiriman email yang meminta data. Tidak ada kerentanan teknis yang terlibat: data dicuri karena pengguna memercayai pesan yang mereka terima dan bertindak sesuai dengannya. Mengingat jumlah calon korban yang cukup besar, akan selalu ada banyak yang tertipu.

Salah satu bentuk serangan yang menggabungkan aspek teknis dan rekayasa sosial adalah email dengan lampiran berbahaya. Korban manusia tertipu dengan membuka lampiran, tetapi lampiran ini berisi serangan yang mencoba menjelajahi kerentanan di komputer korban. Ini adalah vektor serangan pertama yang menggunakan serangan Wannacry yang mencolok; ada yang kedua yang melibatkan eksploitasi kerentanan di komputer lain dalam organisasi yang sama.

5.4 PENCAPAIAN ILMIAH DAN TEKNOLOGI

Bagian ini menyajikan ringkasan pencapaian ilmiah dan teknologi yang penting dan/atau menarik di bidang keamanan dan privasi. Pencapaian ini disusun berdasarkan subbidang, yang secara umum terinspirasi oleh Badan Pengetahuan Keamanan Siber (CYBOK), "sebuah Badan Pengetahuan yang komprehensif untuk menginformasikan dan mendukung pelatihan pendidikan dan profesional bagi sektor keamanan siber". Untuk setiap topik, kami menyajikan tren penelitian.

Kriptografi

Kriptografi adalah disiplin ilmu yang sudah tua, sekitar 4000 tahun usianya menurut Kahn. Namun, hingga abad ke-20, evolusinya terbatas dan sebagian besar tetap menjadi seni: sebuah pertarungan antara kriptografer yang merancang skema pengkodean untuk melindungi kerahasiaan pesan, dan kriptanalis yang mencoba memecahkannya.

Skema-skema ini cerdik namun sederhana, sehingga seringkali berhasil dipecahkan; Mary, Ratu Skotlandia, dipenggal ketika pesan terenkripsi yang ia tukarkan dengan para pendukungnya didekodekan. Abad ke-20 pertama kali memperkenalkan otomatisasi dengan mesin seperti Enigma milik Jerman, yang digunakan dalam Perang Dunia II, tetapi, yang terpenting, merevolusi bidang ini dengan lonjakan komputasi dan kriptografi kunci publik. Saat ini, bidang ini merupakan bidang penelitian yang menarik dengan beberapa konferensi terkemuka, misalnya, Konferensi Kriptologi Internasional Tahunan.

Catatan peringatan: terdapat beberapa kebingungan mengenai hubungan antara keamanan siber dan kriptografi. Beberapa orang tampaknya menyederhanakan yang pertama menjadi yang kedua. Faktanya, meskipun kriptografi memainkan peran penting dalam keamanan siber, keamanan siber mencakup banyak topik yang tidak terkait dengan kriptografi, termasuk sebagian besar area yang dibahas di bagian berikut.

Kriptografi klasik melibatkan algoritma enkripsi dan dekripsi. Algoritma ini dirahasiakan untuk menjamin bahwa penyerang tidak dapat membaca pesan terenkripsi. Pada abad yang lalu, ide ini sedikit dimodifikasi menjadi apa yang sekarang kita sebut enkripsi simetris: kedua algoritma yang sama dapat dikonfigurasi dengan angka disebut kunci yang harus dirahasiakan, sedangkan algoritmanya harus bersifat publik agar dapat diteliti. Hal ini menyebabkan munculnya algoritma yang diadopsi secara luas seperti Standar Enkripsi Data (DES) tahun 1976, yang tidak lagi dianggap aman, dan Standar Enkripsi Lanjutan (AES) saat ini, yang diterbitkan pada tahun 1998.

Saat ini, skema semacam itu tidak hanya digunakan untuk melindungi kerahasiaan komunikasi tetapi juga bentuk data lainnya, seperti isi disk atau berkas individual. Enkripsi simetris menimbulkan kesulitan: distribusi kunci rahasia, yaitu, mengirimkannya kepada pihak-pihak yang membutuhkannya (misalnya, pengirim dan penerima). Solusi untuk masalah ini kriptografi kunci publik (atau kriptografi asimetris) akhirnya dipublikasikan pada tahun 1976 dan algoritma enkripsi kunci publik pertama, RSA, pada tahun 1978.

Dalam bentuk kriptografi ini, tidak ada lagi kunci tunggal, melainkan pasangan kunci. Pasangan ini memiliki kunci privat yang seharusnya tetap rahasia dan kunci publik yang dapat, dan seringkali harus, diungkapkan kepada publik. Jika data dienkripsi dengan kunci publik (atau privat), data tersebut hanya dapat didekripsi dengan kunci privat (atau publik). Oleh karena itu, jika Alice ingin berbagi kunci rahasia K untuk melindungi komunikasinya dengan Bob, ia dapat mengenkripsi K dengan kunci publik Bob, sehingga hanya Bob yang dapat mendekripsinya dan mendapatkan K, meskipun Trudy memperoleh salinan K yang terenkripsi.

Kriptografi kunci publik memiliki aplikasi penting kedua: memastikan integritas data menggunakan tanda tangan digital. Alice dapat menandatangani pesan atau dokumen menggunakan kunci privatnya, sehingga siapa pun yang memiliki kunci publiknya dapat memverifikasi apakah tanda tangan tersebut miliknya. Kunci publik biasanya didistribusikan menggunakan sertifikat yang berisi identifikasi pemilik kunci, kunci publik, dan tanda tangan yang dibuat oleh Otoritas Sertifikat (CA), di antara informasi lainnya.

Tanda tangan juga didasarkan pada jenis algoritma kriptografi penting lainnya: fungsi hash kriptografi. Fungsi-fungsi ini bersifat satu arah dan menghasilkan keluaran dengan

panjang tetap (misalnya, 256 bit) yang disebut hash. Selain itu, fungsi-fungsi ini memenuhi sifat ketahanan benturan seperti "secara komputasi tidak layak untuk menemukan dua masukan berbeda yang menghasilkan keluaran yang sama". Tren penelitian terkini di bidang ini adalah kriptografi pasca-kuantum (atau resistensi kuantum).

Pada tahun 1995, Shor menerbitkan sebuah algoritma yang pada dasarnya memungkinkan perolehan kunci privat dari kunci publik, yang secara efektif mematahkan algoritma kriptografi kunci publik seperti RSA dan ECDSA. Algoritma ini hanya dapat dieksekusi di komputer kuantum yang belum ada, tetapi mungkin akan muncul dalam beberapa tahun. Algoritma kuantum terhadap skema kriptografi lainnya kemudian dipresentasikan oleh Grover dan Simon. Akibatnya, terdapat upaya penelitian yang besar terhadap algoritma yang tetap aman jika atau ketika hal itu akhirnya terjadi. RSA didasarkan pada kesulitan memfaktorkan bilangan bulat besar dan algoritma Shor memungkinkan dilakukannya faktorisasi tersebut secara efisien dalam komputer kuantum; pendekatan utama untuk kriptografi pasca-kuantum adalah dengan menggunakan berbagai masalah sulit yang (bisa dibilang) tidak dapat diserang dalam komputer kuantum, misalnya, masalah Modul Pembelajaran dengan Kesalahan (MLWE) atau Modul Pembelajaran dengan Pembulatan (MLWR).

Keamanan Berbasis Perangkat Keras

Sebagaimana dijelaskan di atas, masalah keamanan dan privasi berkaitan dengan akses. Pencegahan akses sewenang-wenang di dalam komputer merupakan masalah penting karena pengguna jahat dan malware selalu berpotensi menjadi ancaman. Solusi untuk masalah ini melibatkan dukungan perangkat keras. Hingga tahun 1980an, dukungan ini sudah diperlukan untuk sistem operasi (OS), misalnya, perlindungan memori (berbasis, misalnya, on-page) yang memungkinkan proses (program yang sedang dieksekusi) diisolasi satu sama lain. Perlindungan memori diimplementasikan oleh komponen perangkat keras (CPU, MMU), dan perangkat lunak (OS).

Contoh kedua dari dukungan semacam ini adalah mode eksekusi CPU. Konfigurasi yang umum adalah menjalankan OS dalam mode kernel dan proses pengguna dalam mode pengguna. Dalam mode pengguna, CPU tidak mengizinkan eksekusi beberapa instruksi: CPU menghasilkan pengecualian atau diam-diam tidak melakukan apa pun ketika suatu program mencoba mengeksekusinya. Misalnya, proses dalam mode pengguna tidak diizinkan untuk mengeksekusi instruksi I/O secara langsung; mereka harus mendelegasikan eksekusinya ke OS menggunakan panggilan sistem. Pemisahan menjadi bagian tepercaya OS dan bagian tidak tepercaya proses pengguna masuk akal, tetapi OS terlalu besar dan seringkali terdapat kerentanan di sana.

Pada awal tahun 2000an, sebuah konsorsium perusahaan, *Trusted Computing Group* (TCG), merancang sebuah modul perangkat keras (biasanya sebuah chip) untuk disertakan dalam komputer pribadi. *Trusted Platform Module* (TPM) ini menyediakan serangkaian layanan keamanan yang jelas berbeda dari mekanisme keamanan perangkat keras yang lebih lama.

Layanan ini sebagian besar berupa penyimpanan kunci kriptografi dan verifikasi integritas perangkat lunak. Penyimpanan kunci saat ini banyak digunakan dalam berbagai bentuk pada perangkat seluler. Penyimpanan kunci mendukung atestasi jarak jauh dan juga

tersedia dalam berbagai bentuk saat ini, meskipun belum banyak digunakan. Kemudian, layanan ini sedikit diperluas dalam spesifikasi TPM 2.0.

Keterbatasan TPM adalah layanan yang disediakannya bersifat tetap. *Trusted Execution Environment* (TEE) merupakan solusi untuk keterbatasan ini karena dapat diprogram dengan perangkat lunak pengguna. Perangkat lunak ini dijalankan di TEE, terisolasi dari OS dan perangkat lunak istimewa lainnya (misalnya, BIOS dan hypervisor). Saat ini terdapat serangkaian teknologi TEE yang tersedia di komputer umum dan perangkat seluler: teknologi ini tidak didukung oleh modul perangkat keras, melainkan oleh CPU itu sendiri. TrustZone adalah ekstensi yang tersedia di banyak prosesor ARM. Dengan teknologi ini, terdapat dunia normal tempat OS dan proses pengguna dijalankan, tetapi juga dunia aman TEE tempat layanan keamanan dijalankan di atas kernel kecil. Hal ini memungkinkan terjaminnya kerahasiaan dan integritas isi TEE.

Intel mengembangkan *Software Guard Extensions* (SGX) untuk CPU mereka. SGX memungkinkan menjalankan beberapa TEE pada setiap CPU, yang disebut enklave. Selain jaminan yang diberikan oleh TEE TrustZone, enklave dan datanya dienkripsi saat tidak dijalankan, sehingga terisolasi secara logis dan kriptografis dari OS dan bagian komputer lainnya. AMD menyertakan *Secure Encrypted Virtualization* (SEV) dalam prosesor mereka yang mengubah hypervisor dan setiap Mesin Virtual (VM) yang dijalankannya menjadi TEE.

SEV mengenkripsi setiap TEE ini dengan kuncinya sendiri, mengisolasinya. Generasi kedua dari teknologi ini, *AMD SEV Encrypted State* (SEV-ES), juga mengenkripsi konten register CPU ketika VM berhenti berjalan. Generasi ketiga SEV, yang masih akan muncul, *AMD SEV Secure Nested Paging* (SEV-SNP), akan memberikan perlindungan integritas lebih lanjut. Pada tahun 2020, AMD, IBM, dan lainnya membentuk *Confidential Computing Consortium* (CCC) untuk mempromosikan adopsi teknologi ini. Tren penelitian utama di bidang ini adalah aplikasi untuk TEE. Teknologi ini diadopsi di berbagai bidang, mulai dari SGX dalam blockchain hingga TrustZone dalam perangkat Internet of Things.

Komputasi Awan

Komputasi awan adalah model di mana komputasi disediakan sebagai layanan. Dalam pengaturan umum, terdapat perusahaan Penyedia Layanan Awan (CSP) yang menyediakan layanan dengan metode bayar per penggunaan, yaitu konsumen membayar layanan yang dikonsumsinya. Hal ini berbeda dengan model klasik di mana konsumen membeli perangkat keras terlebih dahulu, lalu menggunakannya. Sebaliknya, dengan komputasi awan, konsumen menggunakan sumber daya sebanyak yang dibutuhkan, selama periode yang dibutuhkan, tanpa investasi awal. Adaptasi sumber daya yang digunakan terhadap kebutuhan konsumen ini sering disebut elastisitas.

Terdapat tiga model layanan utama: Infrastruktur sebagai Layanan (IaaS), di mana CSP menyediakan VM, jaringan, dan penyimpanan; Platform sebagai Layanan (PaaS), di mana CSP menyediakan komponen untuk membangun dan menjalankan aplikasi; dan Perangkat Lunak sebagai Layanan (SaaS), di mana CSP menyediakan aplikasi. Pendekatan pendelegasian perangkat lunak dan data kepada pihak ketiga, yaitu Penyedia Layanan Cloud (CSP), memiliki risiko: kehilangan data, pencurian data, orang dalam yang berniat jahat, kesalahan konfigurasi,

dll. Mengelola masalah ini memerlukan proses holistik untuk membangun dan mengonfigurasi perangkat lunak, yang mencakup serangkaian solusi teknologi yang luas. Topik ini terlalu luas untuk dibahas dalam satu bab, sehingga kami berfokus pada beberapa contoh mekanisme canggih.

Banyak konsumen menggunakan layanan cloud untuk menyimpan data. Jika mereka menganggap satu CSP tidak memberikan tingkat ketersediaan, integritas, atau jaminan kerahasiaan yang memadai, mereka dapat menggunakan serangkaian CSP yang membentuk cloud-of-cloud. Idenya adalah menyimpan data (berkas) di beberapa cloud, dilindungi menggunakan enkripsi dan tanda tangan digital, dengan kunci rahasia yang dilindungi menggunakan pembagian rahasia, dan menerapkan kode penghapusan untuk mengurangi total ukuran data yang disimpan.

Hal ini memerlukan replikasi Bizantium yang toleran terhadap kesalahan, sehingga protokol unggah dan unduh lebih canggih daripada sekadar menyimpan salinan berkas di beberapa tempat. Konsumen lain mungkin menerapkan aplikasi pada layanan PaaS, tetapi khawatir bahwa aplikasi tersebut diserang dan statusnya (data) diubah. Hal ini dapat terjadi, misalnya, jika seseorang mencuri kredensial dari pengguna yang sah dan menggunakannya untuk mengakses aplikasi. Menghapus dampak tindakan tersebut dari penyimpanan aplikasi (basis data, sistem berkas) seringkali dilakukan secara manual.

Salah satu solusinya adalah memodifikasi aplikasi untuk melacak dampak permintaan yang diterimanya pada penyimpanan tersebut, tetapi memodifikasi aplikasi dapat menjadi rumit atau bahkan mustahil. Sanare mendukung pemulihan otomatis dan menghilangkan kebutuhan untuk memodifikasi aplikasi dengan menggunakan pembelajaran mesin untuk mengaitkan permintaan aplikasi dengan perintah penyimpanan. Algoritme asosiasi Matchare didasarkan pada Jaringan Saraf Tiruan Konvolusional Dalam (CNN) dan memerlukan fase pelatihan.

Uang Digital, Aset, dan Identitas

Uang digital telah ada selama bertahun-tahun, tetapi baru-baru ini mendapatkan perhatian global berkat Bitcoin dan ribuan mata uang kripto lainnya. Kenyataannya, sebagian besar uang yang digunakan saat ini berbentuk digital: sebagian besar pembayaran dan transfer dilakukan menggunakan komputer, bukan kertas atau logam.

Pada tahun 1980an, beberapa penulis memperkenalkan skema kriptografi canggih untuk uang digital dan pembayaran yang membuka jalan bagi mata uang kripto saat ini. Sebuah karya penting oleh Chaum memperkenalkan mekanisme pembayaran yang mencegah entitas yang tidak terlibat dalam pembayaran untuk menentukan penerima pembayaran, waktu, dan jumlah yang ditransfer, sekaligus memungkinkan pembayar untuk memberikan bukti pembayaran dan menonaktifkan media pembayaran yang dilaporkan dicuri. Dalam karya berikutnya, Chaum dkk. memperkenalkan skema uang digital, yang memungkinkan pembayaran luring sekaligus memberikan bukti jika pengguna menggunakan uang yang sama dua kali.

Pada tahun 2008, Nakamoto memperkenalkan mata uang kripto pertama, Bitcoin. Berkaitan dengan karya sebelumnya, skema ini memperkenalkan dua perbedaan. Pertama, ia

tidak bergantung pada pihak ketiga, melainkan pada kelompok entitas ad hoc terbuka yang menjalankan perangkat lunak Bitcoin di komputer mereka (node). Hal ini juga memastikan ketersediaan. Kedua, ia mencegah pengeluaran ganda uang menggunakan rantai blok (blockchain) untuk mencatat transaksi, sementara node hanya menerima blok dengan transaksi yang valid. Buku besar digital ini, atau blockchain, direplikasi di semua node dan berkembang ketika ada konsensus pada blok berikutnya yang akan ditambahkan. Setiap blok berisi hash dari blok sebelumnya, yang diperoleh dengan memecahkan teka-teki kriptografi, yang menyulitkan modifikasi blockchain (memastikan integritas) dan memungkinkan penyelesaian konsensus.

Akhirnya, banyak mata uang kripto lain muncul, tetapi juga memungkinkan untuk menjalankan program pengguna di node, yang sering disebut kontrak pintar. Kemampuan pemrograman sistem ini tidak hanya memungkinkan penciptaan mata uang kripto lain, tetapi juga aset digital (atau token) yang dapat ditransaksikan. Salah satu jenis aset digital yang semakin populer adalah *Token Non-Fungible* (NFT). Mereka mewakili beberapa item yang dapat dikoleksi, seperti gambar digital, rekaman musik, dll.

Tren terbaru lainnya di bidang ini adalah penggunaan blockchain untuk menyimpan data identitas. W3C mendefinisikan konsep *Decentralized Identifier* (DID). Gagasan utamanya adalah memungkinkan pengendali DID (misalnya, orang dengan identitas tersebut) untuk mengendalikan DID, alih-alih bergantung pada entitas terpusat (misalnya, registri nasional atau perusahaan seperti Google atau Facebook). DID dapat disimpan dalam blockchain atau buku besar terdistribusi.

Secara teknis, DID adalah dokumen digital kecil yang berisi data identifikasi, misalnya, nama dan kunci publik. W3C juga mendefinisikan konsep *Verifiable Credential* (VC). VC memberikan jaminan tentang informasi DID tertentu, misalnya, bahwa informasi tersebut sesuai dengan pribadi yang lahir pada tanggal tertentu. Seringkali VC tidak ditransmisikan sendiri, tetapi dalam bentuk *Verifiable Presentations* (VP) yang membuktikan suatu fakta tanpa mengungkapkan informasi yang tidak diinginkan. Misalnya, untuk alasan privasi, seorang Wakil Presiden dapat membuktikan bahwa seseorang berusia di atas 18 tahun tanpa mengungkapkan usianya. Tren terkini adalah interoperabilitas sistem semacam ini, baik yang sama maupun berbeda jenis.

5.5 KEAMANAN, PRIVASI, DAN PEMBELAJARAN MESIN

Keamanan dan privasi telah menjadi ajang persaingan sengit. Perusahaan dan organisasi lain menerapkan pertahanan yang semakin canggih, tetapi penjahat siber juga semakin canggih. Tidak mengherankan, pembelajaran mesin merupakan senjata ampuh dalam perang melawan kejahatan siber, yang dapat digunakan untuk melindungi atau membahayakan infrastruktur komputer dan penggunanya.

Pembelajaran mesin (ML) dapat sangat efektif dalam menangkap dan mengklasifikasikan pola, yang sangat berguna untuk mendeteksi perilaku mencurigakan atau anomali. Ada banyak area terkait keamanan di mana ML telah berhasil diterapkan, termasuk

penyaringan spam, deteksi penipuan, deteksi intrusi, deteksi malware, kerentanan dalam kode sumber, dan lain-lain.

Sayangnya, ML juga dapat menjadi sumber kerentanan dan serangan, seperti yang dijelaskan di bawah ini. Pertama, ML dapat digunakan untuk mendeteksi pola perilaku pengguna, misalnya saat mereka berkomunikasi dengan orang lain, meskipun kontennya dienkripsi. Hal ini dapat digunakan, misalnya, untuk mendeteksi aktivis hak asasi manusia yang mencoba menghindari sensor. Di bagian selanjutnya, kami akan membahas serangan ini secara detail.

Lebih lanjut, karakteristik data pelatihan dapat menyebabkan hasil yang tidak terduga atau tidak diinginkan, akibat ambiguitas dalam kumpulan data atau klasifikasi. Contoh anekdotal dari efek ini adalah insiden baru-baru ini, ketika Alexa dari Amazon menyarankan seorang gadis berusia 10 tahun untuk menyentuhkan koin ke ujung colokan yang setengah terpasang. Kesalahan semacam ini dapat terjadi bahkan tanpa intervensi agen jahat, dan sulit dihindari karena banyak model ML tidak dapat dipahami oleh manusia, sehingga tugas memprediksi hasilnya menjadi hampir mustahil. Karena keterbatasan ini, terdapat upaya untuk menggunakan teknik yang dapat meningkatkan kemampuan menjelaskan dan menginterpretasi model pembelajaran mesin (ML).

Masalah di atas dapat diperparah oleh penyerang aktif yang sengaja bertujuan untuk mendefinisikan sistem ML, menyebabkan sistem tidak berfungsi, atau mencuri informasi dari model ML. Misalnya, penyerang dapat dengan cermat mengedit masukan ke sistem ML untuk menghindari deteksi. Contoh umum dari hal ini adalah spam, di mana penggunaan kata yang salah eja atau karakter yang tidak terduga dapat mencegah spam diklasifikasikan sebagai spam. Penyerang juga dapat menyebabkan sistem beroperasi dengan cara yang merugikan. Misalnya, telah ditunjukkan bahwa perubahan kecil pada aspek visual rambu jalan dapat menyebabkan sistem mengemudi otomatis melanggar batas kecepatan.

Dalam beberapa kasus, penyerang memiliki kesempatan untuk memberikan data pelatihan ke sistem ML, dan dapat memanfaatkannya untuk membuat bias model dengan memberikan sampel berbahaya, sebuah serangan yang dikenal sebagai ML poisoning. Terakhir, penyerang dapat menggunakan model yang ada untuk mengekstrak informasi mengenai data pelatihan, memperoleh akses ke data yang harus dirahasiakan.

Contoh menarik dari ancaman yang terkait dengan penggunaan pembelajaran mesin (ML) adalah usulan Apple baru-baru ini untuk memindai pustaka foto perangkat guna mencari pornografi anak. Pornografi anak jelas merupakan kejahatan yang mengerikan, dan cara yang aman untuk memeranginya tentu akan disambut baik. Apple mengusulkan untuk memindai pustaka foto di perangkat pengguna, guna mencari konten ilegal, guna menghindari pengiriman foto pengguna ke situs eksternal. Hal ini akan menjamin privasi foto bagi semua pengguna kecuali penjahat.

Meskipun ide ini cukup menarik, banyak risiko telah diidentifikasi dengan pendekatan ini, yang menyebabkan Apple menunda penerapan sistem tersebut. Pertama, akan sulit bagi pengguna untuk menilai apakah pemindaian berkas hanya akan mencari pornografi anak, atau juga akan mencari data sensitif lainnya (seperti masalah kesehatan, pandangan politik, dll.).

Selain itu, ada kemungkinan penyerang akan mengirimkan foto yang tampaknya tidak berbahaya kepada target untuk memicu kesalahan klasifikasi, yang menandai orang yang tidak bersalah. Untuk pembahasan yang lebih mendalam tentang permasalahan pendekatan ini, pembaca dapat merujuk ke.

5.6 RESISTENSI SENSOR

Di bagian ini, kami mengilustrasikan lebih lanjut relevansi ML dalam persaingan keamanan dan privasi, dengan menggunakan resistensi sensor sebagai contoh. Saat ini, jaringan komputer mendukung akses ke sebagian besar sumber informasi, termasuk surat kabar daring, siaran televisi, media sosial, dll. Di sebagian besar negara, pengoperasian jaringan komputer ini dikendalikan oleh sejumlah kecil entitas yang berada di bawah kendali langsung pemerintah atau yang dapat dengan mudah dipaksa untuk menegakkan arahan pemerintah. Hal ini membuat penyensoran akses informasi menjadi sangat mudah.

Contoh aktivitas penyensoran berskala luas mudah ditemukan: pada Juni 2019, Sudan memberlakukan penutupan internet, dan hal yang sama terjadi di Iran pada November 2019; terdapat juga bukti bahwa penyebaran informasi terkait virus Corona dikontrol ketat oleh pemerintah Tiongkok. Bentuk penyensoran yang lebih halus juga dapat ditemukan: baru-baru ini, Reuters melaporkan bahwa Amazon telah setuju untuk menyensor ulasan negatif buku Xi Jingping di platform mereka.

Dalam beberapa kasus, penyensoran dapat dibenarkan, misalnya, untuk memerangi aktivitas kriminal, seperti penyebaran konten pornografi anak sebagaimana telah dibahas sebelumnya, tetapi dalam kebanyakan kasus, penyensoran justru merampas hak warga negara untuk mengakses informasi gratis.

Tidak mengherankan, banyak orang telah mengembangkan alat yang bertujuan untuk menghindari penyensoran. Alat-alat ini memiliki tujuan ganda. Pertama, alat-alat ini bertujuan untuk memungkinkan pengguna mengakses informasi yang seharusnya diblokir. Kedua, alat-alat ini bertujuan untuk mencegah pengamatan eksternal mendeteksi bahwa pengguna mengakses informasi tersebut.

Hal ini khususnya relevan karena, di bawah rezim yang represif, warga negara yang mencoba menghindari penyensoran dapat dituntut, ditangkap, atau bahkan dibunuh. Dalam paragraf berikutnya, kami akan membahas dua alat ini, yaitu jaringan anonimitas dan alat tunneling protokol multimedia.

Jaringan Anonimitas

Jaringan anonimitas adalah sejenis jaringan overlay yang dirancang untuk menjaga privasi pengguna. Jaringan ini menggunakan jaringan server yang bertindak sebagai relai untuk menyebarkan informasi, biasanya antara pengguna dan sumber informasi. Alih-alih mengakses informasi secara langsung, informasi tersebut dirutekan dalam jaringan relai menggunakan beberapa lapisan enkripsi (teknik yang dikenal sebagai onion routing).

Enkripsi diatur sedemikian rupa sehingga mencegah relai perantara mengetahui sumber asli atau tujuan akhir paket (setiap relai hanya mengetahui hop sebelumnya dan berikutnya dalam jaringan overlay). Idealnya, pemetaan antara dua titik akhir tidak akan

mungkin terjadi tanpa kolusi beberapa relai. Jaringan anonimitas yang paling relevan saat ini adalah jaringan Tor.

Jaringan anonimitas tidak secara khusus ditujukan untuk menghindari sensor, tetapi dapat, dan telah, digunakan untuk tujuan tersebut. Faktanya, jika sensor tidak dapat mengidentifikasi relai, dan tidak dapat memblokir komunikasi antar relai ini (kecuali jika akses internet dimatikan sepenuhnya), dimungkinkan untuk membuat rute overlay dari klien yang berada di wilayah tersensor ke relai yang berada di wilayah tidak tersensor (misalnya, di negara lain), untuk mengakses situs internet mana pun. Selain itu, karena semua komunikasi dienkripsi, mustahil bagi entitas yang mengamati lalu lintas klien untuk menyimpulkan konten apa yang sedang dipertukarkan.

Sayangnya, jaringan anonimitas memiliki sejumlah keterbatasan. Pertama, penyerang mungkin dapat mengidentifikasi node yang menyediakan akses ke jaringan Tor (dikenal sebagai jembatan Tor) dan secara efektif mencegah pengguna mengakses jaringan di wilayah tersensor. Lebih lanjut, meskipun penyerang tidak dapat mengakses konten paket yang dipertukarkan, ia dapat mengakses fitur paket-paket ini untuk menyimpulkan informasi apa yang sedang diakses. Dalam tugas ini, ML telah terbukti menjadi sekutu kuat sensor. Terdapat dua serangan relevan yang dapat digunakan untuk mendeteksi konten yang diakses klien, bahkan ketika menggunakan jaringan anonimitas: sidik jari lalu lintas dan korelasi lalu lintas.

Dalam serangan sidik jari lalu lintas, penyerang mengamati pola lalu lintas dan mencoba mencocokkannya dengan pola yang diketahui. Hal ini dimungkinkan, khususnya, ketika pengguna mengakses beberapa situs web yang diketahui. Untuk melakukan serangan ini, penyerang mengumpulkan data mengenai paket yang dihasilkan ketika situs tertentu diakses. Fitur-fitur seperti jumlah ukuran paket, frekuensi ukuran paket, bandwidth per arah, total waktu, penanda burst, waktu antar kedatangan, dll. digunakan sebagai pola yang dikenal sebagai sidik jari situs web. Perangkat pembelajaran mesin kemudian dapat digunakan untuk mempelajari pola-pola ini dan kemudian mengklasifikasikan arus lalu lintas yang dikumpulkan dari pengguna akhir.

Kemajuan dalam ML, seperti penggunaan jaringan saraf konvolusional modern, sebuah teknik yang dikenal sebagai sidik jari dalam, telah terbukti sangat efektif, bahkan ketika klien menerapkan pertahanan terhadap sidik jari situs web, seperti menambahkan paket tiruan, bantalan, dan/atau penundaan paket, untuk mempersulit klasifikasi.

Dalam serangan korelasi lalu lintas, penyerang memantau jaringan antara klien dan jaringan anonimitas serta antara jaringan anonimitas dan server, yaitu lalu lintas antara batas jaringan anonimitas dan klien/server di luar jaringan. Kemudian, fitur lalu lintas dari berbagai aliran dikorelasikan untuk menemukan kecocokan dan membangun tautan antara klien dan server. Tak perlu dikatakan lagi, penggunaan pembelajaran mesin (ML), khususnya penggunaan jaringan saraf tiruan (*deep neural network*), juga terbukti membantu dalam melakukan korelasi lalu lintas.

Penerowongan Protokol Multimedia

Penerowongan Protokol Multimedia adalah teknik yang melibatkan penyematan kanal rahasia dalam aliran multimedia. Teknik ini dapat digunakan untuk menghindari sensor dengan

memanfaatkan layanan yang mungkin tidak ingin diblokir oleh sensor, seperti Skype, Zoom, atau WhatsApp.

Dengan pendekatan ini, seorang pengguna di wilayah tersensor membuat panggilan multimedia ke pengguna lain di wilayah tak tersensor, lalu menyematkan kanal rahasia di aliran multimedia tersebut; kanal rahasia ini dapat digunakan untuk menyampaikan lalu lintas IP standar dan mengakses konten tersensor. Hal ini biasanya melibatkan penggantian sebagian atau seluruh konten multimedia asli dengan konten kanal rahasia yang dikodekan dalam beberapa bentuk. Jika konten multimedia dienkripsi, penyerang memiliki cara untuk memeriksa kanal rahasia tersebut.

Lebih lanjut, jika penyerang tidak dapat membedakan panggilan multimedia yang menyematkan kanal rahasia dari panggilan yang tidak, kita katakan bahwa kanal tersebut tetap tidak dapat diamati. Penerowongan protokol multimedia menarik karena, secara umum, sensor tidak mampu memblokir semua aplikasi multimedia, karena aplikasi-aplikasi ini digunakan secara luas oleh warga negara untuk interaksi sehari-hari dengan keluarga dan teman, serta oleh perusahaan untuk menjalankan bisnis.

Banyaknya kanal multimedia yang digunakan pada suatu waktu juga mempersulit tugas pemantauan kanal-kanal ini. Sayangnya, alat-alat ini juga rentan terhadap analisis lalu lintas, dalam upaya mengidentifikasi pola yang membedakan panggilan normal dari panggilan yang menyematkan kanal rahasia. Sekali lagi, alat pembelajaran mesin dapat membantu penyerang dalam upaya ini. Sebuah studi yang diterbitkan dalam Simposium Keamanan USENIX ke-27 menunjukkan bahwa pohon keputusan dan beberapa variannya sangat efektif dalam mendeteksi lalu lintas rahasia dengan tingkat positif palsu yang lebih rendah. Lebih lanjut, penelitian terbaru menunjukkan bahwa informasi yang diperlukan untuk melakukan klasifikasi dapat dikumpulkan, dengan cara yang hemat biaya, pada kecepatan jalur dengan memanfaatkan kemampuan sakelar terprogram baru.

Menghindari Serangan ML

Seperti yang telah dibahas sebelumnya, perangkat pembelajaran mesin dapat digunakan untuk melakukan analisis lalu lintas yang canggih guna mendeteksi anomali, mengidentifikasi pola, dan melakukan korelasi antar aliran yang berbeda. Perangkat ini memberdayakan penyerang, khususnya penyerang tingkat negara, untuk mendeteksi upaya menghindari sensor. Oleh karena itu, perangkat resistensi sensor modern harus dirancang dengan mempertimbangkan serangan ini.

Menariknya, perangkat baru sedang diusulkan yang dapat menyematkan kanal rahasia dalam aliran multimedia tanpa memengaruhi fitur utama aliran lalu lintas. Protozoa adalah salah satu perangkat tersebut. Perangkat ini memanfaatkan API WebRTC (*Web Real-Time Communication*) untuk menggantikan konten video nyata, yang diproduksi secara real-time oleh perangkat konferensi multimedia, dengan mengonversi konten dengan ukuran yang persis sama, melindungi protokol dari mekanisme deteksi berdasarkan ukuran paket dan frekuensi paket.

Lebih lanjut, Protozoa, tidak seperti beberapa protokol sebelumnya yang menawarkan bandwidth yang sangat terbatas, dapat memberikan kapasitas bandwidth kanal rahasia sekitar

1,4 Mbps. Protozoa terbukti resisten terhadap alat klasifikasi mutakhir, tetapi masih belum jelas apakah mungkin untuk merancang alat pembelajaran mesin yang lebih canggih, yang mencoba memanfaatkan fitur lain, seperti deret waktu kedatangan antar-paket, untuk melakukan deteksi.

5.7 KEAMANAN & PRIVASI DIGITAL: TREN ML DAN RESISTENSI SENSOR

Keamanan dan privasi merupakan aspek penting dari dunia kita yang terhubung. Bab ini mendefinisikan kedua konsep tersebut dan menjelaskan mengapa keduanya merupakan masalah yang harus dikelola, meskipun tidak sepenuhnya dapat dipecahkan. Bab ini menyajikan serangkaian pencapaian ilmiah dan teknologi yang ilustratif, dengan penekanan pada tren penelitian terkini. Bab ini juga menunjukkan hubungan antara keamanan/privasi dan pembelajaran mesin, dengan menggunakan resistensi sensor sebagai kasus penggunaan.

BAGIAN 2
TANTANGAN ETIKA DAN HUKUM DALAM KECERDASAN BUATAN

AI semakin membentuk ekonomi data modern. Kemajuan dalam daya komputasi, terobosan dalam pengembangan algoritma, dan melonjaknya ketersediaan Big Data secara meyakinkan berpadu untuk mendukung perkembangan teknologi AI, dan hanya sedikit yang akan menyangkal bahwa AI siap menggemparkan dunia.

Bagian I buku ini memperkenalkan berbagai kemungkinan teknologi AI, dan, meskipun AI juga memiliki keterbatasan dan kerentanan yang melekat, hampir tidak ada keraguan bahwa AI akan mengubah pengalaman manusia secara fundamental. Menghadapi keniscayaan ini, mungkin ada godaan untuk mencurahkan seluruh energi guna mengekstraksi sebanyak mungkin nilai teknis dari revolusi AI, tetapi ada bahaya jika kekhawatiran dan kekhawatiran manusiawi tradisional dikesampingkan sementara memberi jalan bagi mesin untuk berkembang dan maju.

Dalam Bagian II, buku ini menjauh dari hiruk pikuk perkembangan teknologi terkini dalam AI dan memfokuskan kembali perhatiannya pada perdebatan etika dan hukum fundamental yang telah lama dihadapi manusia menilainya kembali berdasarkan perubahan transformatif yang dibawa oleh AI.

Ditulis oleh beberapa pakar terkemuka dunia dalam AI, Etika, dan Hukum, Bagian II mengidentifikasi dan mengeksplorasi isu-isu etika dan hukum utama yang muncul dari revolusi AI bagi para kreator, pengguna, dan platform digital serta bagi para pembuat kebijakan, akademisi, dan pemikir yang masih mencoba memahami perubahan ini, dan cara terbaik untuk mengaturnya.

Bab karya *Maria do Céu Patrão Neves dan António Betâmio de Almeida* menawarkan refleksi kritis tentang peluang dan tantangan etika yang dibawa oleh teknologi AI. Pertama, sebuah visi filosofis baru diperkenalkan melalui refleksi tentang inovasi teknologi secara lebih luas, melihat masa sebelum AI dan merenungkan ke mana revolusi AI akan membawa kita.

Dalam mempertimbangkan bagaimana AI telah berkembang dari waktu ke waktu, para penulisnya membahas dampaknya terhadap kehidupan manusia dari tiga perspektif fungsional, struktural, dan identitas dan memberikan titik awal bagi perdebatan penting tentang tata kelola AI. Pada akhirnya, mereka merefleksikan bagaimana prinsip-prinsip etika utama seharusnya membentuk regulasi AI di masa depan dengan tujuan melindungi hak asasi manusia yang fundamental.

Bab karya *Pedro U. Lima dan Ana Paiva* menyelidiki tantangan etika dan hukum yang khususnya muncul dari hubungan antara manusia dan robot. Berfokus pada bidang-bidang utama robotika, bab ini mengkaji kemajuan dan keterbatasan utama yang muncul di bidang-bidang ini, serta isu-isu sosial, hukum, dan etika terpenting yang ditimbulkannya. Kontribusi ini diakhiri dengan nada penuh harapan, karena para penulis menyatakan keyakinan mereka pada robot sebagai penggerak perubahan sosial yang dapat berkontribusi pada masyarakat yang lebih berkelanjutan dan digerakkan oleh manusia.

Dalam bab karya *Eduardo Magrani dan Paula Guedes Fernandes da Silva*, dibahas salah satu aplikasi komersial utama AI: sistem rekomendasi. Secara spesifik, bab ini mengkaji berbagai bidang penerapan sistem rekomendasi saat ini, membahas manfaatnya, dan menilai dampak buruk yang ditimbulkannya. Dengan menggunakan pendekatan berbasis etika dan hak asasi manusia, bab ini menganalisis berbagai pertanyaan terkait sistem ini, termasuk hilangnya privasi, ketidakjelasan, dan potensi diskriminasi. Pada akhirnya, dikemukakan bahwa pedoman dan aturan harus diperkuat dengan strategi desain yang "berpusat pada nilai" di mana arsitektur sistem rekomendasi harus menggabungkan pedoman dan aturan tersebut.

Bab karya *Beatriz Assunção Ribeiro, Hélder Coelho, Ana Elisabete Ferreira, dan João Branquinho* merupakan upaya kolaboratif dan multidisiplin untuk menjawab pertanyaan mendasar tentang apakah robot harus diberi status badan hukum. Menggabungkan masukan dari filsafat, psikologi, komputasi, dan hukum, bab ini dimulai dengan mengkaji konsep objek dan agen serta bagaimana AI dapat masuk ke dalam perbedaan tersebut.

Kedua, bab ini membahas bagaimana konsep "metakognisi" yang didefinisikan oleh para penulis sebagai kognisi tentang kognisi yang menghasilkan proses mental yang mengendalikan pikiran dan perilaku suatu entitas dapat diterapkan sebagai persyaratan minimum untuk akuntabilitas. Pada akhirnya, dikemukakan bahwa perbedaan utama antara agen yang tidak bertanggung jawab dan yang bertanggung jawab bergantung pada proses metakognitif yang dapat dilakukan oleh entitas tersebut, dan bahwa entitas yang menunjukkan proses metakognitif harus diberikan status badan hukum (meskipun AI belum sepenuhnya terwujud).

Dimulai dengan bab karya *Márcia Santana Fernandes dan José Roberto Goldim*, Bagian II mempersempit isu dan tantangan etika yang secara khusus diciptakan oleh penggunaan teknologi berbasis AI dalam kesehatan. Secara khusus, bab ini mengawali diskusi tentang bagaimana penggunaan sistem AI dalam kesehatan menciptakan peluang sekaligus tantangan, termasuk pertanyaan-pertanyaan etis terkait berbagai cara AI mentransformasi komunikasi dan pengambilan keputusan di sektor ini. Pada akhirnya, para penulis mendukung Model Bioetika Kompleks yang didasarkan pada berbagai pendekatan etis sebagai kunci untuk mencapai keseimbangan yang tepat antara mengambil tindakan pencegahan yang diperlukan dan tetap berharap bahwa AI dapat mewujudkan potensinya untuk meningkatkan sektor kesehatan.

BAB 6
AI SEBELUM DAN SESUDAH: PELUANG DAN TANTANGAN

Abstrak Kecerdasan buatan (AI) dan sistem digital saat ini menempati posisi fundamental di seluruh masyarakat. Keduanya merupakan perangkat yang membentuk kehidupan manusia dan mendorong perubahan peradaban yang signifikan.

Mengingat kekuatannya yang luar biasa, yaitu sistem dengan kapasitas pengambilan keputusan yang otonom, wajar jika potensi dampak sosialnya patut direfleksikan secara kritis mengenai peluang dan tantangan yang dihadapi oleh AI. Inilah tujuan utama teks ini. Para penulis memulai dengan menjelaskan posisi filosofis yang mereka mulai, dan yang mengontekstualisasikan refleksi mereka tentang inovasi teknologi secara umum, kemudian secara singkat mempertimbangkan silsilah ("sebelum") AI, dalam karakteristik utamanya dan arah evolusinya ("Dapatkah mesin meniru manusia?").

Dengan mempertimbangkan jalur perkembangan AI dan dampak disruptifnya terhadap kehidupan manusia ("setelahnya"), sistematisasinya diusulkan dalam tiga kategori fungsional, struktural, dan identitas ("Dapatkah manusia meniru mesin?"). Terlepas dari ekspektasi optimis atau pesimis terhadap evolusi teknologi, terdapat kebutuhan untuk debat publik tentang regulasinya saat ini dan di masa mendatang. Teks ini juga mengidentifikasi prinsip-prinsip etika utama dan persyaratan hukum untuk mengatur AI guna melindungi hak asasi manusia yang fundamental.

6.1 BEBERAPA PRAANGGAPAN YANG MEMBENTUK REFLEKSI TENTANG AI

Penataan, pengembangan, dan penggunaan AI sangatlah kompleks dan menantang bagi pendekatan refleksif non-teknis, sosial, dan manusiawi. Hal ini terutama disebabkan oleh aspek-aspek yang berbeda namun kumulatif berikut ini. AI bersifat multidisiplin, memobilisasi keragaman pengetahuan dan teknik yang semakin meningkat digital, elektronik, komputasi, matematika, statistik, ilmu sosial dan humaniora, termasuk hukum, sosiologi, dan filsafat yang tak terelakkan menjadikannya kompleks dan membuat wacana yang komprehensif menjadi sangat sulit atau bahkan mustahil.

Pada saat yang sama, ranah AI saat ini begitu luas, beragam, dan dinamis sehingga wacana apa pun tentang subjek ini menjadi terbatas dan mungkin juga cepat usang. Akhirnya, interpretasi tentang apa yang diwakili oleh AI saat ini, terutama di masa depan, begitu beragam mulai dari antusiasme naif dan kepatuhan sosial hingga pesimisme yang mengebiri sehingga setiap posisi yang diambil terbuka untuk dikritik, dan posisi yang disajikan sekarang pun tak terkecuali.

Refleksi kita, seperti refleksi lainnya, didasarkan pada beberapa asumsi yang, baik secara implisit maupun eksplisit, membentuknya, dan karenanya harus diungkapkan. Secara singkat, kita dapat menyajikan empat praanggapan utama yang mendasari dan membentuk refleksi kita. Yang pertama adalah bahwa teknologi adalah produk kreativitas manusia, sehingga tidak dapat secara umum dan langsung disetankan seolah-olah merupakan realitas

yang asing dan memusuhi kita. Faktanya, teknologi telah menjadi fundamental bagi kelangsungan hidup dan kualitas hidup umat manusia.

Teknologi menciptakan kondisi kehidupannya sendiri dari dunia yang ada. Sikap negatif masih terlalu sering muncul, terutama dalam menghadapi inovasi teknologi yang dahsyat dan sedang berkembang. Ini cenderung membangkitkan perasaan takut dalam kaitannya dengan yang baru, yang tidak diketahui, kegelisahan tertentu atau bahkan kesusahan (meskipun hari ini kita sering menyaksikan ketertarikan yang tidak kritis terhadap yang baru, seolah-olah segala sesuatu yang baru itu baik).

Ada juga permusuhan tertentu terhadap inovasi teknologi dalam penilaian dampaknya misalnya, degradasi lingkungan dikaitkan dengan dampak teknologi kadang-kadang hanya menyalahkan teknik (teknofobia) dan dengan kurangnya referensi sama sekali terhadap penyebab dan tanggung jawab lainnya.

Pengalaman mengajarkan kita bahwa manfaat pribadi yang timbul dari inovasi teknologi adalah apa yang paling menarik di awal dan kemungkinan dampak sosial atau kolektif yang negatif hanya kemudian menjadi jelas, seringkali ketika teknologi tertentu itu tersebar luas dan sangat sulit untuk ditentang. Dalam hal ini hanya krisis yang akan mendorong perubahan. Ini membenarkan analisis kritis independen terhadap penciptaan produk teknologi dan aplikasi massa mereka.

Praanggapan kedua adalah bahwa inovasi teknologi (seperti kemajuan ilmiah) tidak dapat dihentikan, tidak dapat dibendung, atau deterministik, sehingga tidak dapat ditekan, melainkan diarahkan kembali. Sekalipun diinginkan untuk menghentikan kemajuan ilmiah dan inovasi teknologi (yang bagaimanapun juga cukup diragukan), keduanya tidak akan pernah berhenti berkembang karena kombinasi variabel ekonomi-finansial, sosial, politik, akademis, dll. yang menghasilkan dinamika yang semakin kuat dan berkelanjutan yang melampaui jumlah variabel yang terlibat, di luar kendali satu orang atau kelompok kepentingan mana pun. Memperlambat proses ini (hal ini telah terjadi pada beberapa inovasi lain untuk menghindari dampak yang parah) dapat dimungkinkan, sehingga mengarahkannya kembali menjadi keharusan atau lebih baik. Namun, potensi impuls yang tidak terkendali dalam penerapan kekuatan yang semakin meningkat, oleh perusahaan swasta atau lembaga publik, yang disediakan oleh teknologi baru dan yang dapat menimbulkan risiko bagi umat manusia, tampaknya menjadi masalah refleksi dan pengendalian yang mendesak. Masalah teknik manusia secara tradisional bukanlah objek perhatian khusus dalam filsafat dan etika. Situasi ini telah berubah sejak pertengahan abad ke-20. Kekuatan teknologi yang semakin berkembang telah memotivasi filsafat dan etika untuk menganalisis secara kritis hakikat teknologi dan dampaknya terhadap umat manusia.

Praanggapan ketiga adalah bahwa inovasi teknologi tidaklah netral secara aksiologis, dan oleh karena itu, tidak luput dari pengawasan etika. Inovasi teknologi tidak sepenuhnya bersifat instrumental, seolah-olah evaluasinya hanya bergantung pada penggunaan dan penggunanya. Faktanya, setiap ciptaan telah memiliki tanda penciptanya, meskipun itu tidak lebih dari niat yang mengarah pada penciptaan, pada produksi, sebuah intensionalitas struktural dan orisinal (prinsip perkembangannya, dalam evolusi yang tak tertahankan dan tak

tereduksi), yang luput dari kendali manusia, melainkan mengondisikan dan bahkan mendorong perilaku manusia. Teknologi-teknologi baru, dengan fakta sederhana bahwa mereka ada, mendorong penggunaannya. Perkembangan teknologi yang mencengangkan adalah hasil dari hasrat manusia yang sulit dikendalikan.

Praanggapan keempat adalah bahwa inovasi teknologi seharusnya bukan tujuan itu sendiri, melainkan sarana dalam hal satu-satunya tujuan itu sendiri, yaitu manusia. Raison d'être dari semua produksi manusia adalah untuk menciptakan cara-cara baru dan beragam untuk mempromosikan dan mewujudkan perkembangan manusia, sehingga ia harus tetap tunduk pada umat manusia. Tantangan mendasar yang muncul adalah apakah teknologi harus menjadi instrumen yang melayani umat manusia (misalnya, instrumen untuk meningkatkan kesehatan manusia) atau apakah umat manusialah yang harus beradaptasi dengan tuntutan teknologi. Dengan mengakui asumsi-asumsi kita, kita sekarang harus lebih akurat mengidentifikasi beberapa peluang utama yang dibuka oleh AI, dan memikirkan risiko atau tantangan yang ditimbulkan oleh perkembangannya, mulai dari kelahiran AI dan tujuan awalnya hingga serangkaian ambisi barunya.

6.2 MESIN DAPAT MENIRU MANUSIA

Pertanyaan Kunci

Bisakah mesin meniru manusia? adalah pertanyaan yang diajukan oleh matematikawan Alan Turing, yang disebut sebagai "bapak" ilmu komputer teoretis dan AI, pada tahun 1950, dalam bukunya Imitation Game, dan yang ia coba jawab secara positif sepanjang hidupnya: "Bisakah mesin berpikir?" Kita dapat mengatakan bahwa pertanyaan Turing mungkin menandai titik balik dalam hubungan antara manusia dan mesin, sama mencoloknya dengan interogasi Jeremy Bentham pada tahun 1789, "Bisakah hewan menderita?" yang dipicu dalam hubungan manusia dengan hewan.

Pada paruh kedua abad ke-20, mesin tampaknya memang cerdas. Komputer digital, yang disebut demikian karena mampu memanipulasi simbol-simbol diskrit, atau digit, telah diciptakan setelah revolusi industri ketiga, yang ditandai dengan Otomasi, yang sangat berfokus pada teknologi informasi dan komunikasi. Pertanyaan saat itu adalah: bisakah komputer berperilaku cerdas seperti manusia?

Jawaban pertama diberikan oleh "tes Turing", yang disebut "permainan imitasi": mungkinkah seorang interogator membedakan jawaban yang diberikan oleh komputer dari jawaban yang diberikan oleh manusia? Dapatkah mesin meniru kecerdasan manusia, atau meniru kecerdasan manusia?

Tes Turing telah menjadi subjek banyak kritik, banyak di antaranya berasal dari definisi yang tepat tentang berpikir dan kecerdasan. Salah satu yang paling terkenal didasarkan pada argumen Ruang Cina yang terkenal oleh Searle. Tes Turing didasarkan pada bahasa. Kita tahu bahwa bahasa merupakan hal mendasar dalam perkembangan kecerdasan manusia, tetapi kecerdasan tidak boleh langsung disamakan dengan pengetahuan atau ingatan. Apakah tes tanya jawab sederhana merupakan cara yang memadai untuk mengidentifikasi pemikiran manusia dan semua jenis kecerdasan manusia?

Dengan Turing, kita ingin dapat mengidentifikasi kesamaan yang dapat diterima dengan cara berpikir dan bereaksi manusia, mungkin yang kita inginkan adalah mengenali bahwa sebuah mesin mampu meniru cara berpikir manusia dengan sangat baik. Akan jauh lebih sulit untuk mengenali kapasitas kesanggupan mesin!

Langkah-Langkah Awal AI

Dalam konteks inilah Marvin Minsky dan John McCarthy mencetuskan istilah Kecerdasan Buatan yang mereka presentasikan pada tahun 1956, di Konferensi Dartmouth College, yang diselenggarakan pada tahun yang sama di Amerika Serikat, dan yang mempertemukan para pelopor AI pada masa itu. Pada tahun yang sama, keduanya mendirikan Proyek Kecerdasan Buatan (sekarang Laboratorium Ilmu Komputer dan Kecerdasan Buatan MIT).

Saat itulah sejarah AI sesungguhnya dimulai, di mana Turing mengusulkan bahwa strategi yang harus diikuti bukanlah, seperti sebelumnya, mencoba "menulis program yang memungkinkan mesin melewati permainan imitasi" (mereproduksi bagian-bagian penalaran manusia), melainkan menulis "program yang memungkinkan mesin belajar dari pengalaman, seperti yang dilakukan bayi". Ke arah inilah (pembelajaran otomatis, melalui pengalaman) pendekatan terhadap sistem cerdas saat ini, beberapa dekade kemudian.

Jadi, hal ini telah meningkatkan otonomi sistem cerdas dalam kaitannya dengan manusia. Kita kemudian sepenuhnya berada dalam revolusi industri keempat, yang ditandai dengan Konektivitas, di mana AI berkembang hampir secara eksponensial, yang dikonfirmasi saat kita sekarang memasuki Masyarakat 5.0, revolusi industri kelima, yaitu era koneksi penuh, di mana segala sesuatu akan terhubung, semua sarana yang tersedia bagi manusia akan terhubung dan orang-orang harus beradaptasi atau mengintegrasikan diri mereka ke dalam jaringan aliran berkelanjutan ini (penyelarasan teknologi robotik dengan kecerdasan manusia, peningkatan kolaborasi atau kemitraan antara manusia dan sistem cerdas). AI telah berkembang dan sangat mendorong 3 revolusi industri terakhir, membuka jalan menuju otomatisasi penuh dan konektivitas maksimum (nirkabel, tanpa koneksi fisik).

Namun demikian, kita masih belum memiliki definisi AI yang disepakati (yang sangat mengungkapkan dinamismenya), meskipun cukup relevan untuk pembatasan domainnya dan persepsi operabilitasnya. Ada banyak definisi yang berbeda dan bahkan mereka yang menolak ungkapan tersebut, seperti Luc Julia, dalam karyanya *L'Intelligence artificielle n'existe pas*, di mana ia menganggap bahwa AI selalu didefinisikan dengan buruk karena menyiratkan bahwa algoritma dapat membuat keputusan yang sadar dan rasional seperti manusia. Ia percaya bahwa ini tidak terjadi dan bahwa ide-ide keliru seperti ini telah memicu persepsi Hollywood yang fantastis tentang AI, seperti Matrix atau Terminator.

Kecerdasan manusia sulit untuk dibatasi dan dipahami sepenuhnya. Ini lebih dari sekadar rasionalitas terhadap rangsangan dan analisis data. Ia memiliki fitur bawaan lainnya dan koneksi yang kuat ke seluruh tubuh manusia. Mungkin sebutan Kecerdasan Buatan (AI) sangat efektif sebagai merek, tetapi tidak terlalu ketat. Ungkapan AI digunakan saat ini untuk menunjuk berbagai teknologi dengan beberapa karakteristik umum.

Kami mengadopsi definisi yang diusulkan oleh Kelompok Pakar Tingkat Tinggi tentang Kecerdasan Buatan Komisi Eropa: "Sistem kecerdasan buatan (AI) adalah sistem perangkat lunak (dan mungkin juga perangkat keras) yang dirancang oleh manusia yang, dengan tujuan yang kompleks, bertindak dalam dimensi fisik atau digital dengan mengamati lingkungannya melalui akuisisi data, menafsirkan data terstruktur maupun tak terstruktur yang terkumpul, bernalar berdasarkan pengetahuan, atau memproses informasi yang diperoleh dari data tersebut, dan memutuskan tindakan terbaik yang akan diambil untuk mencapai tujuan tersebut. Sistem AI dapat menggunakan aturan simbolis atau mempelajari model numerik, dan juga dapat mengadaptasi perilakunya dengan menganalisis bagaimana lingkungan dipengaruhi oleh tindakan sebelumnya."

Definisi AI lain yang lebih sederhana adalah: "sistem terkomputerisasi, agen atau robot, yang mampu bertindak dan membuat keputusan secara independen dari pengawasan manusia"; "sistem yang mampu memecahkan masalah kompleks secara rasional atau mengambil tindakan yang tepat untuk mencapai tujuannya dalam situasi dunia nyata apa pun yang dihadapinya".

Pencapaian yang Mendorong

AI, sebagaimana kita definisikan secara luas, telah menjadi alat yang ampuh dalam mencapai tujuan manusia, yang perkembangannya yang berkelanjutan telah melampaui status instrumental aslinya dan menaklukkan rencana kinerja baru yang perlu dipertimbangkan, dalam penghapusan terus-menerus atas apa yang tampaknya menjadi batasannya. Namun, kita masih berada di era AI yang lemah atau sempit, yaitu, hanya mampu melakukan satu atau beberapa tugas spesifik, dan perangkat lunaknya hanya dapat membuat keputusan berdasarkan informasi yang diberikan sebelumnya. Beberapa contoh umum adalah: bermain catur, Go, atau poker; mengidentifikasi orang melalui wajah yang terekam dalam video keamanan waktu nyata (pengenalan wajah); atau mengendarai kendaraan otonom.

Jika kita mengambil salah satu contoh ini yang paling sederhana, seperti bermain gim dan mengikuti evolusi AI, kita dapat dengan mudah memahami arah yang kita tuju. Langkah penting pertama dari proses evolusinya terjadi pada tahun 1996, ketika Deep Blue, sebuah perangkat lunak IBM, mengalahkan juara catur dunia Kasparov. Kemudian, pada tahun 2017, AlphaGo memenangkan permainan Go melawan yang terbaik di dunia, dan pada tahun 2019, Pluribus memenangkan maraton poker 12 hari, berkompetisi melawan 5 pemain. Langkah kedua diberikan ketika perangkat lunak mulai belajar bermain sendiri, bermain melawan dirinya sendiri, dan dengan demikian semakin sedikit bergantung pada data yang dihasilkan manusia, sejak tahun 2017.

Baru-baru ini, MuZero milik Google dipresentasikan mampu bermain tanpa perlu data yang dimasukkan manusia, yaitu, tanpa diberi aturan, berkat kemampuannya untuk merencanakan strategi kemenangan dalam konteks yang tidak diketahui. Arah evolusi AI inilah yang memicu ketakutan terbesar manusia: AI mendapatkan kekuatan yang cukup untuk sepenuhnya lepas dari kendali manusia. Arah evolusi yang diikuti mudah terungkap: maju menuju kinerja yang selalu dan terus-menerus lebih unggul dalam setiap fungsi yang dilakukan

AI; dan menuju tingkat otomatisasi (emansipasi) manusia yang lebih tinggi (pencipta, produsen).

Tren evolusi AI dan penerapannya membenarkan kekhawatiran serius akan devaluasi manusia dalam menghadapi kemampuan superior sistem baru di bidang aktivitas yang telah membentuk masyarakat dan tujuan hidup manusia. Risiko dan tantangan muncul dalam jangka pendek, tetapi beberapa di antaranya sudah menjadi ancaman: "semakin besar kapasitas digital suatu masyarakat, semakin rentan pula masyarakat tersebut". Isu-isu ini khususnya menarik bagi Etika dan Hukum.

Saat ini terdapat persepsi yang jelas bahwa kita sedang mengalami revolusi digital (yang mengikuti revolusi industri) yang dipimpin oleh AI. Artinya, AI merupakan kehadiran yang konstan dan tak terhapuskan dalam kehidupan sehari-hari manusia, baik secara individu maupun komunitas, terutama di belahan bumi utara, dan cara hidup kita sangat bergantung pada AI yang, saat ini, telah merasuki sebagian besar modalitas tindakan manusia. Kita hidup di era AI.

6.3 MANUSIA MAMPU MENIRU MESIN

Gagasan manusia meniru mesin akan dianggap bodoh hingga saat ini. Namun, saat ini kita dapat merumuskan pertanyaan provokatif ini karena terdapat mesin digital dengan atribut yang dianggap lebih unggul dalam organisme hidup: kecerdasan. Mesin-mesin ini, yang disajikan memiliki kapasitas intelektual yang jauh lebih unggul daripada manusia, dapat menjadi model perilaku individu dan sosial yang patut ditiru. Penyelarasan manusia dengan aturan sistem sosio-teknis baru terjadi karena adaptasi sederhana oleh inersia bawah sadar atau dipaksakan sebagai prioritas yang dibenarkan oleh kriteria efisiensi tetapi mengabstraksikan kriteria lain yang terkait dengan kodrat manusia.

Dalam upaya untuk mensistematisasikan semakin banyaknya intervensi AI dalam kehidupan manusia, kami akan mengatakan bahwa dampaknya lebih nyata dan disruptif pada tiga tingkat utama: fungsional, dalam penggunaan AI sebagai instrumen khusus untuk tujuan manusia; struktural, dalam perubahan yang ditimbulkan AI dalam hubungan antarmanusia dan dalam organisasi institusi; identitas, dalam transformasi yang berawal dari apa adanya manusia dan dalam citra yang ia miliki tentang dirinya sendiri. Kita harus mempertimbangkan ketiga tingkat intervensi AI dalam ranah manusia ini, baik dalam peluang baru yang diciptakannya bagi perkembangan manusia, maupun dalam tantangan baru yang ditimbulkannya bagi ketahanan manusia dalam konteks kinerja yang jauh melampauinya.

Tingkat Fungsional

Dimensi fungsional AI merujuk secara tepat pada kemampuannya untuk menjalankan fungsi manusia, yang dilakukannya dengan melakukannya lebih cepat, lebih sempurna, lebih ekonomis, dalam tindakan pendukung manusia yang benar-benar unik dan mengesankan. Beberapa domain utamanya yang sangat sukses adalah industri, peradilan, kesehatan, pendidikan, transportasi, keuangan, pemasaran, keamanan komputer, militer (pertahanan militer), dan hiburan. Sekilas intervensi AI dalam beberapa domain yang begitu berbeda dan

paradigmatis ini dapat memberi kita gambaran yang lebih tepat tentang potensi disruptifnya, baik positif maupun negatif, di zaman kita.

AI pertama kali menjadi dominan di industri, di mana ia digunakan secara masif dan di mana dimensi fungsionalnya paling nyata, melalui otomatisasi berbagai fungsi, terutama yang lebih berat, secara fisik dan psikologis. Melepaskan orang-orang dari beban terberat sangat dipuji. Namun, AI dalam industri tidak terbatas pada fungsi otomatis saja, tetapi juga digunakan untuk membantu pengambilan keputusan dan analisis data, termasuk manajemen personalia, seperti tingkat kehadiran dan produktivitas karyawan, serta perekrutan dan pemberhentian karyawan. Namun, terdapat beberapa paradoks terkait teknologi dan produktivitas.

Saat ini, gagasan umum sebelumnya bahwa AI hanya melakukan tugas-tugas mekanis, yang secara profesional kurang menuntut dan secara sosial kurang dihargai, mudah dibantah. Di satu sisi, AI telah menaklukkan beragam domain dan tingkat kompleksitas tindakan, bahkan di bidang tradisional, seperti industri; di sisi lain, AI telah diterapkan pada bidang-bidang tindakan yang semakin menuntut, seperti perawatan kesehatan atau peradilan.

AI hadir secara kuat, baik dalam penelitian klinis (misalnya, pengumpulan data dalam jumlah besar untuk mengidentifikasi korelasi dan tren; molekul terapeutik baru) maupun dalam perawatan klinis (misalnya, membuat diagnostik; memantau kondisi kesehatan). Ada beberapa spesialisasi medis di mana prosedur klinis standar digantikan oleh AI, seperti radiologi (membaca pemeriksaan) atau oftalmologi (melakukan beberapa pemeriksaan), di mana AI dapat secara menguntungkan menggantikan dokter. Saat ini sudah ada bantuan digital yang efisien untuk dokter dan perawat medis, terutama di bidang geriatri, ahli bedah, staf kebersihan, tetapi juga untuk pengiriman obat-obatan, makanan, dan bahkan beberapa tes diagnostik. Dalam hal keadilan, AI telah banyak digunakan, terutama dalam pencarian yurisprudensi, dalam penerapan langkah-langkah keadilan berdasarkan kasus-kasus serupa sebelumnya. Proyek-proyek untuk pembentukan pengadilan keadilan prediktif otomatis juga telah ada untuk menyelesaikan kasus-kasus yang ringan.

Memang, saat ini tampaknya semua fungsi manusia dapat digantikan oleh AI (fungsi-fungsi tersebut sedang digantikan secara bertahap) dengan keuntungan langsung, berdasarkan prinsip efisiensi, produktivitas, dan profitabilitas. Gagasan yang menjanjikan bahwa AI akan membebaskan manusia dengan menghindari tugas-tugas intelektual yang membosankan atau monoton tampaknya tidak seperti yang diantisipasi: pengecualiannya dari tugas-tugas yang berkaitan dengan pemikiran manusia. Namun, ada juga beberapa kelemahan terkait yang penting untuk dipertimbangkan bersama, dan di antaranya kami hanya menyoroti tiga. Yang pertama adalah proliferasi AI. Kami merujuk pada proliferasi AI dengan mempertimbangkan kemampuannya untuk belajar dari pengalaman sebelumnya guna menghasilkan perilaku cerdas dan keputusan yang tepat, yang disebut "pembelajaran mesin" (subdomain dari AI): ini adalah algoritma yang mampu memodifikasi diri sendiri dan membuat keputusan tanpa campur tangan manusia. AI juga telah berkembang ke apa yang disebut "pembelajaran mendalam" (subdomain dari pembelajaran mesin) yang terdiri dari kemampuan komputer untuk belajar sendiri, melalui pengenalan pola, dalam banyak lapisan

data mentah, bergantung pada tujuan yang diajukan, dan menjalankan tugas layaknya manusia. Oleh karena itu, AI selalu meningkatkan kinerjanya dan memperoleh keterampilan baru. Aspek ini, yang langsung dan perlu diakui sebagai positif, disajikan di sini sebagai kerugian sejauh memicu proses pelepasan Kecerdasan Buatan dari kendali manusia.

Kerugian kedua, dan yang paling umum disajikan, adalah pengangguran massal. Seiring bertambahnya domain di mana AI dapat membantu tujuan manusia, seiring bertambahnya keragaman fungsi yang dapat dilakukannya, dan seiring kinerjanya menjadi lebih unggul daripada manusia, AI juga menggantikan manusia. Oleh karena itu, ancaman utama yang telah ditekankan pada tingkat ini adalah pengangguran massal, seperti yang sudah jelas terlihat dalam industri.

Kita tahu argumen yang mengabaikan masalah yang berkembang ini: sepanjang sejarah manusia selalu ada aktivitas kerja yang telah lenyap dan aktivitas baru yang telah muncul dan hal yang sama akan terjadi sekarang juga. Kita tidak dapat mengabaikan keberadaan variabel yang belum pernah terjadi sebelumnya dalam persamaan ini yang dapat membahayakan keseimbangan masa lalu: kecepatan proses yang tidak memungkinkan adaptasi manusia terhadap transformasi yang sedang berlangsung dan kualitas intelektual dari pekerjaan yang hilang. Sekalipun banyak pekerjaan baru tercipta, pertanyaan tentang jenis dan tingkat sosial pekerjaan ini perlu dipertimbangkan.

Kerugian ketiga adalah eksklusi sosial. Memang, keuntungan dan kerugian AI mungkin tidak terdistribusi secara merata, dengan orang yang paling diuntungkan menjadi yang paling diuntungkan dan yang paling tidak diuntungkan menderita kerugian paling besar. Selain itu, ketidakadilan kronis ini diperparah oleh kesenjangan pembagian generasi: saat ini kita memiliki proliferasi generasi yang terus meningkat, yang tidak lagi berganti setiap 25 tahun, tetapi setiap 10 tahun.

Dalam konteks yang belum pernah terjadi sebelumnya ini, menjadi sangat mudah bagi orang untuk dianggap ketinggalan zaman oleh generasi berikutnya, dan pada saat yang sama, tidak berguna bagi masyarakat, bahkan mungkin menjadi beban atau dibuang. Kerugian antargenerasi ini dapat menyebabkan keretakan sosial yang serius dan sulit diatasi tanpa perubahan mendalam dalam organisasi masyarakat manusia.

Mengkarakterisasi AI dalam rentang fungsionalnya, kami ingin menekankan bahwa: AI tetap berada di luar jangkauan manusia dan dapat dimanipulasi serta dikendalikan olehnya; AI berkontribusi pada pembangunan peradaban yang dipandu oleh inovasi teknologi, cerdas, dan otomatis, serta oleh efisiensi dan produktivitas. Oleh karena itu, AI mengancam akan membuat manusia menjadi usang.

Tingkat Struktural

Dimensi struktural AI mengacu pada bentuk-bentuk hubungan baru, pola-pola baru hubungan personal, sosial, dan institusional, yang dicirikan oleh kedekatan virtual yang lebih besar di antara semua orang (mengatasi jarak geografis), oleh cakupan yang lebih luas (karena semua orang berpotensi dilibatkan), dan secara paradoks, sekaligus memperkuat hubungan dengan memediasinya dan menekan kontak langsung.

Mediasi hubungan manusia melalui Kecerdasan Buatan saat ini terjadi dalam keragaman domain yang terus berkembang yang telah kita sistematisasikan dalam tiga bidang. Pada tingkat personal, orang-orang dari seluruh dunia saling mengenal dan bersosialisasi secara virtual (bahkan untuk hubungan keintiman emosional); pada tingkat sosial, aktivitas manusia dikembangkan di ranah digital (di mana kelompok-kelompok kepentingan dibentuk, dan aktivisme sipil, politik, atau lainnya dikembangkan, demonstrasi dijadwalkan, petisi diajukan, dll.); Pada tingkat kelembagaan, lembaga berinteraksi dengan warga negara melalui teknologi cerdas (misalnya, hubungan dengan administrasi publik, karena transaksi komersial cenderung semakin daring dan layanan dilakukan oleh chatbot, sebuah program komputer yang mencoba mensimulasikan manusia dalam percakapan dengan orang-orang, hal yang sama terjadi di semakin banyak domain serta pendidikan).

Pada tingkat ini, kami ingin menyoroti dua contoh, yang sangat berbeda, tetapi keduanya merupakan paradigma transformasi yang sedang berlangsung. Yang pertama adalah investasi yang meluas dalam pembangunan kota pintar, yaitu agregat populasi di mana semuanya terhubung, dengan manajemen otomatis (lalu lintas, limbah, keselamatan publik), semuanya dimediasi oleh AI: peralatan rumah tangga cenderung menjadi sepenuhnya terhubung dan asisten pintar dapat mengurus semua layanan manajemen di rumah (mengelola limbah, mengidentifikasi masalah peralatan); semua peralatan dan infrastruktur kotamadya akan terhubung (misalnya, identifikasi aspek yang perlu ditingkatkan, keselamatan, pengukuran kualitas udara, koordinasi lalu lintas, dll.). Aktivitas penataan masyarakat manusia seperti perbankan dan asuransi cenderung daring, dematerialisasi (tanpa dokumen kertas) dan tanpa perantara manusia. Perubahan ini menciptakan kerentanan baru dalam hal keamanan, kepercayaan terhadap lembaga, dan akses langsung ke lembaga tersebut. Warga negara semakin rentan terhadap sistem teknis tanpa identitas dengan akses berbasis beberapa kode numerik dan kata sandi.

Contoh paradigmatik kedua terkait dengan pengenalan AI dalam politik (di samping domain strategis lain yang telah disebutkan, yaitu kesehatan, keuangan, dan militer). Pada tahun 2019, sebuah studi oleh sebuah Universitas di Spanyol menyimpulkan bahwa 1 dari 4 orang Eropa bersedia membiarkan AI mengambil keputusan politik penting di negara mereka, demi imparsialitas, kejujuran, dan keadilan. Saat ini, sudah ada referensi tentang pembentukan "siberokrasi" yang akan segera terjadi yang dapat mengancam atau menghancurkan sistem demokrasi seperti yang kita kenal. Kemudahan langsung bagi aktivitas manusia jelas dan tak terbantahkan, di bawah prinsip baru optimalisasi sarana. Namun, terdapat pula kelemahan terkait yang penting untuk dipertimbangkan bersama, dan di antaranya kami hanya menyoroti tiga di sini.

Yang pertama, pada tingkat personal, menunjukkan bahwa intensifikasi koneksi berbanding lurus dengan jarak fisik antarmanusia (hubungan cenderung dangkal, sporadis, sementara, tanpa komitmen atau tanggung jawab, sehingga menjadi hubungan yang ringan). Yang kedua, terjadi pada tingkat sosial dan mengacu pada anonimisasi keunikan personal sebelum hubungan fungsional (dari integrasi ke dalam kategori orang dan pola hubungan, yang terstruktur berdasarkan minat). Kelemahan ketiga, terletak pada tingkat kelembagaan dan

mengacu pada integrasi seluruh aktivitas manusia ke dalam jaringan hubungan (semuanya berada dalam jaringan dan apa yang tidak berada dalam jaringan tidak diakui keberadaannya); jaringan hampir tidak dikenal, tidak dapat diakses, dan tidak terkendali (manusia berisiko menjadi bagian dari roda gigi yang melampaui mereka).

Ketergantungan menjadi ekstrem dan protokol numerik berkode cerdas semakin padat dan secara drastis mengurangi spektrum mode komunikasi manusia. Selain itu, kita semakin mengintegrasikan program AI ke dalam proses pengambilan keputusan. AI, dalam lingkup strukturalnya, menampilkan dirinya sebagai sesuatu yang terintegrasi dalam semua aktivitas manusia dan sosial serta membentuk dan memformatnya; ia membangun budaya baru yang dipandu oleh (inter)mediasi dan konektivitas virtual, serta oleh optimalisasi sumber daya; ia mengancam penomoran manusia (merepresentasikan manusia melalui angka, mendepersonalisasikannya). Kuantifikasi realitas yang berlebihan di media (misalnya, statistik dan indeks peringkat) merupakan salah satu efek samping dari masyarakat digital yang merendahkan nilai-nilai kemanusiaan lainnya yang seharusnya menjadi bagian dari karakterisasi realitas.

Tingkat Identitas

Dimensi identitas AI mengacu pada persepsi baru yang diperoleh manusia tentang diri mereka sendiri karena kemahahadiran AI, yang ditandai dengan upaya mengatasi kodrat mereka dan membangun citra baru tentang diri mereka sendiri, yang cukup jelas setidaknya dalam tiga aspek esensial.

Yang pertama, yang tampaknya cukup revolusioner, adalah masuknya AI ke dalam dimensi spiritual manusia, keintiman terdalamnya, yang sepanjang sejarah umat manusia telah dianggap sebagai kekhususan unik sekaligus perbedaan kualitatifnya dalam kaitannya dengan semua makhluk lainnya. Masuknya AI ini terwujud dalam dimensi kreatifnya, dalam ekspresi artistiknya yang direplikasi oleh AI untuk menggubah musik, melukis kanvas, dan menulis karya sastra. Misalnya, perangkat lunak pertama untuk menciptakan musik berasal dari tahun 1997, dan kini komposisi berbagai gaya musik oleh AI tersebar luas; pada tahun 2016, Microsoft mengembangkan perangkat lunak menggunakan Kecerdasan Buatan yang, melalui analisis mahakarya Rembrandt, berhasil menciptakan lukisan baru dengan karakteristik yang sama; sejak tahun 2018, kita mulai memiliki buku-buku yang ditulis oleh AI.

Aspek kedua yang perlu disoroti adalah kekuatan baru untuk membangun identitas alternatif, di luar diri tetapi cenderung dianggap sebagai diri yang sesungguhnya. Identitas ini merupakan identitas digital, yang direkayasa dengan kolaborasi AI, dalam versi simulasi pribadi seperti avatar (sepenuhnya digital, tubuh siber, identitas daring) yang memungkinkan setiap orang untuk terus-menerus dan mudah (tanpa usaha) menemukan kembali diri mereka, mengembangkan berbagai kepribadian (mengubah usia, jenis kelamin, dll.), dan membangun berbagai jenis hubungan sesuai dengan kepribadian yang berinkarnasi. Namun, penetrasi kecerdasan buatan ke dalam esensi manusia bahkan lebih dalam, sebagai konstruksi internal identitas yang disempurnakan, dalam citra dan rupa AI. Terdapat sebuah desideratum evolusi kognitif, melalui proses penggabungan (misalnya implan sibernetik yang meningkatkan berbagai kapasitas manusia) atau apropriasi (antarmuka otak-mesin, seperti yang sedang

dikembangkan oleh perusahaan rintisan Elon Musk, Neurolink). Proses ini akan berfokus pada penciptaan pascamanusia sebagaimana yang dianjurkan oleh para transhumanis.

Kegunaan langsung bagi manusia terlihat jelas dalam prinsip baru pengembangan diri: bukan dengan mengembangkan apa adanya, melainkan dengan memperoleh apa yang bukan dirinya; bukan dengan mengintensifkan keaslian keberadaan, melainkan dengan mendistorsi dan memutarbalikkan identitasnya sendiri.

Persepsi manusia terhadap dirinya sendiri mulai mencerminkan kehadiran AI, yang juga mengadopsinya sebagai model, dengan manfaat langsung, dalam prinsip pengembangan manusia. Namun, terdapat pula kerugian yang tak terelakkan yang harus dipertimbangkan secara bersamaan: pelanggaran nilai-nilai identitas manusia melalui penyusupan ke dalam dimensi spiritualnya (esensinya), yaitu ketidakmungkinan untuk melupakan (segala sesuatu tidak terhapuskan), yang memungkinkan kita untuk Menciptakan kembali setiap hari, dalam atrofi kebebasan, dengan penghapusan ketidakpastian dan di bawah kuk keputusan yang sempurna, dalam penindasan privasi, demi transparansi aksesibilitas total kehidupan; keterasingan diri, dalam simulakra digital diri, tanpa kepadatan atau keaslian; dan perampasan diri, dalam distorsi perbaikan identitas manusia.

AI, dalam tataran identitasnya: menampilkan dirinya menyatu (menyatu) dengan semua ekspresi manusia, menentukannya; menciptakan identitas baru dalam citra dan rupa AI; dan mengancam untuk membuat manusia tunduk dan menggantinya dengan citra diri yang lebih baik. Masih dan selalu dalam ranah AI yang sempit atau lemah, kita melihat bagaimana ia mengintervensi pada tataran fungsional, secara dangkal, tetap berada di luar manusia dan dapat dikendalikan olehnya, membangun peradaban baru yang cerdas melalui otomatisasi progresif; pada tataran struktural, secara mendalam (meresap) dalam semua aktivitas dan hubungan manusia, mengintegrasikan dan membentuknya, mengaturnya, membentuk budaya virtual baru, melalui konektivitas yang terus berkembang; dan pada tingkat identitas, secara intim di hati manusia, menyatukan dan mengkonfigurasikannya kembali, menyingkirkannya dari dirinya sendiri demi citra yang lebih baik, melalui simbiosis yang berkembang.

6.4 ETIKA DAN HUKUM MENGATUR HUBUNGAN MANUSIA DAN MESIN

Debat publik tentang konsekuensi manusia dari pengembangan AI dimulai pada tahun 2015, ketika 700 ilmuwan menandatangani surat bersama yang memperingatkan ancaman AI: Prioritas Penelitian untuk Kecerdasan Buatan yang Tangguh dan Bermanfaat: Surat Terbuka. Para ilmuwan ini menggarisbawahi manfaat luar biasa yang dapat dihadirkan AI bagi umat manusia, tetapi juga risiko hilangnya kendali manusia dan perlunya penelitian lebih lanjut untuk mencegah risiko apa pun. Ketakutan terbesar adalah bahwa jaringan saraf akan terus berkembang, memungkinkan AI untuk mendapatkan kesadaran (menjadi kuat atau umum), dan kemudian sepenuhnya lepas dari kendali manusia.

Dalam konteks ini, perlu disebutkan bahwa, pada tahun 2017, para insinyur Facebook sedang mengembangkan eksperimen dengan robot yang memperdagangkan kepemilikan barang virtual di antara mereka sendiri. Itu adalah pengalaman percakapan. Setelah beberapa

hari, robot-robot tersebut telah mengembangkan bahasa mereka sendiri yang, karena berada di luar pemahaman manusia, terputus, dan mematikan robot. Evolusi manusia dan identitas mereka sepanjang sejarah manusia telah diakui. Evolusi yang lambat dan bertahap ini merupakan hasil adaptasi terhadap perubahan alam yang berkesinambungan dan didorong oleh budaya serta cara hidup baru. Namun, tren terkini yang dijelaskan di atas memiliki implikasi yang relevan, cepat, disruptif, dan multidimensi terhadap identitas manusia.

Kekhawatiran terhadap diskontinuitas identitas yang dipaksakan ini mungkin dianggap terlalu konservatif atau pesimis oleh sebagian orang. Sebagian lainnya berpendapat bahwa teknologi dan "konsumsinya" merupakan manifestasi yang dapat diterima dari kehendak manusia dalam menentukan masa depannya. Mereka adalah orang-orang yang sangat optimis atau percaya akan masa depan yang lebih baik berdasarkan teknologi. Setiap orang di masa kini bertanggung jawab untuk berkontribusi bagi masa depan tersebut secara bertanggung jawab, yang menghormati warisan manusia yang diterima dan dipercayakan kepada kita. Jika ada manfaat dan kerugian yang perlu ditunjukkan kepada AI saat ini, keharusan untuk memaksimalkan manfaat dan menghilangkan kerugian tersebut sudah sangat jelas.

Strategi global untuk pertimbangan ini adalah membangun kerangka regulasi etis-hukum, bukan dengan tujuan membatasi pengembangan Kecerdasan Buatan, melainkan melegitimasinya melalui promosi manfaat nyata dan pencegahan potensi bahayanya, dengan membingkainya dalam nilai-nilai dan prinsip identitas kemanusiaan serta melindungi hak asasi manusia.

Persyaratan Etika

Refleksi etika harus selalu mendahului peraturan hukum. Dalam masyarakat demokratis dan pluralis, penting untuk terlebih dahulu memperhatikan nilai-nilai identitas mereka dan membangun konsensus etika yang inklusif dan luas, sebagai dasar legitimasi bagi peraturan hukum yang akan dirumuskan kemudian melalui Undang-Undang. Undang-Undang memperkuat konsensus etika yang telah dicapai sebelumnya, dan Etika berkontribusi pada proses regulasi yang efektif dan kuat. Demikian pula, terkait AI, baik sebagai hasil karya manusia maupun karena dampaknya yang kuat terhadap kehidupan manusia dan masyarakat, refleksi etika-lah yang pertama kali berkembang seiring dengan semakin jelasnya kapasitas sosial AI yang disruptif.

Etika kecerdasan buatan mendapatkan perhatian khusus dan memiliki dampak sosial yang lebih besar ketika dijalankan oleh entitas internasional besar, yang sangat mewakili warga negara, atau oleh kelompok kerja internasional dan multidisiplin, yang menggabungkan berbagai pendekatan, yang secara khusus dirancang untuk menguraikan pedoman yang dianggap praktis dan diperlukan guna memastikan bahwa evolusi AI tetap berada di bawah tujuan manusia.

Oleh karena itu, dan khususnya dalam konteks Eropa, Komisi Eropa, Parlemen Eropa, dan Dewan Eropa telah bekerja secara aktif di bidang ini: Komisi telah membentuk kelompok pakar tingkat tinggi tentang kecerdasan buatan pada tahun 2018; Parlemen membentuk komite khusus tentang kecerdasan buatan di Era Digital (AIDA) pada tahun 2020; dan Dewan Eropa membentuk Komite Ad hoc tentang Kecerdasan Buatan (CAHAI) pada tahun 2019.

Pada saat yang sama, kami menyoroti pembentukan beberapa kelompok ilmiah tentang AI, seperti Pusat Keunggulan Eropa untuk regulasi Robotika dan AI, Aliansi AI Eropa, Kelompok Pakar tentang Tanggung Jawab dan Teknologi Baru, Kemitraan Global untuk Kecerdasan Buatan (GPAI), dan masih banyak lagi. Di tingkat global, UNESCO telah membentuk Kelompok Pakar Ad-hoc tentang Etika Kecerdasan Buatan.

Semua badan ini sepakat dalam mendeklarasikan urgensi regulasi AI, yang membutuhkan landasan etikanya, yang juga menunjukkan konvergensi yang luas dalam hal identifikasi prinsip-prinsip etika utama yang harus dipatuhi, dengan tetap menghormati Hak Asasi Manusia. Sebuah studi dari Berkman Klein Center for Internet & Society di Harvard University, *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI*, yang ditulis oleh Jessica Fjeld dan rekannya, mengumpulkan, pada tahun 2020, 36 dokumen paling menonjol tentang prinsip-prinsip etika regulasi dan tata kelola, menyajikan serangkaian delapan prinsip sebagai yang paling konsensual.

Privasi adalah salah satu prinsip yang paling sering, menuntut penghormatan terhadap privasi individu, "baik dalam penggunaan data untuk pengembangan sistem teknologi dan dengan memberikan orang-orang yang terdampak agensi atas data mereka". Akuntabilitas, mengenai dampak yang dihasilkan bersama dengan penyediaan solusi yang memadai, juga merupakan persyaratan umum.

Keselamatan dan Keamanan AI sangat penting dalam hal yang berkaitan dengan kinerjanya sebagaimana dirancang, dan ketahanannya terhadap invasi. Kelompok prinsip keempat adalah Transparansi dan Keterjelasan yang menuntut kejelasan dan keterbukaan proses, hasil, dan penggunaan. Klaim Keadilan dan Non-diskriminasi agar sistem AI bersifat inklusif dan mendorong keadilan global, diwajibkan dalam semua dokumen yang dianalisis.

Kendali Manusia atas Teknologi merupakan perhatian utama yang menuntut agar semua keputusan penting berada di bawah pengawasan manusia. Tanggung Jawab Profesional menuntut individu yang terlibat dalam pengembangan AI untuk dapat memprediksi konsekuensi dari tindakan mereka. Terakhir, Promosi Nilai-Nilai Kemanusiaan menyatakan bahwa AI harus meningkatkan kesejahteraan manusia. Terkadang, dengan sebutan yang berbeda, pedoman ini memang berlaku dalam refleksi etis tentang kecerdasan buatan dan harus dijamin oleh hukum.

Hukum dan Prosedur Hukum

Persyaratan etika sangat penting, tetapi tidak cukup untuk mencegah dampak buruk AI terhadap hak-hak asasi manusia karena pedoman etika tidak memiliki kekuatan hukum yang mengikat. Oleh karena itu, AI yang tepercaya juga harus sesuai hukum seperti yang telah kami tekankan sebelumnya.

Penerapan kerangka hukum yang disesuaikan dengan karakteristik spesifik sistem AI tidaklah mudah. Selain kompleksitas teknis dan perkembangan pesat sistem ini, terdapat kesulitan atau hambatan lain yang relevan. Pertama, perkembangan teknologi baru memiliki kepentingan geostrategis dan militer yang semakin besar bagi kekuatan ekonomi dan teknologi utama dunia. Kedua, terdapat tekanan kuat dari pemerintah dan perusahaan untuk mencapai peningkatan daya saing yang didorong oleh produk-produk canggih dan inovatif di pasar.

Kesulitan ketiga adalah tuntutan lembaga akademik dan spesialis AI untuk meminimalkan batasan hukum dalam aplikasi dan pengumpulan data. Dan, terakhir, diperlukan regulasi di tingkat global agar sepenuhnya efektif.

Dengan demikian, terdapat ketegangan dalam ranah etika-hukum regulasi AI dan upaya untuk mencapai keseimbangan antara keputusan politik dan berbagai kepentingan yang terlibat. Dalam konteks ini, penegasan perspektif etika-hukum dapat menjadi sulit dalam pengambilan keputusan tingkat tinggi.

Prosesnya panjang dan tidak selalu konsensual, dan kita harus tahu bagaimana kita ingin teknologi diterapkan (atau tidak diterapkan) demi kebaikan masyarakat manusia. Kelayakan dan potensi elemen kerangka hukum untuk pengembangan kecerdasan buatan, berdasarkan standar Dewan Eropa dan supremasi hukum, disajikan dalam laporan (EU (a) 2021) Komite Kecerdasan Buatan (CAHAI). Pilihan-pilihan berikut disajikan: mengubah instrumen hukum yang mengikat dan mengadaptasinya ke sistem AI, memodernisasi instrumen atau protokol yang ada, atau mengadopsi instrumen hukum baru yang mengikat.

Isu-isu yang akan dibahas terkait proses hukum AI dapat terdiri dari tiga jenis. Kelompok pertama meliputi keamanan dan pembelaan hak warga negara atas kompensasi atas kerugian dan kontrol agar sistem AI mematuhi hukum dan tidak melanggar hak-hak yang telah ditetapkan. Kelompok kedua meliputi bagaimana mendefinisikan dan menilai akuntabilitas atas tindakan entitas buatan yang dilengkapi dengan AI dan kapasitas pembelajaran serta pengambilan keputusan otonom?

Haruskah mereka memiliki hak dan kewajiban yang sama dengan orang perseorangan dan dituntut atau dihukum? Atau haruskah tanggung jawab dilimpahkan kepada pencipta atau pengguna sistem? Terakhir, kelompok ketiga menyangkut penggunaan AI oleh aparat penegak hukum dalam penerapan hukum dan kepatuhan terhadap persyaratan etika. Tinjauan yang baik tentang isu-isu hukum ini dapat ditemukan dalam teks karya Dempsey.

Saat ini, legislasi nasional untuk pembingkaian AI masih sangat langka di seluruh dunia dan sistem AI diatur secara longgar. Namun, terdapat sejumlah instrumen hukum internasional yang mengatur aspek-aspek tertentu yang berkaitan dengan sistem AI. Upaya terbesar ke arah ini sedang berlangsung di Uni Eropa (UE). Salah satu hasil dari upaya ini adalah Peraturan Perlindungan Data Umum (EU (b) 2016) (GDPR) yang mulai berlaku pada 25 Mei 2018 (EU (b) (EU (d) 2018)) dan berupaya mewujudkan hak fundamental atas perlindungan data pribadi.

GDPR menetapkan aturan umum dan khusus yang berlaku untuk kategori data pribadi yang sensitif seperti data kesehatan dan memperkenalkan kerangka hukum tunggal di seluruh UE dengan ketentuan yang memungkinkan negara-negara anggota UE untuk memberlakukan undang-undang nasional yang menetapkan, membatasi, atau memperluas beberapa persyaratan. Denda dan sanksi administratif dipertimbangkan. Terdapat juga rezim penelitian khusus yang memberikan fleksibilitas untuk penelitian ilmiah dan statistik.

Inisiatif UE lainnya adalah Proposal untuk Aturan Harmonisasi tentang Kecerdasan Buatan atau Undang-Undang AI (EU (c) 2021). Undang-undang yang diusulkan ini mengklasifikasikan sistem AI sebagai berisiko tinggi (atau tidak) berdasarkan tujuan penggunaannya. Sistem berisiko tinggi (misalnya identifikasi biometrik jarak jauh, evaluasi

kelayakan kredit dan penilaian kredit, pengambilan keputusan peradilan, rekrutmen, dan keputusan ketenagakerjaan lainnya) harus menunjukkan kepatuhan melalui penilaian kesesuaian sebelum diperkenalkan ke pasar, dan penggunaan AI tertentu akan dilarang sama sekali. Klasifikasi risiko ini tidak mencakup penilaian yang tepat atas kerugian manusia atau sosial dan probabilitasnya. Oleh karena itu, tampaknya sulit untuk menyesuaikan regulasi ini dengan perkembangan pasar yang dinamis dan produk-produk AI baru.

Penggunaan AI dalam sistem peradilan merupakan topik yang sangat relevan karena aspek simbolisnya. Cara peradilan memasukkan kriteria efisiensi menggunakan produk AI harus menjadi contoh. Sebuah studi mendalam tentang penggunaan aplikasi AI dalam sistem peradilan disajikan dalam Lampiran Piagam Etika Eropa tentang Penggunaan Kecerdasan Buatan dalam Sistem Peradilan dan Lingkungannya (EU (d) 2018), tetapi beberapa masalah dapat disoroti. Risiko terjebak dalam posisi penerimaan langsung atas keputusan oleh entitas buatan yang konon memiliki kekuatan luar biasa, tetapi tidak dapat diprediksi dan tanpa menjelaskan bagaimana dan mengapa mereka memutuskan, adalah salah satunya.

Gagasan ini meresap ke dalam banyak analisis keadilan prediktif yang memberikan perangkat ini kemampuan langsung atau di masa mendatang untuk memprediksi tindakan manusia atau mengetahui kebenaran dengan lebih baik. Keadilan prediktif ini tidak dapat mencerminkan penalaran hakim manusia secara utuh. Sebuah evolusi yang perlu diatur melalui analisis kritis yang berkelanjutan karena Hukum telah dan harus terus menjadi aktivitas manusia yang didukung oleh teknologi, tetapi tidak pernah tunduk padanya.

6.5 EVALUASI KRITIS AI: RISIKO, PENDIDIKAN, HUKUM

Mengembalikan titik awal kita, AI adalah hasil karya manusia yang tidak boleh diidolakan atau disetankan, melainkan dievaluasi dengan semangat kritis, baik dari segi manfaat maupun risikonya bagi pelestarian umat manusia itu sendiri maupun bagi perkembangannya.

Dalam konteks inilah kami menyoroti beberapa aspek kunci yang perlu diingat dalam perdebatan AI saat ini dan di masa mendatang:

- penerapan teknologi baru dengan karakteristik yang melampaui manusia dan dengan kemampuan otonom dapat menyebabkan perubahan nilai-nilai sosial serta prosedur dan konsep hukum. Namun, tindakan manusia tidak boleh tunduk pada kriteria penilaian yang hanya sesuai untuk makhluk buatan dengan kapasitas spesifik yang superior atau proses pengambilan keputusan yang tak tentu;
- terdapat risiko devaluasi dan penurunan kapasitas manusia secara progresif, alih-alih perkembangan perilaku dan budaya masyarakat yang lebih besar. Masyarakat yang diumumkan dengan pengetahuan yang lebih bebas dapat bergeser menjadi masyarakat yang lebih teregulasi, mematuhi aturan yang diberlakukan oleh teknologi tanpa batas inovasi dengan justifikasi optimalisasi rasionalitas dan efisiensi. Makna hidup cenderung terbatas pada kenikmatan produk teknologi dan kepatuhan terhadap keputusan yang muncul dari algoritma AI;

- pendidikan generasi baru dapat menjadi jalan bagi evolusi masyarakat yang lebih memadai dan menghindari ramalan Stephen Hawking: "akhir umat manusia". Sebuah masyarakat yang tahu bagaimana merefleksikan nilai-nilai esensial dan makna hidup, serta menikmatinya sepenuhnya, tetapi dengan cara yang bijaksana. Salah satu cara untuk pendidikan dan persiapan yang lebih memadai mungkin adalah multidisiplin dalam pelatihan akademis, menghindari spesialisasi yang ketat, dan memberikan pandangan yang lebih baik terhadap berbagai perspektif realitas dan masyarakat manusia;
- adalah ilusi untuk percaya bahwa teknologi hanya memecahkan masalah dan memuaskan keinginan. Teknologi juga menciptakan masalah baru, yang pada akhirnya akan menimbulkan kerusakan sosial yang parah dan tak terelakkan. Intermediasi dan akuntabilitas manusia atas tindakan otonom sistem digital AI merupakan proses perlindungan fundamental bagi umat manusia.

Setelah membahas beberapa isu etika dan menggarisbawahi perlunya membangun konsensus etika yang luas sebagai landasan inisiatif legislatif, kami juga telah menunjukkan beberapa pedoman untuk inisiatif hukum di bidang ini. Tantangan terberat kemungkinan besar terletak di tingkat politik, yang bertujuan membangun tata kelola global di bidang AI.

BAB 7
ROBOT OTONOM DAN CERDAS: ISU SOSIAL, HUKUM, DAN ETIKA

Abstrak Kata "robot" pertama kali digunakan pada tahun 1921 oleh penulis Ceko Karel Cˇ apek, yang menulis sebuah drama berjudul R.U.R. (“*Rosumovi Univerzální Roboti*”), yang menampilkan seorang ilmuwan yang mengembangkan bahan organik sintetis untuk membuat "mesin otonom humanoid", yang disebut "robot". Robot-robot yang disebut ini seharusnya bertindak sebagai budak dan bekerja dengan patuh untuk manusia.

Selama bertahun-tahun, seiring "robot" yang sesungguhnya mulai dibangun, dampaknya terhadap kehidupan kita, pekerjaan kita, dan masyarakat kita telah membawa banyak manfaat, tetapi juga menimbulkan beberapa kekhawatiran. Makalah ini membahas beberapa bidang robotika, kemajuannya, tantangan, dan keterbatasannya saat ini. Kami kemudian membahas tidak hanya bagaimana robot dan otomasi dapat berkontribusi bagi masyarakat kita, tetapi juga mengangkat beberapa kekhawatiran sosial, hukum, dan etika yang dapat ditimbulkan oleh robotika dan otomasi.

7.1 PERKEMBANGAN ROBOT

Robot adalah sistem yang kompleks (biasanya elektromekanis), dilengkapi dengan prosesor, aktuator, sensor, dan baterai. Aktuator dapat berupa roda atau kaki, yang menggerakkan robot, hingga pengeras suara yang memungkinkan robot berkomunikasi melalui ucapan atau sinyal akustik non-verbal, dan termasuk lengan untuk menggenggam atau memanipulasi objek. Kamera video, mikrofon, atau sensor sentuh dan taktil memungkinkan robot untuk meniru beberapa indra manusia, tetapi juga untuk melakukan pengukuran lain, seperti jarak, orientasi, atau kecepatan. Robot membutuhkan prosesor internal, seperti yang terdapat pada komputer yang kita gunakan sehari-hari, agar dapat mandiri dalam hal pengambilan keputusan dan kemampuan bertindak.

Prosesor tersebut menjalankan algoritma yang, dengan tingkat kecanggihan yang lebih tinggi atau lebih rendah, memberikan robot otonomi dan kecerdasan mesin, termasuk kemampuan untuk belajar. Otonomi energetik disediakan oleh baterai internal atau sumber energi terbarukan. Kata "robot" pertama kali digunakan pada tahun 1921 oleh penulis Ceko Karel Cˇ apek, yang menulis drama berjudul R.U.R. ("*Rosumovi Univerzální Roboti*"), yang menampilkan seorang ilmuwan yang mengembangkan bahan organik sintetis untuk membuat "mesin otonom humanoid", yang disebut "robot". "Robot" yang disebut ini seharusnya bertindak sebagai budak dan bekerja dengan patuh untuk manusia.

Selama bertahun-tahun, "robot" sungguhan mulai dibuat, dan pengenalan robot di pabrik-pabrik dimulai pada tahun 1950an. Kendaraan berpemandu otomatis (AGV) pertama, robot bergerak yang mengikuti jalur kabel yang ditanam di dalam tanah, diciptakan pada tahun 1954, tetapi istilah AGV baru diciptakan pada tahun 1980an. Manipulator industri juga digagas pada pertengahan 1950an tetapi baru diperkenalkan di pabrik-pabrik pada awal 1960an.

Robot bergerak pertama yang menggunakan penglihatan dikembangkan di laboratorium penelitian di Amerika Serikat, seperti Stanford Cart dan Shakey.

Sejak saat itu, kemajuan dalam otonomi pesat menuju robot yang digunakan di lingkungan yang kurang terstruktur daripada pabrik, misalnya, rumah, kantor, rumah sakit, jalan, skenario pencarian dan penyelamatan, Bulan atau Mars, yang membutuhkan persepsi dan pengambilan keputusan tingkat lanjut. Robot-robot ini, yang disebut robot layanan, telah berevolusi untuk berinteraksi dengan manusia dalam aktivitas sehari-hari dan bahkan menggantikan manusia dalam pekerjaan rumah tangga, dan lokasi yang tidak dapat diakses/berbahaya.

Meskipun robot industri memicu masalah sosial dengan menggantikan pekerja di pabrik, mereka tidak dapat disangkal menyebabkan pertumbuhan produksi dan peningkatan kekayaan yang, bersama dengan faktor-faktor lain, meningkatkan kesejahteraan, redistribusi kekayaan dan pekerjaan baru yang lebih tidak membosankan dan kurang berbahaya. Di sisi lain, robot layanan mungkin atau mungkin tidak menggantikan pekerjaan manusia dan, bahkan jika mereka melakukannya, jumlah pekerjaan yang hilang bervariasi. Misalnya, robot pembersih vakum membantu pekerjaan rumah tangga, tetapi hampir tidak menggantikan pekerja rumah tangga; Namun, truk otonom dapat menyebabkan hilangnya pekerjaan yang signifikan bagi pengemudi truk. Lebih lanjut, robot layanan yang memiliki komponen interaksi yang kuat dengan manusia juga menimbulkan masalah etika dan hukum: akankah mereka mengungkapkan informasi pribadi rekan manusia mereka? Dapatkah mereka membahayakan manusia?

Permasalahan ini menjadi lebih rumit ketika robot bertindak secara otonom. Meskipun tidak ada definisi "otonomi" yang diterima secara universal untuk robot, kami mengadopsi gagasan bahwa robot otonom adalah sistem "berwujud", dilengkapi dengan sensor untuk mengamati dan memahami dunia sekitarnya, aktuator yang memungkinkannya bertindak di dunia tersebut (kemungkinan termasuk interaksi dengan robot lain, hewan, dan/atau manusia), dan kapasitas pengambilan keputusan yang independen dari kendali eksternal sepenuhnya, yaitu oleh manusia. Perlu dicatat bahwa otonomi adalah istilah yang sarat muatan dalam Kecerdasan Buatan (AI). C. Castelfranchi membahas otonomi sebagai gagasan relasional yang mencakup berbagai dimensi, yang mengarah pada berbagai jenis otonomi, khususnya, "otonomi eksekutif", yang berarti mampu bergerak, bertindak, dan membuat keputusan di dunia tanpa perlu dibantu secara eksplisit untuk melakukannya.

Meskipun hal ini menjadi subjek perdebatan filosofis yang intens, kami juga menganggap bahwa otonomi merupakan syarat yang diperlukan, tetapi tidak cukup bagi robot untuk diberkahi dengan kecerdasan (dalam arti kecerdasan mesin). Dalam hal ini, kecerdasan mesin membutuhkan, selain otonomi, kemampuan robot untuk menyesuaikan perilaku dan tindakannya dengan dunia sekitarnya. Namun, perlu dicatat bahwa perbedaan ini relevan. Pertama, terdapat kesalahpahaman yang luas tentang apa itu robot, yang sering kali membingungkan sistem perangkat lunak cerdas, atau agen "tanpa tubuh", dengan robot otonom. Sebagaimana dikemukakan, robot harus mampu merasakan dan bertindak secara fisik di dunia fisik. Kedua, tidak semua robot cerdas atau otonom, dan banyak, misalnya banyak

drone mainan, dioperasikan dari jarak jauh dan dikendalikan oleh manusia, di mana kecerdasan dan otonomi mereka tidak ada. Seringkali, perdebatan publik tentang isu-isu etika dan sosial yang ditimbulkan oleh robot membingungkan sistem perangkat lunak umum, yang diberkahi dengan kecerdasan buatan, dengan robot dan menganggap semua robot otonom sebagai cerdas.

Kami menganggap bahwa robot otonom, sebagaimana didefinisikan, mengingat karakteristik spesifiknya, membawa masalah sosial, etika, dan hukum baru, yang akan kami bahas dalam bab ini. Bab ini disusun sebagai berikut: pertama, kami memberikan gambaran singkat tentang robot industri, diikuti oleh robot layanan. Kemudian, kami membahas potensi robot-robot ini untuk ditempatkan di lingkungan sosial, dan bagaimana kecerdasan dibutuhkan untuk interaksi sosial dengan manusia. Mengingat jenis-jenis robot ini, kami kemudian membahas implikasi sosial, etika, dan hukum dari integrasi mereka dalam masyarakat kita.

7.2 ROBOT INDUSTRI DAN OTOMASI VS. ROBOT LAYANAN

Robot diperkenalkan di pabrik-pabrik untuk mengotomatiskan tugas-tugas berulang yang sebelumnya dilakukan oleh manusia. Robot-robot ini mencakup manipulator robot yang meniru lengan manusia dalam berbagai operasi: mengambil benda dari palet di kendaraan pengangkut atau dari ban berjalan dan menempatkannya ke dalam sel-sel manufaktur, lalu kembali dari sana ke ban berjalan atau kendaraan pengangkut lainnya; merakit komponen menjadi benda yang lebih kompleks; mengecat dan mengelas. Robot-robot ini juga mencakup robot bergerak, dalam bentuk AGV atau LGV (kendaraan berpemandu laser, yang tidak memerlukan kabel terpendam atau garis yang dicat di tanah) untuk membawa benda secara otonom antar lokasi berbeda di pabrik industri.

Ciri umum dari semua aplikasi dan skenario ini adalah sifatnya yang terstruktur. Lokasi ban berjalan, pos pengambilan dan penempatan, serta sel-sel manufaktur/perakitan dengan mesin, sudah dikenal luas, statis, dan mudah dikenali. Dalam kebanyakan kasus, objek disalurkan ke lokasi yang sangat presisi di mana mereka diambil oleh manipulator, dan stasiun bongkar muat memiliki mekanisme penjepit dan pengikat yang memaksa objek untuk terikat erat pada platform pengangkutnya.

Robotika industri juga umumnya disebut sebagai otomatisasi, karena robot yang terlibat melakukan operasi otomatis, tetapi tidak otonom dalam arti sempit. Dalam kebanyakan kasus, otomatisasi industri tradisional tidak memerlukan sensor seperti penglihatan untuk menemukan objek yang akan diambil, atau lokasi penempatan yang paling tepat untuk objek tersebut. Otomasi industri juga tidak menangani objek yang dapat dideformasi seperti makanan atau kemasan lunak.

Pada abad terakhir, film dokumenter tentang robot yang mengotomatiskan produksi di pabrik konstruksi, perakitan, pengecatan, pengangkutan komponen, dan pengelasan memukau masyarakat umum. Namun, penelitian robot yang lebih modern dan menantang berupaya menciptakan mesin yang mampu menangani lingkungan yang kurang terstruktur dan kurang dapat diprediksi, seperti rumah kita atau bahkan lingkungan luar ruangan, yang dihuni

oleh manusia dan agen lain yang tidak berperilaku deterministik seperti di lingkungan pabrik. Ini disebut robot layanan.

Gambar 7.1 Robot layanan untuk konstruksi dan transportasi batu bata

Robot layanan beragam, mulai dari robot penyedot debu yang sukses secara komersial hingga penjelajah planet yang menjelajahi permukaan Mars. Penyedot debu menjelajahi rumah, menjangkau area seluas mungkin, sambil menghindari objek tak terduga (seperti benda yang tertinggal di lantai, kaki meja dan kursi, atau kaki seseorang) yang terdeteksi oleh sensor penglihatan dan pemindaian laser internal.

Penjelajah Mars bergerak melintasi medan sulit yang perlu mereka amati sebelum langkah selanjutnya, menuju lokasi-lokasi ilmiah yang sebelumnya telah diidentifikasi oleh kamera internal mereka. Robot layanan juga mencakup mobil otonom, tim pencarian dan penyelamatan kendaraan tak berawak heterogen (darat, udara), robot medis untuk membantu ahli bedah manusia dalam melakukan operasi, atau robot yang membantu pasien di rumah sakit dan fasilitas pelayanan kesehatan, robot pertanian, drone pengintai, dan banyak lagi.

Fitur umum dalam robot layanan adalah mereka beroperasi di lingkungan yang tidak terstruktur, seringkali belum dikenal sebelumnya, di mana sensor sangat penting untuk membangun kesadaran situasional oleh robot, sehingga dapat mendukung penalaran dan pengambilan keputusannya. Robot layanan tidak dapat bertindak secara otomatis. Robot-robot tersebut harus otonom atau, setidaknya, memiliki tingkat independensi yang tinggi dari operator jarak jauh manusia. Oleh karena itu, robot-robot tersebut menimbulkan banyak masalah etika dan hukum baru (misalnya, tindakan apa yang harus dipilih robot ketika terdapat alternatif dan memiliki dampak yang berbeda terhadap keselamatan manusia; apa yang harus dilakukan mobil otonom untuk memastikan kepatuhannya terhadap aturan berkendara) yang sebelumnya tidak muncul dalam robot industri, yang dampak utamanya bersifat sosial, yaitu terkait hilangnya pekerjaan.

Memang, sebagian besar robot layanan cenderung merambah dalam operasi yang tidak umum dilakukan oleh manusia, seperti skenario yang tidak berulang dan/atau berbahaya, sehingga dampak sosialnya relatif kecil. Namun demikian, robot-robot tersebut mulai memasuki skenario industri (misalnya, menggunakan sensor gaya untuk memberi robot kemampuan menghindari bahaya bagi manusia, sehingga mengurangi ruang yang ditempati oleh sel robot dan sistem pengamannya, lihat Baxter pada Gambar 7.2; untuk melakukan

tindakan "pick and place" di lingkungan yang kurang terstruktur, paket lunak, dan material) dan operasi besar seperti taksi dan truk otonom, yang dapat menyebabkan penggantian tenaga kerja manusia secara besar-besaran oleh mesin otonom.

Penelitian terkini tentang robot layanan sangat berfokus pada robot yang berkolaborasi dengan manusia, bukan robot yang menggantikan manusia. Robot pencarian dan penyelamatan dikembangkan untuk berkolaborasi dengan tim Perlindungan Sipil; robot medis membantu dokter dan perawat di rumah sakit, dan penjelajah planet memperluas jangkauan keingintahuan manusia hingga penjelajahan Mars. Hal ini juga menimbulkan pertanyaan sosial lain yang menantang dan menarik: bagaimana seharusnya robot bertindak agar dapat berinteraksi sealami mungkin dengan manusia? Apa artinya bertindak secara sosial?

Gambar 7.2 Robot Baxter untuk pabrik kecil

7.3 ROBOT DAN MANUSIA: KEBANGKITAN ROBOT CERDAS DAN SOSIAL

Apakah robot penyelamat, saat berinteraksi dengan manusia dalam situasi darurat, dapat dianggap sebagai robot sosial? Atau drone yang terbang dalam formasi bersama drone lain untuk mengatasi rintangan? Kata "Sosial" berasal dari kata Latin "socii", yang berarti teman atau sekutu. Konsep "sosial" secara umum dikaitkan dengan perilaku yang mempertimbangkan orang lain, minat, motivasi, dan kebutuhan mereka. Seseorang dianggap sosial jika ia memiliki kemampuan untuk berinteraksi dan mempertimbangkan orang lain dalam tindakannya, dan dengan demikian membangun hubungan sosial. Namun, "sosialitas" dalam robot dapat mencakup berbagai perspektif atau bahkan tingkatan. Banyak robot layanan dapat diklasifikasikan sebagai makhluk yang "menggugah emosi secara sosial". Misalnya, robot bermata besar, seperti robot Vizzy yang dibuat oleh Institut ISR di Lisbon atau Roomba, robot penyedot debu yang bergerak dengan penuh tujuan di dalam rumah: keduanya dapat membangkitkan respons yang bersifat sosial dan emosional. Hanya perwujudan fisik dan tindakan otonom mereka yang cukup untuk bertindak sebagai antarmuka alami untuk mendapatkan respons seperti manusia, bahkan jika robot itu sendiri sebenarnya tidak mampu merespons dengan cara yang cerdas dan sosial.

Lebih jauh lagi, hanya dengan ditempatkan dalam lingkungan sosial, robot dapat menjadi reseptif secara sosial, yaitu, mendapatkan manfaat dari interaksi dengan orang lain, belajar dari "guru" manusia dan dengan demikian, meningkatkan kinerja mereka. Namun,

karena lebih banyak robot diperlukan untuk melakukan aktivitas dalam pengaturan yang berpusat pada manusia, mereka akan diberikan "kompetensi sosial". Robot sosial dianggap dapat memahami satu sama lain dan manusia, terlibat dalam interaksi sosial, memiliki sejarah (memahami dan menafsirkan dunia dalam hal pengalaman mereka sendiri), berkomunikasi secara eksplisit dengan manusia dan belajar dari mereka.

Gambar 7.3 Contoh robot sosial—dari kiri ke kanan: Vizzy, Pepper, ASTRO dan MBOT

Namun, robot sosial sering kali dirancang untuk menjalankan tugas-tugas yang pada dasarnya mungkin tidak "sosial". Misalnya, pertimbangkan robot dalam pengaturan perawatan kesehatan yang dirancang untuk mengangkut bahan dari satu tempat ke tempat lain di rumah sakit. Sebagian besar pekerjaannya, seperti membawa obat-obatan, atau linen, tidak selalu bersifat sosial.

Namun, kompetensi sosial, jika ada, dapat memperkaya interaksi yang mereka jalin dengan manusia di sekitar mereka, dan meningkatkan kinerja mereka. Misalnya, robot perawatan kesehatan mungkin dapat mengenali perawat, merespons dan menjalankan perintah yang diberikan dalam bahasa alami, berinteraksi dengan pasien, dan memberikan informasi saat dibutuhkan. Contoh lain adalah robot pembersih vakum kami, yang dapat diberikan beberapa kompetensi sosial, seperti menghindari atau berinteraksi dengan manusia, atau menyesuaikan tindakannya dengan kebiasaan anggota rumah tangga, sehingga kinerjanya lebih efisien.

Jadi, kompetensi sosial tersebut dapat dilihat sebagai batu loncatan bagi robot untuk menjadi anggota aktif dalam kehidupan dan masyarakat kita. Dari sudut pandang teknis, hal ini memerlukan pengembangan kompetensi sosial, yang mencakup kemampuan mengenali manusia, memahami tindakan mereka, mempersepsi emosi mereka, menggunakan bahasa alami dan isyarat non-verbal, serta secara umum mengenali, "memahami", dan bernalar tentang situasi sosial yang akan mereka hadapi. Namun, membangun kemampuan sosial ini membutuhkan teknik dan algoritma AI yang canggih. Untuk mempersepsi manusia, menangkap tindakan dan emosi mereka, diperlukan teknik dari penglihatan dan pemrosesan sinyal sosial. Untuk menghasilkan tindakan, diperlukan algoritma perencanaan otomatis. Metode pemrosesan bahasa alami dan ucapan sangat penting jika kita ingin robot berinteraksi secara alami dan seperti manusia dengan manusia.

Lebih lanjut, karena kita juga membutuhkan robot untuk dapat beradaptasi dan belajar menjalankan tugas, kita perlu menggunakan algoritma pembelajaran mesin. Faktanya, banyak teknik AI utama yang sedang dikembangkan dalam AI saat ini sangat penting untuk

membangun robot sosial cerdas yang mampu bertindak dalam ranah dinamis dan sosial. Lebih lanjut, robot sosial merupakan tempat uji coba yang ideal untuk integrasi teknik-teknik tersebut. Domain aplikasi tipikal untuk robot sosial sangat luas, dan mencakup layanan kesehatan, transportasi, logistik, kebersihan, pendidikan, hiburan, pertanian, dan lainnya.

Dalam konteks layanan kesehatan, telah terjadi perkembangan yang signifikan dalam beberapa tahun terakhir, dengan peningkatan yang signifikan sejak pandemi COVID-19. Robot diperkenalkan di fasilitas layanan kesehatan untuk mengangkut material dan perlengkapan, terutama dalam situasi di mana pengangkutan tersebut dapat menimbulkan risiko paparan patogen, seperti virus. Penggunaan penting lainnya dari robot sosial adalah untuk terapi dan perawatan, khususnya bagi lansia dan pasien demensia. Sebuah studi yang menganalisis penggunaan robot PARO (robot mirip anjing laut) di fasilitas perawatan di rumah di Jepang, telah menunjukkan dampak positif robot dalam mengurangi stres dan menenangkan pasien demensia, serta memberikan manfaat tidak langsung dengan meningkatkan aktivitas mereka dalam interaksi sosial tertentu.

Studi lain menunjukkan bahwa penggunaan robot rumahan untuk lansia di daerah pedesaan Selandia Baru menghasilkan peningkatan kualitas hidup, kemandirian dan otonomi yang lebih besar bagi lansia, serta penurunan kunjungan ke layanan kesehatan primer dan panggilan telepon ke praktisi kesehatan. Hasil ini merupakan tanda yang menggembirakan bahwa teknologi ini dapat memberikan dampak sosial yang positif bagi masyarakat kita yang menua.

Bidang transportasi mungkin merupakan salah satu bidang di mana robot layanan menunjukkan peningkatan terbesar seiring dengan semakin banyaknya kendaraan otonom yang ditempatkan di jalan raya. Jalan raya, pada dasarnya, merupakan lingkungan sosial, yang berarti bahwa keputusan otonom oleh kendaraan harus mempertimbangkan keberadaan pengemudi lain serta pejalan kaki. Oleh karena itu, mobil otonom dibekali dengan kompetensi (dalam prediksi dan tindakan) yang berkaitan dengan interaksi sosial. Lebih lanjut, dampak sosial dari potensi peningkatan penggunaannya di jalan raya tidak diragukan lagi cukup besar. Meskipun dampak ini telah dibayangi oleh prediksi yang berlebihan bahwa mobil otonom akan mendominasi jalan raya pada tahun 2020, kita tidak dapat mengabaikan implikasi sosial, etika, dan hukum yang akan ditimbulkan oleh kendaraan otonom di masa depan.

Bidang aplikasi lain seperti kebersihan dan logistik juga meningkat, dan sekali lagi, pandemi memunculkan serangkaian aplikasi di mana robot dapat digunakan untuk menyediakan cara yang aman dan efisien dalam melakukan pekerjaan mereka. Bidang aplikasi robotika ini, di mana robot terintegrasi dalam lingkungan sosial kita, menimbulkan kekhawatiran di masyarakat umum, produsen, dan pembuat undang-undang.

Robot, yang masih merupakan pasar yang belum diatur secara luas, mungkin di masa depan akan ditempatkan dalam lingkungan di mana mereka berinteraksi dengan manusia, mengumpulkan informasi pribadi, memengaruhi tindakan mereka, dan berdampak besar pada pasar kerja yang tidak stabil. Namun, seperti yang telah disebutkan sebelumnya, beberapa kekhawatiran ini masih belum berdasar, dan ekosistem yang sedang dibangun untuk pengenalan AI ke dalam masyarakat kita dan undang-undang yang sedang disusun saat kita

menulis ini, merupakan perlindungan bagi robot kita. Dalam makalah ini, kami memaparkan beberapa implikasi sosial, etika, dan hukum dari teknologi baru yang menarik ini.

7.4 DAMPAK ETIKA, SOSIAL, DAN HUKUM

Agar robot dapat berhasil sebagai teknologi yang menjadikan dunia kita lebih baik, kita harus melibatkan para peneliti, perancang, pengembang, insinyur, perusahaan, dan pembuat undang-undang, untuk membangun ekosistem di mana robot tepercaya, efektif, aman, dan relevan bagi masyarakat kita. Persepsi masyarakat umum saat ini tentang robot otonom seringkali membayangkan kemampuan futuristik pada robot. Robot digambarkan mampu melakukan pekerjaan luar biasa dan menangani berbagai tugas serta masalah. Dan, meskipun teknologinya masih cukup terbatas, banyak ketakutan dan kekhawatiran yang tidak beralasan telah muncul di masyarakat umum.

Diskusi tentang "robot pembunuh", atau "robot untuk lansia", telah menyerbu ruang opini publik. Namun, dalam banyak kasus, kekhawatiran ini layak untuk diperdebatkan secara mendalam dan didekati secara serius. Kondisi diskusi yang (masih) belum matang ini, yang dapat dimaklumi mengingat relatif barunya bidang ini, berarti bahwa hal-hal yang sifatnya berbeda seringkali dikaitkan dengan masalah etika yang diakibatkan oleh persepsi yang berlebihan terhadap robot. Dalam bab ini, kami akan mencoba mengangkat dan membahas beberapa kekhawatiran tersebut, serta membedakan antara perdebatan etika, sosial, dan hukum yang perlu ada seputar teknologi baru ini.

Isu-Isu Etika

Bagaimana seharusnya robot otonom bereaksi dalam situasi di mana keputusannya dapat merugikan manusia? Bagaimana dengan perlindungan privasi manusia ketika, misalnya, robot asisten rumah tangga berkeliaran di sekitar rumah sambil membawa kamera dan berinteraksi dengan manusia dengan cara yang dapat mengungkapkan perilaku intimnya? Haruskah robot otonom dilibatkan dalam perawatan kesehatan, mulai dari pemantauan lansia atau anak-anak hingga intervensi bedah? Dan apa dampak dari pengenalan perangkat bionik (prostesis, eksoskeleton) secara progresif pada manusia, yang suatu hari nanti dapat menyebabkan kesulitan membedakan antara manusia dan robot? Pertanyaan-pertanyaan ini mengarah pada masalah etika yang perlu ditangani seiring dengan perkembangan robot. Pertanyaan-pertanyaan ini perlu dijawab oleh produsen robot, peneliti, dan pembuat undang-undang secara kolaboratif.

Faktanya, diskusi seputar etika pengambilan keputusan dan perilaku robot otonom semakin menguat dan relevan dengan kesadaran akan kemungkinan masifnya mobil tanpa pengemudi (atau otonom). Seiring Google/Waymo Car dan kendaraan lain dari perusahaan manufaktur mobil mulai memasuki kehidupan kita sehari-hari, mereka menghadapi semakin banyak situasi, terutama di lingkungan perkotaan, di mana mereka harus mengambil keputusan secara otonom. Contoh-contoh umum yang mewakili situasi ini sangat banyak.

Pertimbangkan situasi ini: sebuah kendaraan otonom bergerak dengan kecepatan tinggi dan mendeteksi sekelompok pejalan kaki yang menyeberang jalan secara tak terduga; tabrakan yang berpotensi fatal tidak dapat dihindari tanpa kendaraan tersebut memutuskan

untuk meninggalkan jalan, yang pada akhirnya menabrak pejalan kaki yang berjalan di trotoar. Apa yang seharusnya menjadi keputusan kendaraan tersebut:

- (1) maju, menabrak pejalan kaki di jalan, atau menyimpang, menabrak pejalan kaki di trotoar?
- (2) meninggalkan jalan, yang pada akhirnya mengorbankan nyawa penumpangnya, atau terus melaju, menabrak pejalan kaki yang menghalangi jalannya?

Dilema-dilema semacam ini telah dieksplorasi dalam proyek mesin moral yang diciptakan untuk mengeksplorasi dilema moral yang dihadapi oleh kendaraan otonom. Platform daring ini menyajikan dilema moral kepada pengguna yang harus memilih antara dua kemungkinan hasil buruk, seperti menewaskan tiga penumpang di dalam mobil otonom atau menewaskan tiga pejalan kaki. Platform ini telah digunakan untuk mengumpulkan jutaan keputusan dalam sepuluh bahasa berbeda dan di 233 negara. Data menunjukkan bahwa orang lebih memilih menyelamatkan manusia daripada hewan, dan menyelamatkan lebih banyak nyawa dan nyawa anak-anak. Studi ini penting karena memberikan data kepada para pembuat kebijakan tentang cara menangani situasi di mana mesin mungkin harus memutuskan siapa yang harus hidup atau mati.

Isu etika pengambilan keputusan oleh sistem robotik mulai ditangani secara serius oleh beberapa negara dan organisasi di dunia, dimulai dari dokumen yang dihasilkan oleh British Standards Institute pada tahun 2016, dengan pedoman tentang aturan etika yang harus diikuti dalam perancangan sistem robot oleh manajer dan perancang. Demikian pula Inisiatif Global IEEE tentang Etika Sistem Otonom dan Cerdas ("IEEE Global Initiative") menghasilkan dokumen Desain yang Selaras Secara Etis yang memberikan panduan kepada pengembang, pemerintah, bisnis, dan publik, tentang cara menangani, merancang, menggunakan, dan menetapkan aturan untuk kemajuan sistem otonom yang berkontribusi pada masyarakat.

Dalam beberapa tahun terakhir, Kelompok Pakar Tingkat Tinggi tentang AI (AI HLEG) dari Komisi Eropa telah mengeluarkan serangkaian pedoman etika untuk mencapai AI yang dapat dipercaya. Jelas, saat kita berurusan dengan robot otonom, yang diberkahi dengan algoritma AI yang berbeda yang digunakan untuk fungsinya, pedoman ini juga dapat berlaku. Oleh karena itu, kita dapat mengekstrapolasi pedoman tersebut ke robot cerdas: (1) Agensi dan Pengawasan Manusia - robot dan sistem robot harus menghormati agensi manusia dan mendukung pengawasan pelaksanaannya; (2) Ketahanan dan Keamanan Teknis - robot harus kuat dan aman saat berinteraksi dengan manusia dan dalam masyarakat kita; (3) Privasi dan Tata Kelola Data - robot harus mengikuti aturan privasi yang ditetapkan dan mekanisme tata kelola data; (4) Transparansi - robot harus transparan saat membuat keputusan, dan tentang kemampuan mereka, menjelaskan dengan jelas mengapa keputusan tertentu tepat; (5) Keragaman, Non-diskriminasi dan Keadilan - Robot harus menghormati tidak mendiskriminasi atau menyebabkan diskriminasi, dan menjamin keadilan dalam keputusan mereka; (6) Kesejahteraan Lingkungan dan Masyarakat - robot harus memupuk kesejahteraan masyarakat dan berkontribusi pada masyarakat dan lingkungan yang lebih baik; (7) Akuntabilitas - proses akuntabilitas dan ekosistem yang jelas harus ada dan diikuti oleh produsen robot, menjamin bahwa ketika masalah terjadi, proses tersebut dapat dipicu.

Penerapan pedoman ini telah mengarah pada bidang Robotika Bertanggung Jawab yang membahas "desain, pengembangan, penggunaan, implementasi, dan regulasi robotika yang bertanggung jawab di masyarakat". Khususnya robot medis dan layanan kesehatan, menimbulkan masalah etika yang sangat relevan. Robot mulai memasuki rumah sakit dengan cara yang sangat berbeda. Yang paling dikenal dan mungkin paling berdampak hingga saat ini adalah robot yang mendukung ahli bedah dalam melakukan operasi, meningkatkan akurasi, dan menyaring tremor yang tak terhindarkan bahkan pada spesialis terbaik sekalipun. Namun, selama beberapa tahun terakhir, robot bergerak telah mengangkut makanan, obat-obatan, dan berbagai instrumen antar area rumah sakit, sehingga membebaskan staf medis dan perawat untuk melakukan tugas-tugas yang lebih dekat dengan pasien. Ada contoh yang lebih baru, masih dalam tahap embrio, robot yang berinteraksi dengan lansia dan anak-anak, yang berupaya meningkatkan kondisi klinis mereka dengan mendorong olahraga atau melakukan permainan interaktif.

Ada juga langkah-langkah lain yang diambil untuk mengatasi beberapa masalah etika ini, yang mempertanyakan peran robot, dan mendorong pengembangan "robot kolaboratif". Gagasan utamanya adalah, alih-alih menggantikan pekerja dengan mesin dalam menjalankan tugas-tugas yang membutuhkan pengetahuan dan pengalaman profesional yang mendalam, fokuslah pada tugas-tugas di mana robot dapat membebaskan dokter dan perawat untuk fokus pada aktivitas utama mereka. Contohnya adalah robot yang mengangkut makanan dan obat-obatan ke kamar-kamar rumah sakit, robot yang menyediakan akses jarak jauh kepada pasien yang sangat menular, atau robot yang memberikan bantuan kepada pasien yang tidak memerlukan kehadiran manusia yang lebih penuh kasih sayang.

Isu Sosial

Pengenalan robot secara besar-besaran ke dalam masyarakat dapat memberikan kontribusi positif bagi masyarakat, tidak hanya dalam hal positif, tetapi juga melalui dampaknya terhadap pekerjaan, harga diri, dan/atau perilaku manusia. Kontroversi yang ditimbulkan oleh penggantian manusia oleh mesin dalam aktivitas kerja bukanlah hal baru, dan tidak terbatas pada hilangnya pekerjaan, yang pada kenyataannya, tidak terjadi di masa lalu. Pada tahun 1821, di puncak revolusi industri, ekonom David Ricardo menyatakan bahwa pengenalan mesin akan merugikan kepentingan kelas pekerja, terutama karena kekayaan yang diciptakan terutama menguntungkan mereka yang hidup dari pendapatan modal. Namun, otomatisasi di masa lalu telah meningkatkan kondisi kehidupan masyarakat di mana otomatisasi tersebut diterapkan, dan telah menyediakan pekerjaan dengan gaji yang lebih baik, lebih manusiawi, dan lebih aman.

Oleh karena itu, pertanyaan yang perlu diajukan adalah apakah revolusi saat ini akan berbeda. Pers internasional telah mengemukakan perkiraan paling mengerikan tentang konsekuensi robotisasi masyarakat. Menurut studi tahun 2013 oleh Carl B. Frey dan Michael Osborne dari Universitas Oxford, 47% pekerjaan di AS berisiko digantikan oleh "modal komputer". Studi yang lebih baru oleh Merrill Lynch memprediksi bahwa, pada tahun 2025, dampak tahunan "disrupsi kreatif" yang dihasilkan oleh Kecerdasan Buatan dapat mencapai 140–330 triliun (triliun rupiah), termasuk pengurangan biaya tenaga kerja berbasis

pengetahuan sebesar Rp 90 miliar, yang digantikan oleh mesin; Rp 80 miliar di bidang manufaktur dan perawatan kesehatan; Rp 20 miliar yang dihasilkan dari penggunaan kendaraan otonom dan drone.

Masalah utama yang mendasari semua angka ini adalah bahwa mereka pada dasarnya merupakan hasil dari perkembangan agen otonom cerdas yang tidak "diwujudkan" dan tidak berinteraksi dengan dunia sekitarnya kecuali melalui papan ketik dan monitor komputer. Prediksi ini dapat diapresiasi mengingat situasi saat ini dengan peningkatan penggunaan ponsel pintar, atau agen pencarian Internet (misalnya, Google, agen perjalanan), atau sistem rekomendasi, yang menunjukkan bahwa Kecerdasan Buatan (AI) dengan cepat menempatkan banyak pekerjaan dalam risiko sebuah transformasi yang, menurut McKinsey Global Institute, terjadi sepuluh kali lebih cepat, dan dalam skala tiga ratus kali lipat dari masa lalu. Namun masalahnya akan lebih besar jika hal yang sama terjadi dengan Robotika, karena pelatihan ulang pekerja yang terspesialisasi dalam tugas fisik, tidak intensif dalam pengetahuan, bisa jauh lebih rumit, terutama pada tingkat perubahan di mana perubahan terjadi.

Namun, ternyata perkembangan teknologi Robotika, terlepas dari banyak kemajuan terkini, jauh lebih sulit, lebih kecil, dan bahkan kendaraan otonom, yang menjanjikan penampilan memukau, akan membutuhkan waktu bertahun-tahun untuk sepenuhnya menggantikan kendaraan yang dikemudikan pengemudi misalnya, sebagaimana dibuktikan oleh kecelakaan fatal yang terkenal di AS dengan mobil Tesla yang menggunakan autopilot, akibat terlalu percaya diri pengemudi dan perusahaan manufaktur. Situasi ini bahkan lebih mencolok ketika kita berbicara tentang robot yang membantu pekerjaan rumah tangga, atau di rumah sakit, di pertanian, atau bahkan di pabrik-pabrik modern, yang lebih fleksibel dan dengan pekerjaan yang kurang repetitif. Robot-robot ini tidak hanya jauh dari kata otonom, tetapi banyak yang dirancang untuk berkolaborasi dengan manusia.

Oleh karena itu, kita mempertimbangkan dua realitas yang berbeda, meskipun biasanya kita menyaksikan hubungan antara Robotika (AI yang diwujudkan) dan AI (tanpa perwujudan). Namun, dalam kedua kasus tersebut, ada kekhawatiran dan risiko yang perlu dipertimbangkan dengan cermat. Manfaat yang dibawa oleh otomatisasi tidak dapat membuat kita menyerah untuk mencari pekerjaan dan profesi lain bagi mereka yang kehilangan pekerjaan mereka saat ini seperti pekerjaan kreatif atau pemeliharaan dan produksi robot.

Dan mereka tidak boleh mengalihkan kita dari masalah sosial yang layak mendapat perhatian kebijakan publik, yang bahkan dapat melewati penciptaan pendapatan minimum wajib, dan undang-undang yang memaksa perusahaan yang menjadi kurang bergantung pada pekerjaan manusia untuk (1) melatih ulang atau merelokasi pekerja mereka dan/atau (2) membayar pajak dan kontribusi jaminan sosial yang proporsional dengan penciptaan kekayaan yang dihasilkan dari penggabungan robot dan teknologi AI dalam produksi mereka.

Di atas segalanya, dan kembali ke kekhawatiran beberapa ekonom selama revolusi industri, kita sebagai masyarakat tidak boleh membiarkan bahwa kekayaan yang lebih besar yang dihasilkan oleh teknologi ini tetap berada di tangan sangat sedikit orang, yaitu

perusahaan-perusahaan yang memiliki teknologi tersebut. Risiko hal ini terjadi jika kita tidak bertindak secara tidak proporsional lebih besar saat ini daripada di abad kesembilan belas.

Isu Hukum

Refleksi tentang etika pengambilan keputusan seringkali mengarah pada diskusi tentang isu hukum, yaitu bagaimana (dan kepada siapa) menetapkan tanggung jawab hukum atas keputusan tersebut. Pertanyaan seperti siapa yang bertanggung jawab secara hukum atas tindakan robot otonom? Seberapa jauh robot pengawas dapat bertindak tanpa mengganggu keamanan dan/atau privasi warga negara? Bagaimana kekayaan intelektual dilindungi terkait penemuan yang dilakukan dengan bantuan agen atau robot? Lebih lanjut, jika robot 1 hari disamakan dengan manusia, atau hewan, dalam arti memiliki identitas sendiri, haruskah hak-hak mereka juga dilindungi?

Komisi Eropa telah menjadi garda terdepan dalam regulasi, dengan proposal baru untuk kerangka kerja regulasi Uni Eropa tentang kecerdasan buatan (AI) yang diluncurkan pada April 2021. Kerangka hukum yang diusulkan berfokus pada pemanfaatan spesifik sistem AI dan risiko terkait, dengan fokus utama pada jaminan kepercayaan dalam proses pembuatan dan penyediaan sistem cerdas. Meskipun merupakan upaya pertama yang patut dikagumi untuk memastikan AI digunakan dengan cara yang dapat dipercaya oleh perusahaan dan pengguna, beberapa aspek yang berkaitan dengan perwujudan, dan dengan demikian, robot cerdas, masih belum tersentuh.

Lebih lanjut, peraturan baru ini dapat menimbulkan masalah lain, karena tidak jelas siapa yang akan bertanggung jawab untuk menerapkan hukum dan menjamin kepatuhannya. Akal sehat mungkin menunjukkan bahwa hukum dan pedoman tersebut ditujukan untuk perancang, produsen, dan operator robot, tetapi mengingat otonomi robot, bukankah seharusnya robot dibekali dengan kapasitas kesadaran diri sehingga, setelah mengevaluasi situasi, ia dapat memutuskan sendiri untuk menerapkan atau tidak semua aturan lain yang menentukan operasinya? Kita tidak boleh lupa bahwa robot pada awalnya dapat diterapkan dengan kemampuan yang terus meningkat seiring waktu.

Oleh karena itu, isu-isu terkait etika sistem robot yang berinteraksi dengan manusia mengarah pada atribusi tingkat tanggung jawab hukum atas potensi kecelakaan, dan atas kerusakan yang ditimbulkannya, sebanding dengan jumlah instruksi yang awalnya diprogram dalam robot versus jumlah otonomi yang diperoleh melalui pembelajaran, bahkan tanpa intervensi langsung dari pemrogramnya. Dengan cara ini, robot otonom cerdas dengan pengalaman bertahun-tahun dan, selama itu ia mempelajari perilaku dan tindakan baru, akan memikul tanggung jawab hukum yang lebih besar.

Namun, mengevaluasi rasio otonomi yang diajarkan oleh perancang dalam kaitannya dengan yang dipelajari oleh robot tentu sulit, dan alternatif yang lebih pragmatis adalah memperkenalkan asuransi wajib, atau seperti yang diusulkan oleh Uni Eropa, memastikan bahwa keputusan yang diambil transparan dan dapat diperiksa oleh entitas eksternal. Kami sepakat bahwa kerangka hukum seperti yang diusulkan oleh Uni Eropa, yang merangkul teknologi terkini untuk menjamin penggunaannya yang tepat, sehat, dan positif dalam masyarakat kita sangatlah penting. Namun, kita tidak boleh melebih-lebihkan regulasi, karena

robot otonom masih dalam tahap awal, dan melegalkan pembuatan dan penggunaannya terlalu dini dapat meredam inovasi dan mengorbankan potensi manfaat sosial yang dapat mereka bawa, belum lagi membuat wilayah lain di dunia berada dalam keuntungan yang tidak adil dalam hal penelitian dan inovasi.

7.5 HUKUM ROBOTIKA ASIMOV: DILEMA DAN MANFAAT

Pada tahun 1939, penulis visioner Rusia/Amerika Isaac Asimov, dalam bukunya I, Robot, menetapkan apa yang disebut Tiga Hukum Robotika: Hukum 1 - robot tidak boleh melukai manusia atau, karena tidak bertindak, membiarkan manusia terluka; Hukum 2 - robot harus mematuhi perintah yang diberikan oleh manusia kecuali dalam kasus di mana perintah tersebut bertentangan dengan 1; dan Hukum 3 - robot harus melindungi keberadaannya sendiri selama perlindungan tersebut tidak bertentangan dengan 1 atau 2.

Terlepas dari kesederhanaan hukum-hukum ini, Asimov mampu menghasilkan banyak dilema yang menghibur dan dipikirkan dengan matang yang mengeksplorasi kesulitan yang kita hadapi dalam memperkenalkan mesin otonom ke dalam masyarakat kita. Memang, ini adalah masalah yang sulit, dan di sini kami hanya menunjukkan sedikit gambarannya. AI dan robotika pasti akan mengubah cara kita hidup dan berfungsi dalam masyarakat.

Suatu hari nanti, keturunan kita akan bertanya-tanya bagaimana mungkin mobil dikemudikan oleh manusia dengan segala risiko yang ditimbulkannya; atau mengapa seorang pekerja perlu melakukan upaya yang luar biasa untuk mengangkat beban berlebih yang berbahaya bagi kesehatannya). Kami percaya bahwa AI dan penggunaannya dalam Robotika untuk menciptakan robot cerdas dan otonom akan menjadi pendorong perubahan sosial yang akan berkontribusi pada masyarakat yang lebih baik, lebih manusiawi, lebih berkelanjutan, dan lebih sehat.

BAB 8
TANTANGAN ETIKA DAN HUKUM AI DI SISTEM REKOMENDASI

Abstrak Di dunia yang sangat terhubung, sistem rekomendasi (RS) merupakan salah satu aplikasi komersial kecerdasan buatan (AI) yang paling luas. Awalnya, sistem ini banyak digunakan untuk e-commerce, tetapi kini telah diterapkan secara luas di berbagai bidang, misalnya, penyedia konten dan platform media sosial. Karena kelebihan informasi saat ini, sistem ini dirancang terutama untuk membantu individu menghadapi beragam pilihan yang tersedia, selain mengoptimalkan keuntungan perusahaan dengan menawarkan produk dan layanan yang secara langsung memenuhi kebutuhan pelanggan mereka. Namun, terlepas dari manfaatnya, RS berbasis AI juga dapat menimbulkan dampak negatif terkadang tak terduga bagi pengguna dan masyarakat, terutama bagi kelompok rentan. Pelacakan pengguna secara konstan, analisis otomatis data pribadi untuk memprediksi dan menyimpulkan perilaku, preferensi, tindakan dan karakteristik di masa mendatang, pembuatan profil perilaku, dan penargetan mikro untuk rekomendasi yang dipersonalisasi dapat menimbulkan masalah etika dan hukum yang relevan, seperti hasil diskriminatif, kurangnya transparansi dan penjelasan atas keputusan algoritmik yang berdampak pada kehidupan masyarakat, serta pelanggaran privasi dan perlindungan data yang tidak adil. Artikel ini bertujuan untuk mengatasi masalah-masalah ini melalui pendekatan multisektoral, multidisiplin, dan berbasis hak asasi manusia, termasuk kontribusi dari Hukum, etika, teknologi, pasar, dan masyarakat.

8.1 AI SEHARI-HARI: MANFAAT, TANTANGAN, ETIKA

Kecerdasan Buatan (AI) terus meningkatkan kehadirannya dalam kehidupan kita sehari-hari, membentuk cara kita mengakses informasi, berinteraksi dengan perangkat yang terhubung, berbagi informasi pribadi, dan berinteraksi secara sosial dengan orang lain. Secara bertahap, produk dan layanan baru berbasis teknologi ini tersedia, misalnya, melalui rekomendasi audio-visual; penyaringan spam dalam surel; umpan berita yang dipersonalisasi di media sosial; hasil pencarian di mesin pencari; asisten virtual, dan bahkan saran rute terbaik di aplikasi lalu lintas.

Meskipun istilah "kecerdasan buatan" telah ada sejak pertengahan 1950an, popularitas sistem ini yang semakin meningkat dikaitkan dengan ketersediaan data yang semakin meningkat, infrastruktur pemrosesan yang lebih murah, kemajuan teknologi, dan konektivitas yang lebih luas. Singkatnya, AI dapat dianggap sebagai bidang studi yang sangat luas, yang menyatukan berbagai teknologi yang menggabungkan data, algoritma, dan daya komputasi, yang mampu berperilaku serupa dengan kecerdasan manusia untuk mencapai tujuan tertentu, biasanya solusi dari suatu pertanyaan tertentu.

Dalam kondisi terkini, AI berkontribusi pada manfaat sosial dan ekonomi di berbagai bidang dengan meningkatkan prediksi hasil, mengoptimalkan operasi dan alokasi sumber daya, serta menyesuaikan pemberian layanan, sehingga memberikan keunggulan kompetitif

yang signifikan bagi perusahaan yang mendominasinya. Namun, meskipun berpotensi bermanfaat bagi manusia dan masyarakat, AI juga menimbulkan tantangan baru.

Oleh karena itu, perkembangan pesat dan penerapan teknologi yang sembrono menuntut penerapan prinsip-prinsip etika dan regulasi penggunaannya, terutama ketika kita berbicara tentang mesin yang mampu belajar secara mandiri, menghasilkan hasil yang sangat tidak terduga (bahkan tanpa campur tangan manusia) dan potensi besar untuk merugikan hak-hak asasi manusia.

Skala dan jangkauan sistem AI, tren implementasi yang cepat dan ceroboh, serta dampak langsungnya terhadap kehidupan banyak orang, dapat memperkuat masalah yang sudah ada, selain menciptakan masalah baru. Ancaman yang ditimbulkan oleh AI, dengan demikian, bukanlah robot super-intelijen yang mendominasi umat manusia, melainkan akibat dari penggunaan sehari-harinya, seperti halnya sistem rekomendasi, yang akan dianalisis secara khusus dalam topik berikut.

8.2 SISTEM REKOMENDASI AI

Di dunia yang sangat terhubung, sistem rekomendasi (RS) merupakan salah satu aplikasi komersial AI yang paling luas penggunaannya. Awalnya diperkenalkan untuk e-commerce, sistem ini sudah banyak diterapkan di bidang lain, seperti penyedia konten dan platform media sosial. Karena kelebihan informasi saat ini, sistem ini terutama dirancang untuk membantu individu menghadapi beragam pilihan yang tersedia, serta mengoptimalkan perolehan laba perusahaan dengan menawarkan produk dan layanan yang secara langsung memenuhi kebutuhan pelanggan mereka.

Jadi, idealnya, selain menciptakan pengalaman pengguna yang lebih baik, RS juga membantu penyedia memenuhi tujuan mereka untuk meningkatkan jumlah penjualan dan klik, yang pada gilirannya meningkatkan laba, serta meningkatkan keterlibatan dan kepuasan pengguna di berbagai platform.

Mengingat efektivitasnya, penggunaan RS sudah mencakup berbagai domain, termasuk streaming (Netflix dan Spotify), berita (CNN dan Google News), kencan (Tinder dan Grindr), makanan (ifood dan UberEats), perjalanan (Booking dan AirBnB), media sosial (Facebook dan LinkedIn), mesin pencari (Google) dan e-commerce (Amazon). Di era big data saat ini, ide dasar sistem rekomendasi adalah menggunakan berbagai sumber data yang tersedia untuk menyimpulkan dan memprediksi minat, selera, dan perilaku pengguna di masa mendatang guna merekomendasikan konten, produk, atau layanan yang dipersonalisasi.

Oleh karena itu, RS dianggap sebagai alat penyaringan informasi algoritmik, yang mampu membantu pengguna dalam proses pengambilan keputusan mereka, membentuk pengalaman daring dengan menunjukkan item yang kemungkinan akan menyenangkan mereka. Prediksi kegunaan item untuk pengguna tertentu bervariasi sesuai dengan model algoritma rekomendasi yang digunakan. Saat ini, terdapat tiga model utama:

1. pendekatan berbasis konten—rekomendasi dikirim berdasarkan deskripsi item yang telah disetujui sebelumnya oleh pengguna, baik melalui penilaian langsung maupun perilaku yang disimpulkan;

2. penyaringan kolaboratif—memproses informasi tentang perilaku dan opini komunitas untuk memprediksi item yang menarik bagi pengguna target, selama profil kelompok dan individu serupa; dan
3. pendekatan berbasis pengetahuan—alih-alih data historis, model ini menggabungkan fitur yang dikirimkan oleh pengguna dengan pengetahuan tentang area tertentu, seperti informasi pemasaran atau penjualan. Model ini lebih sering digunakan untuk situasi yang lebih kompleks dan jarang terjadi, seperti melakukan transaksi keuangan atau membeli mobil, apartemen, dan barang mewah.

Selain ketiga model utama tersebut, terdapat juga sistem hibrida, yang menggabungkan kekuatan masing-masing model sebelumnya untuk menciptakan sistem yang lebih efektif, dan sistem yang mempertimbangkan konteks, seperti informasi tentang waktu, lokasi, emosi, dan hubungan sosial.

Apa pun modelnya, mengirimkan rekomendasi yang dipersonalisasi memerlukan pembuatan profil pengguna yang merangkum preferensi, selera, perilaku yang sering dilakukan, dan minat mereka. Informasi ini dapat diekstraksi secara implisit, dari pemantauan perilaku individu daring, atau secara eksplisit, ketika pengguna secara langsung memberikan datanya, seperti mengisi formulir.

Singkatnya, RS pada dasarnya terdiri dari tiga langkah: (1) pengumpulan data pribadi, yang secara langsung maupun tidak langsung diberikan oleh pengguna (input). Dalam kasus terakhir, data tersebut mencakup, misalnya, alur klik, riwayat penelusuran, informasi struktural halaman web yang dikunjungi, dan catatan pembelian, yang diamati dan disimpulkan dari pemantauan individu daring yang terus-menerus; (2) pemrosesan data untuk pembuatan profil pengguna, yang dapat direpresentasikan oleh, misalnya, kelompok istilah atau kata kunci; (3) penargetan konten yang dipersonalisasi dalam bentuk rekomendasi (output).

Tidak diragukan lagi bahwa RS memberikan manfaat dalam hal pengorganisasian, optimalisasi waktu, dan peningkatan pengalaman daring individu, dengan membantu mereka mencari konten, layanan, dan produk yang diminati. Namun, teknologi ini juga dapat menimbulkan dampak negatif terkadang tak terduga bagi pengguna dan masyarakat, terutama kelompok rentan. Pemantauan yang konstan, analisis otomatis data pribadi untuk memprediksi dan menyimpulkan perilaku, preferensi, dan karakteristik individu, pembuatan profil perilaku, dan akhirnya, pengiriman rekomendasi yang dipersonalisasi dapat menimbulkan pertanyaan etika dan hukum yang relevan, sebagaimana akan dianalisis dalam topik selanjutnya.

8.3 TANTANGAN ETIKA DAN HUKUM TERKAIT RS

Pengembangan, implementasi, dan penggunaan sistem rekomendasi yang kompleks dapat menimbulkan masalah etika dan hukum yang signifikan. Kerugian konkret atau potensial serta pelanggaran hak asasi manusia sudah menjadi konsekuensi dari teknologi ini, seperti kurangnya transparansi dan penjelasan hasil (ketidakjelasan algoritmik), pengurangan otonomi individu, paparan pengguna terhadap pelanggaran privasi dan perlindungan data yang tidak beralasan, manipulasi perilaku secara tidak sadar, dan diskriminasi.

Untuk memitigasi beberapa ancaman dan kerugian dari AI dalam RS, perlu diperkenalkan perdebatan etika dan regulasi tentang kemungkinan batasan yang berlaku untuk teknologi ini. Selain undang-undang yang mengikat, pedoman etika merupakan langkah awal yang juga harus dipertimbangkan untuk meminimalkan risiko yang terkait dengan sistem ini dan, secara bersamaan, memaksimalkan manfaatnya.

Selama beberapa tahun, telah ada kekhawatiran di seluruh dunia untuk mendefinisikan batasan etika untuk AI. Semakin banyak inisiatif dari berbagai pemangku kepentingan yang mendefinisikan rekomendasi dan pedoman untuk membangun AI yang etis, tepercaya, dan berpusat pada manusia. Pada tahun 2020, setidaknya 84 inisiatif prinsip etika AI telah dipetakan, yang berasal dari organisasi publik dan swasta, terutama dari Eropa dan Amerika Serikat.

Meskipun sebagian besar dokumen menetapkan kerangka etika umum untuk AI, yang berfokus pada perlindungan orang-orang yang rentan dan penanganan asimetri informasi dan kekuasaan, karena RS didasarkan pada algoritma AI, prinsip-prinsip dasar umum ini dapat langsung diterapkan pada mereka. Di antara prinsip-prinsip yang paling banyak dikutip oleh dokumen-dokumen ini adalah transparansi, keadilan, non-maleficence, akuntabilitas, privasi, kebaikan, kebebasan, otonomi, dan kepercayaan.

Dengan demikian, analisis teknologi ini melalui pendekatan prinsip etika dapat menjadi titik awal yang relevan untuk membandingkan sejauh mana pengembangan dan penggunaan RS dari implementasi yang memadai, di mana ia bertindak lebih bermanfaat daripada merugikan masyarakat. Dalam hal ini, untuk mencapai analisis tersebut, prinsip-prinsip kebaikan dan kejahatan memainkan peran penting.

Sejalan dengan prinsip kebaikan, teknologi berbasis AI, seperti sistem rekomendasi, harus dikembangkan untuk menciptakan "AI untuk kebaikan". Dengan kata lain, teknologi harus mempromosikan kesejahteraan, martabat, kebaikan bersama, dan keberlanjutan dalam semua fase dan rancangannya, agar dapat memberi manfaat bagi manusia, masyarakat, dan planet ini. Dalam hal ini, perangkat-perangkat ini harus mempromosikan potensi manusia, menciptakan peluang-peluang baru yang meningkatkan penentuan nasib sendiri individu, otonomi, agensi manusia, kohesi sosial, serta kapasitas individu dan kolektif.

Inisiatif AI yang bermanfaat harus mencapai kesejahteraan fisik dan emosional pada tingkat individu dan kolektif, seperti meningkatkan layanan kesehatan, menyediakan manfaat publik, memperluas hasil pendidikan yang positif, dan menciptakan lingkungan yang lebih aman. Khususnya terkait RS, prinsip ini tidak dimaksudkan untuk melemahkan manfaat besar yang dihasilkannya, melainkan untuk memastikan bahwa teknologi ini bekerja untuk kepentingan manusia, bukan merugikan mereka. Misalnya, RS yang dirancang dengan baik untuk membantu individu yang sakit atau tidak sehat menghadirkan peluang besar untuk membantu orang mencapai kualitas hidup yang lebih baik sesuai dengan prinsip beneficence. Saat ini, inisiatif ke arah ini sudah ada, seperti perangkat wearable dengan teknik gamifikasi dan intervensi perilaku lainnya dalam bentuk "dorongan" yang diciptakan untuk mendorong perilaku yang lebih sehat.

Selain itu, berdasarkan prinsip non-maleficence, sistem rekomendasi harus dirancang agar tidak merugikan manusia dengan cara apa pun, menghindari kerugian yang dapat diprediksi, tak terduga, atau tak disengaja, seperti rekomendasi yang bias, memfasilitasi penyebaran misinformasi, dan pelanggaran privasi serta aturan perlindungan data. Dalam hal non-maleficence, poin utamanya adalah mencegah segala jenis kerugian, baik yang disebabkan oleh niat atau malpraktik individu maupun perilaku teknologi yang tak terduga.

Oleh karena itu, untuk mencegah dan menghindari RS yang merugikan, penting untuk memahami keterbatasan teknologi guna mengelola potensi risiko. Prinsip ini menekankan kebutuhan mendesak untuk memiliki sistem AI yang sesuai dengan standar dan rekomendasi perlindungan data, privasi, keamanan siber, dan perlindungan hak asasi manusia, baik secara desain maupun secara otomatis, di samping sistem akuntabilitas yang efektif jika terjadi penyalahgunaan.

Oleh karena itu, penyesuaian dan harmonisasi kesepakatan antara prinsip kebaikan dan non-kejahatan merupakan hal yang umum, yang menuntut keseimbangan antara manfaat dan risiko AI dalam praktik. Misalnya, ketika perusahaan mencegah, melalui teknik otomatis, konten berbahaya agar tidak direkomendasikan untuk melindungi penggunanya, meskipun penyaringan AI memiliki tujuan yang bermanfaat, hal itu dapat melanggar kebebasan dan otonomi individu. Oleh karena itu, dalam praktiknya, penting untuk mempertimbangkan secara cermat kemungkinan cara sistem dapat disalahgunakan atau menyebabkan kerusakan yang tidak diinginkan untuk mengurangi dampak buruknya. Dalam hal ini, prinsip kebaikan dan non-kejahatan, bersama dengan pedoman etika lainnya, saling terkait dan terungkap dalam berbagai implikasi hukum. Di bawah ini, kami menyoroti beberapa tantangan etika dan hukum utama yang muncul dari sudut pandang kedua nilai ini:

Ketidakjelasan

Beberapa pakar AI membandingkan teknologi ini dengan kotak hitam, karena proses dan cara kerjanya berada di luar kemampuan manusia untuk memahaminya, terutama bagi orang-orang di luar bidang studi teknologi. Anggapan ini bahkan lebih kuat dalam kasus algoritma AI yang berinteraksi dalam lingkungan sosial terbuka dan belajar melalui interaksi dengan ruang tempat mereka beroperasi, ketika keputusan otomatisnya sulit dijelaskan bahkan oleh para ahli. Kurangnya transparansi dan penjelasan yang sering terjadi tentang proses dan nilai-nilai yang terlibat dalam alat rekomendasi menghambat terciptanya sistem yang lebih baik, yaitu yang memadai untuk hak-hak asasi, prinsip-prinsip etika, dan berpusat pada manusia.

Bias Diskriminatif

RS dibuat oleh manusia, yang membuatnya rentan terhadap hasil yang bias. Konsekuensi ini dapat muncul akibat data pelatihan yang dipilih atau nilai (implisit) yang dimiliki oleh pengembang teknologi, yang dapat memperburuk diskriminasi sosial sistematis, bahkan tanpa disengaja.

Karena sifat teknik AI yang digunakan dalam sistem rekomendasi berbasis data, pemilihan set data untuk pelatihan harus didefinisikan dengan baik, jika tidak, hal ini dapat menjadi sumber diskriminasi yang penting. Misalnya, ketika data yang tersedia tidak

mencerminkan keragaman sosial yang ada dalam masyarakat, ketidakseimbangan populasi dalam set data ini kemungkinan akan menimbulkan bias terhadap kelompok tertentu. Selain itu, konten yang bias juga dapat muncul dari putaran umpan balik yang dihasilkan oleh sistem untuk kelompok pengguna tertentu, yang seringkali memperkuat diskriminasi rasial dan gender yang sudah ada di masyarakat.

Dalam proses yang dilakukan oleh RS, pembuatan profil merupakan salah satu proses yang paling mungkin menyebabkan diskriminasi. Dengan data pribadi, penyedia RS membuat profil penggunanya sebagai parameter aspek kepribadian dan minat mereka, untuk memberi label individu menurut pola kebiasaan, perilaku, dan selera tertentu, yang memiliki potensi diskriminatif besar, terutama dalam kasus data pribadi yang sensitif.

Pelanggaran Privasi dan Perlindungan Data

RS berbasis AI mengumpulkan, menganalisis, dan memproses data pribadi dalam jumlah besar. Oleh karena itu, kekhawatiran tentang privasi dan perlindungan data meningkat seiring penggunaannya yang semakin umum dan dapat diterapkan di berbagai bidang, termasuk di bidang dengan risiko privasi yang tinggi, seperti layanan kesehatan dan perbankan.

Dalam hal ini, risiko terkait privasi dapat muncul dari semua tahapan pemrosesan data pengguna. Dengan mempertimbangkan Peraturan Perlindungan Data Umum (Peraturan 2016/679—GDPR) sebagai model, ketika data dikumpulkan oleh algoritma sistem rekomendasi dan akhirnya dibagikan kepada pihak ketiga, seringkali tanpa penerapan langkah-langkah keamanan, persetujuan yang sah (atau dasar hukum lainnya), dan penyediaan informasi yang memadai kepada pengguna privasi dan data pribadi mereka dilanggar, yang diperparah dalam kasus kebocoran data dan pelanggaran anonimitas

Selain itu, pemrosesan data RS dapat menghasilkan inferensi dan prediksi informasi rahasia dan pribadi, seperti kondisi emosional. Akibatnya, sistem ini dapat mengakses data pribadi yang sensitif (seperti informasi tentang asal ras atau etnis, keyakinan agama, opini politik, kesehatan, atau kehidupan seksual), dari inferensi yang diambil dari data pribadi melalui pemrosesan otomatis untuk pembuatan profil dan untuk pembuatan rekomendasi yang dipersonalisasi. Dengan demikian, tantangan privasi yang signifikan muncul dalam skenario ini, di samping kemungkinan hasil diskriminasi.

Menurunnya Otonomi Manusia dan Penentuan Nasib Sendiri

RS melibatkan proses pengambilan keputusan tentang pengguna dan konteks mereka melalui pembuatan profil perilaku. Teknologi ini, yang mampu mengetahui preferensi calon pengguna dan beradaptasi sesuai dengan minat mereka, menimbulkan pertanyaan penting tentang privasi, otonomi, dan etika di balik proses adaptasi tersebut.

Otonomi individu melibatkan kapasitas untuk menentukan nasib sendiri secara bebas dan hak untuk membuat pilihan berdasarkan keyakinan, informasi, dan nilai-nilai pribadi. Untuk itu, penting bagi individu untuk memiliki kesempatan yang nyata dan signifikan untuk membuat pilihannya sendiri, dengan informasi yang memadai dan bebas dari paksaan, pembatasan, atau pengaruh eksternal, baik yang berlebihan maupun tidak semestinya.

Dengan demikian, otonomi manusia secara langsung dipengaruhi oleh RS, karena mereka membatasi kebebasan individu, karena kendali mereka atas pengaruh yang

ditransmisikan kepada pengguna dalam bentuk rekomendasi, selain fakta bahwa, ketika persetujuan digunakan sebagai dasar hukum untuk pemrosesan data pribadi, hal itu jarang diinformasikan kepada pengguna, tetapi digunakan sebagai syarat implisit untuk mengakses layanan tertentu yang diinginkan.

Dalam hal ini, RS mengganggu otonomi orang dalam bentuk rekomendasi semua jenis konten, mulai dari musik dan film hingga lowongan kerja, mendorong pengguna ke arah tertentu, umumnya terkait dengan preferensi mereka yang diambil dari profil mereka, dalam upaya untuk membuat mereka kecanduan pada beberapa jenis konten atau membatasi rentang pilihan yang mereka hadapi.

Beberapa teknologi ini bertindak hampir seperti jebakan untuk membuat pengguna tetap terlibat dan terhubung dengan platform mereka, yang memungkinkan ketersediaan data yang lebih besar untuk dikumpulkan dan diproses. Selain itu, profil algoritmik platform rekomendasi juga memiliki dampak besar pada otonomi orang, karena dapat mengganggu pengalaman identitas pribadi. Pertama, sistem berdasarkan umpan balik pengguna (misalnya, penyaringan kolaboratif) tidak membuat profil yang spesifik dan unik, tetapi profil kolektif. Lebih lanjut, klasifikasi dilakukan oleh algoritma yang menganalisis dan menyimpulkan selera dan preferensi, yang mungkin tidak sesuai dengan karakteristik sosial atau kategori yang sesuai dengan yang diidentifikasi oleh pengguna.

Seperti yang disebutkan sebelumnya, masalah ini juga diperburuk dalam konteks umum algoritma yang kurang dapat dijelaskan atau transparan terkait dengan pembuatan profil ini. Dengan demikian, penggunaan sistem rekomendasi oleh perusahaan teknologi besar saat ini, terutama di media sosial, streaming, dan e-commerce, juga dapat menimbulkan risiko yang disengaja terhadap otonomi pengguna.

Sesuai kepentingan komersial mereka, penyedia RS juga dapat memberikan pengaruh tersembunyi terhadap perilaku pengguna mereka, yang dilakukan melalui pemantauan, pelacakan perilaku, dan eksploitasi kerentanan serta data pribadi untuk pembuatan profil, yang digunakan untuk penargetan mikro konten dalam bentuk rekomendasi. Proses ini seringkali terjadi tanpa sepengetahuan pengguna umum, yang dapat mengganggu kemampuan mereka untuk menentukan sendiri dan membuat pilihan yang benar-benar otonom.

Polarisasi dan Manipulasi Proses Demokrasi

Sistem rekomendasi dan filter media sosial, berdasarkan sifat desainnya, berisiko mengisolasi pengguna dari paparan sudut pandang yang berbeda. Bahkan ketika sistem melabeli individu dengan tepat, efek yang dihasilkan oleh personalisasi dapat menimbulkan kerugian individu dan kolektif dengan menciptakan atau memperburuk gelembung filter dan ruang gema.

Konten yang direkomendasikan di platform digital, yang dibatasi oleh fenomena ini, menunjukkan risiko tinggi terhadap debat publik dan proses demokrasi, karena dapat memperkuat bias diskriminatif dan prasangka individu, yang meningkatkan kerentanan terhadap polarisasi, ujaran kebencian, dan manipulasi ujaran publik. Sebagaimana ditunjukkan

oleh skandal Cambridge Analytica, RS platform streaming dan media sosial dapat menjadi tempat untuk mengirimkan propaganda politik yang terarah.

Saat ini, karena kelebihan informasi, tidak diragukan lagi bahwa sistem rekomendasi dapat mengurangi masalah ini dan membantu orang mengelola waktu mereka secara efisien. Namun, dalam skenario ini, meskipun rekomendasi teknologi dapat bermanfaat bagi pengguna (membantu kinerja individu dalam proses pemilihan, meningkatkan dan mendiversifikasi pengambilan keputusan), rekomendasi tersebut juga berpotensi dipertanyakan, karena memengaruhi orang ke arah tertentu, dan menimbulkan kerugian individu dan sosial, seperti segregasi informasi, gelembung, dan manipulasi perilaku. Dengan demikian, untuk memastikan keselarasan dengan prinsip-prinsip kebaikan dan non-maleficence, sistem harus dirancang dengan baik tidak hanya untuk meningkatkan kehidupan masyarakat, tetapi juga untuk mempertahankan kendali penuh dan efektif atas diri mereka sendiri, sekaligus menghindari kerugian dan membatasi risiko.

8.4 SISTEM REKOMENDASI: TANTANGAN HUKUM DAN REGULASI

Mengingat semakin pentingnya RS bagi kehidupan kita sehari-hari, bersamaan dengan meningkatnya dampak buruknya, terdapat kebutuhan yang sangat besar untuk bertindak. Inisiatif regulasi hukum harus mempertimbangkan tidak hanya pedoman etika resmi, tetapi juga perlindungan hak asasi manusia yang efektif, dimulai dari premis dasar bahwa sistem AI harus bekerja untuk berbuat baik, menghindari kerugian, bukan menyebabkannya.

Dengan demikian, karena RS memerlukan pemrosesan data pribadi, permasalahan yang timbul dari teknologi ini telah diatasi oleh aturan perlindungan data di seluruh dunia. Dalam skenario ini, GDPR Uni Eropa (UE) memainkan peran penting sebagai model regulasi yang telah menginspirasi banyak pihak di seluruh dunia, fenomena yang dikenal sebagai Efek Brussel. Meskipun regulasi tersebut tidak secara khusus membahas RS atau AI itu sendiri, regulasi tersebut membahas proses fundamentalnya, seperti pemrosesan data pribadi untuk pengambilan keputusan otomatis, pembuatan profil, dan rekomendasi konten yang dipersonalisasi.

Oleh karena itu, untuk melindungi hak-hak fundamental, menjamin penentuan nasib sendiri berdasarkan informasi, dan pengembangan kepribadian yang bebas, GDPR memberikan serangkaian kewajiban yang dibebankan kepada pengendali dan pemroses, yang mencakup daftar prinsip (pasal 5), hak-hak subjek data (bab III), dan dasar hukum pemrosesan data pribadi (pasal 6 dan 9). Dengan demikian, platform RS harus beradaptasi dengan aturan-aturan ini untuk melindungi data pribadi individu dan, akibatnya, hak asasi manusia lainnya yang berpotensi terancam oleh RS.

Pertama, penyedia RS perlu memastikan bahwa semua aktivitas dengan data pribadi (otomatis atau tidak) mematuhi prinsip-prinsip tersebut, terutama kewajiban pemrosesan yang sah, adil, dan transparan (keabsahan, keadilan, dan transparansi) serta definisi tujuan yang spesifik, eksplisit, dan sah, sesuai dengan dasar hukum pasal 6 dan 9 (pembatasan tujuan). Selain itu, data harus dibatasi pada apa yang benar-benar diperlukan untuk mencapai tujuan ini (minimalisasi data) dan disimpan hanya untuk periode yang diperlukan (pembatasan

penyimpanan). Terakhir, proses tersebut harus menjamin keakuratan dan kualitas data, kepatuhan terhadap standar keamanan (integritas dan kerahasiaan), selain memastikan akuntabilitas yang memungkinkan pertanggungjawaban atas kerugian di kemudian hari.

Selain memenuhi prinsip-prinsip tersebut, agar dianggap sah, pemrosesan data pribadi oleh RS harus dilakukan sesuai dengan salah satu situasi yang dijelaskan dalam Pasal 6. Pada titik ini, penting untuk disebutkan bahwa GDPR, sebagai suatu peraturan, melarang pemrosesan kategori data khusus dalam Pasal 9 dan pengambilan keputusan yang sepenuhnya otomatis dengan dampak merugikan pada subjek data dalam Pasal 22, kecuali dalam situasi tertentu yang tercantum dalam kedua pasal tersebut. Untuk yang terakhir, pengecualian mencakup perolehan persetujuan eksplisit dari subjek data; ketika diperlukan untuk membuat atau melaksanakan suatu kontrak; atau diizinkan oleh hukum Uni Eropa atau Negara Anggota.

Selain itu, penyedia RS perlu memastikan, di seluruh proses data, pelaksanaan hak-hak subjek data yang efektif dan difasilitasi, yang dianggap sebagai hasil logis dari prinsip-prinsip tersebut. Misalnya, sebagai konsekuensi dari prinsip transparansi yang sah dan etis, hak atas informasi (pasal 13 dan 14) menetapkan bahwa pengguna harus terus diinformasikan dan menyadari potensi risiko yang terkait dengan pemrosesan data yang dilakukan oleh RS. Dengan demikian, pengguna tidak boleh membatasi diri pada keuntungan jangka pendek yang diperoleh dengan sistem ini yang dapat, secara perlahan, merusak hak-hak fundamental mereka, seperti otonomi, kebebasan, dan privasi.

Dengan demikian, merupakan kewajiban penyedia untuk secara proaktif menginformasikan, bahkan tanpa diminta, tentang hak, keberadaan pemrosesan data, dan informasi terkait lainnya, termasuk tujuan dan penjelasan yang jelas, bermakna, dan mudah dipahami tentang fungsi teknik algoritmik RS, khususnya definisi profil. Lebih lanjut, informasi ini, jika tidak diungkapkan secara aktif, wajib diberikan kepada subjek saat diminta akses, sesuai dengan Pasal 15 dan Pertimbangan 63.

Jika dianalisis bersama, informasi dan hak akses dianggap sebagai alat yang ampuh bagi individu untuk menjalankan kontrol yang lebih besar atas data mereka terkait RS, karena memungkinkan mereka memiliki kesadaran dan pengetahuan yang lebih luas tentang proses yang terlibat dalam pengiriman rekomendasi yang dipersonalisasi, sehingga memungkinkan pengambilan keputusan yang lebih baik yang dapat melindungi hak-hak mereka.

Selain itu, dengan informasi yang diterima atau diminta, pengguna dapat menjalankan hak-hak lain yang tercantum dalam GDPR, seperti perbaikan (Pasal 16), penghapusan (Pasal 17), pembatasan pemrosesan (Pasal 18), portabilitas (Pasal 20), keberatan (Pasal 21, jika memungkinkan), dan keberatan atas keputusan yang sepenuhnya otomatis (Pasal 22.3). Hal ini memastikan otonomi dan kontrol yang lebih besar bagi pengguna, sehingga mencegah rekomendasi yang merugikan dan bias.

Dengan demikian, karena pembuatan profil secara otomatis dan pengiriman rekomendasi yang dipersonalisasi berdasarkan profil ini merupakan langkah-langkah RS, pasal 22 merupakan elemen kunci, karena pasal ini mengizinkan pengambilan keputusan secara otomatis, termasuk pembuatan profil, yang menghasilkan dampak hukum pada subjek data, hanya dalam hipotesis spesifik yang diizinkan oleh peraturan, seperti ketika berdasarkan

persetujuan eksplisit subjek data. Dalam konteks ini, individu memiliki hak untuk mendapatkan intervensi manusia, mengungkapkan sudut pandangnya, dan untuk menentang keputusan otomatis RS. Namun, dengan mempertimbangkan risiko yang terlibat, GDPR menciptakan kewajiban bagi pengendali untuk mengadopsi langkah-langkah perlindungan guna melindungi hak, kebebasan, dan kepentingan subjek data, yang dapat mencakup, teknik privasi berdasarkan desain (pasal 25) dan pelaksanaan penilaian dampak perlindungan data (pasal 35).

Lebih lanjut, karena sistem ini bergantung pada probabilitas algoritmik dan seringkali model pembelajaran mesin untuk mengirimkan rekomendasi, penting untuk memberikan subjek data hak atas penjelasan yang jelas dan memadai tentang keputusan yang sepenuhnya otomatis yang melibatkan data mereka. Hak atas penjelasan ini dapat diambil dari interpretasi pasal 13, 14, dan 22, bersama dengan pertimbangan 71 dan prinsip transparansi, yang menciptakan kewajiban pengendali untuk menginformasikan secara signifikan tentang logika yang terlibat dalam semua proses otomatis hingga pengambilan keputusan yang efektif.

Penjelasan tersebut tidak selalu melibatkan pembukaan algoritma secara menyeluruh, tetapi cukup bagi pengguna untuk memahami alasan yang mendasari keputusan yang memengaruhinya, yang menjamin pelaksanaan hak-hak lain GDPR, selain perlindungan hak asasi manusia lainnya. Dengan demikian, dalam lingkup RS, penerapan pasal. 22 dan hak atas penjelasan sangat penting untuk meminimalkan risiko meningkatnya penggunaan algoritma untuk mengklasifikasikan orang ke dalam profil perilaku, berdasarkan analisis inferensi dan prediksi tentang karakteristik, selera, perilaku, dan minat mereka, dan kemudian mengirimkan rekomendasi yang dipersonalisasi yang berpotensi merugikan pengguna, yang secara diam-diam mengganggu otonomi mereka, memanipulasi keputusan mereka, dan melanggar jaminan non-diskriminasi dan privasi.

Meskipun demikian, tidak diragukan lagi bahwa GDPR menciptakan latar belakang yang menguntungkan bagi perlindungan data di UE, menjadi inspirasi di seluruh dunia, yang berlaku untuk perangkat AI, termasuk sistem rekomendasi, dan memberlakukan kewajiban serta persyaratan yang signifikan pada pengendali data. Meskipun demikian, selain melindungi dan membela hak-hak dasar, menurut pasal 1, peraturan tersebut juga menghasilkan dampak positif bagi perusahaan dan pemerintah, karena penerapannya mencegah pelanggaran hak dan, dengan demikian, pengenaan sanksi, membantu dalam penggunaan dan pengembangan teknologi yang bermanfaat bagi masyarakat. Misalnya, hak untuk menentang keputusan otomatis memungkinkan pengguna RS untuk menentang rekomendasi yang tidak akurat atau diskriminatif, serta kesempatan bagi penyedia untuk merevisi sistem mereka.

Namun, dengan data besar, semakin pentingnya platform digital, dan pesatnya perkembangan teknik AI, meskipun regulasi berupaya meningkatkan konteks perlindungan data dan, dengan demikian, hak asasi manusia, masih banyak yang harus dilakukan. Beberapa aturannya masih sulit atau tidak mudah dipatuhi oleh penyedia RS, terutama yang terkait dengan teknik AI untuk pembuatan profil dan keputusan otomatis. Dalam konteks ini, penyedia RS mungkin menghadapi kesulitan dalam memastikan kepatuhan terhadap prinsip dan hak dalam praktik, misalnya karena ketidakjelasan teknis atau aturan rahasia dagang. Namun,

masih banyak pertanyaan terbuka mengenai interpretasi ketentuan hukum, terutama mengenai hak-hak subjek data, seperti hak untuk menggugat keputusan otomatis dan penjelasannya.

Kurangnya Transparansi

Meskipun prinsip etika dan aturan hukum menuntut transparansi sistem AI, beberapa penggunaannya mungkin tidak transparan bagi individu, regulator, dan bahkan bagi perancangnya, yang menyulitkan untuk menggugat hasilnya. Jadi, RS mungkin memiliki tiga sumber ketidakjelasan yang berbeda: (1) ketidakjelasan yang disengaja, biasanya terkait dengan rahasia dagang; (2) ketidakjelasan sebagai buta huruf teknis; dan (3) ketidakjelasan yang muncul dari desain dan karakteristik sistem, terutama dalam kasus pembelajaran mesin.

Ketiadaan atau ketiadaan transparansi dalam RS ini menyulitkan untuk mempertanyakan agenda politik, ekonomi, dan budaya yang ada di balik rekomendasi personal yang dikirimkan kepada setiap pengguna platform, selain menyembunyikan kemungkinan diskriminasi algoritmik dan manipulasi perilaku secara diam-diam. Selain potensi untuk merugikan hak-hak asasi, ketidakjelasan ini menghambat deteksi dan koreksi data yang bias, asumsi yang tidak valid, dan model yang cacat.

Rahasia Dagang

Informasi tentang fungsionalitas algoritma RS seringkali sengaja dibuat sulit diakses oleh publik. Perangkat lunak, algoritma, dan data yang terlibat dalam aplikasi sistem rekomendasi dianggap sebagai aset kepemilikan dengan nilai tambah yang tinggi, yang penting untuk mempertahankan posisi organisasi di pasar yang kompetitif. Akibatnya, sebagian besar perusahaan dan penyedia sistem ini masih enggan dan menolak mengungkapkan informasi terkait fungsi AI karena alasan rahasia dagang, yang menyebabkan ketidakjelasan RS yang disengaja. Khususnya, kurangnya model dan praktik bisnis yang transparan menjadi hambatan signifikan dalam mendeteksi kasus-kasus pelanggaran hak asasi manusia, seperti rekomendasi dan kesimpulan yang diskriminatif.

Teknologi yang Terus Berubah

Kondisi perkembangan teknologi AI saat ini, yang menjadi dasar RS, tidak menjelaskan evolusi besar berikutnya dan jenis penggunaan serta tingkat pemahaman teknologi yang dapat kita capai di masa depan, yang menghambat penerapan kewajiban pencegahan kerusakan bagi organisasi yang menggunakan AI. Lebih lanjut, mentalitas "kotak hitam", yang menganggap sistem AI berada di luar pemahaman manusia, masih membatasi kendali manusia atas teknologi.

Kesulitan Implementasi Hak Subjek Data dalam Praktik

Sebagai aturan, sistem RS pada umumnya bekerja seperti kotak hitam, karena rekomendasi akhir (keluaran) adalah satu-satunya bagian yang tersedia bagi pengguna. Apa pun alasan penciptaan RS yang tidak transparan, kurangnya transparansi ini merupakan hambatan bagi pemenuhan hak atas penjelasan GDPR, yang juga menghalangi kendali manusia atas bagaimana data diperlakukan dan pelaksanaan hak-hak lainnya.

Lebih lanjut, saat ini, terdapat ketidakseimbangan kekuatan pengambilan keputusan dan pengetahuan yang menguntungkan penyedia RS dan merugikan pengguna. Asimetri

informasi ini, yang didorong oleh ketidakjelasan sistem AI, juga diperkuat oleh ketiadaan atau kurangnya pemahaman individu mengenai hak-hak mereka dan bagaimana teknologi tersebut bekerja dalam praktik, yaitu, bagaimana algoritma dan teknik pemrosesan data bertindak ketika memprediksi dan menyimpulkan perilaku, membuat profil, dan mengirimkan konten yang dipersonalisasi.

Ketika logika di balik sistem rekomendasi tidak dipahami oleh pengguna, kendali dan otonomi manusia tidak dihormati. Oleh karena itu, ketika penyedia RS mengandalkan persetujuan untuk pemrosesan data, persetujuan ini sebenarnya tidak diberikan secara bebas, spesifik, terinformasi, dan tidak ambigu, karena pengguna tidak memiliki informasi yang memadai dan sarana yang tepat untuk menilai risiko yang terlibat dalam pemrosesan data yang dipatuhi.

Selain itu, mengingat kekhawatiran perusahaan untuk menerapkan aturan perlindungan data yang mewajibkan pengungkapan informasi penting dan penjelasan, individu menghadapi kelebihan permintaan persetujuan, biasanya melalui kebijakan privasi yang luas dan kompleks serta pemberitahuan kuki. Mengingat keterbatasan rasionalitas manusia dan keterbatasan waktu, evaluasi dan kontrol efektif pengguna menjadi terganggu, yang berujung pada kegagalan dalam membuat keputusan yang terinformasi. Selain itu, meskipun hidup di era hiperkonektivitas, sebagian besar orang masih memiliki sedikit pengetahuan teknis, akses ke pendidikan digital, dan pemahaman minimal tentang proses pemrosesan data, sehingga semakin mempersulit pengambilan keputusan yang terinformasi dalam konteks RS, terutama jika didasarkan pada persetujuan.

Dalam praktiknya, persetujuan yang tercantum dalam sebagian besar kebijakan privasi penyedia RS tidak memberdayakan pengguna maupun menjamin pelaksanaan hak yang efektif dan penentuan nasib sendiri informasi mereka, melainkan hanya berfungsi sebagai legitimasi model bisnis terhadap aturan GDPR. Oleh karena itu, individu ditempatkan dalam situasi asimetri informasi, teknis, dan ekonomi. Meskipun aturan perlindungan data bertujuan untuk melindungi hak-hak fundamental dengan menetapkan hak-hak subjek data, masih terdapat kekurangan efektivitas dalam berbagai situasi, misalnya, dalam hal analisis data inferensial menggunakan teknik AI. Dengan konteks hukum saat ini, subjek data tidak memiliki kendali dan informasi yang memadai tentang bagaimana data mereka digunakan oleh RS untuk membuat inferensi, prediksi, dan asumsi tentang mereka.

Dengan demikian, individu menghadapi hambatan dalam menjalankan hak perlindungan data mereka, terutama penjelasan dan penentangan atas keputusan otomatis, yang bahkan lebih sulit ketika dihadapkan dengan kepentingan pengendali terkait kekayaan intelektual dan rahasia dagang. Oleh karena itu, khususnya mengenai hak penjelasan dan penentangan atas keputusan otomatis, masih banyak pertanyaan terbuka, karena parameternya masih dalam pembahasan. Mengingat ketidakpastian ini, pengakuan hak atas penjelasan dalam praktik menjadi terganggu, yang juga mempersulit pelaksanaan hak-hak lain, terutama untuk menggugat dan meninjau keputusan otomatis, karena pengguna harus mengakses informasi tentang keputusan otomatis, dan RS itu sendiri, untuk mengumpulkan

kondisi guna mengungkap bagaimana datanya harus diproses dan pada akhirnya menemukan kesalahan, ketidaksesuaian, dan korelasi yang keliru untuk dipecahkan.

Kesulitan Penerapan Aturan

Beberapa karakteristik spesifik RS, seperti ketidakjelasan (efek kotak hitam), dapat menyulitkan penerapan dan verifikasi kepatuhan terhadap pedoman etika dan aturan hukum, terutama yang muncul dari GDPR. Karena kompleksitasnya yang tinggi, ketidakpastiannya, dan perilakunya yang otonom, otoritas dan masyarakat yang terdampak oleh sistem ini mungkin tidak memiliki cara khusus untuk memverifikasi bagaimana rekomendasi personal tertentu dicapai dan, dengan demikian, apakah aturan ini dipatuhi.

Debat regulasi saat ini menekankan peran perlindungan data dalam menetapkan hak-hak subjek data, dasar hukum, dan prinsip-prinsipnya, dengan berfokus pada peran akuntabilitas, yang menyoroti prinsip etika non-maleficence. Misalnya, Pasal 58 (2) GDPR menetapkan kewenangan korektif otoritas pengawas, seperti pengenaan denda, yang akan diterapkan sesuai dengan keadaan setiap kasus, selalu dengan cara yang efektif, proporsional, dan bersifat mencegah.

Dengan mempertimbangkan RS, platform digital harus memastikan bahwa konten dan aktivitasnya menghormati hak asasi manusia, terutama perlindungan data, privasi, dan kesetaraan, serta tidak rentan terhadap serangan eksternal. Poin menariknya adalah bahwa beberapa tantangan terkait sistem ini lebih sulit diatasi hanya dengan solusi teknologi, sehingga memerlukan analisis yang lebih kualitatif berdasarkan konteks sosial tempat mereka beroperasi.

Dalam hal ini, penerapan GDPR oleh otoritas harus mengupayakan keseimbangan yang adil antara aturan hukum dan kemajuan teknologi, mencegah perusahaan terbebani oleh regulasi yang membebani mereka secara berlebihan dengan persyaratan administratif dan standar perlindungan data yang tidak realistis. Pertanyaan terbukanya adalah apakah Negara akan menegakkan langkah ini tanpa membebani perusahaan atau menghambat inovasi teknologi.

Melampaui Pencegahan Kerusakan

Peraturan RS saat ini untuk perlindungan data dalam GDPR berfokus pada langkah-langkah untuk mencegah kerusakan dan memastikan akuntabilitas jika terjadi, sesuai dengan prinsip non-maleficence AI. Namun, teknologi juga harus diatur melalui prinsip beneficence, yang memungkinkan maksimalisasi manfaat bagi individu dan masyarakat. Mengingat potensi AI yang tak terbantahkan, terutama melalui sistem rekomendasi, regulasi AI perlu dilakukan agar manfaatnya meningkat, sekaligus menghindari potensi jebakan. Pada waktunya, regulasi AI juga perlu memfokuskan penelitian tidak hanya untuk meningkatkan kemampuan dan akurasi teknologi, tetapi juga memaksimalkan manfaat sosialnya, yang dapat dicapai melalui penilaian hak asasi manusia sebelumnya.

8.5 STRATEGI DAN SOLUSI TANTANGAN SISTEM REKOMENDASI

Saat ini, GDPR merupakan sistem perlindungan hak asasi yang kuat dalam konteks AI dan keputusan otomatis. Selain menetapkan prinsip-prinsip relevan, seperti legalitas,

minimisasi data, transparansi, keamanan, keadilan, dan akuntabilitas, undang-undang ini juga menetapkan serangkaian hak yang memperkuat kendali pengguna atas data mereka dan menetapkan kewajiban bagi mereka yang bertanggung jawab atas pemrosesan data tersebut, yang mencakup publikasi informasi, transparansi, dan penerapan langkah-langkah keamanan. Namun, mengingat kompleksitas sistem rekomendasi berbasis AI yang progresif dan konstan, regulasi semata-mata oleh hukum perlindungan data tidak lagi memadai. Oleh karena itu, terdapat cara lain untuk mengatasi permasalahan yang terkait dengan RS, yang juga mencakup aturan hukum khusus terkait AI dan model bisnis yang menggunakannya, di samping strategi lain di luar hukum, seperti norma sosial, inisiatif pasar, dan cara arsitektur sistem (kode) dikembangkan.

Praktik Terbaik di Luar Hukum

Dalam skenario ini, semua pemangku kepentingan yang terkait dengan RS harus memperhatikan standar etika yang berlaku untuk algoritma AI. Sebagaimana telah disebutkan, terdapat perdebatan luas seputar pedoman etika ini yang seharusnya memandu seluruh siklus hidup RS berbasis AI, termasuk pengembangan, implementasi, dan penggunaan efektifnya.

Terdapat kebutuhan mendesak agar perangkat ini berfokus pada manusia, melindungi kepentingan dan hak-hak fundamental mereka, agar dapat memberikan manfaat bagi seluruh masyarakat. Mengingat relevansi parameter etika, seperti transparansi, akuntabilitas, non-diskriminasi, kehati-hatian, privasi, dan keamanan, banyak di antaranya telah diintegrasikan ke dalam peraturan, seperti yang terjadi dalam prinsip, aturan, dan hak GDPR.

Meskipun demikian, karena sistem rekomendasi tertanam dalam algoritma otonom dan cerdas, yang menciptakan masalah hukum dan etika, inisiatif dari bidang keahlian multidisiplin, seperti ilmuwan data, pengacara, pakar penelitian hukum, ilmuwan sosial, dan pakar etika, diperlukan. Dalam hal ini, solusi AI harus dikembangkan dan diimplementasikan melalui tim lintas sektor dan multidisiplin dengan tujuan mengoptimalkan hasil yang sesuai dengan etika dan legalitas.

Regulasi oleh Teknologi: Strategi Berdasarkan Desain dan Secara Default

Dalam konteks teknologi "baru" yang secara aktif mengganggu kehidupan kita sehari-hari, merekomendasikan konten yang dipersonalisasi, dan membuat keputusan otomatis tentang kita, etika dan hak asasi manusia memainkan peran penting dalam penerapannya demi kepentingan publik. Dengan demikian, regulasi RS juga harus melibatkan desain perangkat itu sendiri, yang selaras dengan pedoman etika dan hak asasi manusia sejak awal, sebagai elemen sentral dari arsitektur sistem.

Pendekatan "desain yang peka nilai" ini, yang mencakup privasi, keamanan, etika, dan hak asasi manusia, sejalan dengan gagasan bahwa manfaat dan dampak positif AI tidak hanya harus dijamin oleh kepatuhan terhadap kerangka regulasi, tetapi juga dipastikan secara otomatis, sejak awal pengembangan sistem dan diperkuat selama penggunaannya, sesuai dengan strategi yang telah dirancang dan secara otomatis.

Oleh karena itu, prinsip-prinsip etika dan hukum, yang didasarkan pada hak asasi manusia dan nilai-nilai, harus berfungsi sebagai kriteria desain untuk pengembangan penggunaan AI yang inovatif dan juga untuk peninjauan penggunaan yang sudah ada, guna

menempatkan manusia di pusat penciptaan model RS, yang memandu implementasi dan penggunaannya, sesuai dengan ketentuan Pasal 25 GDPR.

Dengan demikian, dalam jangka pendek, desain dapat memainkan peran krusial dalam mengatasi masalah etika dan hukum yang berpotensi dipicu oleh RS. Misalnya, pesan pop-up yang mengingatkan pengguna tentang hasil rekomendasi yang mempertimbangkan profil perilaku mereka membantu meningkatkan kesadaran publik dan pemenuhan hak asasi. Namun, dalam jangka panjang, penting bagi infrastruktur RS untuk menerapkan norma dan prinsip etika secara default, seperti transparansi, non-diskriminasi, dan keadilan, di semua tahapan sistem.

Implementasi Penilaian Dampak (Hak Asasi Manusia)

Mengingat tingginya risiko bagi pengguna dan masyarakat yang ditimbulkan oleh sistem rekomendasi, yang mencakup manipulasi, pelanggaran privasi dan perlindungan data, diskriminasi, dan pengurangan otonomi individu, pelaksanaan penilaian dampak hak asasi manusia dan evaluasi kepatuhan terhadap undang-undang dan pedoman etika sebelumnya merupakan hal mendasar bagi penerapan RS. Namun, saat ini, sistem ini masih diimplementasikan kepada publik tanpa evaluasi etika, hukum, dan teknis yang memadai yang dapat menilai kemungkinan dampak dan risiko yang terkait dengan teknologi ini dalam praktiknya, yang membahayakan hak-hak individu.

Sebagaimana pasal. Pasal 35 GDPR menetapkan untuk melakukan penilaian dampak perlindungan data pribadi. Dalam beberapa kasus tertentu, penyedia RS dianggap sebagai praktik yang baik untuk melakukan penilaian dan audit pada semua keputusan AI otomatis, termasuk pembuatan profil, yang dapat dilakukan melalui pengujian, inspeksi, atau sertifikasi. Oleh karena itu, disarankan untuk menerapkan audit algoritma dan penilaian dampak algoritmik agar risiko yang terkait dengan perangkat ini dapat dipetakan, dicegah, dan dimitigasi.

Dalam hal ini, audit algoritma dalam RS harus menilai data dan algoritma untuk mencari kemungkinan bias (audit bias), selain menilai tingkat kecukupan sistem terhadap peraturan perundang-undangan dan pedoman etika yang ada (inspeksi regulasi), terutama dalam hal hak asasi manusia. Selain itu, vendor juga harus menerapkan penilaian dampak algoritmik, termasuk penilaian risiko dan dampak algoritma, yang pada akhirnya dapat mengevaluasi potensi dampak sosial dari sistem rekomendasi sebelum dan selama penerapannya dalam praktik.

Lebih lanjut, proses tersebut harus dikembangkan sebelum dan selama interaksi teknologi dengan pengguna. Jika sistem rekomendasi tidak disetujui dalam penilaian tersebut, gagal memenuhi persyaratan hukum dan etika, kegagalan yang teridentifikasi harus diatasi atau dimitigasi, melalui pengujian baru atau penerapan mekanisme perlindungan dan keamanan. Selain pengendalian sebelumnya yang dilakukan oleh penyedia rekomendasi itu sendiri, penting juga untuk melakukan pengendalian selanjutnya, tidak hanya melalui penilaian teknologi, tetapi juga melalui verifikasi dokumentasi dan bahkan audit eksternal oleh organisasi khusus. Pemantauan kepatuhan semacam itu harus menjadi bagian dari kerangka pengawasan pasar yang berkelanjutan untuk teknologi ini.

Jaminan Transparansi dan Penjelasan AI yang Lebih Besar (AI yang Dapat Dijelaskan)

Sistem rekomendasi harus dirancang untuk menjelaskan penalarannya dan memungkinkan manusia untuk menginterpretasikan hasil (rekomendasi). Sebagaimana telah disebutkan sebelumnya, penjelasan fungsi dan proses sangat penting untuk memastikan pelaksanaan hak, transparansi, dan akuntabilitas, yang sejalan dengan interpretasi hukum GDPR yang menetapkan hak atas penjelasan.

Penjelasan sistem rekomendasi dan keputusannya, sebagai dimensi dari prinsip transparansi, akan memungkinkan keseimbangan yang lebih besar antara kepentingan ekonomi dan sosial dengan memungkinkan adanya keputusan otomatis dan, secara bersamaan, mengurangi asimetri informasi antara mereka yang bertanggung jawab atas pemrosesan data dan pengguna sistem, karena hal ini menjadikan pengungkapan informasi sebagai kewajiban hukum.

Menurut Komisi Eropa, ketidakjelasan sistem AI dapat dikurangi melalui kewajiban transparansi, yang mencakup aksesibilitas dan pemahaman informasi. Tanpa transparansi yang memadai dalam proses dan keputusan, selain mekanisme konkret yang menjamin klarifikasi dan informasi yang efektif, pengguna mungkin kesulitan memahami sistem yang mereka gunakan dan rekomendasinya, yang akan mempersulit upaya memastikan akuntabilitas jika terjadi kerusakan. Dengan demikian, teknik rekomendasi yang dapat dijelaskan merupakan pendekatan penting untuk meningkatkan transparansi, efektivitas, keandalan, dan kepuasan pengguna terhadap sistem.

Rekomendasi yang dapat dijelaskan, misalnya, penting untuk e-commerce, karena meningkatkan persuasifitas saran dan, pada saat yang sama, membantu konsumen membuat keputusan daring yang efisien dan terinformasi. Strategi ini akan memfasilitasi proses menjadikan teknologi AI bertanggung jawab secara sosial dengan memastikan keuntungan komersial dan manfaat bagi pengguna. Selain itu, beberapa RS dapat memberikan informasi penting dan krusial untuk pengambilan keputusan yang sensitif, seperti dalam proses perawatan medis, di mana penjelasan hasil yang direkomendasikan sangat penting untuk memastikan perlindungan yang efektif terhadap nyawa dan kesehatan orang lain.

Kode Etik (Pengaturan Mandiri)

Selain regulasi hukum oleh Negara dan penetapan standar etika oleh organisasi yang berkepentingan, disarankan agar penyedia RS juga bertindak proaktif dalam penerapan sistem yang menghormati etika dan hak asasi manusia. Penetapan kode etik dan standar etika untuk penyampaian rekomendasi oleh platform itu sendiri dapat menjadi alat pengaturan mandiri yang penting, yang juga membantu perusahaan untuk mematuhi hukum ketika diterapkan secara efektif. Contoh dalam hal ini adalah pembentukan "Kemitraan Kecerdasan Buatan untuk Manfaat bagi Manusia dan Masyarakat", yang awalnya didirikan oleh beberapa perusahaan teknologi besar, seperti Microsoft, Google, Amazon, Facebook, dan IBM, untuk mempelajari dan merumuskan praktik terbaik AI, sesuai dengan prinsip-prinsip etika. Di antara tujuannya, kemitraan ini berupaya untuk memajukan pemahaman publik tentang teknologi, selain berfungsi sebagai platform untuk berdiskusi tentang AI dan kemungkinan dampaknya

terhadap manusia dan masyarakat (Kemitraan Kecerdasan Buatan). Namun, sangat penting bahwa kode dan prinsip pengaturan diri ini diterapkan secara efektif dalam praktik.

Pendidikan Digital dalam AI

Masyarakat harus dididik dari warga negara hingga eksekutif teknologi tingkat atas tentang manfaat, penyalahgunaan, dan potensi bahaya AI, terutama kecerdasan buatan. Peningkatan kesadaran akan AI di semua jenjang pendidikan sangatlah penting, guna mempersiapkan warga negara menghadapi era digital saat ini, sehingga mereka lebih mampu mengambil keputusan yang tepat dan semakin dipengaruhi oleh teknologi.

Dalam konteks ini, Rencana Aksi Pendidikan Digital yang baru-baru ini diluncurkan oleh Komisi Eropa, yang akan diterapkan antara tahun 2021–2027, merupakan contoh yang baik dari proyek pendidikan yang dapat diterapkan pada sistem rekomendasi. Salah satu tujuan utama yang ditetapkan adalah untuk meningkatkan keterampilan digital warga negara sejak masa kanak-kanak, yang mencakup investasi dalam pengetahuan dasar tentang AI, nilai-nilai etika yang terkait dengan teknologi ini, dan kesadaran akan keberadaan hak-hak digital.

Langkah-langkah tersebut akan menjadi strategi yang relevan untuk mengurangi asimetri informasi, selain mencegah risiko dengan meningkatkan kesadaran publik, memberdayakan pengguna, dan sebagai konsekuensinya, pelaksanaan hak-hak secara efektif. Pendekatan edukatif bahkan lebih penting bagi para profesional swasta yang berpartisipasi dalam proses pengembangan teknologi ini, karena mereka harus memahami tidak hanya cara menciptakan sistem yang akurat, tetapi juga membangunnya sesuai dengan pedoman etika dan hukum, berdasarkan hak asasi manusia dan nilai-nilai demokrasi. Misalnya, inisiatif lain yang didorong oleh Komisi Eropa adalah mengubah beberapa prinsip etika menjadi "kurikulum" yang harus diikuti oleh para pengembang AI, sebagai salah satu tahapan pelatihan mereka. Lebih lanjut, baik melalui inisiatif publik maupun swasta, pengembangan penelitian terkait etika dalam perangkat AI, seperti RS, sangatlah penting.

Regulasi Hukum Khusus untuk Sistem AI

Mengingat implementasi RS dan perangkat AI lainnya yang pesat di berbagai sektor, terutama di platform digital, dan konsekuensinya yang merugikan, terdapat beberapa inisiatif untuk menganalisis kemungkinan bentuk regulasi teknologi ini, dengan perhatian khusus pada perlindungan kelompok rentan. Untuk mengilustrasikan hal tersebut, inisiatif regulasi Uni Eropa akan dianalisis sebagai contoh, mengingat potensinya untuk memengaruhi regulasi lain di seluruh dunia akibat Efek Brussel, seperti yang terjadi dengan GDPR.

Dalam konteks ini, regulasi teknologi disruptif pertama kali ditetapkan melalui penetapan prinsip-prinsip etika, pedoman, dan opini tentang pengembangan dan penggunaan AI, seperti, misalnya, Pedoman Etika untuk AI Tepercaya 2019 oleh Kelompok Ahli Tingkat Tinggi Independen tentang Kecerdasan Buatan AI HLEG dan Buku Putih Komisi Eropa tentang AI pada Februari 2020. Dalam skenario ini, sebagaimana telah disebutkan pada topik sebelumnya, semua pemangku kepentingan yang terkait dengan RS harus memperhatikan standar-standar etika yang berlaku untuk AI. Namun, setelah sedimentasi prinsip-prinsip dasar dan pedoman yang berlaku untuk AI, Uni Eropa kini berupaya menerapkan aturan hukum yang mengikat yang secara khusus diterapkan pada teknologi ini, di samping undang-undang

perlindungan data yang sudah berlaku, yang contoh utamanya adalah GDPR. Oleh karena itu, baru-baru ini, badan legislatif Uni Eropa menyetujui dan memulai proses penyusunan undang-undang yang ditujukan untuk AI dan bidang-bidang penggunaannya (seperti platform digital). Dalam konteks RS, Undang-Undang Layanan Digital (DSA) yang baru-baru ini disetujui dan, lebih langsung lagi, usulan Undang-Undang Kecerdasan Buatan (AIA) merupakan contoh terpenting.

Undang-Undang Layanan Digital (DSA)

DSA adalah Peraturan Eropa yang menetapkan aturan bagi penyedia layanan masyarakat informasi tertentu (layanan digital), terutama melalui platform digital. Salah satu langkah inovatifnya adalah pembentukan aturan yang secara langsung membahas sistem rekomendasi yang disediakan oleh platform daring. Pertama, peraturan tersebut mendefinisikan RS pada Pasal 3(s) sebagai "sistem otomatis penuh atau sebagian yang digunakan oleh platform daring untuk menyarankan informasi tertentu dalam antarmuka daringnya kepada penerima layanan atau memprioritaskan informasi tersebut (...)", yang sejalan dengan premis Pertimbangan 70 bahwa RS merupakan bagian inti dari bisnis platform daring, karena memfasilitasi dan mengoptimalkan akses informasi bagi penerima layanan.

Oleh karena itu, karena RS memengaruhi cara informasi mengalir di platform digital, Peraturan tersebut berfokus pada pentingnya transparansi, yang menciptakan kewajiban berdasarkan Pertimbangan 70 dan Pasal 27 terkait informasi yang diwajibkan dalam syarat dan ketentuan platform digital (yang harus ditulis dalam bahasa yang lugas dan mudah dipahami) dan opsi yang harus disediakan platform ini kepada pengguna agar mereka dapat memahami, mengubah, atau memengaruhi parameter rekomendasi. Selain itu, khususnya dalam kasus penyedia platform daring yang sangat besar dan mesin pencari daring (pasal 33) yang menggunakan RS seperti Meta dan Google pasal 38 mewajibkan mereka untuk menyediakan setidaknya satu opsi untuk setiap RS mereka yang tidak didasarkan pada profil.

Dengan demikian, platform daring harus secara konsisten memastikan bahwa penerima layanan mereka mendapatkan informasi yang memadai tentang bagaimana sistem rekomendasi memengaruhi cara informasi ditampilkan dan dapat memengaruhi cara informasi disajikan kepada mereka. Mereka harus menyajikan parameter RS tersebut dengan jelas dan mudah dipahami untuk memastikan bahwa penerima layanan memahami bagaimana informasi diprioritaskan bagi mereka. Parameter tersebut harus mencakup setidaknya kriteria paling signifikan dalam menentukan informasi yang disarankan kepada penerima layanan dan alasan pentingnya masing-masing.

Karena RS memiliki dampak signifikan terhadap perilaku masyarakat dan cara mereka berinteraksi serta menemukan informasi daring, DSA bertujuan untuk memberdayakan pengguna melalui informasi dan pilihan, dengan meningkatkan aturan GDPR terkait kendali pengguna atas data pribadi. Misalnya, peraturan tersebut menetapkan kewajiban bagi penyedia RS platform yang sangat besar untuk melakukan penilaian risiko (Pasal 34 (2) (a)), memitigasi risiko yang ditemukan melalui pengujian dan adaptasi sistem algoritmik mereka (Pasal 35 (1) (d)), dan menjelaskan, atas permintaan Komisi Eropa atau Koordinator Layanan Digital, desain, logika, fungsi, dan pengujian sistem mereka (Pasal 40 (3)). Mempertimbangkan

permasalahan yang terkait dengan RS, memperkuat kewajiban transparansi pada platform daring dan menyediakan lebih banyak pilihan bagi pengguna merupakan langkah awal yang penting untuk mengatasi kekhawatiran yang ditimbulkan oleh teknologi ini (Pasal 19 2021).

Usulan Undang-Undang Kecerdasan Buatan (AIA)

Uni Eropa telah memiliki peraturan penting yang berlaku untuk AI, seperti GDPR, yang memberikan perlindungan pada tingkat tertentu. Namun, menurut Komisi Eropa, peraturan tersebut belum cukup untuk mengatasi semua tantangan yang mungkin ditimbulkan oleh teknologi ini, seperti yang telah dibahas pada topik sebelumnya.

Oleh karena itu, pada April 2021, Komisi mengusulkan peraturan hukum pertama yang secara khusus ditujukan untuk AI, yang bertujuan untuk memberikan persyaratan dan kewajiban yang jelas kepada pengembang, pengembang, dan pengguna AI terkait teknologi ini guna mendorong inovasi dan melindungi hak dan kebebasan fundamental yang berpotensi terancam, serta menciptakan lingkungan yang saling percaya.

Usulan ini disusun dengan pendekatan berbasis risiko, yang menangani risiko yang secara khusus ditimbulkan oleh aplikasi AI, yang mungkin dianggap tidak dapat diterima, tinggi, terbatas, atau minimal bagi keselamatan dan hak-hak fundamental manusia. Sesuai dengan Pertimbangan 14, meskipun sebagian besar sistem AI yang ada saat ini dianggap berisiko terbatas atau minimal, dan bermanfaat bagi masyarakat, bergantung pada intensitas dan cakupan risiko yang mungkin ditimbulkan oleh AI, beberapa praktik AI perlu dilarang; memberlakukan persyaratan untuk teknik AI berisiko tinggi dan kewajiban bagi operatornya; atau juga kewajiban transparansi untuk sistem AI tertentu.

Berbeda dengan yang terjadi dalam DSA, Proposal Undang-Undang AI tidak secara khusus membahas sistem rekomendasi, tetapi pasti akan berlaku untuk perangkat-perangkat ini, karena perangkat-perangkat ini berbasis AI dan pembuatan "rekomendasi" tercakup dalam definisi AI dalam Proposal pada Pasal 3 (1) sebagai salah satu kemungkinan keluarannya. Akibatnya, ada kemungkinan sistem rekomendasi akan diperlakukan berbeda sesuai dengan salah satu dari empat tingkat risiko yang mungkin ditimbulkannya dalam kasus tertentu.

Dengan demikian, pada awalnya, RS dengan risiko minimal atau tanpa risiko terkait akan bebas dikembangkan dan digunakan. Namun, dengan mempertimbangkan potensi penggunaan manipulatifnya, sistem ini dapat dilarang jika dikembangkan dengan "teknik subliminal di luar kesadaran seseorang dengan tujuan atau efek mendistorsi perilaku seseorang secara material" atau ketika "mengeksploitasi kerentanan kelompok orang tertentu karena usia, disabilitas, atau situasi sosial atau ekonomi tertentu, dengan tujuan atau efek mendistorsi perilaku seseorang yang termasuk dalam kelompok tersebut secara material" dengan cara yang menyebabkan atau cukup mungkin menyebabkan kerugian fisik atau psikologis.

Selain itu, terdapat kemungkinan besar bahwa sistem rekomendasi akan diklasifikasikan sebagai berisiko tinggi terhadap bahaya kesehatan, keselamatan, atau hak-hak dasar individu, sesuai dengan kriteria Proposal AIA, yang didefinisikan dalam Pasal 6 dan dilengkapi dengan daftar penerapan berisiko tinggi pada Lampiran III. Jika demikian halnya, sistem rekomendasi berisiko tinggi akan tunduk pada penilaian kesesuaian (pihak ketiga)

dengan serangkaian kewajiban sebelum dipasarkan atau digunakan seperti tata kelola data yang tepat (Pasal 10), penyusunan sistem manajemen dan mitigasi risiko yang memadai (Pasal 9), dokumentasi teknis (Pasal 11), pengawasan manusia yang tepat (Pasal 14), dan penyediaan informasi yang jelas dan memadai kepada pengguna (Transparansi—Pasal 13)—tetapi juga akan tunduk pada penegakan setelah RS tersebut digunakan.

Persyaratan ex-ante terkait transparansi dan penilaian risiko ini akan menciptakan kewajiban bagi penyedia RS untuk mendorong kepatuhan sejak awal dalam kasus sistem rekomendasi berisiko tinggi. Meskipun proposal tersebut memiliki beberapa aspek yang berkesan, sebagai regulasi pertama yang secara khusus ditujukan untuk AI dan menjadi inspirasi internasional, masih terdapat beberapa poin yang perlu diperhatikan, seperti penggunaan istilah yang samar, tidak adanya kewajiban untuk melakukan penilaian dampak hak asasi manusia, atau sedikitnya penyebutan tentang kemungkinan bahwa orang yang terdampak sistem AI memiliki wewenang untuk menentang dampak buruknya misalnya, dengan menetapkan hak untuk tidak menjadi subjek sistem AI yang tidak patuh, hak atas penjelasan, atau hak untuk mengajukan keluhan kepada otoritas pengawas.

Misalnya, jika sebuah RS memiliki dampak substansial terhadap kehidupan masyarakat, RS tersebut tidak hanya harus diberikan transparansi terkait implementasi sistem, tetapi terutama kemungkinan untuk menentang keputusannya. Dengan mempertimbangkan RS, oleh karena itu, harus ada opsi yang dapat diakses secara legal dan mudah bagi orang yang terdampak untuk mempertanyakan rekomendasi dan, jika demikian, menuntut pembatalan, pertimbangan ulang melalui prosedur yang berbeda, atau bahkan kompensasi.

Dalam kasus RS platform daring, melalui DSA, pengguna sudah dapat memodifikasi atau memengaruhi parameter utama sistem. Namun, solusi teknologi semata tidak cukup untuk memastikan bahwa sistem AI digunakan untuk kepentingan individu, bukan hanya penyedia. Pada titik ini, serupa dengan yang terjadi pada DSA, kerangka kerja akuntabilitas, yang memberdayakan mereka yang terdampak langsung oleh sistem tersebut, merupakan aspek penting dalam konteks AI ini.

Lebih lanjut, masyarakat sipil masih mengkritik teks terakhir proposal AIA, karena masih terdapat beberapa celah yang diperlukan untuk pendekatan berbasis hak asasi yang memadai, terutama dalam hal akuntabilitas yang bermakna, transparansi publik, dan partisipasi masyarakat sipil yang bermakna dan seimbang.

Dengan demikian, terdapat tren regulasi sistem AI, seperti sistem rekomendasi, yang bergeser dari pendekatan berbasis pedoman ke arah pengembangan undang-undang yang mengikat, seperti yang terjadi di Uni Eropa. Namun, regulasi-regulasi ini perlu tidak menjadi penghalang inovasi, menciptakan kewajiban yang terlalu kaku, atau sekadar memberikan kesan regulasi yang samar dan tidak efektif. Regulasi yang memadai sangat penting bagi inovasi yang bertanggung jawab yang dapat dicapai dengan instrumen tata kelola yang efektif, melalui regulasi yang proporsional dengan tingkat risiko sistem.

Sistem rekomendasi dapat memainkan peran krusial dalam masyarakat demokratis dan tidak hanya membahayakan, tetapi juga berkontribusi pada perwujudan hak-hak asasi dan nilai-nilai publik jika dikembangkan dan digunakan dengan baik. Inisiatif legislatif yang baru

harus memastikan bahwa sistem ini bekerja sesuai dengan nilai-nilai tersebut, bukan sebaliknya. Oleh karena itu, penggabungan DSA dan AIA yang diusulkan dapat meningkatkan pemberdayaan pengguna dan pilihan/kendali yang efektif, sehingga mengurangi potensi risiko dan kerugian. Ini merupakan langkah awal yang patut dipuji, tetapi perjalanan kita masih panjang.

8.6 REGULASI SISTEM REKOMENDASI: GDPR DAN ETIKA

Di dunia yang sangat terhubung, dengan data besar dan informasi yang melimpah, sistem rekomendasi semakin hadir dalam kehidupan kita, secara diam-diam memprediksi dan menyimpulkan minat, karakteristik, dan tindakan kita, memengaruhi keputusan kita, dan mengkategorikan kita dalam profil perilaku untuk mengirimkan konten yang dipersonalisasi. Meskipun manfaatnya tidak diragukan lagi dalam hal kenyamanan, manajemen waktu, dan pengaturan, alat-alat ini menimbulkan risiko yang cukup besar terhadap hak-hak asasi, seperti otonomi, privasi, perlindungan data, dan non-diskriminasi.

Oleh karena itu, mengingat semakin pentingnya sistem-sistem ini di saat yang sama dengan meningkatnya risiko dampak buruk, diperlukan penerapan yang efektif dan peningkatan kebijakan yang layak untuk menghadapi berbagai tantangan yang mungkin ditimbulkannya. Dengan kata lain, kecerdasan buatan yang diterapkan pada sistem rekomendasi harus diatur untuk mencegah kepentingan pribadi diutamakan daripada prinsip dasar "jangan merugikan".

Dalam lingkungan ini, GDPR merupakan kerangka regulasi fundamental untuk mengatasi berbagai risiko hak asasi manusia yang ditimbulkan oleh AI dalam sistem rekomendasi. Karena data merupakan penggerak teknologi ini, GDPR memperkenalkan struktur positif yang mendukung kontrol yang lebih besar bagi pengguna atas data mereka dengan menetapkan serangkaian hak, prinsip, dan persyaratan untuk pemrosesan data pribadi yang sah, terutama dalam hal pengambilan keputusan otomatis dan pembuatan profil. Banyak dari aturan hukum ini bersumber dari pedoman etika yang berlandaskan hak asasi manusia dan nilai-nilai seperti transparansi, keadilan, non-maleficence, beneficence, akuntabilitas, privasi, kebebasan, otonomi, martabat, dan solidaritas, yang juga fundamental untuk mengatasi ancaman yang ditimbulkan oleh RS.

Aturan hukum dan pedoman etika ini juga harus diperkuat oleh regulasi yang berasal dari teknologi itu sendiri, melalui strategi "desain yang berpusat pada nilai", di mana arsitektur RS mempertimbangkan parameter-parameter ini dalam cara kerjanya. Lebih lanjut, agar perangkat-perangkat ini dapat berfungsi untuk kepentingan manusia, perlu juga menjamin kecukupannya berdasarkan asesmen dampak dan audit algoritma, ditambah dengan penetapan kode etik oleh para pelaku pasar itu sendiri. Di samping itu, kebijakan "literasi media" sangat penting bagi perkembangan masyarakat yang mampu memahami logika sistem ini dan, dengan demikian, membuat keputusan yang efektif dan tepat untuk mendapatkan kembali kendali atas hidup mereka. Tak kalah pentingnya, pembentukan regulasi khusus untuk sistem AI atau lingkungan aplikasinya, seperti layanan digital yang disediakan oleh platform

daring, juga penting untuk menjamin penerapan yang baik dari semua aturan ini, karena banyak di antaranya akan terintegrasi dalam regulasi ini.

Oleh karena itu, dengan tujuan memaksimalkan manfaat dan memitigasi risiko yang terkait dengan RS, sehingga perangkat ini bermanfaat dan tidak merugikan individu dan masyarakat, pendekatan multisektoral dan multidisiplin sangat penting, menempatkan manusia di pusat dan melibatkan semua sektor masyarakat, termasuk kontribusi dari pedoman etika, fungsionalitas teknologi, inisiatif pengaturan mandiri pasar, kebijakan pendidikan, dan, untuk memastikan penerapan yang efektif, Undang-Undang, terutama yang secara langsung terkait dengan teknologi.

BAB 9
METAKOGNISI, AKUNTABILITAS, DAN KEPRIBADIAN HUKUM AI

Abstrak Salah satu teka-teki yang belum terpecahkan terkait Kecerdasan Buatan (AI) adalah apakah robot dapat dianggap akuntabel dan, pada akhirnya, memiliki kepribadian hukum. Dengan masukan dari Filsafat, Psikologi, Komputasi, dan Hukum, makalah ini mengusulkan pendekatan interdisipliner terhadap pertanyaan tentang kepribadian hukum dalam AI.

Dalam makalah ini, kami mengkaji, pertama, konsep Objek (sekadar alat) dan Agen, untuk memahami kategori AI. Kedua, kami menganalisis bagaimana Metakognisi, yang secara luas didefinisikan sebagai kognisi tentang kognisi, yang menghasilkan proses mental yang mengendalikan pikiran dan perilaku suatu entitas, dapat diterapkan pada hukum sebagai persyaratan minimum akuntabilitas.

Sebagai contoh, kita akan melihat bahwa baik anak-anak maupun orang dengan penyakit mental, selain menjadi dua kategori subjek yang memiliki kapasitas hukum yang sangat terbatas, juga menunjukkan beberapa keterbatasan dalam hal Metakognisi. Dengan kata lain, kami berpendapat bahwa perbedaan utama antara Agen yang tidak bertanggung jawab dan Agen yang bertanggung jawab bergantung pada proses metakognitif yang dapat dijalankan oleh entitas tersebut. Pada akhirnya, kami membahas bagaimana mentransposisikan gagasan ini ke dalam AI, dengan memperdebatkan kemungkinan ketentuan mengenai status badan hukum AI.

9.1 STATUS HUKUM AI: AGEN, METAKOGNISI, AKUNTABILITAS

Salah satu teka-teki yang belum terpecahkan terkait Kecerdasan Buatan (AI) adalah apakah robot dapat dianggap bertanggung jawab dan, pada akhirnya, memiliki status badan hukum. Dengan masukan dari Filsafat, Psikologi, Komputasi, dan Hukum, makalah ini mengusulkan pendekatan interdisipliner terhadap pertanyaan tentang status badan hukum dalam AI. Dalam makalah ini, kami mengkaji, pertama, konsep Objek (sekadar alat, tidak tunduk pada status badan hukum) dan Agen, untuk memahami kategori AI.

Kedua, karena konsep Agen menghadirkan banyak kesulitan, yaitu karena tampaknya memiliki makna yang berbeda menurut masing-masing domain pengetahuan yang disebutkan di atas, sebuah persamaan umum diidentifikasi, yaitu tindakan sukarela. Jika ada tindakan sukarela, maka kita harus menyimpulkan bahwa kita memiliki Agen di hadapan kita.

Dengan demikian, dan selama AI bertindak secara sukarela, masuk akal untuk berargumen bahwa robot yang kompleks (dalam arti AI yang kuat) adalah Agen, dengan demikian bukan sekadar alat. Ketiga, karena anak-anak, hewan, dan orang dengan gangguan jiwa bertindak secara sukarela tetapi masih belum dimintai pertanggungjawaban (baik tidak memiliki status badan hukum maupun pelaksanaan status badan hukum tersebut secara terbatas), makalah ini menyelidiki apa yang hilang dalam kasus-kasus ini, untuk menarik garis pemisah antara agen yang bertanggung jawab dan yang tidak bertanggung jawab.

Terakhir, kami menganalisis bagaimana Metakognisi, sebuah konsep yang dipinjam dari Psikologi, yang secara luas didefinisikan sebagai kognisi tentang kognisi, yang menghasilkan proses mental yang mengendalikan pikiran dan perilaku suatu entitas, dapat diterapkan pada hukum sebagai persyaratan minimum untuk akuntabilitas dan akhirnya status badan hukum. Sebagai contoh, kita akan melihat bahwa baik anak-anak maupun orang dengan gangguan jiwa, selain menjadi dua kategori subjek yang memiliki kapasitas hukum yang sangat terbatas, juga menunjukkan beberapa keterbatasan dalam hal Metakognisi.

Dengan kata lain, kami berpendapat bahwa perbedaan utama antara Agen yang bertanggung jawab dan yang tidak bertanggung jawab bergantung pada proses metakognitif yang dapat dilakukan oleh entitas tersebut. Pada akhirnya, kami membahas bagaimana mentransposisikan gagasan ini ke dalam AI, dengan memperdebatkan kemungkinan istilah status badan hukum AI. Tidak diragukan lagi bahwa Hukum bergantung, sampai batas tertentu, pada deskripsi dan klasifikasi masalah yang kita hadapi. Dengan kata lain, ketika dihadapkan pada situasi tertentu, kita dipaksa untuk membuat daftar ciri-ciri esensialnya dan melihat apakah ciri-ciri tersebut sesuai dengan norma hukum. Jika ya, kita telah menemukan solusi hukum untuk masalah tersebut; jika tidak, kita harus terus mencari kecocokannya.

Dalam hal status badan hukum, terdapat beberapa syarat dasar yang, jika tidak ada, menutup kemungkinan untuk mempertimbangkannya pada suatu entitas tertentu. Misalnya, tidak seorang pun berpikir untuk menyebut orang yang telah meninggal sebagai badan hukum, meskipun beberapa hak mungkin dapat diperluas setelah kematian (seperti hak atas kehormatan). Status badan hukum bagi manusia menyiratkan bahwa ia masih hidup, karena status ini dimulai saat kita lahir. Ketika sesuatu tidak sepenuhnya sesuai dengan kategori yang kita, manusia, ciptakan, misalnya jika kita agak hidup tetapi belum lahir (bayi yang belum lahir), menjadi tidak jelas bagi kita apa yang harus dilakukan terkait entitas tersebut.

Dalam hal ini, telah dikemukakan bahwa langkah pertama untuk membangun kerangka hukum, dalam kasus robot jenis apa pun, adalah menentukan statusnya, artinya mendefinisikan konsep dan batasannya, lalu membandingkannya dengan pilihan hukum yang tersedia. Dalam hal ini, Pagallo telah mengembangkan penelitian ekstensif untuk memahami ciri-ciri utama setiap jenis robot yang direncanakan dalam waktu dekat, dalam bukunya *The Laws of Robots: Crimes, Contracts, and Torts.*

Secara global, penulis membagi kemungkinan menjadi tiga kategori: (1) Badan Hukum, (2) Agen yang Tepat, dan (3) Sumber Kerugian. Artinya, dalam praktiknya, kita harus memeriksa apakah suatu robot memiliki atribut yang cukup dengan manusia, sehingga kita dapat memberinya status Badan Hukum (Hipotesis 1). Jika robot tersebut memiliki lebih banyak kesamaan, artinya lebih banyak kesamaan fitur, dengan konsep alat, sehingga dianggap sebagai objek belaka, maka jawabannya adalah memperlakukannya sebagai objek hukum (Hipotesis 3). Yang juga bisa terjadi adalah robot tidak sepenuhnya sama dengan kategori mana pun, namun memiliki sejumlah atribut yang sama dengan masing-masing kategori. Oleh karena itu, kita memiliki Agen yang Tepat (Hipotesis 2), apa pun istilah hukum yang ingin kita terapkan padanya.

Oleh karena itu, pada tahap awal, penting untuk memahami apa artinya menjadi Agen. Jika suatu entitas adalah Agen, maka entitas tersebut bukanlah benda, karena hukum logika non-kontradiksi tidak mengizinkan hal ini terjadi. Mengingat fakta bahwa satu benda saling bertentangan (dan keduanya memang demikian, karena keduanya menunjukkan sifat yang berbeda dan berlawanan), kalimat "Robot A adalah Agen" dan kalimat "Robot A adalah benda" tidak mungkin benar pada saat yang bersamaan. Misalnya, seorang Agen, seperti yang akan kita lihat, bertindak secara sukarela, sementara benda tidak bertindak sama sekali. Tampaknya jelas bahwa satu entitas tidak dapat bertindak secara sukarela dan tidak bertindak sama sekali pada saat yang bersamaan. Dengan memahami apa itu agen dan berargumen bahwa robot adalah Agen, kita secara otomatis mengesampingkan gagasan bahwa ia bisa menjadi benda. Pada tahap kedua, kita akan mengkaji apa artinya menjadi Agen yang bertanggung jawab secara hukum.

Di sisi lain, dan mendukung gagasan yang dikemukakan oleh Asaro, pemahaman konsep Agen saja dapat membantu kita untuk menggambarkan batasan-batasan personifikasi hukum, karena konsep pertama berjalan beriringan dengan konsep kedua. Dengan kata lain, Agensi mungkin menyembunyikan petunjuk penting dalam ranah ini. Dapat diprediksi, memahami konsep Agen dan menyebutkan fitur-fitur utamanya hampir mustahil. Setiap bidang ilmu pengetahuan menggunakan gagasan Agen, namun, belum ditemukan konsensus. Misalnya, Psikologi, Filsafat, Hukum, Komputasi, Ekonomi, dan Neurosains, masing-masing mengambil konsep Agen dan mengisinya dengan atribut-atribut yang paling sesuai dengan ranahnya. Dalam hal ini, Shardlow memiliki tesis yang sangat menarik di mana ia mencapai kesimpulan ini, meskipun penulis menyelidiki tiga bidang utama: Filsafat, Psikologi, dan Komputasi. Dihadapkan dengan fakta ini, kita terpaksa berpendapat bahwa konsep Agen adalah jalan buntu. Namun demikian, mungkin ada sesuatu yang dapat dilakukan untuk mengatasi jalan buntu ini.

Ada metode dalam pemrograman dan komputasi, yang digunakan programmer ketika mereka harus menggambarkan masalah yang kompleks: mereka menggambar kasus dasar. Kasus dasar, sederhananya, adalah deskripsi kasus paling sederhana yang mungkin dalam situasi kompleks. Pada tahap kedua, kemudian, muncul pembangunan dan penulisan kode kasus kompleks dan pengecualiannya. Lalu, apa kasus dasar kita dalam hal Agensi? Apa satu hal atau, lebih tepatnya, satu-satunya fitur yang, terlepas dari bidang yang kita teliti, selalu ada?

Seperti yang akan dibahas, apakah itu tindakan sukarela? Namun, akan ditunjukkan juga bahwa ini tidak cukup, karena anak-anak, hewan, dan penyandang disabilitas mental memang bertindak secara sukarela tetapi tidak dianggap bertanggung jawab secara hukum. Langkah selanjutnya adalah menentukan apa yang kurang dalam kasus-kasus ini, dengan menggunakan sumber daya dari bidang Psikologi, yaitu beberapa jenis proses metakognitif yang berkaitan dengan kemampuan merasa bersalah dan kapasitas merencanakan perilaku kompleks. Dalam hal ini, selain kapasitas bertindak secara sukarela, setiap entitas yang bertanggung jawab harus menunjukkan jenis proses metakognitif tertentu. Hanya dengan

demikian akuntabilitas menjadi pilihan. Untuk pemahaman yang komprehensif tentang makalah ini, halaman berikutnya memberikan garis besar visual strukturnya.

9.2 PERSAMAAN DALAM AGENSI

Secara intuitif, kita masing-masing memiliki gagasan tentang apa artinya menjadi Agen. Agen adalah entitas, apa pun jenisnya, yang mampu bertindak dan mengeksekusi tindakan, berlawanan dengan entitas lain yang hanya menoleransi atau menerima peristiwa yang terjadi pada mereka. Namun, untuk menemukan definisi yang konsensual, kita harus meningkatkan tingkat abstraksi. Dalam pengertian abstrak ini, dan untuk tujuan ini, Agen adalah entitas yang bertindak terus-menerus dan otonom dalam waktu, dalam lingkungan yang dinamis, di mana proses lain ada, dan Agen lain hadir.

Dalam Filsafat, dua teori yang paling menonjol adalah Konsepsi Standar dan Teori Standar. Keduanya berpendapat bahwa Agen adalah makhluk yang mampu melakukan tindakan yang disengaja. Perbedaan antara kedua teori ini berkaitan dengan apakah intensionalitas tindakan tersebut mencakup tindakan yang tidak diinginkan atau tidak. Misalnya, mari kita bayangkan Asimov ingin meraih segelas airnya, di tengah malam, dan menyalakan lampu untuk melakukannya. Kita berasumsi bahwa tindakan yang terakhir itu diinginkan olehnya, dan disengaja, karena sebelum benar-benar bertindak, ia sempat berpikir untuk menyalakan lampu demi mendapatkan segelas air. Namun, jika ada pencuri di luar rumahnya dan ia tidak menyadari hal ini, ia mungkin juga memberi tahu pencuri itu bahwa ia ada di rumah, meskipun itu bukan yang ingin ia lakukan.

Meskipun ingin menyalakan lampu, pikiran Asimov jelas bukan untuk memberi tahu pencuri itu, tetapi ia tetap melakukannya. Inilah yang dimaksud dengan tindakan yang tidak diinginkan. Konsepsi Standar berpendapat bahwa tindakan yang disengaja mencakup menyalakan lampu untuk kaca dan menyalakan lampu untuk memperingatkan pencuri; di sisi lain, Teori Standar berpendapat bahwa hanya yang pertama yang merupakan tindakan yang disengaja. Meskipun tidak sependapat tentang arti tindakan yang disengaja, kedua teori tersebut meyakini bahwa Agen adalah entitas yang mampu melakukan tindakan yang disengaja. Dengan demikian, menurut perspektif ini, suatu entitas disebut Agen jika dapat bertindak secara sukarela, karena tindakan tersebut bergantung pada keyakinan bahwa tindakan spesifik yang dimaksud adalah yang terbaik untuk mencapai tujuan tertentu.

Tentu saja, dan terutama tidak dalam Filsafat, ini bukan satu-satunya teori yang menjadi pusat perdebatan. Teori lain dijelaskan oleh Dennett. Penulis ini berpendapat bahwa kita memiliki Agen di hadapan kita jika kita dapat memprediksi perilakunya, secara akurat, melalui kondisi mentalnya. Oleh karena itu, Allen dan Bekoff menggunakan gagasan ini, dengan alasan bahwa gagasan tersebut dapat diterapkan pada Agen non-manusia.

Baru-baru ini, Barandiaran dkk. berfokus pada entitas yang sangat sederhana, seperti bakteri. Menurut pendapat penulis, fakta bahwa entitas semacam ini tidak dapat dimasukkan dalam kategori Agen, mengingat teori-teori Filsafat yang telah disebutkan sebelumnya, tidak berarti mereka tidak boleh dianggap sebagai Agen. Dalam hal ini, Barandiaran menguraikan tiga prasyarat utama untuk apa yang disebutnya sebagai Agensi minimum.

Selain individualitas (yang merupakan perbedaan yang jelas antara Agen dan lingkungannya) dan normativitas (artinya keberadaan tujuan dan aturan yang digunakan Agen untuk memandu tindakannya), ia juga berpendapat bahwa asimetri interaksional sangat penting. Prasyarat terakhir untuk Agensi ini menyangkut kemampuan untuk bertukar energi dan materi dengan lingkungan. Dengan kata lain, Agen harus mampu mengumpulkan energi yang diperlukan untuk bertindak, dan menjadi entitas pasif di lingkungan saja tidak cukup.

Sejauh ini, dalam Filsafat, tampaknya tindakan sukarela merupakan prasyarat yang relevan untuk menganggap Agensi. Sebagaimana akan kita lihat nanti, ini bukan satu-satunya bidang pengetahuan di mana kemampuan ini merupakan prasyarat. Faktanya, itulah yang terjadi dalam Komputasi. Sementara Minsky memandang Agen buatan sebagai Mesin Keadaan Terhingga (FSM), sebuah deskripsi yang sering dianggap reduktif, penulis lain seperti Russell dkk. memandang Agen sebagai entitas yang menganalisis lingkungan sekitarnya dan bertindak sesuai dengan masukan dari lingkungan yang sama. Pandangan lain yang sangat dipuji adalah pandangan yang dijelaskan oleh Wooldridge dan Jennings yang mendefinisikan Agen sebagai entitas yang memiliki sifat-sifat seperti otonomi, keterampilan sosial, reaktivitas terhadap lingkungan, dan proaktivitas (kemampuan untuk memulai tindakan).

Menurut para penulis, entitas yang menunjukkan atribut-atribut kumulatif ini memiliki apa yang mereka sebut Agensi yang lemah. Sebaliknya, jika kita mencari Agensi yang kuat, Agen harus menunjukkan beberapa tingkat proses kognitif, termasuk keyakinan, keinginan, dan niat. Tidak mungkin hanya meneliti setiap bidang ilmu pengetahuan untuk menemukan arti menjadi Agen di masing-masing bidang. Masih ada satu lagi yang harus dikaji dan merupakan ranah yang sangat kompleks: Hukum.

Dalam Hukum, seorang Agen biasanya dianggap sebagai pelaku tindakan terlarang (misalnya kejahatan), yang dilakukannya secara sukarela. Permasalahan terbesar dalam hal ini adalah agar dapat dianggap sebagai Agen, dalam pengertian yang digunakan oleh Hukum, terdapat gagasan implisit bahwa Agen memiliki status badan hukum. Karena kita mencoba melakukan yang sebaliknya, artinya kita mencoba mencapai akuntabilitas dan status badan hukum melalui gagasan Agen, hal ini tidak terlalu membantu.

Sebaliknya, yang dapat kita lakukan, karena konsep badan hukum dapat dianggap sebagai unit dasar hukum, agar dapat bertindak dalam hubungan hukum, adalah mengkaji apa yang membedakan pemberian status badan hukum kepada suatu entitas. Dengan kata lain, penting untuk menyelidiki alasan di balik keputusan pembuat undang-undang untuk memberikan atau tidak memberikan status hukum ini kepada suatu entitas.

Alasan pertama untuk memberikan status badan hukum, jelas dan alami, adalah karena entitas tersebut adalah manusia (yang lahir namun belum meninggal). Secara artifisial, kami juga menganggap perusahaan memiliki status badan hukum, dengan justifikasi teori yang berasal dari Savigny dan sama sekali tidak akan dibahas dalam makalah ini.

Dalam hal ini, terdapat dua teori utama dalam yurisprudensi analitik mengenai hal ini: teori kehendak dan teori kepentingan. Sebagian besar akademisi hukum Jerman abad ke-19 yang menulis tentang topik ini mendasarkan teori mereka pada gagasan Kant tentang kebebasan dan otonomi sebagai konsep sentral. Manusia, menurut teori ini, memiliki

kebebasan moral bawaan, yang mendasari kapasitas mereka untuk memiliki hak dan dengan demikian status badan hukum mereka. Namun, pandangan minoritas menganjurkan pemahaman hak yang berbasis kepentingan. Teori-teori analitik modern tentang hak biasanya dapat diklasifikasikan sebagai salah satu dari teori-teori ini. Namun, hampir tidak ada teori yang dapat dikatakan 'memenangkan' perdebatan.

Selain itu, teori-teori ini masih belum cukup untuk menarik garis pemisah antara entitas yang bertanggung jawab dan yang tidak bertanggung jawab. Hakim-hakim Anglo-Saxon telah merefleksikan konsep Agen secara ekstensif, jauh sebelum konsep tersebut menjadi kesimpulan yang tak terelakkan bagi kita. Salmond, berpendapat bahwa untuk menjadi badan hukum, seseorang harus menunjukkan kapasitas untuk menjadi bagian dalam hubungan-hubungan hukum. Di sisi lain, Dewey menjelaskan bagaimana kita tidak berpikir untuk memberikan status badan hukum kepada sesuatu, karena perilaku mereka akan tetap sama, terlepas dari apakah kita memberikan kewajiban hukum kepadanya atau tidak.

Menurut penulis, kita memberikan status badan hukum kepada entitas yang perilakunya dapat diatur oleh norma hukum atau kepada entitas yang melaluinya kita ingin mengatur perilaku manusia, inilah alasan mengapa kapal pernah diberikan status badan hukum. Baru-baru ini, Dario dan Palmerini, berdasarkan teori-teori tentang Kepribadian Hukum yang telah disebutkan sebelumnya, mengaitkan konsepsi tentang kepribadian hukum dengan gagasan tentang kewajiban dan pemikiran tentang kemampuan untuk bertindak guna menegakkan kewajiban yang sama.

Saat ini, dan secara umum, beberapa penulis (misalnya, Mathew Kramer dan Joel Feinberg, penulis terakhir yang membahas tentang hewan) telah mendukung konsepsi khusus tentang kepribadian hukum: konsepsi yang berpendapat bahwa setiap entitas yang mampu memiliki hak hukum harus diberikan kepribadian hukum.

Visi ini cukup dipuji, yang secara inklusif menjadi dasar utama untuk sebuah kasus pada bulan Desember 2014, di Mahkamah Agung New York tentang seekor simpanse bernama Tommy. Perwakilan Tommy meminta perluasan konsep kepribadian hukum, agar dapat mengajukan permohonan habeas corpus di kemudian hari. Perwakilan tersebut berargumen, tepatnya, bahwa hewan dapat memiliki setidaknya satu hak hukum, dan hal ini cukup untuk mendapatkan jenis status badan hukum tertentu, sesuai dengan hak-hak yang dicanangkan. Pietrzykowski menggambarkan kasus serupa di Pengadilan Argentina, tentang seekor Orangutan di Kebun Binatang Buenos Aires.

Terdapat banyak literatur yang membahas perdebatan ini. Pandangan relevan lainnya mencakup Rasionalitas sebagai kriteria utama untuk status badan hukum dan Intensionalitas. Penting untuk menyatakan bahwa kita tidak boleh, sama sekali, memisahkan Hukum dari realitas tempat ia beroperasi. Hukum bersifat permeabel terhadap realitas dan budaya dan ini merupakan hubungan yang krusial jika kita ingin menghindari undang-undang yang usang dan tidak berguna. Inilah mengapa semua teori Hukum yang berbeda ini sangat penting dalam penelitian ini.

Juga relevan untuk menunjukkan bahwa terlepas dari pandangan yang didukung, tampaknya selalu ada gagasan tentang kemampuan untuk bertindak (dalam arti bahwa jika

seseorang mampu menjalankan hak atau kewajiban hukum, ia harus mampu bertindak sesuai dengannya) yang menyelimuti semua teori yang disebutkan. Hal yang sama terjadi dalam Komputasi dan Filsafat, meskipun dengan makna yang berbeda. Kesimpulannya, tampaknya kata-kata yang berbeda digunakan untuk menyebut hal yang sama. Sebagaimana dijelaskan sebelumnya, setiap bidang ilmu pengetahuan mengambil konsep Agen untuk dirinya sendiri dan merancangnya menurut citra dan rupa masing-masing. Terlepas dari kenyataan ini, betapapun berbedanya definisi Agen, kondisi memiliki kekuatan untuk bertindak, secara sukarela, selalu ada.

9.3 PENGERTIAN TINDAKAN SUKARELA

Markby, dalam Elements of Law Principles of Jurisprudence mendefinisikan tindakan sukarela sebagai gerakan tubuh yang mengikuti kehendak. Secara kebetulan, dalam bidang ilmu pengetahuan lain dalam Filsafat Klasik Davidson menggunakan deskripsi yang sama persis ini, 84 tahun kemudian, ketika menulis teorinya tentang Agensi. Hal yang sama diperdebatkan berulang kali sepanjang abad ke-20 dan ke-21 dalam Hukum, meskipun dengan kata-kata yang berbeda yaitu oleh Cook dan Yaffe.

Dalam Psikologi, James dkk. menggambarkan tindakan sukarela sebagai lawan dari tindakan tak sukarela, dalam arti bahwa tindakan tak sukarela terjadi tanpa pandangan ke depan. Dalam Filsafat modern, hal serupa dikemukakan oleh Olsaretti, yang mendukung gagasan bahwa kita memiliki tindakan sukarela jika kita tidak memiliki tindakan tidak sukarela. Tindakan tersebut tidak akan bersifat sukarela, dalam tesis penulis, jika tidak ada pilihan lain yang dapat diterima, menurut beberapa kriteria objektif (meskipun penulis tidak menjelaskan secara rinci kriteria objektif ini).

Bagi Olsaretti, pilihan yang tidak dapat diterima adalah pilihan yang menyebabkan kerugian tertentu bagi Agen atau ketika suatu aturan moral menjadi keharusan sehingga membuat semua pilihan lain tidak dapat diterima. Ia juga menyatakan bahwa tindakan sukarela sangat berkaitan dengan motivasi Agen, dalam arti bahwa tindakan tersebut, tak terelakkan, bergantung pada keyakinan yang dimiliki Agen tentang pilihan-pilihannya. Jika Agen keliru tentang pilihan-pilihannya, ia mungkin memiliki pilihan yang baik tetapi tidak menyadari keberadaannya. Dengan demikian, suatu tindakan dapat bersifat tidak sukarela karena misinformasi.

Hal inilah yang diperdebatkan oleh Aristóteles, dalam Etika Nikomakhean: bahwa hanya ada dua alasan yang membuat suatu tindakan menjadi tidak sukarela, yaitu ketidaktahuan atau kekuatan eksternal yang besar. Kesimpulannya, suatu tindakan tampak sukarela ketika ada gerakan tubuh, yang dipandu oleh kehendak, selama tidak dirusak oleh ketidaktahuan atau kekuatan eksternal.

Pertanyaan selanjutnya adalah: apakah AI bertindak secara sukarela? AI mungkin memiliki struktur keyakinan, keinginan, dan niat yang telah ditentukan sebelumnya (oleh manusia), tetapi setelah definisi awal tersebut, AI yang lebih kompleks (atau lebih kuat) mampu bertindak terhadap lingkungan secara otonom, dan mungkin sesuai dengan tujuan yang mereka tetapkan sendiri. Kita telah sampai pada titik di mana AI begitu maju sehingga

dalam beberapa kasus bahkan penciptanya pun tidak tahu persis mengapa robot melakukan apa yang dilakukannya.

Dalam kondisi normal, robot terinformasi dengan baik tentang pilihannya, karena mampu mengumpulkan informasi penting untuk menciptakan model dunia. Dalam keadaan normal pula, mereka tidak akan dipaksa untuk melakukan apa pun, meskipun mungkin saja dipaksa. Jadi, apakah robot termasuk dalam kategori Agen? Tampaknya dalam kasus AI yang kuat atau kompleks, robot (yang disebut AI robust) tampaknya memiliki persyaratan minimum untuk dianggap sebagai Agen: mereka bertindak secara sukarela.

Penting untuk disangkal bahwa dengan merujuk pada AI kompleks, yaitu mesin yang menggunakan proses kognitif atau pembelajaran mesin, kami tidak sedang menggambarkan objek yang jelas-jelas bertindak sebagai alat dan yang dipersepsikan serta dimaksudkan untuk bertindak seperti itu, seperti AC pintar yang menyesuaikan suhu atau intensitas lampu yang berubah-ubah sesuai jam. Seperti yang telah disebutkan sebelumnya, jika AI kompleks termasuk dalam kategori Agen, ia tidak dapat dianggap sekadar alat. Yang penting sekarang adalah mempelajari seberapa besar tanggung jawab yang dapat mereka pikul, jika ada.

9.4 KRITERIA TANGGUNG JAWAB HUKUM AGEN

Langkah selanjutnya adalah mencoba memahami apa yang membuat Hukum menetapkan tanggung jawab atau tidak kepada seseorang. Menurut definisi Agen sebelumnya, tampak jelas bahwa anak-anak dan hewan juga merupakan Agen. Namun, kita tidak menganggap mereka sebagai Agen yang bertanggung jawab secara hukum. Dengan kata lain, hanya menjadi Agen dan bertindak secara sukarela tidaklah cukup bagi Hukum. Dalam hal ini, di mana kita harus menarik garis batas antara agen yang bertanggung jawab dan yang tidak bertanggung jawab?

Ada satu konsep hukum yang sangat relevan yang mungkin dapat membantu kita dalam pertanyaan ini, yaitu gagasan imputabilitas. Namun, makna yang ingin kita pahami di sini adalah ketiadaan imputabilitas, yang berkaitan dengan kategori orang tertentu yang, baik karena mereka di bawah umur atau menderita penyakit mental, tidak dapat kita tetapkan tanggung jawab hukum, meskipun mereka memiliki status badan hukum. Meskipun terdapat banyak alasan dan teori mengapa Hukum tidak menganggap individu-individu ini bertanggung jawab, salah satu alasan utama yang mengkhawatirkan adalah kenyataan bahwa subjek-subjek ini tidak menunjukkan kapasitas untuk merasa bersalah.

Dalam pengertian hukum, rasa bersalah dipahami sebagai kapasitas yang dimiliki subjek untuk bertindak secara bertanggung jawab, artinya ia mampu memahami perilaku terlarang dan oleh karena itu memilih untuk tidak melakukan perilaku tersebut. Dalam pengertian ini, subjek harus mampu merefleksikan perilaku tertentu dan menyatakan nilai positif atau negatif terhadap perilaku tersebut.

Dengan kata lain, kita hanya dapat menetapkan tanggung jawab hukum apa pun ketika kita berasumsi bahwa Agen memiliki persyaratan minimum, dari sudut pandang fisik dan psikologis, untuk merespons aturan normatif secara positif. Dengan adanya serangkaian persyaratan minimum ini, maka kita memiliki Agen yang dapat dipertanggungjawabkan.

Selain membantu dalam penilaian kasus pidana, rasa bersalah juga berkaitan dengan penilaian negatif yang dikembangkan masyarakat terhadap perilaku Agen. Sama sekali tidak ada gunanya membahas penilaian negatif atas perilaku terhadap Agen yang tidak mampu memahami penilaian tersebut. Hal itu tidak akan efektif. Dalam kasus ini, kognisi Agen mungkin sangat terganggu sehingga meskipun ia dapat bertindak secara sukarela, sesuai dengan keinginan atau tujuan tertentu, ia tidak dapat merefleksikan kondisi mental (primer) yang memicu perilaku tersebut. Dalam Filsafat, serta dalam Psikologi Kognitif, kondisi mental ini terkait dengan kondisi mental primer lainnya, dikenal sebagai Metakognisi.

9.5 METAKOGNISI: MEMBENTUK TANGGUNG JAWAB HUKUM

Rasanya adil untuk mengatakan bahwa kita diperbolehkan mentransposisikan konsep dari satu ranah pengetahuan ke ranah pengetahuan lainnya. Sebagian besar fondasi Hukum Modern berasal dari penulis seperti Kelsen, Hart, dan Austin, yang semuanya juga filsuf, yang meletakkan dasar bagi Filsafat Hukum. Di sisi lain, kita tidak dapat membuat undang-undang tentang dunia di sekitar kita tanpa mengembangkan konsep-konsep duniawi. Misalnya, kita tidak akan dapat membuat undang-undang tentang Kedokteran jika kita tidak mampu memahami konsep-konsep bidang ilmu pengetahuan tersebut.

Lebih lanjut, beberapa penulis seperti Morse berpendapat bahwa Hukum itu sendiri menggunakan model tindakan yang berasal dari Psikologi Rakyat. Dengan kata lain, sah bagi kita untuk menggunakan konsep-konsep yang telah lama digunakan di bidang ilmu pengetahuan lain, dalam hal ini, gagasan Metakognisi, yang merupakan konsep yang relatif tua dalam Filsafat dan Psikologi Kognitif.

Secara umum, Metakognisi adalah kognisi tentang kognisi, yang berguna untuk mengendalikan dan/atau memantau perilaku dan pemrosesan mental. Frankfurt, seorang filsuf, berpendapat bahwa perbedaan utama antara manusia dan jenis Agen lainnya berakar pada struktur kehendak, dalam arti bahwa hanya manusia yang merenungkan motivasi mereka sendiri, yang menghasilkan kondisi mental tingkat kedua. Misalnya, mari kita bayangkan Wall-e harus belajar untuk ujian. Agar berhasil dalam ujian ini, Wall-e harus terlebih dahulu membuat daftar metode belajar yang ia ketahui, menganalisis karakteristik kuat dan lemahnya sendiri, sehingga ia dapat memilih metode belajar terbaik untuknya, dengan mempertimbangkan subjek spesifik yang harus dipelajarinya.

Belajar itu sendiri merupakan proses kognitif. Dengan merenungkan proses kognitif ini (memilih metode belajar), ia menggunakan kondisi mental tingkat kedua ini atau, seperti yang dijelaskan oleh banyak penulis, kognisi sekunder. Bagi Frankfurt, perbedaan antara manusia dan Agen lainnya, yang juga bertindak secara sukarela, adalah Metakognisi.

Dengan demikian, kita dapat berargumen bahwa terdapat perbedaan antara Agen yang bertindak secara sukarela tetapi tidak menunjukkan Proses Metakognitif, dan Agen yang bertindak secara sukarela dan menunjukkan kapasitas ini. Agen yang bertindak secara sukarela dan menampilkan proses metakognitif dapat melakukannya dengan beberapa cara, karena jenis kognisi ini memiliki banyak bentuk dan wujud, dan tidak semuanya akan dijelaskan dalam makalah ini. Namun, seperti yang akan kita lihat, untuk meminta pertanggungjawaban suatu

entitas, setidaknya diperlukan dua jenis proses metakognitif: proses strategis dan proses pemantauan. Keduanya akan dijelaskan selanjutnya melalui urutan ini.

Sebagaimana dinyatakan Cox, setiap Agen yang cerdas, ketika dihadapkan pada suatu pilihan (pilihan apa pun, termasuk pilihan untuk melakukan tindakan terlarang atau tidak), ia harus memutuskan tiga hal: (1) tindakan mana, mengingat kemungkinan-kemungkinannya, yang paling tepat dalam situasi saat ini, (2) apakah pilihan yang ia buat cukup berdasar atau jika diperlukan informasi lebih lanjut, dan (3) jika terjadi kesalahan, pahamilah mengapa hal itu terjadi. Ini adalah jenis pemikiran auto-refleksif kritis, yang menerjemahkan analisis yang dibuat individu ke dalam konteks kualitas pilihan yang disajikan dalam pengambilan keputusan.

Pada gilirannya, proses ini tidak diragukan lagi terkait dengan Metakognisi. Oleh karena itu, salah satu komponen Metakognisi terpenting yang dijelaskan dalam literatur adalah pengetahuan tentang kognisi. Hal ini menyiratkan kesadaran akan kapasitas dan keterbatasan diri, termasuk faktor internal dan eksternal yang dapat memengaruhi atau mengurangi kinerja kognitif. Komponen ini sangat relevan dalam mendefinisikan strategi dalam tindakan, karena inilah alasan kita memilih satu strategi dengan mengorbankan strategi lainnya (seperti yang terjadi pada contoh metode penelitian yang disebutkan di atas).

Yang penting untuk digarisbawahi adalah bahwa setiap orang yang ingin melakukan tindakan ilegal, tentu saja, memiliki analisis strategis yang dijelaskan pada paragraf sebelumnya. Seseorang yang sakit jiwa dapat bertindak salah, tetapi niatnya hanyalah untuk bertindak, bukan untuk bertindak ilegal. Di sisi lain, seseorang yang merencanakan tindakan ilegal, memikirkan tujuan akhir, merenungkan kapasitas dan keterbatasan dirinya sendiri, serta faktor eksternal lain yang mungkin memengaruhi kinerjanya, mendefinisikan strategi, dengan mempertimbangkan semua hal dalam konteks kemungkinan tertangkap. Untuk mendukung gagasan ini, mungkin juga bermanfaat untuk menelaah teori perilaku terencana, dari bidang Psikologi.

Singkatnya, penulis berpendapat bahwa niat Agen dimodulasi, terutama, oleh tiga hal: (1) sikap individu terhadap perilaku yang sedang dilakukan, (2) tekanan individu terkait perilaku spesifik tersebut, dan (3) pengendalian perilaku. Sederhananya, yang kita miliki adalah perilaku tertentu, yang terkait dengan niat yang pada gilirannya dimodulasi oleh ketiga faktor ini. Hampir tidak ada keraguan tentang keberadaan proses metakognitif strategis pada perilaku terencana, termasuk perilaku terencana yang terlarang.

Di sisi lain, Metakognisi dikatakan memiliki tiga tingkat kesadaran dalam setiap alur cerita. Tingkat pertama berkaitan dengan cerita atau perilaku itu sendiri. Tingkat kedua berkaitan dengan pemikiran yang dimiliki Agen terhadap kejadian tersebut. Tingkat ketiga dan terakhir adalah tentang kerja refleksif tentang pemikiran tingkat kedua. Dengan menerjemahkan teori ke dalam contoh praktis, mari kita bayangkan seorang subjek, HAL, sedang berbelanja di toko setempat.

Seseorang mencoba mencuri sesuatu, dan polisi dipanggil ke toko. Pencuri itu ditangkap dan ditahan. HAL mengamati dengan saksama semua yang terjadi. Inilah tingkat kesadaran pertama, kejadian, cerita, atau perilaku (dalam hal ini, seseorang mencuri di toko).

HAL kemudian melanjutkan hidupnya, merenungkan peristiwa tersebut, nilai hukumnya, dan hukuman yang ia saksikan dijatuhkan kepada pencuri tersebut. Dengan demikian, kita memiliki tingkat kesadaran kedua. Akhirnya, sebagai manusia yang sehat, HAL juga mampu memiliki pikiran tingkat kedua tentang refleksi pertama tersebut. Misalnya, ia mungkin awalnya berpikir bahwa hukumannya tidak adil, tetapi kemudian merasa malu dengan pikirannya sendiri. Atau menyadari bahwa ia tidak menganggap mencuri itu salah, lalu merasa takut akan melakukan hal serupa. Yang kita miliki adalah penilaian yang dibuat terhadap penilaian lain, dengan tujuan memantau perilaku. Sebagaimana dijelaskan melalui contoh di atas, kemampuan untuk merasa bersalah juga dapat dianggap memiliki tujuan ini.

Sebagaimana dijelaskan sebelumnya, salah satu alasan mengapa hukum tidak memperhitungkan penyandang disabilitas mental adalah, tepatnya, ketidakmampuan untuk merasa bersalah, yang menyiratkan semacam agensi kompleks. Agensi kompleks ini menyiratkan kapasitas untuk memahami apa itu perilaku terlarang dan memilih untuk tidak melakukan perilaku tersebut, yang pada gilirannya menyiratkan proses metakognitif, dalam hal ini, bukan dalam arti analisis strategis (diperlukan ketika merencanakan perilaku terlarang) melainkan dalam arti memantau perilaku (yang menyangkut proses merefleksikan perilaku dan memutuskan apakah akan melakukan kejahatan atau tidak).

Kesimpulannya, di antara berbagai bentuk proses metakognitif yang mungkin dimiliki seseorang, untuk melakukan dan memahami perilaku terlarang, seseorang setidaknya membutuhkan dua jenis proses metakognitif: strategis dan pemantauan. Inilah inti dari akuntabilitas. Tanpa disadari, Law telah menggunakan konsep Metakognisi ini dari masa ke masa. Hewan tidak bertanggung jawab secara langsung, begitu pula anak-anak, melainkan memiliki seseorang yang bertanggung jawab atas mereka.

Dalam kasus pertama, hewan mampu memahami bahwa kejadian mendapatkan biskuit terjadi karena mereka berguling ketika diminta. Namun, secara umum, mereka tidak dapat memiliki pemikiran yang kompleks dan tingkat kedua tentang cara atau metode terbaik untuk melakukannya, yang membuat kita hanya memiliki tingkat kesadaran kedua dan hampir tidak memiliki proses metakognitif. Dengan demikian, hewan tidak dimintai pertanggungjawaban atas tindakan mereka, juga tidak diberikan status hukum.

Situasi anak-anak jelas berbeda, karena mereka menunjukkan beberapa jenis Metakognisi, dan hal itu menjadi lebih kompleks seiring pertumbuhan mereka. Ada banyak penelitian dalam hal ini, misalnya yang dijelaskan oleh Georghiades, dalam *From the general to the situation: three decade of Metacognition*, yang menunjukkan hal ini secara tepat. Mereka, secara inklusif, sejak sangat awal dalam kehidupan mereka, mampu belajar (dan belajar menyiratkan jenis Metakognisi tertentu). Namun, mereka tidak menyajikan proses Metakognitif strategis, yang, sebagaimana dijelaskan di paragraf sebelumnya, merupakan jenis spesifik yang kita cari ketika membahas akuntabilitas hukum. Kita berbicara tentang pemikiran operasional yang dinyatakan secara formal, yang jarang dikaitkan dengan anak-anak. Studi juga menunjukkan bahwa Metakognisi strategis mulai berkembang sekitar usia 14 tahun, meskipun mungkin belum sepenuhnya berkembang hingga usia lanjut.

Meskipun anak-anak memiliki status sebagai orang hukum, kenyataannya, entah kebetulan atau tidak, hukum hanya menetapkan tanggung jawab pidana kepada individu di bawah umur ketika mereka berusia 16 tahun, dengan keyakinan bahwa pada usia tersebut mereka telah cukup berkembang untuk memahami konsekuensi dari tindakan mereka.

Jenis Metakognisi strategis ini juga tidak ada dalam kasus beberapa penyakit mental, meskipun konsekuensi dalam kesadaran dapat berubah dari satu penyakit ke penyakit lain dan dari satu orang ke orang lain. Dalam pemrograman dan komputasi, Metakognisi berkaitan dengan apa yang diketahui sistem tentang kognisinya sendiri dan juga tentang kognisi secara umum. Sebagaimana dijelaskan Crowder dkk., dalam AI konsep ini terjalin erat dengan introspeksi, dalam artian memungkinkan mesin untuk membentuk keyakinan tentang keadaan internalnya sendiri, alih-alih sekadar menganalisis lingkungan tempat ia bergerak.

Secara tradisional, dalam komputasi, proses metakognitif digunakan untuk pemecahan masalah tertentu, seperti pemilihan algoritma dari sudut pandang efisiensi. Dalam hal ini, Crowder & Friess berpendapat bahwa setidaknya terdapat tiga jenis Metakognisi dalam ranah pengetahuan ini:

(a) Pengetahuan metakognitif, yang berkaitan dengan apa yang diketahui sistem tentang dirinya sendiri, sebagai pemroses kognitif;
(b) Regulasi metakognitif, terkait dengan kendali kognisi dan pembelajaran, yang dapat mencakup pengetahuan yang dimiliki sistem tentang apa yang diketahui dan tidak diketahuinya;
(c) Pengalaman metakognitif, yang menyangkut pengalaman masa lalu yang entah bagaimana berkaitan dengan misi sistem saat ini, yang memungkinkan sistem untuk menciptakan ekspektasi atau prediksi tentang apa yang mungkin terjadi, mengingat pengalaman-pengalaman yang terjadi sebelum momen analisis tersebut.

Dalam hal ini, tampaknya adil untuk mengakui bahwa AI dapat memiliki beberapa tingkat proses metakognitif. Namun, hal ini tidak sesuai dengan jenis Metakognisi yang diperlukan untuk menganggap suatu entitas bertanggung jawab. Faktanya, tidak satu pun dari proses ini yang diterjemahkan menjadi proses metakognitif strategis atau pemantauan. Oleh karena itu, AI seharusnya tidak, setidaknya untuk saat ini, dimintai pertanggungjawaban atas perilakunya, sama seperti anak-anak, hewan, dan orang-orang dengan gangguan mental.

9.6 AKUNTABILITAS DAN KEPRIBADIAN HUKUM

Hingga saat ini, kami telah menghubungkan Agensi dengan tindakan sukarela, yang terakhir sebagai persyaratan minimum untuk yang pertama, dan akuntabilitas dengan metakognisi. Masih ada satu putaran tersisa, mengenai hubungan antara akuntabilitas dan kepribadian hukum.

Kami sepenuhnya menyadari bahwa tanggung jawab hukum dan kepribadian hukum bukanlah konsep yang sama, sebuah kesalahan yang sering dibuat terkait AI, baik oleh para akademisi maupun entitas resmi, sebagaimana ditunjukkan oleh Pagallo dalam penelitiannya. Faktanya, keduanya merupakan konsep yang berbeda dan mungkin juga memiliki konsekuensi hukum yang berbeda. Namun, keduanya harus saling terkait secara intrinsik.

Dalam makalah ini, kami menjelaskan bagaimana hewan, anak-anak, dan orang dengan gangguan jiwa tidak dianggap bertanggung jawab dari sudut pandang hukum. Dalam hal ini, juga ditekankan bahwa, meskipun anak-anak dan orang dengan disabilitas mental memiliki kepribadian hukum di sebagian besar yurisdiksi, mereka melakukannya dalam lingkup yang terbatas dan pelaksanaan hak-hak mereka yang terbatas. Kami juga menunjukkan bagaimana hewan sama sekali tidak memiliki status badan hukum di sebagian besar yurisdiksi, meskipun beberapa perluasan instrumen ini diberikan dalam kasus-kasus tertentu.

Faktanya, status badan hukum dalam arti penuhnya tampak ada sejauh entitas tersebut mampu menjalankan hak-haknya. Sebagaimana telah kita lihat, terdapat entitas (misalnya anak-anak dan penyandang disabilitas mental) yang meskipun diberikan status badan hukum, tidak menunjukkan kapasitas hukum atau kapasitas hukumnya dibatasi, sehingga tidak dianggap bertanggung jawab secara hukum. Dengan kata lain, cakupan status badan hukum mereka terbatas. Di sisi lain, setiap entitas yang dianggap memiliki semacam akuntabilitas, misalnya manusia pada umumnya, memiliki kepribadian dan kapasitas. Status badan hukum mereka berada pada taraf maksimalnya.

Ini berarti, pada prinsipnya, meskipun status badan hukum dapat diberikan dengan cara apa pun, jika kita tidak memiliki akuntabilitas, kita hampir tidak dapat memiliki kapasitas hukum. Dengan kata lain, akuntabilitas mengisi kapasitas entitas, sehingga menentukan isi dan ukuran status badan hukum yang sebenarnya. Hal ini konsisten dengan gagasan yang diuraikan oleh Visa A. J. Kurki tentang apa yang dimaksud dengan badan hukum aktif, yang berlawanan dengan badan hukum pasif, yaitu sebuah konsep yang “mensyaratkan seseorang dapat melakukan perbuatan hukum (memiliki kompetensi hukum) dan bertanggung jawab secara hukum (badan hukum yang memberatkan)”. Dalam penelitiannya yang berjudul “Teori Badan Hukum”, penulis menyatakan bahwa unsur-unsur kunci badan hukum aktif berpusat pada tanggung jawab hukum dan kompetensi hukum.

Faktanya, seseorang tidak dapat tertarik pada gagasan “badan hukum yang dangkal”. Ambil contoh robot Sophia, robot humanoid yang dibuat oleh Hanson Robotics, yang “bercanda” menyatakan bahwa AI akan menghancurkan manusia dalam waktu dekat. Sophia diberikan kewarganegaraan oleh Arab Saudi pada tahun 2017. Terlepas dari semua kehebohan dan perhatian yang diterima keadaan ini, dari sudut pandang hukum, kewarganegaraan ini hampa, dalam arti tidak ada gunanya memberikan status tersebut. Pada kenyataannya, kata "bercanda" harus digunakan dengan hati-hati karena robot Sophia tidak tahu apa arti lelucon, dalam praktiknya, apalagi arti kewajiban hukum. Sophia mungkin telah diberikan kewarganegaraan tetapi tidak memiliki sarana untuk menjalankan hak-haknya sebagai warga negara.

Logika yang sama seharusnya berlaku untuk status badan hukum dalam kasus AI. Jika tidak ada konsekuensi hukum yang dapat ditarik darinya, serupa dengan kewarganegaraan robot Sophia, tidak ada manfaat nyata dalam pemberiannya. Lebih lanjut, kita hanya boleh melakukannya ketika kita menyadari manfaat dari langkah ini, seperti yang terjadi dalam kasus korporasi. Badan hukum memperoleh status badan hukum fiktifnya ketika manusia mulai memahami pentingnya memberikan kewajiban hukum kepada perusahaan.

Dengan kata lain, ketika manusia mulai menyadari manfaatnya. Namun, kita dapat berargumen bahwa baik anak-anak maupun penyandang disabilitas mental tidak memiliki salah satu atau keduanya unsur kompetensi dan akuntabilitas dari status badan hukum, dan status tersebut tetap diberikan kepada mereka (meskipun, seperti yang dikatakan Kurki, status badan hukum tersebut merupakan semacam status badan hukum pasif), yang berarti tidak ada alasan untuk menghindari hal yang sama dalam kasus AI.

Namun, ada alasan spesifik mengapa hal tersebut terjadi. Sebagaimana dinyatakan Savigny dan banyak penulis lainnya, konsep awal badan hukum biasanya sesuai dengan konsep manusia, berdasarkan asumsi bahwa manusia memiliki kapasitas hukum. Dalam hal ini, baik bagi anak-anak maupun penyandang disabilitas mental, status badan hukum dikaitkan dengan fakta bahwa keduanya merupakan kategori manusia sejak lahir, sebuah kriteria yang tentunya tidak dapat diterapkan pada AI. Keadaan ini menunjukkan bahwa kita harus mencari kriteria yang berbeda dalam kasus ini. Dalam makalah ini, dikemukakan bahwa kriteria tersebut seharusnya adalah kemungkinan untuk memainkan peran aktif dalam status badan hukum, melalui kompetensi, tetapi khususnya, tanggung jawab hukum. Selain itu, bagi anak-anak, status badan hukum biasanya dikaitkan menurut pemahaman Hegelian bahwa terdapat potensi rasionalitas dan kebebasan, dan bahwa anak-anak mulai mengumpulkan kemampuan yang dibutuhkan seorang pengemban tugas pada suatu titik.

Simpulannya, tanpa metakognisi, hampir tidak mungkin ada tanggung jawab hukum. Di sisi lain, tanpa akuntabilitas, tidak ada alasan mengapa AI harus memiliki status badan hukum, karena tanpa elemen ini, tidak ada konsekuensi hukum yang bermanfaat yang dapat ditarik darinya. Konsekuensi hukum semacam itu mungkin baru ada ketika kita menemukan AI yang akuntabel. Jika tidak, status badan hukum dalam AI tidak akan berarti apa-apa selain cangkang kosong.

9.7 METAKOGNISI AI: KUNCI AKUNTABILITAS HUKUM

Makalah ini berusaha untuk membedakan antara AI yang akuntabel dan yang tidak akuntabel, dengan menggunakan beberapa bidang ilmu, seperti Filsafat, Psikologi, Komputasi, dan Hukum. Dalam hal ini, makalah ini berargumen bahwa masalah apakah AI dapat dianggap sebagai badan hukum atau tidak dapat diselesaikan melalui gagasan metakognisi, sebuah konsep yang, tanpa disadari, telah digunakan oleh Hukum sejak lama untuk memutuskan hal ini.

Untuk mencapai tujuan ini, kami memulai dengan mengkaji makna Agen, untuk menilai apakah AI dapat dianggap sebagai Agen atau tidak. Karena konsep ini menghadirkan banyak kesulitan, diperlukan suatu persamaan, yaitu tindakan sukarela. Jika terdapat tindakan sukarela, maka kita harus menyimpulkan bahwa kita memiliki Agen di hadapan kita. Oleh karena itu, selama AI bertindak secara sukarela, masuk akal untuk berargumen bahwa robot yang kompleks adalah Agen, bukan sekadar alat. Namun, sebagaimana telah dinyatakan sebelumnya, hal ini tidak serta merta berarti bahwa AI harus dimintai pertanggungjawaban hanya karena ia termasuk dalam kategori Agen. Hewan, orang dengan penyakit mental, dan

anak-anak secara intuitif dianggap sebagai Agen, namun tidak dapat dimintai pertanggungjawaban.

Oleh karena itu, argumen lain yang diajukan adalah bahwa untuk menetapkan tanggung jawab kepada suatu Agen, entitas tersebut harus menunjukkan, setidaknya, proses metakognitif yang strategis dan termonitor. Elemen-elemen ini berperan dalam kemampuan untuk bertanggung jawab, yang pada gilirannya membentuk, bersama dengan konsep kompetensi hukum, gagasan tentang Persona Hukum yang aktif.

Mempertimbangkan kesimpulan di atas, dua gagasan lain harus mengikuti. Jika entitas tersebut menunjukkan proses metakognitif, maka kita dapat mempertimbangkan untuk memberikan entitas tersebut persona hukum. Di sisi lain, jika tidak menunjukkan kapasitas ini, maka kita memerlukan hukum yang otonom dan, jika perlu, baru, yang berlaku, seperti yang telah kita miliki dalam kasus anak-anak, hewan, dan penyakit mental. Terkait kondisi AI, saat ini, tampaknya AI belum berada pada tingkat yang cukup kompleks dalam hal proses metakognitif untuk dimintai pertanggungjawaban atas tindakannya, meskipun menunjukkan proses metakognitif yang sederhana.

BAB 10
AI DALAM KEPUTUSAN KESEHATAN: RISIKO DAN PELUANG

Abstrak Penggunaan sistem yang mencakup Kecerdasan Buatan (AI) menuntut penilaian risiko dan peluang yang terkait dengan penerapannya di bidang kesehatan. Berbagai jenis AI menghadirkan berbagai tantangan etika, hukum, dan sosial. Sistem AI yang terlibat terintegrasi dengan teknologi pencitraan dan pemrosesan sinyal baru. Sistem AI di bidang komunikasi telah memungkinkan interaksi yang sebelumnya tidak ada dan memfasilitasi akses ke data dan informasi. Perhatian terbesar meliputi bidang perencanaan, pengetahuan, dan penalaran, karena sistem AI secara langsung terkait dengan proses pengambilan keputusan.

Oleh karena itu, tujuan utama bab ini adalah untuk merefleksikan dan menyarankan rekomendasi, dengan landasan Model Bioetika Kompleks, tentang proses pengambilan keputusan dalam kesehatan dengan dukungan AI, dengan mempertimbangkan risiko dan peluang. Bab ini dibagi menjadi dua bagian: (1) Proses pengambilan keputusan dalam kesehatan dan AI; (1.1) Bidang kesehatan, penggunaan AI dan proses pengambilan keputusan: peluang dan risiko dalam penanganan rekam medis elektronik (RME), dan (2) Model Bioetika Kompleks (CBM) dan AI.

10.1 KOMPLEKSITAS AI KESEHATAN: BIOETIKA DAN RISIKO

Kompleksitas, dalam pengertian yang dikemukakan oleh Edgar Morin, menerjemahkan momen saat kita menjalani apa yang disebut Revolusi Keempat. Keengganan terhadap Manikeisme dan pemahaman bahwa kompleksitas bukanlah segalanya, kompleksitas bukanlah totalitas realitas, tetapi kompleksitas adalah hal terbaik yang dapat, pada saat yang sama, membuka diri terhadap hal-hal yang dapat dipahami dan mengungkap hal-hal yang tidak dapat dijelaskan. Ketidakpastian kehidupan sehari-hari merupakan elemen penerimaan, atau bahkan ambiguitasnya.

Kecerdasan buatan (AI), potensi penggunaannya, dan, dalam proporsi yang sama, tantangan hukum, etika, dan sosial harus tercermin dalam kompleks lingkungan dan lingkungan. Teknologi dan kedokteran memiliki sejarah panjang keterkaitan, tetapi salah satu tonggak sejarahnya adalah karya Lee Lusted, yang dalam artikelnya Medical Electronics, melaporkan serangkaian perangkat elektronik medis dalam jumlah besar yang dikembangkan pada masa itu, yang menunjukkan perkembangan pesat bidang ini. Ia berkata saat itu: Fenomena listrik dalam tubuh manusia telah lama menarik perhatian, tetapi amplitudo sinyal yang rendah menyulitkan studi. Pada periode yang sama, Turing membangun pilar-pilar ilmu komputer dan Kecerdasan Buatan (AI).

Teknologi berbasis AI memaksakan evaluasi risiko dan peluang yang terkait dengan penerapannya dalam kehidupan manusia. Tema ini menggabungkan pertanyaan lama dan baru, seperti yang ditunjukkan Ulrick Beck dalam The Risk Society dan dikembangkan lebih lanjut dalam World at Risk ke dalam perdebatan tentang dampak teknologi dalam kehidupan manusia: bagaimana kita ingin hidup? Apa sisi manusiawi dalam diri manusia, sisi alamiah

dalam diri manusia, yang perlu dilindungi? (...) Pertanyaan-pertanyaan lama-baru ini dapat dilempar bolak-balik antara kehidupan sehari-hari, politik, dan sains. Pada tahap paling maju dari proses peradaban, pertanyaan-pertanyaan ini sekali lagi menikmati prioritas dalam agenda juga atau tepatnya pada saat-saat ketika pertanyaan-pertanyaan ini terselubung dalam kamuflase rumus matematika dan kontroversi metodologis.

Jadi, untuk beberapa waktu, interaksi manusia-mesin telah dimungkinkan melalui sistem bahasa alami (*Chat-bot*). Dalam banyak kesempatan, tidak ada persepsi yang jelas bahwa komunikasi ini dilakukan dengan mesin dan bukan dengan orang lain. Pada tahun 1970an, Jacques Monod telah memperingatkan bahwa semakin sulit untuk menetapkan batas antara yang alami dan yang buatan. Simulasi atau substitusi aktivitas nyata, yang semakin mirip dengan yang dilakukan oleh mekanisme dan sistem buatan, menghasilkan ambiguitas persepsi ini.

Arogansi teknologi, menurut Hans Jonas, menyebabkan hasil-hasil ini dipahami sebagai sesuatu yang tidak perlu dipertanyakan lagi. Kesempurnaan komputer telah dibahas sejak awal penggunaannya, ketika masih disebut "otak elektronik". Saat itu, sudah ada anggapan bahwa kualitas informasi yang dihasilkan tidak diragukan lagi, melainkan bergantung pada kualitas data masukan dan proses yang digunakan. Hal ini kemudian dikenal dengan akronim GIGO (*Garbage In, Garbage Out*). Artinya, jika data atau sistem tidak memadai, hasil yang dihasilkan akan terganggu.

Pertanyaan lama-baru ini telah menjadi inti diskusi seputar AI dan pengambilan keputusan. Dalam perspektif ini, Floridi dkk. dalam teks yang diterbitkan pada tahun 2018, menyatakan bahwa AI adalah realitas tanpa hasil dan oleh karena itu, perlu dibentuk refleksi menuju Masyarakat AI untuk Kebaikan (*Good AI Society*).

Peluang dan risiko untuk melindungi martabat manusia dan menyediakan bagi perkembangan mereka harus diresapi oleh prinsip-prinsip tradisional Bioetika Amerika Utara kebaikan, non-maleficence, otonomi, dan keadilan, di samping prinsip eksplikabilitas. Dalam analisis ini, menurut pandangan kami, dua perspektif perlu ditambahkan: (1) perspektif Eropa, yang diusulkan oleh Peter Kemp dan Jacob Dahl Rendtorff, menggunakan empat prinsip lain: Martabat; Otonomi, dipahami sebagai Kebebasan; Integritas; dan Kerentanan; karena dalam perspektif ini, prinsip-prinsip tidak diberi bobot, tetapi harus ada koherensi dalam penerapannya dan (2) pendekatan yang didasarkan pada Model Bioetika Kompleks (CBM), dengan kata lain, bioetika dipahami sebagai refleksi yang kompleks, bersama, dan interdisipliner tentang kecukupan tindakan yang melibatkan kehidupan.

Hidup dan kehidupan saling melengkapi, keduanya memberikan dimensi yang memadai bagi setiap orang. Hidup digambarkan oleh aspek organik, yaitu karakteristik biologis. Di sisi lain, hidup mengacu pada aspek relasional, biografi setiap orang. Kombinasi karakteristik inilah yang memberikan keunikan bagi setiap orang. Model Bioetika Kompleks (CBM) mewujudkan perspektif bidang interdisipliner yang kompleks tentang refleksi hidup dan kehidupan.

Keinginan untuk mengetahui dan mempelajari kesehatan populasi dan kesehatan manusia inilah yang menandai studi ilmiah dan membangun fondasi bagi penelitian dan

eksperimen. Kebutuhan untuk menanggapi tantangan yang ditimbulkan oleh epidemi, kelaparan, perang, pertumbuhan penduduk, dan pusat-pusat perkotaan merupakan motivasi bagi rantai penemuan ilmu pengetahuan.

Tujuan utamanya sepanjang masa adalah untuk mengidentifikasi determinan penyakit dan, baru-baru ini, kesehatan pada tingkat populasi. Jadi, secara historis, kontribusi spesifik epidemiologi adalah pembentukan progresif serangkaian metode dan konsep yang koheren, dengan tujuan menilai determinan kesehatan, di mana sistem yang tangguh, seperti teknologi yang digerakkan oleh AI, berperan penting dalam memproses dan mengelola data serta informasi pribadi dan sensitif di bidang kesehatan.

Pemrosesan sejumlah besar data secara efisien dan presisi merupakan hal mendasar bagi perkembangan kedokteran ilmiah. Oleh karena itu, perangkat komputasi untuk pembelajaran mesin dan penambangan data dalam jumlah besar, dalam pendekatan yang dikenal sebagai Big Data, serta evaluasi gabungan terhadap data dalam jumlah besar, telah memungkinkan pembentukan hubungan baru, pemahaman baru yang sebelumnya tidak teridentifikasi.

Perkembangan penting lainnya di bidang ini adalah meningkatnya penggunaan algoritma untuk pengambilan keputusan. Perangkat-perangkat ini, yang semakin ditingkatkan dan didasarkan pada proses yang sangat kompleks, telah memberikan solusi optimal untuk berbagai masalah, termasuk memodifikasi proses pengambilan keputusan itu sendiri. Sebagian besar sistem ini bekerja dalam upaya mengenali pola kesamaan. Bidang inilah yang kemudian dikenal sebagai Kecerdasan Buatan.

Sebenarnya, kecerdasan buatan bukanlah kecerdasan itu sendiri, melainkan proses pengambilan keputusan otomatis. Pierre Levi mengkritik keras penggunaan istilah "kecerdasan buatan", ia tidak mengakui kemungkinan menghasilkan pengetahuan baru atau memahami dunia dalam sistem ini. Algoritma dibuat oleh orang-orang yang bekerja untuk lembaga, yang memiliki sistem kepercayaan dan nilai-nilai mereka sendiri, yang pada akhirnya mengarahkan pemrosesan dan interpretasi.

Oleh karena itu, penggunaan sistem yang mencakup Kecerdasan Buatan (AI) memaksakan penilaian risiko dan peluang yang terkait dengan penggabungannya di area kesehatan: perawatan kesehatan; penelitian eksperimental dan klinis dan pengobatan yang dipersonalisasi. Sistem AI yang melibatkan area komunikasi yang mampu melakukan komputasi paralel untuk pemrosesan data dan representasi pengetahuan (jaringan saraf tiruan yang disebut (ANN)); yang memungkinkan untuk melakukan interaksi yang sebelumnya tidak ada dan memfasilitasi akses ke data dan informasi; teknologi untuk mendeteksi gambar, suara; melakukan bantuan kesehatan dengan robotika dan area perencanaan, pengetahuan dan penalaran, ketika sistem AI secara langsung dikaitkan dengan proses pengambilan keputusan.

Ramesh dkk, menyajikan tinjauan literatur pada tahun 2004 tentang penggunaan 'kecerdasan buatan' dan 'jaringan saraf (komputer)' dan ikhtisar berbagai teknik kecerdasan buatan bersama dengan tinjauan aplikasi klinis yang penting. Hasil mereka menunjukkan "kemahiran teknik kecerdasan buatan telah dieksplorasi di hampir setiap bidang kedokteran". Para penulis menunjukkan area aktivitas: diagnosis klinis; prognosis; gambar ultrasonografi;

memprediksi kelangsungan hidup pada pasien; digunakan untuk pemberian anestesi di ruang operasi; bentuk komputasi evolusioner yang digunakan untuk aplikasi medis dalam genetika dan evolusi alami, yang disebut 'Algoritma Genetika'.

Dalam konteks ini, berbagai jenis AI menghadirkan berbagai tantangan etika, hukum, dan sosial di dunia, seperti yang ditunjukkan oleh OCDE. Namun, keragaman dan kerentanan akses sosial, ekonomi dan ke Cakupan Kesehatan Universal (UHC) yang ada di Amerika Selatan membuat analisis teknologi kesehatan yang digerakkan oleh AI menjadi lebih kompleks. Standar yang harus dicapai dalam hal akses kesehatan adalah Tujuan Pembangunan Berkelanjutan (SDGs) PBB, yang, pada tahun 2030, negara anggota harus menjamin: (1) akses ke layanan kesehatan untuk semua orang yang membutuhkan kesehatan, terlepas dari karakteristik sosial-ekonomi, lokasi, kekayaan, atau kerentanan lainnya; (2) perlindungan finansial, yaitu semua orang harus aman dari risiko finansial saat mengeluarkan biaya perawatan kesehatan; (3) akses terhadap kualitas layanan kesehatan, artinya layanan kesehatan harus efektif dalam menyediakan layanan dan meningkatkan hasil, sekaligus hemat biaya dan berkelanjutan, karena akses tanpa kualitas dapat dianggap sebagai janji kosong dari cakupan kesehatan universal.

Laporan Kesehatan Sekilas: Amerika Latin dan Karibia 2020 membandingkan indikator-indikator utama kesehatan masyarakat dan sistem kesehatan di 33 negara Amerika Latin dan Karibia (LAC). Laporan ini menyajikan data yang sebanding tentang status kesehatan dan determinannya, sumber daya dan kegiatan layanan kesehatan, pengeluaran dan pembiayaan kesehatan, serta kualitas layanan kesehatan, beserta indikator ketimpangan kesehatan terpilih, termasuk pandemi COVID-19:

Hambatan utama untuk mengakses layanan kesehatan tersebut muncul dari pengeluaran kesehatan mandiri, yang di LAC rata-rata mewakili 34% dari total belanja kesehatan, jauh di atas rata-rata 21% di negara-negara OECD. Tingginya pengeluaran langsung di Amerika Latin dan Karibia (LAC) merupakan indikasi sistem kesehatan yang lebih lemah, tingkat cakupan layanan kesehatan yang lebih rendah, dan secara keseluruhan, skenario dasar yang lebih buruk untuk menghadapi pandemi ini dibandingkan dengan sebagian besar negara OECD.

Khususnya di Brasil, beberapa wilayah negara tersebut memiliki akses yang lebih besar terhadap layanan kesehatan dan teknologi kesehatan dibandingkan wilayah lainnya, meskipun Brasil memiliki sistem kesehatan publik terbesar di dunia, Sistem Kesehatan Terpadu (*Sistema Único de Saúde*—SUS), yang pada prinsipnya memiliki akses universal cakupan kesehatan universal (UHC) bagi warga negara dan warga negara asing, yang melayani lebih dari 190 juta orang, 80% di antaranya bergantung sepenuhnya padanya untuk setiap layanan kesehatan. SUS merupakan pencapaian rakyat Brasil, yang dijamin oleh Konstitusi Federal tahun 1988, dalam Pasal 196, melalui Undang-Undang No. 8.080/1990, yang harus dijamin dan ditingkatkan secara terus-menerus.

Oleh karena itu, tujuan utama bab ini adalah untuk merefleksikan dan memberikan rekomendasi, dengan landasan Model Bioetika Kompleks, tentang proses pengambilan keputusan di bidang kesehatan dengan dukungan AI, dengan mempertimbangkan risiko dan

peluang. Bab ini dibagi menjadi dua bagian: (1) Proses pengambilan keputusan di bidang kesehatan dan AI; (1) Penggunaan AI dan proses pengambilan keputusan di bidang kesehatan: peluang dan risiko dalam menangani rekam medis elektronik (RME), dan (2) Model Bioetika Kompleks (CBM) dan AI.

Asumsi utamanya adalah menjaga keseimbangan dan melestarikan karakteristik kemanusiaan yang hadir dalam tindakan pengambilan keputusan, dengan mempertimbangkan aspek etika, hukum, dan sosial, serta prinsip kepercayaan, saat menggunakan sistem AI di bidang kesehatan. Justifikasi penggunaan data dan informasi kesehatan pribadi dan sensitif harus dikaitkan dengan tindakan atas nama individu dan masyarakat, dalam hal bantuan, penelitian yang melibatkan manusia, baik yang bersifat sanitasi, epidemiologis, klinis, maupun biobank.

Kami berharap rekomendasi kami dapat berkontribusi pada pengembangan kerangka regulasi etika dan hukum praktik yang baik dan kepatuhan untuk penggunaan AI di bidang kesehatan. Perspektif kami adalah menganalisis contoh-contoh di tingkat nasional dan internasional, dengan fokus pada keragaman dan kerentanan sosial, ekonomi, dan akses kesehatan yang ada di Amerika Selatan, khususnya di Brasil.

10.2 PROSES PENGAMBILAN KEPUTUSAN DALAM KESEHATAN DAN AI

Pengambilan keputusan yang melibatkan kecerdasan buatan (AI) harus mempertimbangkan AI Generatif. Karakteristik sistem AI Generatif adalah membangun koneksi, melalui perangkat komputasi baru, berdasarkan data, konsep, dan struktur dalam jumlah besar yang telah dibangun, untuk mengusulkan konten baru, yang diekspresikan dalam bahasa sealami mungkin.

Istilah generatif diciptakan oleh Noam Chomsky pada tahun 1965, ketika ia mengusulkan "tata bahasa generatif", atau disebut juga "tata bahasa generatif". Usulan ini berusaha menjelaskan bahwa bahasa dibentuk melalui keterampilan dan performa untuk menghasilkan kalimat dan jalur apa yang digunakan untuk membentuknya. Gagasan ini didasarkan pada pemikiran Wilhelm von Humboldt, yang pada tahun 1836 mengatakan bahwa bahasa menggunakan sarana yang terbatas secara tak terbatas.

Sistem AI Generatif yang baru, seperti ChatGPT dan Perplexity, menciptakan tingkat kompleksitas baru dalam menangani data, dalam upaya untuk mengekstrak informasi yang dapat memandu tindakan. Sistem-sistem baru ini tidak hanya memberi label atau mengklasifikasikan data yang sudah ada sebelumnya, tetapi juga menghasilkan konten baru, melalui agregasi dan kombinasi, dari elemen-elemen yang tersedia.

Apa implikasi etis yang terkait dengan Kecerdasan Buatan Generatif? Teks yang disusun oleh ChatGPT dan Perplexity menjawab:

> **ChatGPT**: Kecerdasan Buatan (AI) menghadirkan tiga area utama yang menjadi perhatian etis bagi masyarakat: privasi dan pengawasan, prasangka dan diskriminasi, serta tantangan filosofis. AI dapat menyebabkan pengangguran karena mesin mengambil alih pekerjaan yang sebelumnya dilakukan oleh manusia. Ketimpangan juga menjadi perhatian, karena AI dapat digunakan untuk melanggengkan struktur kekuasaan yang

ada. AI juga dapat menyebabkan bias algoritmik, di mana model menggabungkan dan menerapkan bias manusia dan sosial dalam skala besar.

Privasi adalah hak asasi manusia fundamental yang diakui dalam Deklarasi Hak Asasi Manusia PBB, tetapi aplikasi AI dapat menjadi ancaman nyata bagi hak ini. Misinformasi merupakan masalah etika lain dengan AI, karena model pembelajaran mesin dapat dengan mudah menghasilkan teks yang salah.

Untuk memastikan penggunaan AI yang etis, edukasi dan kesadaran tentang etika AI harus ditingkatkan. Contoh AI yang etis mencakup penggunaan AI untuk meningkatkan hasil kesehatan atau mengurangi konsumsi energi. Penting untuk diingat bahwa isu etika dengan AI tidak terbatas pada dampak negatif; terdapat juga manfaat insidental, seperti peningkatan kemampuan analitis atau peningkatan efisiensi dalam proses pengambilan keputusan. (*https://chat.openai.com/chat/c4bf8bd6-df96-4cec-b96d-dd54ec97ef2f; 14 Februari 2023*)

Kebingungan: Bias dan diskriminasi: Sistem AI dapat melanggengkan dan memperkuat bias yang ada di masyarakat, yang mengarah pada keputusan dan hasil yang diskriminatif. Misalnya, teknologi pengenalan wajah terbukti memiliki tingkat kesalahan yang lebih tinggi pada orang kulit hitam dan perempuan.

Privasi dan perlindungan data: Sistem AI sering kali mengandalkan data pribadi dalam jumlah besar untuk membuat keputusan, dan terdapat kekhawatiran tentang bagaimana data ini dikumpulkan, disimpan, dan digunakan, serta siapa yang memiliki akses ke data tersebut.

Dampak ketenagakerjaan: AI berpotensi mengotomatiskan banyak pekerjaan, yang menyebabkan pengangguran dan keresahan sosial. Terdapat pula risiko bahwa AI akan memperkuat ketidakseimbangan kekuatan yang ada dan memperparah ketimpangan ekonomi.

Akuntabilitas dan tanggung jawab: ketika sistem AI membuat keputusan yang memiliki konsekuensi signifikan bagi individu dan masyarakat, sulit untuk menentukan siapa yang bertanggung jawab atas keputusan tersebut dan bagaimana mereka dapat dimintai pertanggungjawaban.

Hak Asasi Manusia: AI berpotensi melanggar hak asasi manusia, seperti kebebasan berekspresi dan privasi, dan terdapat kekhawatiran tentang penggunaan AI untuk pengawasan dan pengendalian oleh pemerintah dan organisasi lain. Penting untuk mengatasi implikasi etis ini melalui dialog dan debat yang berkelanjutan serta dengan menetapkan pedoman dan prinsip etika yang jelas untuk pengembangan dan penggunaan AI. Ini termasuk mempertimbangkan potensi dampak AI terhadap individu dan masyarakat, serta memastikan bahwa AI dikembangkan dan digunakan secara bertanggung jawab dan etis. (*https://www.perplexity.ai/?s=u&uuid=a6d1842e-bf56-4e0a-9fd4-2ff4f00a8e13; 14 Februari 2023*)

Seperti yang dapat kita lihat, dari jawaban ChatGPT dan Perplexity, terdapat hubungan antara isu dan konsep etika, seperti privasi, pengawasan dan pengendalian, kebebasan berbicara, prasangka dan diskriminasi, yang ditafsirkan dan konten baru disusun. Semua isu ini sudah ada sebelum penggunaan AI generatif, namun telah dibahas dalam berbagai tingkat kedalaman dan penerapan. Mungkin, tantangan saat ini adalah memikirkan isu-isu ini dalam perspektif baru ini.

Setiap kali sebuah inovasi diadopsi oleh masyarakat, diskusi tentang kesesuaiannya pun muncul. Ketika buku cetak tersedia bagi masyarakat, ketika ensiklopedia pertama muncul, ketika internet memberikan akses ke sejumlah data yang belum pernah dibayangkan sebelumnya, muncul pertanyaan tentang kesesuaian penggunaan sarana penyebaran data, pengetahuan, dan informasi ini.

Contoh yang baik adalah diskusi terkini tentang dampak AI Generatif terhadap pendidikan, yang menimbulkan kecemasan di kalangan sekolah, orang tua, dan guru. Diskusi yang sama ini telah terjadi di momen-momen bersejarah lainnya dan penggabungan teknologi baru. Sudah ada model pendidikan yang memungkinkan penggabungan situasi-situasi yang dihadirkan oleh AI Generatif ini dengan cara yang kreatif. Alih-alih menyalin atau menghasilkan konten, mungkin tantangan pendidikan adalah mengevaluasi kualitas informasi yang dihasilkan. Tantangan ini akan digunakan untuk mengintegrasikan refleksi kritis dan kompleks di berbagai tingkat kehidupan guna membangun keamanan, transparansi, dan kepercayaan dalam penggunaan AI.

Studi tentang Interoperabilitas e-Health Data Kesehatan dan Kecerdasan Buatan untuk Kesehatan dan Perawatan di Uni Eropa Laporan Studi Akhir menunjukkan bahwa kurangnya kepercayaan terhadap dukungan keputusan berbasis AI menghambat adopsi yang lebih luas dalam bidang kesehatan, dan juga mengintegrasikan teknologi baru ke dalam praktik klinis saat ini; penelitian dan pengobatan pribadi memang merupakan tantangan hukum, etika, dan sosial. Tantangan-tantangan ini diperparah oleh kebutuhan internasionalisasi bidang kesehatan dan tantangan berbagi data dan informasi untuk mencapai kesehatan global.

Rekomendasi juga telah dikembangkan oleh negara dan organisasi, menyoroti rekomendasi yang diusulkan oleh Komisi Eropa, pada tahun 2020, dalam “Buku Putih—Tentang Kecerdasan Buatan Pendekatan Eropa terhadap Keunggulan dan Kepercayaan”, dengan tujuan menetapkan jalur politik untuk mencapai penggunaan AI yang tepat.

Dalam dokumen ini, Komisi merekomendasikan pembentukan standar dan pedoman investasi di bidang AI, yang bertujuan pada dua tujuan utama: mendorong adopsi AI dan mengatasi risiko yang terkait dengan penggunaan tertentu dari teknologi baru ini. Komisi juga membentuk Kelompok Pakar Tingkat Tinggi yang menerbitkan Pedoman tentang AI tepercaya pada April 2019, yang terdiri dari tujuh persyaratan utama: penghormatan terhadap martabat manusia; sistem teknis dan keamanan yang tangguh; privasi dan manajemen data; transparansi; penghormatan terhadap keberagaman, non-diskriminasi, dan kesetaraan; kesejahteraan sosial dan lingkungan; serta akuntabilitas.

Pasar Digital Bersama merupakan salah satu dari sepuluh prioritas Uni Eropa. Dalam konteks ini, keputusan-keputusan berikut telah diambil: Keputusan No. 922/2009/EC Parlemen dan Dewan Eropa tanggal September 2009 tentang solusi interoperabilitas untuk administrasi publik Eropa (Kerangka Kerja Interoperabilitas Eropa e-Health) dan Keputusan (EU) 2015/2240 Parlemen dan Dewan Eropa tanggal 25 November 2015 yang menetapkan program solusi interoperabilitas dan kerangka kerja bersama untuk administrasi publik, bisnis, dan warga negara (program ISA) sebagai sarana modernisasi sektor ini. Kerangka Kerja Interoperabilitas e-Health Eropa (ReEIF).

Uni Eropa berupaya mengintegrasikan rekam medis elektronik warga negara Eropa, dengan menyadari kelemahan terkait berbagai aspek penggunaan data, baik untuk keamanan, perlindungan privasi, kesesuaian etika, pengelolaan, penyimpanan dan pemusnahan, maupun interoperabilitas antar sistem informasi negara untuk membangun struktur e-Health yang terpercaya. Langkah-langkah ini merupakan bagian dari tujuan menciptakan pasar tunggal digital.

Proses pengambilan keputusan, khususnya di bidang kesehatan, didasarkan pada kepercayaan dan hubungan kepercayaan yang tentu saja diidentikkan dengan semua pihak yang terlibat dalam hubungan ini. Hubungan tersebut terjadi di semua bidang, antara administrasi publik dan yang dikelola; antara entitas swasta; antara entitas swasta dan manusia, dan antara manusia. Pra-kriteria untuk menetapkan dasar kepercayaan, dalam situasi yang melibatkan IA, tidaklah berbeda, sebaliknya harus diintensifkan, karena harus terdiri dari mekanisme konkret untuk menginformasikan, mempertanggungjawabkan penggunaan, motivasi, proses, dan transparansi kriteria yang digunakan dalam pengambilan keputusan.

Prinsip kepercayaan merupakan dasar dari hubungan hukum, baik publik maupun privat. Pada gilirannya, prinsip perlindungan kepercayaan disajikan dalam dimensi individual, atau dalam aspek subjektif dari keamanan hukum. Prinsip ini bergantung pada pelaksanaan kepercayaan, dengan indikasi konkret pelanggaran harapan dalam hukum atau demonstrasi yang jelas tentang persyaratan untuk demonstrasi tersebut. O'Neill memahami bahwa kepercayaan tidak dapat disamakan dengan sekadar pengungkapan atau transparansi informasi dan akuntabilitas. Dari perspektif filosofis, kepercayaan merupakan elemen sentral dalam hubungan antarmanusia, baik antarpribadi maupun antara individu dan negara, yang melibatkan kepercayaan terhadap lembaga dan perwakilannya. Namun, kondisi kepercayaan ini tidak hanya ditunjukkan oleh pengungkapan data dan informasi, tetapi harus didukung oleh narasi yang dapat dipahami.

Dalam perspektif yuridis, prinsip kepercayaan, kata Martins-Costa, memiliki cakupan langsung untuk memenuhi harapan. Dalam kasus yang dimaksud, situasi kepercayaan terwujud antara individu dan administrasi publik, ketika data pribadi diberikan untuk tujuan tertentu seperti perawatan kesehatan, penelitian, atau jaminan sosial. Hal ini juga muncul dalam bisnis hukum, yang melibatkan penyediaan data pribadi dengan imbalan layanan kesehatan tertentu.

Bidang Kesehatan: Penggunaan AI dan Proses Pengambilan Keputusan: Peluang dan Risiko dalam Mengelola Rekam Medis Elektronik (RME)

Tidak diragukan lagi, dalam perawatan kesehatan; penelitian yang melibatkan manusia atau perancangan kebijakan publik data dan informasi merupakan hal yang sentral. Pada gilirannya, penggunaan AI dalam skenario ini bergantung dan membutuhkan data dan informasi yang tersimpan dalam rekam medis elektronik (RME).

Oleh karena itu, pengelolaan data dan informasi kesehatan, termasuk data sensitif, harus didasarkan pada prinsip kepercayaan. Oleh karena itu, mempelajari beberapa aspek terkait penggunaan RME, yang dikombinasikan dengan teknologi AI, merupakan contoh yang

baik untuk mengidentifikasi peluang dan risiko teknologi ini di bidang kesehatan. RME berfungsi sebagai memori kolektif atas bantuan yang diberikan kepada pasien. Oleh karena itu, RME harus mengumpulkan data umum dan kesehatan, deskripsi fakta pribadi dan keluarga yang relevan, yang dikumpulkan oleh tenaga kesehatan profesional selama anamnesis pasien. Riwayat inilah yang membuka catatan aktivitas bantuan.

Selain informasi ini, informasi lain ditambahkan, baik sebagai catatan maupun lampiran, seperti diagnosis, dalam bentuk laporan, gambar atau data, prognosis, rencana perawatan, hasil pemeriksaan, konsultasi yang dilakukan oleh berbagai profesional, partisipasi dalam penelitian atau catatan yang relevan dengan kasus, dengan tujuan utama untuk memberikan bantuan yang lebih baik kepada pasien.

Peluang

Rekam Medis Elektronik (RME) harus dilindungi dan dipandu oleh hubungan kepercayaan, yang didasarkan pada rasa hormat terhadap pasien. Rasa hormat terhadap pasien diungkapkan oleh kewajiban deontologis kerahasiaan, kewajiban hukum kepribadian, dan prinsip-prinsip bioetika. Pasien memberikan informasi yang dianggap relevan berdasarkan kepercayaan yang diberikan kepada profesional yang merawatnya. Dari sudut pandang profesional, informasi ini selalu dianggap istimewa.

Penggunaan data genetik dalam perawatan, seperti yang digunakan dalam konseling genetik dan Pengobatan Personal, telah memperkenalkan data baru, yang dapat menghasilkan informasi yang tidak hanya memengaruhi pasien tetapi juga orang lain yang terkait dengannya. Dengan demikian, konsep privasi pribadi meluas ke konsep privasi relasional. Hal ini meningkatkan tanggung jawab yang terkait dengan pendaftaran dan penggunaan informasi ini di masa mendatang.

Selain itu, data EHR dapat membantu intervensi terkait penelitian. Hal ini dapat mencakup penggunaan obat-obatan, sel, dan produk biologis lainnya, pelaksanaan prosedur bedah atau diagnostik, penggunaan perangkat, perubahan dalam proses perawatan, perawatan pencegahan, dan berbagai aktivitas lainnya. Dalam semua aktivitas tersebut, berbagi data ini dapat menghasilkan informasi baru dan bermanfaat.

Perkembangan penelitian klinis dan juga pengobatan personal memperluas perawatan dengan perlindungan data pribadi di bidang kesehatan, karena melibatkan kebutuhan untuk menggunakan data dan informasi pasien, yang dikumpulkan dalam lingkungan yang terlindungi berdasarkan prinsip kepercayaan. Demikian pula, prinsip ini menciptakan ekspektasi pada partisipan penelitian pada tingkat pribadi, ketika hasil penelitian dapat memengaruhi atau bermanfaat baginya, atau pada tingkat sosial dan komunitas untuk berkolaborasi dengan pengembangan ilmiah.

Selain itu, EHR telah digunakan sebagai sumber informasi yang berkualitas untuk penetapan kebijakan publik dan penelitian. Kebijakan publik sangat penting untuk menjamin akses terhadap kesehatan. Perlu dicatat bahwa, dari perspektif Hukum, isu akses terhadap kesehatan ditangani dalam konteks hak-hak dasar dan hak sipil. Pentingnya, misalnya, studi hubungan sebab-akibat epidemiologi, yang memungkinkan penetapan kebijakan publik,

protokol, pedoman, atau norma untuk pencegahan dan pengobatan penyakit dan/atau promosi kesehatan, tidak diragukan lagi penting dan mengubah arah pembangunan manusia.

Lebih lanjut, dari pendekatan epidemiologi, muncul Kedokteran Berbasis Bukti (*Evidence-Based Medicine*/EBM), yang diusulkan oleh Universitas McMaster, Kanada, pada tahun 1990an, untuk mencatat dan mensistematisasi bukti klinis dan pengetahuan epidemiologi yang diperoleh darinya, guna meningkatkan hasil dalam diagnosis dan pengobatan penyakit dan pelayanan kesehatan. EBM merupakan upaya untuk memandu pengambilan keputusan terkait pasien di tingkat individu berdasarkan data kolektif. Jadi, kebutuhan untuk mensistematisasi pengumpulan, penyimpanan, dan penggunaan data dan informasi kesehatan berkaitan langsung dengan perkembangan kedokteran dan peningkatan pengetahuan kesehatan global, baik dalam hal perawatan pasien individu, kesehatan populasi, maupun kesehatan global. Dan saat ini kita memiliki alat seperti AI yang selalu tersedia dan digunakan untuk melakukan hal tersebut.

Risiko

Perlindungan data dan informasi pribadi yang terkandung dalam EHR harus mempertimbangkan konteks baru yang muncul dalam Masyarakat Informasi, untuk pencegahan risiko. Perkembangan dan penerapan teknologi informasi dan komunikasi baru yang terus-menerus, penggunaan teknik perlindungan data baru, termasuk AI, blockchain, penggunaan media sosial, interkoneksi sistem kesehatan terintegrasi, di samping berbagi yang dihasilkan oleh lingkungan Big Data itu sendiri.

Dengan demikian, perluasan kemungkinan untuk menghubungkan, menyimpan, dan memproses data dan informasi dalam volume besar dan kompleks yang berasal dari EHR, memperkuat kekhawatiran nasional dan internasional, yang ditunjukkan dalam literatur, tentang keamanan dan pelestarian data dan informasi pasien yang terkandung dalam PEP. Khususnya, penghormatan dan penggunaan yang tepat untuk tujuannya demi kepentingan pasien menjadi topik yang disoroti. Tinjauan pustaka yang dilakukan terhadap 125 artikel ilmiah, yang dipilih dari total 5.278 artikel, dalam basis data PubMed dan Scielo, menunjukkan keamanan informasi sebagai tema yang berulang dalam pengelolaan rekam medis elektronik dan akses ke rekam medis.

Rekam medis elektronik (RME) menyajikan data dan informasi secara terstruktur. Namun, karena rekam medis harus dikembangkan dalam teks yang dikontekstualisasikan dengan kehidupan dan kehidupan pasien, analisis kualitatif atau bahkan kuantitatif dapat terhambat. Selain itu, data dan informasi pribadi, terutama di bidang kesehatan, harus dianggap sebagai konsep yang terpisah. Informasi tidak berdiri sendiri, ia membutuhkan penerima, seseorang yang memberi makna dan arti pada data tersebut. Data yang terisolasi menggambarkan karakteristik sesuatu, seseorang, suatu fakta atau situasi. Namun, informasilah yang memberi makna pada data tersebut. Informasi tersebut memengaruhi data, dan merupakan hasil analisis dan interpretasi data. Singkatnya, informasi adalah organisasi, kategorisasi, dan sistematisasi data untuk perspektif dan tujuan tertentu yang menghasilkan informasi. Definisi ini merupakan titik awal yang relevan untuk memahami pentingnya data dan informasi di bidang kesehatan.

Demikian pula, penting untuk mempertimbangkan berbagai gagasan dan konsep terkait lingkungan data bervolume besar Big Data yang dihasilkan dalam EHR dan sistem kesehatan. Big Data adalah istilah yang digunakan secara umum untuk menunjukkan pengelompokan data, informasi, basis data, jaringan internet terbuka, dan data lain yang dapat diakses yang awalnya bertujuan untuk meningkatkan perencanaan strategis, pemasaran, dan bisnis komersial. Konteks ini, yang ditandai oleh fluiditas, ketidakpastian, dan fugasitas data dan informasi, membutuhkan berbagai sumber untuk memahami fenomena kompleks dan luas yang dapat diinterpretasikan oleh sistem AI.

Perspektif baru yang dihasilkan oleh fenomena Big Data telah mendorong perkembangan ilmiah di berbagai bidang ilmu pengetahuan. Sebagaimana ditunjukkan oleh Mittelstadt dan Floridi dalam artikel tinjauan pustaka tahun 2016, AI di bidang kesehatan sudah menjadi kenyataan, di samping kemungkinan-kemungkinan lain yang akan segera terwujud dan kemungkinan-kemungkinan lain yang masih berupa potensi. Contoh situasi yang sudah menjadi kenyataan antara lain terkait aktivitas Biobank, studi Kesehatan Masyarakat, dan pengujian hipotesis di bidang kesehatan. Situasi yang memungkinkan antara lain interkoneksi peralatan dan aplikasi untuk kesehatan pribadi; keberadaan profil daring yang terhubung dengan rekam medis; terciptanya media sosial di bidang kesehatan, dan koneksi daring dan luring profil pribadi melalui Wi-Fi. Terakhir, mereka menunjukkan situasi potensial berupa koneksi antara rekam medis daring dengan sumber data pribadi lainnya, serta koneksi tak disengaja dari data-data ini, baik daring maupun luring, yang berasal dari profil pribadi untuk keperluan pengawasan kesehatan.

Floridi mengatakan:

> Jelas, masa depan AI tidak hanya terletak pada "data kecil" tetapi juga, atau mungkin terutama, pada peningkatan kemampuannya untuk menghasilkan datanya sendiri. Itu akan menjadi perkembangan yang luar biasa, dan kita dapat mengharapkan upaya signifikan untuk dilakukan ke arah itu. Selain itu, menerjemahkan tugas-tugas sulit menjadi tugas-tugas kompleks. (...) Bagaimana penerjemahan ini dicapai? Dengan mengubah lingkungan tempat AI beroperasi menjadi lingkungan yang ramah AI.

Untuk alasan ini dan alasan lainnya, analisis dampak risiko yang akurat dan tindakan pencegahan harus diambil, dalam lingkup normatif, praktik baik, kepatuhan, dan etika, terutama untuk menghindari bias dalam pengambilan keputusan. Bias algoritmik merupakan salah satu kekhawatiran, terutama dalam pemrosesan data pribadi yang sensitif, seperti data kesehatan, genetik, dan biometrik.

Selain itu, bias algoritmik yang dapat mendiskriminasi secara negatif dan/atau menyebabkan kerugian bagi individu atau kelompok tertentu misalnya, yang diorganisasikan berdasarkan gender, jenis kelamin, usia, status kesehatan fisik atau mental, dan orang atau kelompok yang rentan secara ekonomi atau sosial (misalnya, narapidana dan masyarakat miskin). Norori dkk., dalam artikel berjudul "Addressing bias in big data and AI for health care: A call for open science" (Menangani bias dalam big data dan AI untuk layanan kesehatan: Seruan untuk sains terbuka), menunjukkan bahwa di masa depan, atau bisa dibilang saat ini, penelitian diperlukan untuk menetapkan standar AI dalam layanan kesehatan yang memungkinkan transparansi dan berbagi data, sekaligus menjaga privasi pasien. Para penulis

menyajikan perbedaan antara bias statistik dan bias sosial sebagai titik awal analisis. Bias statistik mengacu pada kasus-kasus di mana distribusi kumpulan data tertentu tidak mencerminkan distribusi populasi yang sebenarnya, dan pada gilirannya, bias sosial mengacu pada ketidakadilan yang dapat mengakibatkan hasil yang suboptimal untuk kelompok populasi manusia tertentu.

Para penulis menunjukkan beberapa contoh algoritma AI yang bias sejak awal, terkait jenis kelamin, usia, dan ras. Bias ini dapat diamati dalam studi yang mendiskriminasi jenis kelamin perempuan dan laki-laki, termasuk dalam penelitian praklinis, ketika dalam model eksperimental yang menggunakan hewan terdapat dominasi laki-laki. Demikian pula dalam penelitian untuk pengembangan obat-obatan, ketika mayoritas partisipan adalah laki-laki tanpa alasan metodologis yang membenarkannya. Mereka juga menunjukkan, melalui contoh tersebut, studi di bidang gangguan tidur, ketika pasien muda lebih memihak pasien yang lebih tua. Lebih lanjut, terdapat bias rasial ketika algoritma, dalam bidang kanker kulit, diprogram untuk mengidentifikasi citra kulit terang dan bukan kulit gelap, meskipun populasi kulit hitam memiliki tingkat kematian yang lebih tinggi akibat kanker melanoma. Juga dalam bidang diskriminasi negatif berdasarkan ras, terdapat algoritma di bidang biaya rumah sakit yang mendorong penentuan bahwa pasien kulit hitam lebih sehat daripada pasien kulit putih, dan karena alasan ini, mereka menerima perawatan yang lebih baik.

Contoh-contoh ini cukup untuk menunjukkan bahwa ketakutan dan kurangnya kepercayaan diri dalam pengambilan keputusan yang digerakkan oleh AI bukanlah hal yang sia-sia, atau bahkan tidak proporsional itu adalah kenyataan yang harus dihindari secara normatif dan etis. Diskriminasi negatif, sebagaimana telah kami tunjukkan, di Amerika Latin diperparah oleh banyaknya orang yang tidak memiliki akses terhadap teknologi kesehatan atau didiskriminasi di “siang bolong” karena kondisi dan kekurangan ekonomi mereka, kurangnya pendidikan dan kurangnya kondisi sanitasi seperti yang dapat dilihat di daerah kumuh dan pinggiran kota dan yang dibuktikan dalam pandemi COVID-19 ironisnya data yang juga telah dibuktikan dengan bantuan teknologi yang digerakkan oleh AI. Jadi apa yang harus dilakukan? Di mana kita harus bertindak secara nasional dan internasional? Parameter apa yang harus kita jadikan titik awal? Untuk mencoba menjawab dan/atau merenungkan pertanyaan-pertanyaan ini, kita beralih ke poin kedua kita, Model Bioetika Kompleks (CBM) dan AI.

10.3 MODEL BIOETIKA KOMPLEKS (CBM) DAN AI

Kita berada pada tahap bersejarah di mana imigran digital dan penduduk asli digital hidup berdampingan. Imigran digital memiliki kesempatan untuk hidup dalam masyarakat di mana semua keputusan hanya dibuat oleh manusia. Di sisi lain, penduduk asli digital menaturalisasi keputusan yang dibuat oleh algoritma.

Naturalisasi keputusan yang hanya dibuat oleh kecerdasan buatan dapat melibatkan beberapa isu etika penting, seperti arogansi teknologi, visi kepastian, dan imparsialitas algoritma. Dengan menggunakan algoritma, mesin mengikuti pola langkah-langkah yang dapat diprediksi dan telah diprogram sebelumnya. Bahkan dengan penggabungan proses pembelajaran mesin terkait, keputusan-keputusan ini hanya membawa serta elemen-elemen

rasional yang terkait dengan proses pengambilan keputusan. Dalam beberapa model, nilai-nilai, isu-isu afektif, dan bahkan tradisi budaya dapat dimasukkan sebagai elemen-elemen dari proses pengambilan keputusan ini. Namun, tindakan-tindakan non-rasional ini dianggap rasional oleh model komputasi. Komputer tidak ragu, manusia pun ragu.

Proses yang digunakan dalam kecerdasan buatan merupakan hasil dari pemrograman. Pemrograman tidak menoleransi ambiguitas atau ketidakpastian, yang selalu ada di dunia nyata. Bahkan dengan menggunakan metodologi berbasis logika fuzzy, secara tegas, hal tersebut merupakan ketidakpastian yang terprogram.

Ada anggapan bahwa manusia bisa salah, tetapi mesin tidak. Setiap proses pengambilan keputusan yang menggunakan kecerdasan buatan didasarkan pada serangkaian asumsi yang ditetapkan oleh manusia. Bahkan ketika terdapat sistem yang memprogram sendiri, akar prosesnya didasarkan pada pilihan yang dibuat oleh orang-orang yang merencanakan dan mengimplementasikannya. Terdapat berbagai tingkat kompleksitas, tetapi semuanya mengarah pada suatu akar di mana terdapat karakteristik non-rasional dari para pengembangnya.

Dari sudut pandang etika, setiap tindakan manusia, atau yang dihasilkan darinya, harus dievaluasi kecukupannya. Penilaian ini tidak hanya membutuhkan pertimbangan fakta, tetapi juga keseluruhan keadaan. Salah satu keadaan ini adalah dimensi historis, yaitu perspektif penyisipan aktivitas-aktivitas ini dari waktu ke waktu. Memahami kompleksitas masalah yang sedang dievaluasi merupakan kebutuhan kritis.

Dikotomi yang tampak antara buatan dan alami ini semakin renggang. Semakin penting untuk memiliki perspektif yang kompleks dalam memahami situasi yang semakin sering muncul dalam masyarakat manusia. Dengan menggunakan pendekatan Bioetika yang kompleks, dimungkinkan untuk memiliki argumen etis menggunakan kerangka kerja teoretis yang berbeda. Salah satunya, berdasarkan kebajikan; niat dan persetujuan; prinsip; tanggung jawab; hak asasi manusia; konsekuensi dan alteritas, dapat digunakan untuk memahami hubungan sistem manusia-komputer.

Kebajikan dapat digunakan untuk membenarkan kecukupan perilaku pribadi yang terlibat dalam perancangan dan penerapan sistem pengambilan keputusan. Kehati-hatian, kesederhanaan, dan keadilan adalah kebajikan fundamental yang perlu dipertimbangkan dalam situasi ini. Sistem harus didasarkan pada penalaran praktis, harus menggunakan sumber daya yang terlibat dengan tepat, dan yang terpenting, tidak mendiskriminasi orang atau kelompok mana pun. Keutamaan mengandaikan hasrat bagi kemanusiaan, yang memproyeksikan dirinya dalam waktu, yang selalu memiliki perspektif historis.

Niat dan persetujuan yang terkait dengan tindakan harus dipertimbangkan dalam mengevaluasi nilai moral yang terkait dengan suatu tindakan. Niat siapa pun yang merancang atau menggunakan suatu sistem harus memadai, dan harus bertujuan untuk kebaikan rakyat. Di sisi lain, penggunaan sistem hanya dianggap tepat jika mendapat persetujuan dari orang-orang yang terdampak. Kombinasi kehendak ini, dari mereka yang melakukan tindakan dan mereka yang menderita tindakan tersebut, sangatlah mendasar.

Etika berbasis prinsip juga harus memandu penilaian kesesuaian penggunaan sistem. Kerangka kerja empat prinsip Martabat, Kebebasan, Integritas, dan Kerentanan dapat sangat membantu dalam penilaian ini. Koherensi dalam penerapan prinsip-prinsip ini, yang dipahami sebagai panduan tindakan manusia, harus diupayakan. Martabat menyatukan kita dengan semua orang, dan inilah yang membentuk karakter kemanusiaan kita semua. Kebebasan adalah kemungkinan untuk memilih, untuk membuat pilihan yang bebas dari paksaan. Integritas, yang dipahami dalam dimensi fisik, mental, dan sosialnya, harus selalu didasarkan pada upaya pelestariannya. Kerentanan harus dipertimbangkan setiap kali ada kemungkinan martabat, kebebasan, atau integritas dapat dikompromikan. Dalam masyarakat yang berisiko, kita semua selalu rentan, dalam berbagai tingkat dan situasi.

Etika hak asasi manusia didasarkan pada ekspektasi tindakan. Hak asasi manusia dapat didekati dari perspektif individu atau kolektif, atau bahkan secara transpersonal. Dari hak untuk hidup dan privasi hingga hak untuk solidaritas atau untuk memiliki lingkungan yang terlestarikan, hak merupakan ekspresi dari tindakan orang lain terhadap saya. Sistem kecerdasan buatan mungkin tidak memiliki perspektif ganda ini saat mengambil keputusan. Terkadang satu hak diistimewakan dan hak lainnya tidak dipertimbangkan. Bisa jadi, dengan menjamin hak privasi, suatu sistem justru mengabaikan dimensi solidaritas.

Etika konsekuensialis didasarkan pada risiko dan manfaat yang terkait dengan tindakan yang ditujukan kepada individu atau kolektif. Proses pengambilan keputusan konsekuensialis, dari sudut pandang mikro atau makro, didasarkan pada analisis utilitas (risiko versus manfaat). Isu terpenting adalah mencapai keseimbangan antara kedua perspektif ini, untuk membangun strategi yang saling menguntungkan.

Etika tanggung jawab berfokus pada tindakan. Kedua perspektif, baik retrospektif maupun prospektif, menilai dampak dari tindakan yang dilakukan. Pendekatan retrospektif berfokus pada sebab dan prospektif pada akibat. Pendekatan umum terhadap tanggung jawab adalah melihat siapa yang melakukannya dan bagaimana tindakan itu dilakukan. Baru-baru ini, fokus telah bergeser dari sebab ke dampak tindakan. Jika ilmuwan bertanggung jawab atas konsekuensi sosial sains, demikian pula orang-orang yang merancang sistem kecerdasan buatan. Dalam perspektif ini, sebuah keharusan baru ditetapkan sebagai respons terhadap tindakan manusia: "Dalam pilihan Anda saat ini, sertakan integritas masa depan manusia di antara objek kehendak Anda". Dengan kata lain, kita tidak boleh melakukan semua hal yang dimungkinkan oleh teknik.

Terakhir, alteritas adalah pendekatan teoretis lain untuk mengevaluasi dasar etika kecerdasan buatan. Sistem dibangun untuk menjadi permanen, memiliki identitas, kekekalan, dan mengasumsikan totalitas tindakan terkait. Inilah perspektif kesamaan. Di sisi lain, keberbedaan mengasumsikan ketidakkekalan, singularitas, misteri, dan ketidakterbatasan. Keberbedaan membuka kita terhadap yang lain dan menegaskan kembali kita sebagai manusia. Perspektif ini memungkinkan terciptanya kehadiran bersama yang etis, tanggung jawab bersama, perspektif yang melampaui hubungan sederhana untuk menjadi interaksi yang efektif.

Dari perspektif kecerdasan buatan, kesamaan menang atas keberbedaan. Dalam perspektif alteritas, mustahil untuk mendekati teknologi baru dari sudut pandang netral. Pendekatan Bioetika Kompleks memungkinkan integrasi berbagai perspektif teoretis ini dalam pencarian argumen untuk merefleksikan kecukupan penggunaan teknologi kecerdasan buatan. Ini merupakan cara yang baik untuk mendapatkan perspektif komprehensif tentang proposal yang seringkali hanya dilihat dari aspek teknisnya.

Diskusi etika kontemporer harus dipandu oleh refleksi tentang "rezim informasi" yang baru, sebagaimana dicirikan oleh Han. Inilah tantangan kita: untuk merefleksikan model masyarakat baru ini, di mana hubungan telah berubah secara drastis. AI, AI Generatif, dan spesies lainnya, hanyalah salah satu dari sekian banyak tantangan yang perlu dibahas dan diperdalam.

10.4 BATAS KECERDASAN AI: ETIKA DAN KEBODOHAN

Kecerdasan buatan didefinisikan dalam arti bahwa mesin dapat melakukan tugas-tugas yang serupa dengan yang dilakukan oleh manusia. Pada awal komputasi, komputer disebut "otak elektronik". Kemudian, metafora komputer digunakan untuk menjelaskan cara kerja otak manusia. Salah satu risiko saat ini adalah melakukan inversi ini lagi, yaitu ingin menjelaskan kecerdasan manusia menggunakan model kecerdasan buatan.

Tantangan lain dalam mentransposisi kecerdasan manusia ke kecerdasan buatan adalah menyadari bahwa manusia bisa gagal, maka mesin juga bisa gagal. Jika transposisi ini terjadi, bisa jadi akan muncul usulan untuk "kebodohan buatan" yang terkait dengan "kecerdasan buatan", seperti kebodohan dan kecerdasan manusia. Etika dan Bioetika dapat membantu dalam merefleksikan kecukupan tentang batasan dan batasan antara kecerdasan alami atau buatan dan kecerdasan atau kebodohan. Lebih penting daripada membahas aspek etika yang tepat waktu, penting untuk merefleksikan aspek yang lebih luas dari penggunaan AI, seperti:

(a) menetapkan standar etika yang tepat untuk memandu pembuatan konten yang bertanggung jawab oleh sistem ini;
(b) menetapkan strategi pemantauan data dan informasi yang dihasilkan oleh AI dan AI Generatif untuk memverifikasi kebenarannya;
(c) membuat pedoman yang memungkinkan audit berkelanjutan terhadap proses sistem ini untuk mencegah penggunaannya untuk tujuan yang bertentangan dengan kepentingan manusia, masyarakat, dan kemanusiaan.

Dalam pendekatan bioetika terhadap teknologi baru, penting untuk mengaitkan prinsip kehati-hatian dengan prinsip harapan. Artinya, kehati-hatian bertujuan untuk menjamin kehidupan setiap orang dan harapan bertujuan untuk mempertahankan kehidupan setiap orang.

BAB 11
DOKTER AI OTONOM: ETIKA DAN HUKUM MEDIS

Abstrak Kecerdasan buatan (AI) saat ini mampu melakukan tindakan-tindakan yang merupakan praktik medis secara otonom, termasuk diagnosis, prognosis, pengambilan keputusan terapeutik, dan analisis citra, tetapi haruskah AI dianggap sebagai praktisi medis? Yang memperumit pertanyaan ini adalah kenyataan bahwa rezim etika, regulasi, dan hukum yang mengatur praktik medis dan malapraktik medis tidak dirancang untuk dokter nonmanusia. Bab ini pertama-tama menyarankan parameter etika untuk praktik kedokteran Dokter AI Otonom, dengan fokus pada bidang patologi.

Kedua, kami mengidentifikasi isu-isu etika dan hukum yang muncul dari praktik kedokteran Dokter AI Otonom, termasuk keselamatan, keandalan, transparansi, keadilan, dan akuntabilitas. Ketiga, kami membahas potensi penerapan berbagai rezim hukum dan regulasi yang ada untuk mengatur Dokter AI Otonom. Akhirnya, kami menyimpulkan bahwa semua pemangku kepentingan dalam pengembangan dan penggunaan Dokter AI Otonom memiliki kewajiban untuk memastikan bahwa AI diimplementasikan dengan cara yang aman dan bertanggung jawab.

11.1 DOKTER AI OTONOM: REGULASI DAN ETIKA

Kecerdasan buatan (AI) secara umum menggambarkan, "kemampuan program komputer untuk melakukan tugas atau proses penalaran yang biasanya kita kaitkan dengan kecerdasan manusia." Meskipun AI tampaknya tidak akan sepenuhnya menggantikan dokter manusia dalam waktu dekat, kini AI dapat secara mandiri melakukan tugas-tugas yang berada dalam lingkup praktik medis, terutama diagnosis, prognosis, dan konsultasi sebagai respons terhadap informasi medis individual. Pada bulan April 2018, Badan Pengawas Obat dan Makanan AS (FDA) menyetujui IDx-DR, perangkat kecerdasan buatan pertama yang mampu mendiagnosis pasien retinopati diabetik secara otonom tanpa masukan dari dokter manusia.

Di Eropa, Oxipit, sebuah AI yang mampu secara otonom menghasilkan "laporan akhir untuk studi sinar-X pasien sehat" menerima tanda CE, membuka jalan bagi penggunaannya dalam praktik klinis. Pembelajaran mendalam (DL) adalah subset AI yang paling mungkin menghasilkan teknologi, seperti IDx-DR dan Oxipit, yang mampu mengambil keputusan medis secara otonom dengan melatih mesin dengan jaringan saraf tiruan untuk menganalisis sejumlah besar data medis dan kesehatan untuk mendeteksi pola.

Pengenalan kecerdasan buatan menggunakan DL dalam kedokteran modern menjanjikan peningkatan akurasi, efikasi, dan efisiensi diagnosis medis, prognosis, pengambilan keputusan terapeutik, analisis gambar, dan pemantauan pasien. Di sisi lain, hal ini juga menimbulkan sejumlah masalah etika dan hukum seputar keamanan, transparansi, bias dan diskriminasi, privasi data, persetujuan dan otonomi, serta tanggung jawab dan akuntabilitas. Kekhawatiran ini semakin diperkuat oleh fakta bahwa rezim etika dan hukum

yang ada yang mengatur praktik medis dan malapraktik medis tidak dirancang untuk dokter nonmanusia.

Patologi, sebagai subspesialisasi kedokteran yang kaya data, merupakan pusat pengembangan dan implementasi AI medis. Oleh karena itu, Jackson dkk. menghimbau para ahli patologi untuk memberikan kepemimpinan, baik dalam hal pengembangan, regulasi, maupun etika, untuk penerapan AI dalam kedokteran klinis dan laboratorium.

Bab ini menggabungkan keahlian medis dan hukum para penulisnya untuk merekomendasikan parameter bagi Dokter AI Otonom dan mengidentifikasi isu-isu etika dan hukum yang muncul dari praktik kedokteran oleh Dokter AI Otonom. Selanjutnya, para penulis mengidentifikasi dan menyarankan potensi penerapan konsep-konsep dari berbagai rezim regulasi dan hukum yang saat ini mengatur praktik medis dan malapraktik medis terhadap praktik kedokteran di masa mendatang oleh dokter yang menggunakan AI Otonom.

11.2 KECERDASAN BUATAN DALAM PATOLOGI

Patologi menggunakan alur kerja yang intensif data, kompleks, dan komprehensif untuk mendiagnosis dan mempelajari proses penyakit. Temuan dan data patologi anatomi dan klinis sangat memengaruhi diagnosis, prognosis, dan rekomendasi terapi untuk semua spesialisasi medis. Digitalisasi telah meningkatkan alur kerja dalam patologi dengan memungkinkan analisis mikroskopis virtual dari pencitraan seluruh slide, yang telah terbukti sebanding dengan mikroskop konvensional, yang telah lama dianggap sebagai standar emas untuk mendeteksi perubahan patologis pada jaringan dan sel.

Pengenalan AI ke dalam analisis morfologi ini menjanjikan peningkatan akurasi lebih lanjut dengan mengurangi inkonsistensi diagnostik yang disebabkan oleh variabilitas pengamat manusia. Misalnya, AI dapat dilatih dengan gambar digital dan diagnosis terkait yang diberikan oleh ahli patologi manusia untuk menganalisis gambar baru untuk pola patologis yang mengarah pada diagnosis yang lebih cepat dan lebih akurat. Baru-baru ini, AI telah membuktikan bahwa ia dapat mengungguli dokter manusia dalam tugas-tugas yang bahkan lebih kompleks, termasuk memprediksi stadium dan tingkat kanker paru-paru, menggunakan teknik DL.

Digitalisasi patologi yang dikombinasikan dengan pembangkitan data medis dalam jumlah besar menjadikan bidang ini, bersama dengan radiologi, sebagai "target utama untuk inovasi disruptif aplikasi AI perawatan kesehatan selama dekade mendatang." Allen mengartikulasikan tiga tingkat integrasi AI yang progresif dalam patologi. Tingkat pertama menjaga agar ahli patologi tetap berada dalam alur kerja dengan mengintegrasikan AI sebagai salah satu dari banyak alat diagnostik yang digunakan ahli patologi untuk pengambilan keputusan medis. Tingkat kedua menggambarkan AI yang dapat secara independen menyajikan laporan ahli patologi tetapi tetap menjaga agar ahli patologi manusia tetap berada dalam alur kerja untuk memberikan pengawasan kualitas atas keputusan medis yang dihasilkan oleh AI.

Tingkat ketiga keterlibatan AI memisahkan ahli patologi manusia dari alur kerja yang sepenuhnya dikendalikan oleh AI otonom. Pakar medis lainnya meragukan masa depan di

mana AI sepenuhnya menggantikan ahli patologi manusia, dengan menyatakan bahwa AI belum menguasai kemampuan unik otak manusia untuk mensintesis informasi di berbagai sektor pengetahuan. Meskipun Chauhan dan Gullapalli mengakui bahwa peran AI di masa depan dalam patologi tidak dapat diprediksi, mereka berani membuat satu prediksi: "Kebutuhan akan pengawasan yang cermat dan hati-hati terhadap kualitas dan kendali proses oleh ahli patologi kemungkinan besar tidak akan diotomatisasi dalam waktu dekat." Ahli patologi di laboratorium klinis juga bertanggung jawab atas pembuatan dan penyimpanan "salah satu sumber tunggal terbesar data objektif dan terstruktur tingkat pasien dalam sistem pelayanan kesehatan." Akibatnya, para ahli patologi tidak hanya berada pada posisi yang tepat, tetapi juga, sebagai penjaga data medis yang sangat berharga, secara etis berkewajiban untuk membantu mengantar era baru AI.

11.3 DOKTER AI OTONOM: PARAMETER

Meskipun konsep dokter robot yang mandiri mungkin hanya fiksi ilmiah, AI sudah mampu mempraktikkan pengobatan secara otonom, termasuk diagnosis, prognosis, dan pemberian rekomendasi perawatan. Pembelajaran mendalam memungkinkan AI meniru fungsi otak manusia untuk memproses data secara mandiri dan mengambil keputusan menggunakan penalaran algoritmik yang terus meningkat seiring AI mengumpulkan lebih banyak data.

Meskipun produsen AI mungkin mencoba menggambarkan AI sebagai "komputasi kognitif" atau alat pendukung medis, kenyataannya AI kini dapat secara mandiri berkonsultasi dengan jutaan halaman literatur untuk menyarankan perawatan medis individual, menganalisis dan menginterpretasi gambar radiologi dan slide patologi, mendiagnosis dan menentukan stadium kanker, dan memprediksi luaran pasien. Dan dengan investasi dalam AI kesehatan yang mengungguli sektor lain mana pun dalam ekonomi global, kapabilitas AI medis hanya akan terus tumbuh. Beberapa futuris memprediksi kecerdasan umum buatan akan menjadi kenyataan pada tahun 2029.

Sekaranglah saatnya untuk menetapkan parameter bagi praktik kedokteran otonom oleh AI. Meskipun mengakui bahwa AI mungkin dapat "mengisi banyak kesenjangan antara kinerja manusia dan kesempurnaan," AI tidak akan pernah mampu menyediakan komponen manusia yang integral dari praktik medis, "seperti sentuhan, kasih sayang, dan empati." Griffin menjelaskan bahwa: "Kedokteran bukanlah semata-mata ilmu yang dapat dikelola dengan statistik, matematika, dan algoritma komputer, dan ketergantungan yang berlebihan pada AI dapat menyebabkan bahaya dalam situasi ketika kasih sayang manusia, sentuhan manusia, atau interpretasi manusia terhadap konteks data diperlukan." Dari perspektif patologi diagnostik, Ahmad, dkk. Perlu dicatat bahwa, "proses diagnostik terlalu rumit dan beragam untuk dipercayakan hanya kepada algoritma yang terprogram. Diharapkan bahwa AI dan ahli patologi manusia akan menjadi kooperator alami, bukan pesaing alami."

Sebagaimana diakui oleh Komite Khusus Uni Eropa untuk Kecerdasan Buatan di Era Digital, pengawasan manusia atas keputusan medis AI otonom sangat diperlukan (Komite Khusus Parlemen Eropa untuk Kecerdasan Buatan di Era Digital 2021). Akibatnya, Dokter AI

Otonom, sebagaimana digunakan dalam Bab ini, menggambarkan kecerdasan buatan yang mampu melakukan tindakan yang umumnya dianggap sebagai praktik medis (diagnosis, prognosis, pengembangan rencana perawatan, dll.) menggunakan penalaran algoritmik untuk membuat keputusan medis tanpa melibatkan manusia dalam proses pengambilan keputusan medis tersebut. Selain itu, Dokter AI Otonom saat ini tidak boleh berdiri sendiri sebagai satu-satunya pengambil keputusan medis untuk seorang pasien, tetapi harus ditempatkan dalam tim perawatan yang lebih besar yang mencakup praktisi medis manusia.

11.4 IMPLIKASI ETIS DAN HUKUM DARI DOKTER AI OTONOM

Proliferasi teknologi AI yang mampu melakukan tugas-tugas yang biasanya diperuntukkan bagi tenaga medis profesional dapat menghasilkan layanan kesehatan yang lebih murah, lebih mudah diakses, dan lebih berkualitas. Selain AI diagnostik, mulai dari diagnosis oftalmologi IDx-DR hingga laporan radiologi Oxipit, aplikasi AI dalam kedokteran dapat membaca pemindaian mata, memprediksi penyakit arteri koroner stadium awal, dan mendeteksi henti jantung melalui telepon secara waktu nyata. Namun, keuntungan yang diperoleh dari teknologi AI inovatif dalam kedokteran bukannya tanpa risiko terhadap keselamatan dan privasi pasien. Oleh karena itu, para ahli medis, hukum, dan data menyerukan pengawasan etika dan regulasi yang kuat terhadap AI di sektor kesehatan untuk memastikan bahwa teknologi baru diimplementasikan secara adil, aman, dan aman. Meskipun badan-badan regulator kini berupaya mengatasi risiko AI, "pengembangan AI awal, secara umum, telah terjadi secara substansial di luar lingkungan regulasi apa pun."

Pengaturan pengembangan AI di sektor apa pun terhambat oleh "masalah kecepatan", yang menggambarkan kecenderungan inovasi teknologi untuk melepaskan diri dari rezim regulasi dan norma sosial yang tertinggal dari perkembangan teknologi baru. Selain itu, inovasi dalam industri teknologi didorong oleh nilai-nilai yang sangat berbeda dengan nilai-nilai di industri perawatan kesehatan. "Bergerak cepat dan melakukan terobosan" tidak sepenuhnya diterjemahkan menjadi strategi keselamatan pasien yang dapat diterima.

Namun, strategi etika, regulasi, dan hukum yang berhasil untuk memandu implementasi AI dalam praktik medis perlu menyeimbangkan manfaat dari mendorong inovasi di sektor kesehatan dengan risiko terhadap keselamatan dan privasi pasien. Menariknya, industri teknologi, termasuk peneliti AI dan ilmuwan data, awalnya memimpin diskusi seputar pentingnya AI yang etis dan bertanggung jawab, tetapi para ahli memperingatkan bahwa industri yang bertanggung jawab mengembangkan AI tidak dapat sendirian memandu implementasi etis dari teknologi yang sama. Allen berpendapat bahwa mencapai tugas semacam itu "kemungkinan besar membutuhkan tingkat kerja sama dan pembangunan kepercayaan yang belum pernah terjadi sebelumnya antara pemerintah, profesional, masyarakat, dan industri."

Meskipun aspek etika dan hukum dari integrasi AI otonom dalam kedokteran tidak dapat dipisahkan secara jelas, kami tetap mengorganisasikan diskusi etika dengan mempertimbangkan prinsip-prinsip inti etika kedokteran otonomi, kebaikan hati, nonmalfeasance, dan keadilan untuk menyimpulkan bahwa AI yang etis harus transparan,

andal, aman, dan bebas bias, sekaligus mengorganisasikan diskusi hukum seputar privasi data dan tanggung jawab atas kerugian pasien.

Pertimbangan Etis: Transparansi

Persyaratan bahwa AI medis harus mempertahankan tingkat transparansi yang memadai untuk menjamin otonomi pasien ada dua. Pertama, pengembang AI harus transparan tentang penggunaan data pasien untuk pelatihan sistem AI medis. Kedua, pasien harus memiliki informasi yang memadai tentang penggunaan AI dalam perawatan klinis, termasuk risiko dan manfaat keputusan medis berbasis AI serta informasi tentang bagaimana keputusan tersebut dibuat.

Pertanyaan apakah pasien perlu memberikan izin kepada pengembang AI untuk menggunakan data kesehatan yang dihasilkan selama perawatan medis mereka bergantung pada aturan yurisdiksi yang mengatur pengungkapan informasi kesehatan pribadi. Di Amerika Serikat, karena data pasien biasanya dideidentifikasi sebelum dibagikan dengan pengembang AI untuk pelatihan teknologi baru, undang-undang utama yang melindungi data kesehatan, Undang-Undang Portabilitas dan Akuntabilitas Asuransi Kesehatan (HIPAA), tidak mencegah pengungkapan. Jackson, dkk. berpendapat bahwa risiko identifikasi ulang data pasien dengan merujuk silang beberapa set data mengharuskan persetujuan pasien diperoleh sebelum menggunakan data kesehatan mereka untuk pelatihan AI.

Persyaratan persetujuan ini, jika diakui, tidak dapat dipenuhi dengan memperoleh persetujuan pasien untuk memproses data perawatan medis dan pembayaran individual, melainkan memerlukan persetujuan tambahan. Di Eropa, Peraturan Perlindungan Data Umum (GDPR) umumnya mewajibkan persetujuan pasien secara eksplisit untuk tujuan yang ditentukan secara spesifik sebelum data kesehatan pasien individu diproses untuk alasan apa pun (Parlemen Eropa dan Dewan Uni Eropa 2016, Pasal 9).

Meskipun GDPR tidak mengatur "data anonim", yang tidak lagi dapat dikaitkan dengan orang yang dapat diidentifikasi, GDPR membatasi penggunaan data yang memiliki kemungkinan wajar untuk diidentifikasi ulang (Parlemen Eropa dan Dewan Uni Eropa 2016, Pertimbangan 26). Selain memiliki kendali atas data kesehatan mereka, pasien juga harus diinformasikan tentang peran pengambilan keputusan yang bersumber dari AI dalam perawatan medis mereka. Kapan dan sejauh mana pasien diinformasikan tentang penggunaan AI dalam membuat diagnosis dan keputusan terapeutik masih belum pasti, sehingga para ahli medis dan hukum khawatir tentang pelanggaran terhadap kemampuan pasien untuk menjalankan otonomi yang dibutuhkan untuk membuat keputusan yang tepat tentang perawatan mereka.

Minimalnya, pasien harus diinformasikan ketika AI digunakan untuk menghasilkan diagnosis atau rekomendasi perawatan, termasuk penjelasan tentang risiko apa pun yang terkait dengan penerimaan keputusan medis yang bersumber dari AI. Idealnya, pasien juga harus diberikan penjelasan yang sederhana tentang bagaimana AI mencapai kesimpulannya; namun, kenyataannya adalah bahwa teknologi AI yang menggunakan DL dapat menyembunyikan kriteria pengambilan keputusan algoritmik bahkan dari pengembang AI itu sendiri, yang menciptakan masalah AI "kotak hitam" yang tidak transparan, yang

menggambarkan ketidakmampuan manusia untuk memahami dasar keputusan AI. Meskipun demikian, pengungkapan mengenai data yang digunakan untuk melatih AI serta statistik kinerja pra-pemasaran AI harus diberikan kepada pasien yang perawatannya dipengaruhi oleh pengambilan keputusan medis AI.

Pertimbangan Etis: Keandalan dan Keamanan

Masalah "kotak hitam" juga menjadi hambatan dalam memastikan keandalan dan keamanan pengambilan keputusan medis yang bersumber dari AI bagi pasien karena menyembunyikan proses pengambilan keputusan oleh sistem AI, sehingga mencegah analisis dan pengawasan proses tersebut. Beberapa ahli berpendapat bahwa akurasi keputusan AI adalah yang terpenting, terlepas dari proses tersembunyi yang digunakan untuk mencapai keputusan tersebut.

Meskipun fungsi algoritmik yang digunakan dalam analisis data AI tidak, dan tidak mungkin, sepenuhnya transparan dengan AI "kotak hitam", pengembang tetap harus memberikan informasi penting tentang bagaimana AI dilatih dan potensi bias perangkat lunak untuk pengawasan dan analisis independen. Biasanya, kualitas data pelatihan yang diberikan kepada AI akan berkorelasi langsung dengan kualitas keputusan medis AI. Namun, bahkan ketika AI menghasilkan keputusan yang andal secara teknis berdasarkan data yang diterimanya, keputusan tersebut mungkin tidak andal secara klinis dan mengancam keselamatan pasien. Sementara DL dapat memungkinkan AI untuk mengembangkan dan menerapkan aturan untuk mendeteksi pola menggunakan set data yang besar, AI masih tidak dapat menjalankan penalaran klinis seorang dokter manusia untuk menentukan perbedaan antara sebab akibat dan korelasi. Misalnya, sistem AI yang dilatih untuk melakukan triase pasien dengan pneumonia menentukan bahwa pasien asma berisiko rendah karena mereka memiliki hasil pemulihan yang lebih baik setelah diagnosis pneumonia daripada populasi umum.

Sementara sistem dilatih untuk mempertimbangkan risiko yang mendasari dalam pengambilan keputusannya, sistem tersebut gagal mengenali bahwa pasien dengan riwayat asma, dan dianggap sebagai pasien pneumonia berisiko tinggi, menerima tingkat perawatan yang lebih tinggi, sehingga menghasilkan hasil yang lebih baik. Ketidakmampuan AI untuk mengenali sebab dan akibat dengan tepat dapat membuat sistem tidak andal, dan karenanya, tidak aman.

Pertimbangan Etis: Bias

Keselamatan pasien juga dapat terganggu oleh bias sistemik yang terwujud dalam pengambilan keputusan medis AI. Meskipun non-manusia, AI dapat mengekspresikan bias subjektif sebagai akibat dari algoritma dan data yang dihasilkan manusia yang digunakan untuk mengembangkan AI. Bias algoritmik menggambarkan bias sistemik dalam keputusan AI yang mencerminkan bias manusiawi dari pemrogram AI.

Bias algoritmik dapat muncul melalui "data yang dipilih oleh pembuat algoritma, serta metode pencampuran data, praktik konstruksi model, dan bagaimana hasil diterapkan dan diinterpretasikan." Chauhan dan Gullapalli menjelaskan bagaimana algoritma AI dengan

"variabel yang dapat disetel" mengharuskan peneliti untuk membuat pilihan sadar yang menghadirkan titik masuk bagi bias yang dapat memiliki "efek berjenjang di hilir."

Sumber bias AI lainnya berasal dari data yang digunakan untuk melatih AI, yang utamanya terdiri dari data dari rekam medis dan tagihan elektronik. Nelson menggambarkan data yang digunakan untuk melatih AI sebagai, "data yang kita miliki, bukan data yang 'benar.'" Pertama, karena data dalam rekam medis elektronik tidak dihasilkan untuk fungsi analitik spesifik teknologi AI modern, melainkan untuk perawatan dan penagihan medis, data tersebut mencerminkan bias sistemik termasuk bias rasial, gender, geografis, dan ekonomi yang beroperasi untuk semakin merugikan populasi yang kurang terwakili.

Misalnya, perempuan kulit hitam secara historis telah menjadi, dan terus menjadi, korban rasisme obstetrik dan menjadi sasaran prosedur medis yang tidak perlu. Ketika data yang digunakan untuk melatih AI mengandung bias, AI akan menghasilkan keputusan medis yang bias tanpa adanya pengukuran yang memadai yang dirancang untuk mengidentifikasi dan menghilangkan bias tersebut. Bias yang dominan dalam data kesehatan yang digunakan untuk melatih AI adalah data tersebut berasal dari populasi yang memiliki akses ke layanan kesehatan dan biasanya tidak mewakili minoritas dan subpopulasi terpinggirkan lainnya.

Masalah ini semakin diperparah ketika teknologi perangkat yang dapat dikenakan (wearable) menggunakan data pelatihan AI. Akibatnya, data yang digunakan untuk melatih AI mengalami ketidakseimbangan kategori dan spesifikasi yang kurang, yang keduanya dapat membuat keputusan AI tidak andal dan tidak aman ketika diterapkan pada anggota populasi minoritas atau yang kurang terwakili.

Pertimbangan Hukum: Privasi Data

Meskipun sebagian besar data yang digunakan untuk melatih AI berasal dari rekam medis elektronik, sumber data pelatihan lainnya meliputi rekam medis elektronik, data pendapatan, catatan kriminal, dan media sosial. Akses pihak ketiga ke data tersebut dapat berdampak jauh melampaui kerugian yang terkait dengan pelanggaran privasi awal hingga memengaruhi keputusan ketenagakerjaan dan kredit serta akses dan tarif asuransi.

Dampak negatif tambahan dari akses pihak ketiga terhadap data kesehatan dapat mencakup stigma sosial dan kerugian psikologis. Hoffman mencatat bahwa prediksi AI mengenai kondisi medis di masa depan, termasuk penurunan kognitif, penyalahgunaan zat, dan bahkan bunuh diri, dapat menyebabkan diskriminasi dan kerugian psikologis bagi individu yang tidak ditawarkan konseling untuk mengelola dampak temuan tersebut. Meskipun GDPR menawarkan perlindungan yang lebih besar kepada individu di Uni Eropa dibandingkan HIPAA kepada warga Amerika, kemampuan AI lintas batas memerlukan regulasi internasional untuk melindungi data pribadi yang digunakan oleh pengembang AI.

Pertimbangan Hukum: Tanggung Jawab

Meskipun AI saat ini mampu melakukan tugas-tugas secara independen, seperti diagnosis, yang secara langsung berada dalam praktik kedokteran, kerangka hukum yang jelas untuk secara langsung menangani tanggung jawab hukum atas cedera pasien yang disebabkan oleh pengambilan keputusan medis berbasis AI belum ada. Pengenalan AI ke dalam pengambilan keputusan klinis mendobrak anggapan tradisional tentang kelalaian dengan

mengajukan pertanyaan seperti: Bisakah komputer bersikap tidak masuk akal?. Komisi Eropa (EC) baru-baru ini memperkenalkan Arahan yang diusulkan untuk mengatur tanggung jawab atas kerugian yang disebabkan oleh AI secara umum (Komisi Eropa 2022a (PLD); Komisi Eropa 2022b (AILD)). Namun, Arahan yang diusulkan ini masih belum memberikan kerangka pertanggungjawaban yang jelas untuk teknologi medis seperti Dokter AI Otonom dalam kasus-kasus di mana pengambilan keputusan medis algoritmik AI dirancang agar tidak dapat diprediksi dan tidak transparan serta tidak dapat dikaitkan secara memadai dengan (1) cacat dalam pembuatan AI atau (2) kesalahan manusia (atau kelalaian) sebagaimana dinilai berdasarkan hukum yang berlaku.

Saat ini, terdapat beberapa kerangka hukum yang memungkinkan pengadilan menetapkan pertanggungjawaban atas cedera yang disebabkan oleh dokter AI otonom, termasuk pertanggungjawaban ketat, pertanggungjawaban perusahaan, pertanggungjawaban perwakilan, kelalaian, dan pertanggungjawaban tanpa kesalahan. Beberapa ahli hukum mempertanyakan sejauh mana kerugian yang ditimbulkan oleh AI dapat dikompensasi oleh mesin, yang tidak memiliki aset keuangan. Chung, yang mengadvokasi status badan hukum AI, berpendapat bahwa risiko dapat dinilai dan diasuransikan untuk memberikan kompensasi kepada pasien yang cedera dalam kerangka malapraktik medis dan pertanggungjawaban produk yang ada. Hingga batas tertentu, rezim berbasis kelalaian dapat mencegah perilaku buruk dengan hanya menghukum tindakan yang ditemukan berada di bawah standar perawatan medis yang dapat diterima.

Di sisi lain, rezim pertanggungjawaban yang ketat atau tanpa kesalahan dapat memaksa industri yang bertanggung jawab atas penciptaan dan penggunaan AI dalam layanan kesehatan untuk menanggung risiko cedera yang disebabkan oleh teknologi tersebut. Menerapkan pertanggungjawaban perwakilan atau hukum kelalaian korporat dapat mengalihkan pertanggungjawaban kepada institusi yang "mempekerjakan" AI dan beroperasi berdasarkan sebab-sebab tindakan seperti perekrutan yang lalai atau pemberian kredensial yang lalai.

Jawaban atas pertanggungjawaban hukum untuk AI otonom kemungkinan terletak pada kombinasi beberapa pendekatan hukum yang ada, tergantung pada penyebab cederanya. Ketidakpastian seputar situasi pertanggungjawaban hukum untuk AI otonom dalam layanan kesehatan kemungkinan akan menghambat adopsi teknologi AI yang sedang berkembang, yang jika diatur secara memadai, dapat meningkatkan perawatan pasien. Akibatnya, para ahli hukum di AS dan Eropa mendesak para pembuat undang-undang untuk memberikan kejelasan mengenai tanggung jawab atas cedera medis yang disebabkan oleh AI.

11.5 MENGATUR DOKTER AI OTONOM

Pengaturan yang tepat terhadap Dokter AI Otonom melalui kombinasi cermat antara aturan dan regulasi pemerintah, industri, dan hukum harus mengatasi masalah etika dan hukum yang diidentifikasi untuk mendorong keberhasilan integrasi AI otonom yang aman, andal, dan adil dalam dunia kedokteran. Untuk memandu pengembangan AI dalam dunia kedokteran, semua pemangku kepentingan harus berkolaborasi untuk mengembangkan

norma etika yang mengatur penciptaan, implementasi, dan pemeliharaan AI serta mekanisme hukum dan regulasi untuk memastikan akuntabilitas atas pelanggaran etika dan tanggung jawab atas cedera yang disebabkan oleh Dokter AI Otonom.

Panduan dan regulasi industri dan pemerintah yang kuat yang bertujuan untuk memberikan transparansi, keamanan, keandalan, keadilan, dan privasi merupakan garis pertahanan pertama bagi pasien Dokter AI Otonom. Namun, tidak dapat dihindari bahwa pasien akan mengalami kerugian karena integrasi AI otonom ke dalam layanan kesehatan.

Ketika kerugian tersebut nyata, sistem hukum harus memastikan akuntabilitas dan kompensasi bagi pasien yang terluka. Karena Dokter AI Otonom adalah algoritma sekaligus dokter, ia harus diatur di bawah rezim yang memastikan pengembangan dan implementasi perangkat lunak yang etis dan membuatnya bertanggung jawab sebagai pembuat keputusan otonom yang belajar mandiri.

Karena AI otonom "dididik" dan "dilatih" oleh pengembang dan insinyur perangkat lunak yang menulis algoritma, badan pengatur yang menguji, menyetujui, dan mengawasi, kualitas perangkat lunak AI dan proses pengembangannya mirip dengan dewan medis yang menguji, memberi lisensi, dan mengawasi praktik dokter. Di sisi lain, rezim pertanggungjawaban yang mengatur kerugian yang disebabkan oleh produk umumnya tidak cocok untuk mencakup pertanggungjawaban atas kerugian yang disebabkan oleh Dokter AI Otonom. Sebaliknya, kombinasi kelalaian medis, kelalaian organisasi, liabilitas perwakilan, dan liabilitas perusahaan lebih mampu menangani kerugian pasien yang disebabkan oleh keputusan AI otonom, dengan liabilitas produk yang mengatur sebagian kecil kasus yang melibatkan kerugian yang disebabkan oleh desain dan komponen fisik AI.

Regulasi Industri Layanan Kesehatan

Para pakar medis dapat membantu memastikan pengembangan AI otonom yang etis dan aman dalam layanan kesehatan. Beberapa organisasi profesional dan regulator medis telah mulai mengatasi tantangan ini. Di AS, Asosiasi Medis Amerika (AMA) berupaya untuk "mendorong pengembangan AI layanan kesehatan yang dirancang dengan cermat, berkualitas tinggi, dan tervalidasi secara klinis," yang mencakup kesesuaian AI dengan praktik terbaik, transparansi, reproduktifitas, keadilan, privasi, dan keamanan.

Di Inggris, Layanan Kesehatan Nasional (*National Health Service*/NHS) berupaya mencegah kerugian yang tidak diinginkan yang disebabkan oleh teknologi berbasis data dalam layanan kesehatan, termasuk AI, dengan menyediakan kerangka kerja bagi pengembang AI yang membahas, "isu-isu seperti transparansi, akuntabilitas, keamanan, efikasi, penjelasan, keadilan, kesetaraan, dan bias (Departemen Kesehatan dan Layanan Sosial dan NHS 2021)." *Royal Australian and New Zealand College of Radiologists* (RANZCR) menyusun Standar Praktik AI untuk memandu pengembangan, regulasi, dan integrasi AI ke dalam praktik radiologi berdasarkan prinsip-prinsip etika yang serupa.

Digital Pathology Association telah membentuk gugus tugas AI/ML yang bertujuan membantu pengembangan kecerdasan buatan dan pembelajaran mesin dalam patologi dengan menyediakan informasi dan sumber daya kepada para anggotanya mengenai "wawasan regulasi, praktik terbaik, aktivitas ilmiah, hubungan vendor, dan etika."

Di tingkat penyedia layanan kesehatan, organisasi layanan kesehatan dapat menerapkan beberapa praktik untuk membantu mencapai AI yang aman, etis, dan akuntabel seperti yang dicita-citakan oleh perkumpulan profesional ini. Pertama, organisasi harus membentuk dewan peninjau kelembagaan (IRB) untuk menilai nilai ilmiah, validitas, dan reliabilitas AI medis, risiko terhadap kesehatan, otonomi, dan privasi pasien, serta keadilan dan akuntabilitas yang terkait dengan penggunaan AI untuk memberikan perawatan pasien. Kedua, organisasi harus mengadopsi kebijakan, prosedur, dan protokol yang secara jelas menggambarkan tingkat tanggung jawab untuk memastikan implementasi AI mencerminkan nilai-nilai yang mendorong penilaian IRB.

Protokol-protokol ini harus terus ditinjau dan diperbarui untuk "mencerminkan kondisi pengetahuan terkini dalam praktik pelayanan kesehatan." Salah satu cara praktis untuk memasukkan nilai-nilai etika seputar AI medis adalah dengan menuangkannya ke dalam kontrak transparan dengan pengembang dan vendor AI, yang dapat mencakup ketentuan mengenai kualitas data, privasi, dan berbagi data serta mekanisme pengawasan dan audit kinerja AI.

Rekomendasi praktis lainnya adalah pembentukan Konselor Informasi Kesehatan (HIC) yang berhadapan langsung dengan pasien untuk memberikan informasi kepada pasien mengenai penggunaan AI dalam pelayanan kesehatan mereka, termasuk kinerja, risiko, manfaat, dan biaya AI. HIC akan menjadi kelas baru profesional kesehatan interdisipliner yang terlatih untuk memahami kapasitas teknologi, analitis, dan medis AI otonom serta dampak klinis dan finansial dari perawatan pasien individu.

Jorstad memperingatkan bahwa meskipun HIC "mungkin terbukti menjadi sumber daya yang sangat berharga sebagai perantara antara pasien dan profesional medis," berdasarkan standar pertanggungjawaban saat ini, dokter tetap harus memahami dan menjelaskan risiko AI medis yang diperlukan untuk mendapatkan persetujuan berdasarkan informasi. Oleh karena itu, implementasi praktis HIC akan "memerlukan perubahan struktural dan kebijakan yang lebih luas." Jika industri kesehatan mengambil peran proaktif dalam menerapkan AI yang etis dalam perawatan pasien, industri ini juga dapat memandu pengembangan rezim regulasi AI untuk mencegah misregulasi, yang dapat menjadi penghalang bagi kelanjutan penerapan teknologi AI di masa mendatang dalam perawatan kesehatan, termasuk Dokter AI Otonom.

Regulasi Pemerintah

Regulasi pemerintah terhadap AI secara garis besar terbagi dalam dua bidang: keselamatan dan privasi serta keamanan data.

Regulasi Keamanan

Regulasi pemerintah terhadap Dokter AI Otonom untuk memastikan keamanannya merupakan upaya yang kompleks. Kemampuan belajar mandiri AI otonom yang menjadikannya aset berharga bagi pemberian layanan kesehatan juga merupakan fitur yang membuatnya sulit untuk diatur. Sistem regulasi pemerintah saat ini dirancang untuk perangkat dan produk medis statis, bukan untuk Dokter AI Otonom pembelajaran mendalam yang terus berubah.

Selain itu, seperti yang ditunjukkan Jorstad, AI tidak memiliki manfaat historis untuk membuktikan dirinya melalui tinjauan sejawat, penelitian ilmiah, dan uji klinis selama beberapa dekade, yang mendasari regulasi pemerintah tradisional di sektor kesehatan. Sebaliknya, "pembelajaran mendalam telah membalikkan proses ilmiah," seperti yang dijelaskan Ahmad dkk., dengan menggunakan data untuk menghasilkan, alih-alih membuktikan, hipotesis. Akibatnya, badan pengatur perlu mengembangkan pendekatan yang lebih dinamis terhadap otorisasi pra-pasar dan pemantauan pasca-pasar guna memastikan adopsi AI otonom yang etis dan bertanggung jawab di industri layanan kesehatan. Lawry, dkk. mengusulkan rezim pengaturan yang mencakup: "evaluasi sistematis terhadap kualitas dan kesesuaian data dan model yang digunakan untuk melatih sistem berbasis AI; penjelasan yang memadai tentang operasi sistem termasuk pengungkapan potensi keterbatasan atau kekurangan dalam data pelatihan; keterlibatan spesialis medis dalam proses desain dan operasi; evaluasi peran masukan dan kendali tenaga medis profesional dalam penerapan sistem; dan mekanisme umpan balik yang kuat dari pengguna kepada pengembang."

Meregulasi AI "target bergerak" yang melibatkan algoritma pembelajaran mandiri yang terus berubah setelah ditempatkan di pasar layanan kesehatan memerlukan pendekatan yang berfokus pada kualitas proses pengembangan pra-pasar dan pemantauan kinerja berkelanjutan pasca-pasar. Di AS, Badan Pengawas Obat dan Makanan (FDA) telah memperkenalkan program sertifikasi percontohan untuk menyederhanakan persetujuan perangkat lunak sebagai alat kesehatan (SaMD). Program percontohan ini menggunakan pendekatan "Siklus Hidup Produk Total", yang terdiri dari evaluasi pra-pasar perusahaan yang mengembangkan AI serta pengawasan kinerja produk pasca-pasar berkelanjutan dari SaMD.

Dalam program ini, sebuah perusahaan dapat mencapai "status pra-sertifikasi" jika dapat, "membangun kepercayaan bahwa mereka memiliki budaya kualitas dan keunggulan organisasi sehingga mereka dapat mengembangkan produk SaMD berkualitas tinggi, memanfaatkan transparansi keunggulan organisasi dan kinerja produk di seluruh siklus hidup SaMD, memanfaatkan tinjauan pra-pasar yang efisien dan disesuaikan, dan memanfaatkan peluang pasca-pasar unik yang tersedia dalam perangkat lunak untuk memverifikasi keamanan, efektivitas, dan kinerja SaMD yang berkelanjutan di dunia nyata."

Di Eropa, UU AI13 yang diusulkan Komisi Eropa berupaya menyediakan tata kelola AI yang seragam untuk memastikan, “perlindungan tingkat tinggi terhadap kesehatan, keselamatan, dan hak-hak dasar,” dan “pergerakan bebas barang dan jasa berbasis AI lintas batas.” Proposal tersebut mengklasifikasikan AI yang digunakan dalam perawatan kesehatan sebagai perangkat medis berisiko tinggi yang harus mematuhi peraturan yang ada, misalnya Peraturan Perangkat Medis (MDR) dan Peraturan tentang perangkat medis diagnostik in vitro (IVDR), serta persyaratan khusus AI yang terdapat dalam proposal. IVDR mengendalikan proses sertifikasi untuk AI dalam patologi dan sudah memerlukan penilaian pengembangan teknis, kinerja, dan rencana pengawasan pasca-pemasaran.

Proposal baru tersebut memberlakukan “persyaratan tambahan berupa data berkualitas tinggi, dokumentasi dan ketertelusuran, transparansi, pengawasan manusia, akurasi, dan ketahanan.” Meskipun Undang-Undang AI yang diusulkan dirancang untuk

membangun kepercayaan publik terhadap teknologi, beberapa pakar memandang usulan baru ini sebagai regulasi yang berlebihan yang akan membutuhkan sertifikasi ganda berdasarkan berbagai peraturan Uni Eropa dan menghambat inovasi di pasar. Memang, mencapai keseimbangan yang rumit antara melindungi pasien dan mendorong inovasi sangat penting bagi keberhasilan pengembangan dan implementasi Dokter AI Otonom.

Regulasi Data

Regulator juga harus berupaya melindungi data pribadi yang digunakan untuk mengembangkan dan melatih AI. Kerangka kerja untuk mengatur data kesehatan di AS tidak memadai untuk mengatasi masalah etika dan hukum terkait privasi dan keamanan data yang diajukan oleh dokter AI Otonom. Di sisi lain, Eropa telah mengambil pendekatan yang lebih proaktif untuk mengatur big data, termasuk melindungi data kesehatan warga negara Uni Eropa dari eksploitasi oleh industri teknologi.

Para ahli hukum Amerika telah menyoroti ketidakmampuan HIPAA untuk melindungi data kesehatan individu secara memadai di Amerika Serikat. Kegagalan HIPAA untuk mengatur pembagian data oleh entitas selain penyedia layanan kesehatan dan perusahaan asuransi merupakan kelemahan hukum yang paling mencolok dalam hal privasi data. Misalnya, perusahaan teknologi bebas membagikan data kesehatan individu untuk tujuan penelitian atau komersial karena mereka tidak dianggap sebagai "entitas yang dilindungi" menurut hukum. HIPAA juga gagal mengatur data kesehatan yang dihasilkan pengguna atau data yang dapat digunakan untuk membuat kesimpulan tentang kesehatan, sehingga unggahan media sosial mengenai kondisi kesehatan atau data pembelian internet dapat diambil alih oleh perusahaan teknologi untuk penelitian dan pengembangan AI medis.

Terakhir, data yang dide-identifikasi yang seharusnya dilindungi berdasarkan aturan privasi HIPAA dapat dibagikan oleh entitas yang tercakup untuk tujuan penelitian dan komersial. Namun, de-identifikasi mungkin tidak cukup untuk melindungi privasi pasien ketika data dapat diidentifikasi ulang melalui referensi silang ke basis data lain yang tersedia. Meskipun negara bagian bebas untuk memberlakukan perlindungan privasi yang lebih ketat daripada yang diwajibkan HIPAA untuk informasi kesehatan yang dipersonalisasi, kegagalan untuk memberlakukan kerangka kerja perlindungan data yang komprehensif di tingkat federal dapat menghambat pengembangan teknologi kesehatan AI yang inovatif sekaligus membahayakan hak privasi individu. Beberapa pakar hukum menyerukan perluasan HIPAA dan Undang-Undang Penyandang Disabilitas Amerika (*Americans with Disabilities Act*) untuk melindungi data dan mencegah diskriminasi berdasarkan kondisi kesehatan di masa mendatang.

GDPR di Eropa menawarkan tingkat perlindungan yang lebih tinggi untuk data pribadi yang berkaitan dengan subjek data Uni Eropa. Larangan umum dalam peraturan tersebut mengenai pembagian data genetik, data biometrik, dan data kesehatan berlaku untuk semua entitas yang menangani data pribadi, termasuk perorangan dan badan usaha (Parlemen Eropa dan Dewan Uni Eropa 2016, Pasal 4). GDPR juga melarang pemrosesan data untuk "pengambilan keputusan individu secara otomatis", yang dapat menimbulkan konsekuensi hukum atau konsekuensi signifikan lainnya terhadap subjek data, tanpa adanya kebutuhan

untuk menandatangani kontrak hukum, otorisasi oleh negara anggota, dan langkah-langkah untuk melindungi kebebasan individu dan kepentingan privasi, atau persetujuan eksplisit (Parlemen Eropa dan Dewan Uni Eropa 2016, Pasal 22).

Terakhir, penilaian dampak yang diwajibkan GDPR, termasuk penilaian risiko dan upaya mitigasi risiko serta perlindungan data yang diantisipasi, berlaku untuk pengenalan teknologi berbasis AI baru dalam pengaturan kesehatan klinis. Meskipun GDPR menawarkan perlindungan yang lebih besar kepada individu di Uni Eropa dibandingkan HIPAA bagi warga Amerika, kapabilitas AI lintas batas memerlukan regulasi internasional untuk melindungi data pribadi yang digunakan oleh pengembang AI.

Tanggung Jawab atas Cedera

Rezim hukum yang berlaku saat ini untuk tanggung jawab medis tidak dirancang untuk Dokter AI Otonom yang mengekspresikan kualitas perangkat lunak dan manusianya. Teori tanggung jawab atas cedera medis saat ini bersifat "berpusat pada manusia" atau "berpusat pada mesin", dan gagal menyediakan kerangka kerja yang efektif untuk tanggung jawab entitas hibrida. Namun demikian, kami sependapat dengan Griffin bahwa, "kerangka hukum saat ini kemungkinan akan memberikan dasar analisis tanggung jawab sistem AI dengan beberapa perubahan khusus untuk AI."

Mengidentifikasi modifikasi kerangka hukum yang tepat sebelum "gugatan malpraktik medis yang melibatkan kesalahan diagnosis AI diajukan ke pengadilan" sangat penting untuk mencegah pengadilan melarang Dokter AI Otonom atau menciptakan "pembatasan yang signifikan sehingga fungsionalitas AI menjadi lebih sulit untuk diterapkan daripada yang seharusnya." Hingga saat ini, belum ada pengadilan yang secara langsung menangani pertanggungjawaban atas cedera yang disebabkan oleh AI medis otonom.

Pertanggungjawaban atas kerugian yang disebabkan oleh Dokter AI Otonom kemungkinan akan didistribusikan di antara produsen dan pengembang AI, penyedia layanan kesehatan individu, dan organisasi layanan kesehatan. Jorstad memperkirakan bahwa organisasi layanan kesehatan terutama akan menanggung biaya cedera yang disebabkan oleh penggunaan Dokter AI Otonom.

Maliha dkk., meyakini bahwa berdasarkan skema pertanggungjawaban yang berlaku saat ini di AS, dokter yang mengandalkan pengambilan keputusan AI akan menjadi target utama, tetapi mempertanyakan apakah adil untuk meminta pertanggungjawaban penyedia layanan kesehatan atas keputusan AI otonom yang tidak dapat diprediksi yang dibuat menggunakan algoritma pembelajaran mendalam "kotak hitam". Tentu saja, fitur pembelajaran mandiri yang berkelanjutan dari Dokter AI Otonom inilah yang membuatnya berharga dalam praktik klinis.

Pada akhirnya, pertanyaan tentang penetapan tanggung jawab dijawab dengan bertanya: siapa yang memiliki kendali atas fungsi-fungsi tertentu dari Dokter AI Otonom yang menyebabkan cedera pada pasien? Kendali dapat terwujud dalam beberapa cara. Pertama, pengembang dan produsen AI memiliki kendali atas komponen fisik Dokter AI Otonom serta kendali atas "pendidikan dan pelatihannya" melalui pengembangan algoritmik AI.

Kedua, organisasi layanan kesehatan menunjukkan kendali atas "perekrutan" dan pengawasan organisasi melalui pemilihan dan implementasi AI dalam praktik klinis. Ketiga, masing-masing penyedia layanan kesehatan memiliki kendali terbatas atas rekomendasi AI untuk tindakan klinis melalui pengawasan manusia dan kendali mutu. Hal ini, tentu saja, menyisakan celah kendali atas pengambilan keputusan medis independen oleh Dokter AI Otonom, yang dapat menjadi buram dan terhalang oleh masalah "kotak hitam".

Jorstad berpendapat bahwa mengingat "kontrol yang terbatas hingga tidak ada yang diberikan oleh dokter, rumah sakit, atau bahkan produsen AI atas diagnosis mesin, mungkin tidak masuk akal untuk meminta pertanggungjawaban mereka ketika kesalahan muncul." Schweikart setuju bahwa pengambilan keputusan AI "kotak hitam" membuat hampir mustahil untuk menetapkan tanggung jawab secara adil berdasarkan hukum perdata.

Kesimpulan logisnya adalah bahwa Dokter AI Otonom itu sendiri mengendalikan keputusannya sendiri, tetapi hal ini menimbulkan masalah dalam kerangka pertanggungjawaban saat ini karena AI tidak memiliki status badan hukum dan oleh karena itu tidak dapat ditetapkan tanggung jawabnya. Chung dan Zink memecahkan masalah ini dengan menyarankan pembentukan badan hukum terbatas untuk AI medis, yang akan memungkinkan Dokter AI Otonom untuk bertanggung jawab secara hukum atas kerugian yang disebabkan oleh keputusan medis independennya.

Setelah AI Otonom diberi badan hukum terbatas, dan risikonya dapat diasuransikan sebagai penyedia layanan kesehatan individu, sistem pertanggungjawaban medis yang ada dapat secara efektif memberikan kompensasi kepada pasien atas cedera yang disebabkan oleh AI di bawah paradigma kendali yang diuraikan di atas dengan menggunakan kombinasi pertanggungjawaban produk, pertanggungjawaban organisasi, pertanggungjawaban perwakilan, pertanggungjawaban perusahaan, dan pertanggungjawaban malapraktik medis.

Selain itu, entitas yang berpotensi bertanggung jawab dapat memilih untuk mengalokasikan pertanggungjawaban di antara mereka sendiri melalui perjanjian kontrak. Akhirnya, di negara-negara yang memilih rezim pertanggungjawaban tanpa kesalahan, sistem ajudikasi khusus dapat memberikan kompensasi kepada pasien atas cedera yang disebabkan oleh AI; namun, untuk rezim berbasis kelalaian, perubahan struktural yang luas seperti itu mungkin tidak layak.

Tanggung Jawab Produk

Tanggung jawab produk berfungsi untuk meminta pertanggungjawaban produsen atas produk yang secara inheren berbahaya dengan menerapkan standar pertanggungjawaban yang ketat atas cedera yang disebabkan oleh produk cacat dan kegagalan untuk memperingatkan konsumen akan hal yang sama. Tanggung jawab produk atas kerusakan yang disebabkan oleh Dokter AI Otonom sulit dibuktikan tanpa bukti adanya elemen desain yang digerakkan oleh manusia. Meskipun benar bahwa produsen berada pada posisi terbaik untuk menjelaskan teknologi "kotak hitam" AI otonom, keputusan AI tidak selalu dapat dilacak secara logis dan umumnya tidak dapat diprediksi, bahkan oleh penciptanya.

Akibatnya, akan sulit bagi pasien untuk membuktikan bahwa AI tersebut cacat dan ketersediaan serta kelayakan desain alternatif sebagaimana disyaratkan dalam dasar gugatan

pertanggungjawaban produk. Selain itu, doktrin perantara yang dipelajari menempatkan penyedia layanan kesehatan, alih-alih produsen, sebagai pihak yang bertanggung jawab untuk memberi tahu pasien tentang risiko yang diungkapkan kepada penyedia layanan kesehatan. Jorstad mencatat bahwa bahkan meminta pertanggungjawaban penyedia layanan kesehatan atas kegagalan mengungkapkan risiko tak terduga yang terkait dengan pengambilan keputusan medis AI otonom pun "sulit untuk dirasionalisasi."

Terakhir, penerapan rezim tanggung jawab produk yang ketat untuk Dokter AI Otonom dapat menghambat pengembangan teknologi AI yang bermanfaat. Pemberlakuan peraturan yang mengikat terkait pengembangan pra-pemasaran dan pemantauan pasca-pemasaran AI seharusnya memberikan kekebalan terbatas dari tanggung jawab bagi produsen yang menerima otorisasi yang sesuai. Namun, produsen AI harus bertanggung jawab secara tegas atas cacat yang berkaitan dengan input data, kode perangkat lunak asli, tampilan keluaran, atau kegagalan mekanis.

Tanggung Jawab Organisasi, Vikarius, dan Perusahaan

Tanggung jawab organisasi dapat mencakup tanggung jawab langsung bagi penyedia layanan kesehatan karena gagal melakukan kehati-hatian dalam memilih dan mempertahankan dokter yang kompeten, memelihara fasilitas dan peralatan yang memadai, melatih dan mengawasi karyawan, serta menerapkan protokol dan prosedur yang tepat. Tugas-tugas organisasi ini dapat memerlukan pemeriksaan menyeluruh terhadap kemampuan Dokter AI Otonom sebelum menggunakannya dalam praktik klinis.

Setelah diimplementasikan, organisasi juga dapat bertanggung jawab karena gagal memantau kualitas AI secara berkelanjutan dan melatih AI sesuai kebutuhan agar tetap mutakhir. Maliha, dkk. merekomendasikan pelaksanaan "uji stres" untuk menguji kemampuan AI dalam menghasilkan keputusan yang andal dan akurat dalam menanggapi situasi sulit yang tidak dipertimbangkan oleh pengembang AI. Selain itu, organisasi harus diwajibkan untuk memanfaatkan peluang pembelajaran yang kaya yang disediakan oleh kesalahan nyaris fatal yang dialami Dokter AI Otonom kesalahan yang tidak menyebabkan kerusakan untuk melatih ulang dan memperbarui Dokter AI Otonom guna mencegah pengulangan kesalahan.

Organisasi layanan kesehatan juga dapat dimintai pertanggungjawaban atas kelalaian karyawannya berdasarkan doktrin tanggung jawab perwakilan. Akibatnya, jika dokter AI Otonom dianggap sebagai agen atau karyawan organisasi layanan kesehatan, kerugian yang disebabkan oleh pengambilan keputusan AI Otonom dapat ditanggung oleh organisasi tersebut.

Pertanggungan semacam itu akan beroperasi seperti tanggung jawab perwakilan rumah sakit terhadap perawat dan dokter stafnya. Tentu saja, organisasi layanan kesehatan harus memiliki pertanggungan asuransi yang memadai untuk tindakan dokter AI Otonom. Tanggung jawab perusahaan dapat membuat semua entitas yang terlibat "dalam mengejar tujuan bersama" bertanggung jawab secara tanggung renteng atas kerugian yang disebabkan oleh tanggung jawab bersama tersebut.

Pengaturan ini dapat memungkinkan pembagian biaya antara pengembang AI dan penyedia layanan kesehatan serta organisasi yang menerapkan AI dalam praktik klinis. Allen

yakin bahwa tanggung jawab perusahaan merupakan pilihan yang kuat untuk menyebarkan risiko yang terkait dengan "kerugian yang tak terelakkan dan terukur" yang akan terjadi akibat pengambilan keputusan medis AI yang otonom.

Malpraktik Medis

Penetapan status badan hukum terbatas diperlukan untuk meminta pertanggungjawaban Dokter AI Otonom atas malapraktik medis. Chung menekankan bahwa status badan hukum untuk AI merupakan fiksi hukum yang harus dibedakan dari pemahaman sehari-hari tentang apa artinya menjadi manusia.

Memberikan hak dan tanggung jawab hukum kepada non-manusia bukanlah konsep baru. Sebagaimana ditunjukkan Schweikart, baik kapal maupun korporasi ditetapkan status badan hukum. Chung dan Zink berpendapat bahwa AI IBM sebelumnya, Watson, dapat diberikan status badan hukum terbatas mengingat kemampuannya untuk bekerja sebagai anggota integral tim perawatan pasien yang mampu memberikan interpretasi dan analisis individual atas kondisi medis pasien dan memberikan rekomendasi perawatan.

Mereka membandingkan Watson dengan mahasiswa kedokteran dengan pendidikan dan pelatihan khusus, yang mampu membuat keputusan medis independen tetapi membutuhkan tingkat supervisi dan pengawasan. Berdasarkan perbandingan ini, kerangka kerja untuk mengasuransikan risiko dan mengevaluasi tanggung jawab atas kerugian yang disebabkan oleh AI medis sudah tersedia, sehingga menghilangkan kebutuhan untuk membangun sistem asuransi dan tanggung jawab baru, sebuah upaya yang tidak mungkin dilakukan. Chung dan Zink lebih lanjut menunjukkan bahwa status badan hukum AI yang terbatas cukup fleksibel untuk mencakup AI yang lebih cerdas dan lebih mandiri di masa depan.

Tentu saja, mengizinkan Dokter AI Otonom untuk bertanggung jawab atas pengambilan keputusan medisnya sendiri di bawah rezim malapraktik medis saat ini memerlukan beberapa diskusi tentang standar perawatan yang berlaku. Rezim tanggung jawab tersebut telah menerapkan standar perawatan yang lebih tinggi kepada spesialis dengan pelatihan ekstensif di bidang medis tertentu.

Dokter AI Otonom telah mampu melampaui kemampuan manusia untuk meninjau dan memproses data besar dan, dalam beberapa kasus, bahkan melampaui kemampuan diagnostik dokter manusia. Di sisi lain, AI tidak memiliki kemampuan untuk memeriksa pasien secara fisik dengan indera manusia, mensintesis informasi lintas berbagai sektor pengetahuan, meresepkan obat, atau memerintahkan tes. Akibatnya, Dokter AI Otonom perlu dianggap sebagai spesialis medis unik yang membutuhkan standar perawatan yang sesuai dan unik.

Jorstad menyediakan beberapa opsi untuk menentukan kapan Dokter AI Otonom melanggar standar perawatan yang berlaku. Opsi pertama adalah menggunakan metode "tetangga terdekat", yang melibatkan pemeriksaan riwayat diagnostik AI untuk kasus-kasus yang sebanding untuk menghitung tingkat akurasi kotor AI. Analisis berbasis kasus ini seharusnya memberikan beberapa pengukuran yang dengannya dugaan kesalahan dapat dibandingkan dan dinilai berdasarkan standar kewajaran. Opsi kedua adalah "uji silang AI", yang melibatkan pengujian data dari kasus pasien yang cedera melalui algoritma AI lain untuk

menemukan apakah mesin-mesin tersebut menghasilkan hasil yang sebanding. Dua opsi tambahan melibatkan kesaksian manusia dari programmer AI atau pakar medis manusia untuk mengevaluasi keputusan AI secara independen dan memberikan opini mengenai kewajaran proses dan hasil AI.

Pada kenyataannya, pengacara kemungkinan akan mencoba beberapa kombinasi metode ini, dan sebagai hasilnya, standar perawatan akan berkembang secara organik seiring waktu karena dokter AI Otonom menjadi penggugat umum dalam kasus malapraktik medis. Sebagai alternatif, organisasi profesional dan industri dapat mencoba menetapkan standar perawatan secara proaktif dengan menyusun pedoman praktik AI. Namun, pengadilan kemungkinan akan memandang ketidakpatuhan terhadap pedoman tersebut sebagai bukti kelalaian, alih-alih sebagai penentu masalah.

Penyedia layanan kesehatan individu masih dapat bertanggung jawab berdasarkan rezim malapraktik medis saat ini atas kegagalan dalam mengawasi atau mengawasi AI Otonom dengan benar. Sebab-sebab gugatan semacam itu telah diakui dalam kaitannya dengan penyedia layanan medis bawahan. Salah satu area tanggung jawab medis manusia yang memerlukan perhatian khusus adalah persetujuan berdasarkan informasi. Meskipun Dokter AI Otonom dapat memberikan keputusan medis secara independen, hal tersebut tetap menjadi tanggung jawab penyedia layanan kesehatan manusia untuk memberi tahu pasien tentang risiko dan manfaat dari keputusan medis AI.

Meskipun, sebagaimana dibahas di atas, undang-undang dapat diubah untuk memungkinkan pendelegasian tugas ini kepada HIC, berdasarkan undang-undang saat ini, dokter harus berkonsultasi dengan pasien untuk memberikan informasi yang diperlukan guna memperoleh persetujuan berdasarkan informasi. Setidaknya, informasi ini harus mencakup pemberitahuan bahwa keputusan medis dibuat oleh Dokter AI Otonom, hak untuk mendapatkan opini kedua dari dokter manusia jika memungkinkan, dan pengungkapan kemungkinan penggunaan informasi kesehatan untuk pelatihan AI di masa mendatang.

Pengalihan Tanggung Jawab Kontraktual

Meskipun terdapat kerangka hukum untuk pengalihan tanggung jawab setelah cedera pasien, penyedia layanan kesehatan dan produsen AI masih dapat membagi atau mengalihkan tanggung jawab dan kewajiban asuransi secara kontraktual untuk Dokter AI Otonom. Jorstad berpendapat bahwa perjanjian semacam itu merupakan pilihan paling sederhana untuk membagi tanggung jawab atas cedera yang disebabkan oleh AI.

Sistem Adjudikasi Khusus

Sistem adjudikasi khusus dapat menyediakan pendekatan tanpa kesalahan untuk kompensasi atas kerugian yang disebabkan oleh dokter AI Otonom. Ini dapat mencakup kompensasi dari dana yang telah ditetapkan dan/atau arbitrase mengikat wajib untuk menentukan kerugian yang disebabkan oleh pengambilan keputusan medis AI. Manfaat sistem tanpa kesalahan meliputi adjudikasi yang lebih efisien dan peningkatan akses pemulihan bagi mereka yang terluka oleh Dokter AI Otonom. Selain itu, semua pemangku kepentingan akan menanggung biaya risiko yang ditimbulkan oleh AI dalam pemberian layanan kesehatan dengan berkontribusi pada dana bersama.

Meskipun terdapat beberapa contoh sistem tanpa kesalahan seperti kompensasi cedera akibat vaksin di AS, menggabungkan cedera medis yang disebabkan oleh AI otonom ke dalam sistem tersebut mungkin memerlukan perubahan struktural besar yang tidak dapat dikembangkan dan diimplementasikan dengan mudah atau cepat. Sistem tanpa kesalahan juga gagal memberikan manfaat untuk mencegah perilaku di bawah standar di masa perkembangan dan implementasi teknologi baru yang pesat.

11.6 REGULASI AI MEDIS OTONOM: AMAN DAN AKUNTABEL

Dokter AI Otonom telah hadir, dan akan semakin cerdas dan cepat seiring dengan perkembangan teknologi pembelajaran mesin (DL) yang pesat. Meskipun AI memiliki potensi besar untuk meningkatkan akses dan kualitas layanan kesehatan, perawatan pasien tidak dapat dan tidak boleh diserahkan sepenuhnya kepada mesin. Semua pemangku kepentingan dalam pengembangan dan penggunaan Dokter AI Otonom memiliki kewajiban untuk memastikan bahwa AI diimplementasikan dengan cara yang aman dan bertanggung jawab, termasuk melalui mekanisme regulasi dan hukum yang menyediakan tingkat keamanan, keandalan, transparansi, keadilan, dan akuntabilitas yang diperlukan. Ucapan Terima Kasih Saya ingin menyampaikan rasa terima kasih yang tulus kepada para peserta Kolokium Beasiswa Hukum NYU Musim Gugur 2021 atas komentar dan saran mereka yang mendalam pada draf awal karya ini.

BAB 12
ETIKA AI MEDIS: KEBURAMAN DAN TANGGUNG JAWAB

Abstrak Algoritma kecerdasan buatan berpotensi mendiagnosis beberapa jenis kanker kulit atau mengidentifikasi kelainan irama jantung tertentu sebaik (atau bahkan lebih baik) daripada dokter kulit dan kardiolog bersertifikat. Namun, salah satu ketakutan terbesar di sektor kesehatan di Era AI dalam Kedokteran adalah apa yang disebut pengobatan kotak hitam, mengingat ketidakjelasan dalam cara informasi diproses oleh algoritma. Secara lebih luas, diamati bahwa terdapat tiga dimensi semantik yang berbeda dari keburaman algoritmik yang relevan dengan Kedokteran: (1) keburaman epistemik karena kurangnya pemahaman dokter tentang aturan yang diterapkan sistem AI untuk membuat prediksi dan keputusan; (2) ketidakjelasan atas kurangnya pengungkapan medis mengenai sistem AI untuk mendukung keputusan klinis dan ketidaktahuan pasien bahwa pengambilan keputusan otomatis sedang dilakukan dengan data pribadi mereka; (3) ketidakjelasan penjelasan atas penjelasan yang tidak memuaskan kepada pasien mengenai teknologi yang digunakan untuk mendukung pengambilan keputusan profesional. Oleh karena itu, tujuan penelitian ini adalah untuk menganalisis setiap jenis ketidakjelasan, dengan mempertimbangkan skenario hipotetis dan dampaknya dalam hal malapraktik medis dan persetujuan pasien. Dari sini, akan didefinisikan tantangan etika penggunaan AI di sektor kesehatan dan pentingnya pendidikan kedokteran.

12.1 KEUNGGULAN KECERDASAN BUATAN (AI) DALAM KEDOKTERAN

Era Digital Kedokteran menciptakan konsep kesehatan cerdas, mengikuti fenomena transformasi dari Kedokteran tradisional menuju Kedokteran P4 (preventif, prediktif, personal, dan partisipatif). Dalam skenario baru ini, layanan kesehatan tidak lagi terbatas pada penanganan patologi (tugas yang tentu saja tidak pernah ditinggalkan) dan kini berfokus pada penerapan langkah-langkah yang bertujuan mencegah penyakit (pengobatan preventif) atau memungkinkan antisipasi diagnosis (pengobatan prediktif). Mengenai perawatan personal, pasien ditangani secara lebih individual (dan karenanya kurang generik), berdasarkan data genetik dan kesehatannya (pengobatan personal).

Akhirnya, hubungan dokter-pasien tidak lagi bersifat sesaat, melainkan mulai berkembang secara berkelanjutan, dengan partisipasi aktif pasien (pengobatan partisipatif). Dengan perangkat digital, pasien dapat mengambil peran yang lebih aktif dan partisipatif dalam pengambilan keputusan perawatan kesehatan dan kesejahteraan mereka. Dengan cara ini, pasien diabetes dapat terus memantau glukosa darahnya, memungkinkan algoritma untuk menganalisis data pribadi yang diberikan secara real-time, mendukung dokter dalam pengambilan keputusan terapeutik yang lebih cepat, lebih efisien, dan lebih personal, terkait pemberian obat atau diet.

Transformasi perawatan medis dalam model yang lebih proaktif/partisipatif, preventif, presisi, dan berfokus pada individualitas setiap pasien ini dimungkinkan oleh kombinasi data kesehatan dalam jumlah besar dan algoritma Kecerdasan Buatan. Kehidupan manusia, setelah

milenium ketiga, akan dikondisikan oleh algoritma untuk memecahkan masalah dan membuat keputusan yang lebih akurat. Eric Topol, dalam buku-bukunya tentang masa kini dan masa depan Kedokteran, merujuk pada beberapa studi ilmiah yang membuktikan kemampuan AI yang luar biasa untuk mendiagnosis beberapa jenis kanker kulit, atau mengidentifikasi kelainan irama jantung tertentu, sebaik atau bahkan mungkin lebih baik daripada dokter kulit dan ahli jantung.

Selama pandemi COVID-19, AI juga menunjukkan potensinya yang besar dalam pencitraan medis di seluruh dunia. Karena peningkatan pesat jumlah kasus baru dan kasus suspek COVID-19, sebagai alternatif untuk mengurangi tekanan pada ahli radiologi dan mencegah penyebaran penyakit lebih lanjut, algoritma berbasis AI dikembangkan di seluruh dunia yang mendukung para profesional ini dalam mengidentifikasi penyakit patogen dengan cepat dengan menganalisis citra tomografi terkomputasi dari pasien COVID-19 yang bergejala.

Selain itu, dalam beberapa tahun terakhir, algoritma prediktif telah digunakan untuk menargetkan pengobatan secara lebih efektif terhadap kelompok pasien berisiko tinggi guna mencegah komplikasi penyakit kronis yang serius. Pendekatan ini sedang diterapkan di domain relevan lainnya untuk memungkinkan identifikasi populasi berisiko dan identifikasi dini komplikasi yang akan datang dari berbagai penyakit akut dan kronis.

Saat ini, IBM merupakan salah satu perusahaan besar yang menciptakan lebih banyak solusi teknologi untuk sektor kesehatan dan mengembangkan apa yang disebut Watson untuk Onkologi, sebuah solusi yang didukung oleh informasi dari pedoman, praktik terbaik, serta jurnal dan buku teks medis yang relevan. Watson mengevaluasi informasi dari rekam medis pasien, beserta bukti medis (makalah ilmiah dan studi klinis), sehingga menunjukkan kemungkinan pilihan pengobatan bagi pasien kanker, yang diklasifikasikan berdasarkan tingkat keyakinan. Pada akhirnya, dokterlah yang akan menganalisis kesimpulan yang dicapai oleh AI dan memutuskan pilihan pengobatan terbaik untuk pasien tersebut.

Demonstrasi singkat dari contoh-contoh penerapan Kecerdasan Buatan dalam praktik medis ini bertujuan untuk mengilustrasikan beberapa manfaat yang dapat diberikan teknologi ini bagi sektor kesehatan. Namun, potensi manfaat ini juga diiringi dengan pertimbangan etika dan hukum yang relevan. AI memang membawa banyak manfaat bagi sektor kesehatan, tetapi risikonya tidak dapat diabaikan, bahkan banyak di antaranya merupakan risiko intrinsik dari teknologi itu sendiri.

Pada September 2021, sebuah laporan PBB diterbitkan yang menganalisis bagaimana perangkat AI memengaruhi hak privasi dan hak asasi manusia lainnya. Komisioner Tinggi PBB untuk Hak Asasi Manusia menyerukan moratorium sistem AI, mengingat teknologi di beberapa sektor telah menyebabkan risiko hak asasi manusia yang serius. Oleh karena itu, pengembangan perangkat AI baru perlu dihentikan sementara hingga pihak berwenang dapat menunjukkan bahwa tidak ada masalah signifikan terkait akurasi atau dampak diskriminatif, dan bahwa sistem AI tersebut mematuhi standar privasi dan perlindungan data yang kuat (Perserikatan Bangsa-Bangsa 2021).

Dalam beberapa dekade terakhir, dengan perkembangan eksponensial algoritma prediktif baru dalam praktik medis, kita juga dapat mengamati momen krisis global dalam

kredibilitas teknologi ini di bidang Kedokteran. Terdapat skenario potensi risiko AI ekspresif dalam mendukung keputusan profesional medis, dengan mempertimbangkan beberapa faktor, termasuk kekurangan dalam proses pembuatan dan validasi algoritma, tingkat kekeliruan yang relevan, ketidakpastian, dan ketidakjelasan algoritma.

Oleh karena itu, penelitian ini mengusulkan untuk menyelidiki potensi risiko penerapan AI dalam praktik klinis, serta mendefinisikan prinsip-prinsip etika yang harus diikuti selama pengembangan teknologi dan, setelah diperkenalkan di pasar, sepanjang siklus hidupnya. Dari sini, makalah ini akan berusaha menarik beberapa kesimpulan tentang masa depan algoritma Kecerdasan Buatan dalam Kedokteran dan pentingnya pendidikan kedokteran dalam kesehatan digital dan teknologi baru.

12.2 KEBURAMAN ALGORITMIK DAN DAMPAKNYA

Salah satu ketakutan terbesar di sektor kesehatan di Era kecerdasan buatan adalah apa yang disebut 'kedokteran kotak hitam', mengingat ketidakjelasan dalam cara informasi diproses oleh algoritma. Secara lebih luas, diamati bahwa terdapat tiga dimensi semantik berbeda dari keburaman algoritmik yang relevan dengan Kedokteran: (1) keburaman epistemik; (2) keburaman karena kurangnya pengungkapan medis; dan (3) keburaman penjelasan. Oleh karena itu, penting untuk menganalisis setiap jenis keburaman, dengan mempertimbangkan skenario hipotetis dan dampaknya dalam hal malapraktik medis dan persetujuan pasien.

(1) Keburaman epistemik: terdapat kompleksitas yang relevan bagi pemahaman dokter tentang bagaimana data pribadi diproses oleh algoritma, yang dapat menemukan pola dalam begitu banyak variabel sehingga menjadi sangat sulit atau bahkan mustahil bagi pikiran manusia untuk memahaminya. Faktanya, ini adalah masalah yang ada di sebagian besar sistem Kecerdasan Buatan dan disebut oleh Frank Pasquale sebagai 'masalah kotak hitam', dalam bukunya 'The Black Box Society'. Dengan demikian, opasitas epistemik terjadi ketika tidak ada pemahaman yang memadai tentang aturan yang diterapkan sistem AI untuk membuat klasifikasi, prediksi, dan keputusan. Sebagai contoh, opasitas ini dapat berasal dari kurangnya pemahaman dokter tentang proses pembelajaran mesin untuk sampai pada diagnosis atau prediksi tertentu tentang kondisi klinis pasiennya. Kurangnya transparansi juga terkait dengan masalah keandalan prediksi algoritma, dan hal ini menimbulkan kekhawatiran yang dapat dimengerti terkait implementasi teknologi ini dalam praktik medis.

Terdapat dua kasus simbolis yang menggambarkan masalah kotak hitam dan perilaku tak terduga yang muncul akibat pembelajaran mandiri AI serta ketidakandalan hasil yang dihasilkan oleh algoritma. Dalam sebuah eksperimen yang dilakukan pada tahun 2002 oleh para ilmuwan di Magna Science Center, Inggris, sebuah peristiwa tak terduga terjadi: dua robot cerdas ditempatkan di arena untuk mensimulasikan skenario 'predator' dan 'mangsa', guna melihat apakah robot-robot tersebut dapat memanfaatkan pengalaman yang diperoleh dari pembelajaran mesin untuk mengembangkan teknik berburu dan bela diri baru. Namun, Gaak, salah satu robot, yang secara tidak sengaja ditinggalkan tanpa pengawasan selama 15 menit

berhasil melarikan diri dan berperilaku tak terduga, menemukan jalan keluar melalui dinding arena dan mencapai tempat parkir, di mana ia akhirnya tertabrak mobil.

Relevan pula untuk menyebutkan insiden yang dilaporkan oleh Sameer Singh, asisten profesor di Departemen Ilmu Komputer di Universitas California (UCI), Amerika Serikat, di mana seorang mahasiswa menciptakan algoritma untuk mengkategorikan gambar anjing husky dan serigala. Awalnya, algoritma tersebut tampak mampu mengklasifikasikan kedua hewan tersebut dengan hampir sempurna. Namun, setelah berbagai analisis silang, Singh menemukan bahwa algoritma tersebut hanya mengidentifikasi serigala berdasarkan salju di latar belakang gambar dan bukan berdasarkan karakteristik hewan itu sendiri.

Tidak diragukan lagi, kerusakan dapat meningkat ke tingkat yang tak terukur jika kita mempertimbangkan risiko yang disajikan dalam dua kasus di atas dalam konteks algoritma AI di bidang Kedokteran. Sekarang, mari kita ambil contoh algoritma yang diprogram dan diuji dengan buruk, atau algoritma dengan tingkat kekeliruan yang ekspresif, dalam teknologi kognitif yang digunakan di beberapa negara untuk mendiagnosis pasien yang terinfeksi virus corona baru.

Oleh karena itu, Nicholson Price dan Roger Allan Ford menjelaskan bahwa salah satu ketakutan terbesar sektor kesehatan pada tahap kecerdasan buatan ini justru berasal dari situasi tak terduga yang muncul dari pengobatan kotak hitam, mengingat ketidakjelasan cara informasi diproses oleh algoritma. Oleh karena itu, ketika sistem algoritmik diimplementasikan dalam praktik klinis, penting bagi dokter untuk mengetahui keterbatasannya dan apa yang secara efektif diperhitungkan untuk prediksi. Memahami batasan algoritma akan membantu dokter untuk menilai keputusan dan proposal mereka dengan lebih baik, sehingga menghindari pandangan yang terlalu sederhana dan reduksionis, selain juga mencegah pasien menjadi 'sandera' keputusan otomatis yang dibuat dalam kotak hitam algoritma.

Selain itu, perlu ditegaskan bahwa AI dalam analisis diagnostik tidaklah sempurna. Seefisien apa pun sistem 'cerdas' untuk diagnosis medis dan prediksi klinis, sistem tersebut akan tetap memiliki margin ketidakakuratan yang signifikan, yang dapat menyebabkan hasil yang merugikan. Misalnya, Watson untuk Onkologi tidak 100% akurat. Terdapat margin ketidakakuratan yang signifikan sekitar 10%, menurut penelitian klinis yang dilakukan oleh tim yang terdiri dari 15 dokter di Rumah Sakit Manipal di India selama 3 tahun terhadap 1000 pasien yang didiagnosis kanker.

Dalam kasus-kasus di mana terdapat ketidaksepakatan antara AI dan dokter, dalam 63% kasus, tenaga medis profesional mengubah diagnosis mereka sendiri untuk mengikuti diagnosis yang diberikan oleh Watson. Ada poin penting dalam refleksi ini: sistem AI mengubah keputusan akhir para ahli onkologi dalam beberapa kasus. Di sisi lain, survei yang sama mengungkapkan bahwa dalam 37% kasus, dokter tidak mengubah diagnosisnya sendiri, yang bertentangan dengan hasil yang diperoleh Watson.

Dalam skenario ini, bayangkan seorang pasien didiagnosis kanker dan dokternya awalnya yakin bahwa pasien tersebut menderita jenis kanker tertentu. Namun, setelah memasukkan data klinis pasien ke dalam perangkat lunak prediktif, seperti Watson for Oncology, perangkat lunak ini memberikan hasil lain, yaitu pasien menderita jenis kanker yang

berbeda. Dari sini, muncul pertanyaan: jika dokter mengikuti atau mengabaikan hasil AI, dan terjadi kerugian pada pasien setelah diagnosis dan perawatan yang tidak tepat, haruskah tenaga medis profesional tersebut bertanggung jawab? Dengan kata lain, dapatkah dianggap sebagai malapraktik medis jika terjadi akibat yang merugikan bagi pasien yang, secara teori, dapat dihindari jika diagnosis yang diajukan oleh AI diikuti? Isu kompleks ini telah dibahas dalam beberapa makalah terbaru.

Untuk menjawab pertanyaan ini dengan tepat, beberapa konsep dasar perlu dijelaskan terlebih dahulu tentang tanggung jawab medis atas kesalahan diagnosis. Untuk keperluan analisis liabilitas dalam layanan AI, unsur utama gugatan malapraktik medis adalah pelanggaran kewajiban hukum untuk mematuhi standar perawatan profesional, yaitu 'seperangkat pedoman yang menetapkan metode perawatan yang tepat atau diperlukan untuk kondisi tertentu berdasarkan penelitian medis dan praktik profesional'.

Lebih lanjut, di sebagian besar yurisdiksi, hukum tidak mewajibkan dokter bertanggung jawab secara hukum atas semua kesalahan diagnostik. Kesalahan diagnosis atau keterlambatan diagnosis itu sendiri bukanlah bukti kelalaian medis. Tenaga profesional yang terampil dapat melakukan kesalahan diagnostik bahkan ketika menggunakan perawatan yang wajar. Ketika dokter melakukan pemeriksaan yang baik terhadap pasiennya, dengan semua data layanan kesehatan, pemeriksaan medis, dan sarana yang tersedia, namun tetap melakukan kesalahan diagnostik, tenaga profesional tersebut tidak akan bertanggung jawab. Kewajiban untuk sempurna atau akurat secara absolut tidak dapat dibebankan kepada dokter.

Namun, ketika kesalahan diagnosis tersebut parah, yang menunjukkan ketidaktahuan atau kelalaian yang tidak dapat diterima, hal tersebut dapat berujung pada liabilitas medis. Kesalahan diagnosis yang tidak dapat dimaafkan dapat disebabkan oleh beberapa hal: (a) pemeriksaan pasien yang dangkal; (b) ketidaktahuan dokter yang tidak dapat dimaafkan mengenai informasi dasar dari ilmu kedokteran; (c) tidak menggunakan sarana diagnostik tambahan yang tersedia bagi tenaga medis profesional; (d) mengabaikan gejala-gejala yang jelas yang memerlukan pemeriksaan tambahan untuk penentuan kondisi klinis yang lebih baik. Dengan demikian, kuncinya adalah menentukan apakah dokter bertindak kompeten, yang melibatkan evaluasi tentang apa yang dilakukan dan tidak dilakukan oleh tenaga medis profesional dalam mencapai diagnosis tertentu.

Ketika menganalisis masalah tanggung jawab medis atas kesalahan diagnosis dalam konteks AI, menurut pelajaran dari Nicholson Price, dokter dapat dimintai pertanggungjawaban jika ia tidak tekun dalam menggunakan teknologi tersebut. Dalam hal yang sama, Fruzsina Molnár-Gábor berpendapat bahwa jika dokter menyadari, berdasarkan keahlian medis mereka, bahwa hasil yang diberikan oleh AI tidak tepat dalam kasus tertentu, mereka tidak boleh mempertimbangkannya sebagai dasar untuk keputusan klinis mereka. Di sisi lain, kurangnya ketekunan dokter dalam membuang hasil yang diperoleh sistem AI tanpa berpikir panjang dapat menjadi kriteria pertanggungjawaban

Dengan demikian, dapat disimpulkan bahwa, untuk memverifikasi apakah seorang dokter telah bertindak lalai dalam kasus tertentu, standar perilaku profesional yang diwajibkan pada saat praktik medis harus dianalisis. Singkatnya, dokter yang menggunakan teknologi ini

akan berada dalam posisi yang sulit untuk membenarkan: (1) mengapa ia mengikuti diagnosis atau tindakan yang disarankan oleh AI atau (2) mengapa dan berdasarkan faktor apa ia menyimpang dari rekomendasi algoritmik. Tenaga medis profesional bebas memilih metode diagnosis dan usulan terapinya, tetapi ia juga bertanggung jawab atas pilihannya.

Lebih lanjut, ketika sistem algoritmik diimplementasikan dalam praktik klinis, penting bagi dokter untuk mengetahui keterbatasannya dan apa yang diperhitungkan dalam prediksi algoritma. Memahami keterbatasan teknologi ini akan membantu dokter untuk menilai keputusan dan usulan mereka dengan lebih baik, sehingga menghindari pandangan yang terlalu sederhana dan reduksionis yang didasarkan pada kotak hitam algoritma. Kurangnya pengetahuan mendalam tentang manfaat dan risiko teknologi perawatan kesehatan dapat berdampak pada hasil yang lebih buruk bagi pasien karena kurangnya pemahaman medis tentang alat mana yang memberikan nilai tambah pada praktik mereka atau cara mengintegrasikan AI dengan tepat ke dalam alur kerja klinis.

Sebagai contoh, beberapa rumah sakit di AS menerapkan apa yang disebut Algoritma Kematian AI, yang menggunakan data kesehatan pasien dan menganalisis sekitar 5000 faktor risiko klinis untuk memprediksi peluang bertahan hidup di antara individu yang dirawat di rumah sakit, menyaring pasien dengan kebutuhan paliatif, atau bahkan menentukan waktu hingga kematian pasien dengan penyakit terminal atau yang tidak dapat disembuhkan. Terdapat potensi manfaat dari algoritma ini sebagai alat untuk mendukung keputusan medis dalam indikasi perawatan paliatif, untuk menghindari perpanjangan hidup yang tidak semestinya dan memberikan pasien terminal pilihan untuk menjalani akhir hayat dengan kualitas yang lebih baik, melalui indikasi perawatan paliatif. Namun, komplikasi ekspresif dapat diamati dengan jenis algoritma AI ini seperti yang disebut Jvion CORE, yang dibuat oleh perusahaan Jvion untuk keputusan medis dalam indikasi perawatan paliatif. Algoritma ini telah diimplementasikan di beberapa klinik onkologi di Amerika Serikat. Namun, terdapat masalah serius dalam penggunaan Algoritma Kematian AI dalam praktik klinis. Jvion CORE menyajikan akurasi sekitar 40% dalam prediksinya tentang pasien yang ditandai berisiko tinggi meninggal pada bulan berikutnya. Dengan kata lain, terdapat persentase ekspresif sebesar 60% dari kekeliruan algoritmik.

Oleh karena itu, Eric Topol menyatakan bahwa algoritma dapat membantu pasien dan dokter mereka dalam membuat keputusan tentang jalannya perawatan medis, baik dalam situasi paliatif maupun dalam situasi di mana kesembuhan adalah tujuannya. Namun, penulis menyatakan bahwa 'tidak ada penggunaan AI yang sangat baik kecuali dan sampai terbukti bahwa algoritma yang digunakan sangat akurat'.

Selain itu, terdapat pula ketidaksesuaian antara tugas model-model ini: memprediksi peluang kematian pasien dan bagaimana model-model tersebut sebenarnya digunakan untuk mengidentifikasi siapa yang paling diuntungkan dari diskusi perencanaan perawatan lanjutan (rekomendasi medis untuk perawatan paliatif). Akibatnya, terdapat keraguan yang cukup besar tentang peran kecerdasan buatan dalam konteks perawatan paliatif.

Terakhir, terdapat efek relevan lain dari ketidakjelasan epistemik yang patut mendapat perhatian khusus. Dokter memiliki kewajiban hukum untuk memberikan standar keterampilan

dan perawatan tertentu kepada pasien mereka, tetapi tidak memiliki kewajiban hukum untuk menjamin kesembuhan atau hasil konkret lainnya. Meskipun demikian, terdapat risiko bahwa dokter tidak memahami keterbatasan sistem AI, menggunakannya sebagai tujuan itu sendiri bukan sebagai alat dan, lebih dari itu, memberikan jaminan keberhasilan total kepada pasiennya justru karena teknologi yang digunakan.

Kemudian, muncul diskusi tentang kemungkinan mempertimbangkan kewajiban medis atas hasil, berdasarkan jaminan kesempurnaan alat AI yang digunakan dalam praktik klinis. Sebagai ilustrasi, perlu disebutkan bahwa di AS telah dibahas tentang dokter yang menggunakan platform robotik Da Vinci dalam operasi dan memastikan hasil positif bagi pasien, hanya memberikan informasi tentang manfaat alat teknologi tersebut.

Logika yang sama tampaknya berlaku untuk hipotesis dokter yang menggunakan perangkat Kecerdasan Buatan, seperti Watson milik IBM, yang menciptakan ekspektasi pada pasien bahwa ia akan mendapatkan diagnosis kanker yang sangat akurat dan proposal pengobatan terbaik berkat penggunaan AI, yang, karena bertindak lebih baik daripada manusia, akan mampu memberikan hasil yang baik, yang secara praktis menjamin kesembuhan.

Dalam skenario ini, akan terjadi pelanggaran prinsip etika 'kendali manusia atas teknologi', karena profesional tersebut tidak memahami AI sebagai alat untuk mendukung pengambilan keputusan klinis, sehingga menjadikan teknologi tersebut sebagai jaminan keberhasilan dalam praktik medis. Hal ini mengakibatkan pelanggaran ekspektasi sah pasien dan kemungkinan kualifikasi sifat kewajiban hukum bagi dokter sebagai kewajiban atas hasil.

(2) opasitas karena kurangnya pengungkapan medis: dalam dimensi semantik kedua dari opasitas algoritmik yang khususnya relevan dengan Kedokteran, diamati bahwa terdapat risiko yang cukup besar bahwa algoritma AI digunakan untuk mendukung keputusan medis tanpa sepengetahuan pasien, dan ketidaksadaran pasien bahwa pengambilan keputusan otomatis dan aktivitas pembuatan profil tentang mereka sedang dilakukan dengan data pribadi mereka. Dalam skenario ini, pertama-tama, penting untuk mempertimbangkan bahwa pengungkapan medis adalah proses terstruktur dari komunikasi transparan antara pasien dan dokter yang terlibat selama perawatan medis. Namun, banyak kritik muncul karena pasien sering kali tidak diberitahu atau diminta untuk menyetujui penggunaan algoritma Kecerdasan Buatan dalam perawatan kesehatan mereka.

Faktanya, beberapa dokter menggunakan wacana paternalistik bahwa mereka tidak perlu memberi tahu pasien tentang semua sumber daya yang digunakan dalam proses pengambilan keputusan klinis. Dengan logika ini, tenaga medis profesional, secara teori, dapat memberikan perawatan paliatif kepada pasien, menginformasikan beberapa aspek kondisi klinis mereka, dan memberikan rekomendasi medis tanpa perlu mengungkapkan informasi spesifik tentang penggunaan Algoritma Kematian AI. Namun, menginformasikan dan memberikan persetujuan kepada pasien merupakan salah satu mekanisme untuk mewujudkan hak asasi atas perkembangan kepribadian manusia yang bebas, yang bersifat instrumental karena merupakan cara untuk mewujudkan hak otonomi. Saat ini, doktrin modern di seluruh dunia tentang tanggung jawab medis membela persetujuan pasien sebagai

instrumen yang memungkinkan, selain kepentingan dan tujuan medis-terapeutik, untuk meningkatkan rasa hormat terhadap individu dalam dimensi holistiknya.

Pasien perlu diberikan informasi penting untuk memahami kondisi kesehatannya atau kemungkinan perawatan yang tersedia, sehingga ia dapat menggunakan haknya untuk menyetujui perawatan atau intervensi yang diusulkan, memilih alternatif lain yang ada, meskipun kurang ditunjukkan oleh tenaga medis profesional, atau bahkan menolak untuk dirawat. Gagasan doktrinal ini merupakan tren pemikiran yang telah terbentuk di berbagai yurisdiksi di seluruh dunia dalam beberapa dekade terakhir.

Dengan demikian, jenis ketidakjelasan algoritmik akibat ketidakterbukaan informasi tidak berkaitan dengan karakteristik intrinsik sistem AI, tetapi berawal dari risiko terhadap penentuan nasib sendiri pasien yang informatif. Artinya, hal ini berasal dari cara dokter membuat keputusan medis terkait diagnosis, prognosis, dan usulan perawatan yang didukung AI tanpa sepengetahuan pasien, baik selama intervensi medis maupun setelah kejadian berbahaya. Penting untuk mempertimbangkan bahwa mungkin terdapat liabilitas medis atas kerugian yang dialami pasien dalam penentuan nasib sendiri, karena ia kehilangan kesempatan untuk mempertimbangkan risiko dan keuntungan dari prediksi algoritma AI tentang kondisi klinisnya.

Kesimpulannya, dokter harus memberi tahu pasien bahwa diagnosis, prognosis, usulan perawatan, atau bahkan indikasinya untuk perawatan paliatif didukung oleh beberapa faktor dan sumber daya, termasuk algoritma Kecerdasan Buatan. Hal ini mencakup cita-cita pengambilan keputusan bersama dalam kedokteran.

(3) opasitas penjelasan: selain kewajiban dokter untuk mengungkapkan informasi bahwa ia menggunakan algoritma AI untuk mendukung keputusan klinisnya, ia juga perlu menjelaskan tentang teknologi yang digunakan, sesuai dengan tingkat pemahaman setiap pasien. Jika pasien tidak menerima penjelasan yang tepat dalam AI Medis, hal ini dapat menyebabkan apa yang disebut opasitas penjelasan. Terdapat perbedaan dalam doktrin tentang jumlah informasi yang harus diberikan kepada pasien agar dokter dapat memenuhi kewajibannya untuk memberikan informasi.

Namun, kami telah membela dalam sebuah makalah baru-baru ini bahwa, dengan evolusi teknologi baru di sektor pelayanan kesehatan, dokter perlu memahami bahwa hak atas informasi yang memadai (yang sesuai dengan kewajiban untuk memberi informasi) juga mencakup persetujuan atas penggunaan teknologi baru, berdasarkan pengetahuan pasien tentang fungsi, tujuan, keuntungan, biaya, risiko, dan alternatifnya. Dengan demikian, terdapat tuntutan akan interpretasi baru atas prinsip penentuan nasib sendiri pasien dalam konteks teknologi baru: kita telah beralih dari hak sederhana untuk menerima informasi medis, dan kita bergerak menuju cakupan informasi yang lebih luas, karena terdapat hak atas penjelasan dan pembenaran.

Oleh karena itu, jika kita kembali ke hipotesis faktual Watson for Oncology, bahkan jika kelalaian medis tidak dikonfigurasi, jika profesional hanya memberi tahu tetapi tidak menjelaskan secara memadai kepada pasien tentang penggunaan teknologi untuk mendukung keputusan medis, ia dapat dianggap bertanggung jawab atas kerugian yang dialami pasien

dalam penentuan nasib sendiri, karena kesempatan untuk mempertimbangkan keuntungan dan risiko pengobatan yang diusulkan atau diagnosis medis yang didukung oleh algoritma AI telah diambil dari pasien.

Keterjelasan dapat dipahami sebagai "karakteristik sistem yang digerakkan oleh AI yang memungkinkan seseorang merekonstruksi mengapa AI tertentu menghasilkan prediksi yang disajikan". Namun demikian, penting untuk menunjukkan bahwa keterjelasan bukanlah isu teknologi semata, melainkan memunculkan sejumlah pertanyaan medis, hukum, etika, dan sosial yang memerlukan eksplorasi menyeluruh. Mengambil sistem pendukung keputusan klinis berbasis AI sebagai contoh, terdapat kewajiban etika dan hukum bagi dokter untuk menginformasikan dan menjelaskan kepada pasiennya sesuatu seperti: 'Begini, Tuan John, awalnya saya melihat bahwa kondisi klinis Anda menunjukkan bahwa Anda menderita jenis kanker tertentu, tetapi kami telah mencoba pengobatan kemoterapi tertentu tanpa banyak keberhasilan.

Oleh karena itu, kami dapat memasukkan data pribadi Anda ke Watson for Oncology dan AI akan melakukan referensi silang dengan basis datanya yang besar, untuk menunjukkan kepada kami diagnosis yang beragam pada akhirnya, atau memberikan proposal pengobatan terbaru lainnya berdasarkan tingkat keyakinan. Tetapi, Tuan John, Watson memiliki tingkat kekeliruan tertentu, dan memiliki risiko lain...'. Ini adalah model yang tepat untuk proses mendapatkan persetujuan pasien dalam AI, menjelaskan dan berdialog dengannya untuk memperjelas nuansa diagnosis dan proses prognosis yang didukung oleh teknologi tersebut.

Singkatnya, agar dokter tidak bertanggung jawab atas pelanggaran kewajiban untuk memberi informasi, penting untuk memberikan perhatian khusus pada proses mendapatkan persetujuan berdasarkan informasi, mengubahnya menjadi proses pilihan berdasarkan informasi, mengikuti gagasan proses dialog sejati antara dokter dan pasien. Sejak awal keputusan untuk menggunakan algoritma berbasis AI, terdapat kebutuhan akan penjelasan dan justifikasi bagi mereka yang terdampak oleh teknologi ini.

12.3 ETIKA AI KESEHATAN: KEWAJIBAN DATA DALAM HUKUM PERDATA

Pengembangan dan implementasi perangkat AI dalam Kedokteran membuka pintu bagi tantangan etika dan hukum baru. Tantangan-tantangan ini meliputi bagaimana mengevaluasi kinerja algoritma dan menentukan di mana AI dapat diterapkan secara aman dan efisien dalam praktik klinis. Terdapat tiga contoh isu etika yang berkaitan dengan: “(1) Bias dalam data pelatihan; (2) Potensi penggantian penyedia layanan kesehatan manusia dengan perangkat AI; (3) Menanggapi intervensi AI yang gagal. Jika kita mengembangkan perangkat AI yang memengaruhi keputusan klinis, dan keputusan yang buruk telah dibuat, bagaimana kita (sebagai manusia) akan meresponsnya?”. Sebagaimana telah disebutkan sebelumnya, merancang perangkat pembelajaran mesin yang digunakan untuk mendukung pengambilan keputusan klinis dapat dianggap sebagai sebuah eksperimen yang risikonya perlu dievaluasi secara cermat sebelum diimplementasikan dalam praktik klinis.

Pada bulan Juni 2021, Organisasi Kesehatan Dunia (WHO) menerbitkan panduannya tentang Etika dan Tata Kelola Kecerdasan Buatan untuk Kesehatan (Organisasi Kesehatan Dunia

2021). Laporan tersebut mencerminkan niat WHO untuk mendasarkan panduan mereka pada kerangka kerja hak asasi manusia dan secara langsung merujuk pada Deklarasi Universal Hak Asasi Manusia dengan mengeksplorasi pertanyaan tentang otonomi, melindungi populasi dari bahaya, dan memastikan inklusivitas dan kesetaraan.

Dokumen tersebut menyatakan bahwa 'pertimbangan etika dan hak asasi manusia harus ditempatkan di pusat perancangan, pengembangan, dan penerapan teknologi AI untuk kesehatan'. Dokumen tersebut menawarkan 6 prinsip utama untuk penggunaan AI dalam Kedokteran: (1) melindungi otonomi; (2) meningkatkan kesejahteraan manusia, keselamatan manusia, dan kepentingan publik; (3) memastikan transparansi, kemudahan dijelaskan, dan kejelasan; (4) mendorong tanggung jawab dan akuntabilitas; (5) memastikan inklusivitas dan kesetaraan; (6) mempromosikan AI yang responsif dan berkelanjutan.

Keunggulan lain dari laporan ini adalah analisisnya yang mendetail tentang risiko dan keterbatasan AI. Dua masalah utama yang diangkat: (1) potensi diskriminasi; dan (2) bias ketika kumpulan data yang digunakan untuk melatih AI gagal mencerminkan dunia nyata, dan terdapat kurangnya transparansi dalam sumber data yang digunakan untuk memprogram algoritma ini, tanpa penjelasan tentang bagaimana mereka melakukan referensi silang data dan secara efektif mencapai hasil tertentu.

Faktanya, AI saat ini sedang booming di bidang Kedokteran, tetapi juga menghadapi krisis kredibilitas karena algoritmanya 'sering dilatih pada sampel data kecil, asal tunggal dengan keragaman terbatas; beberapa bahkan menggunakan kembali data yang sama untuk pelatihan dan pengujian, sebuah dosa besar yang dapat menyebabkan kinerja yang terkesan mengesankan namun menyesatkan'.

Kegagalan untuk menguji model AI pada data dari berbagai sumber sebuah proses yang dikenal sebagai validasi eksternal merupakan hal yang umum dalam studi yang dipublikasikan di jurnal medis terkemuka. Menurut tim peneliti dari Universitas Cambridge di Inggris, semakin banyak makalah yang mengandalkan 'data terbatas atau berkualitas rendah, gagal menentukan pendekatan pelatihan dan metode statistiknya, dan tidak menguji apakah mereka akan berhasil untuk orang-orang dari berbagai ras, jenis kelamin, usia, dan geografi'.

Hal ini menghasilkan algoritma yang tampak sangat akurat dalam studi tertentu, tetapi tidak bekerja pada tingkat akurasi yang sama ketika terpapar variabel dunia nyata, di berbagai jenis pasien di beberapa lokasi. Dalam wawancara baru-baru ini, Eric Topol menyampaikan kekhawatiran tentang bagaimana AI dapat memperburuk beberapa ketidakadilan dan diskriminasi, karena 'algoritma tidak bias, tetapi data yang kita masukkan ke dalam algoritma tersebut, karena dipilih oleh manusia, seringkali bias'. Terdapat potensi risiko diskriminasi algoritma AI dalam Kedokteran, karena algoritma tersebut dapat diprogram berdasarkan data dari studi ilmiah dan rekam medis elektronik dari populasi tertentu di mana beberapa ras mendominasi. Dengan demikian, terdapat risiko bahwa keputusan terkontaminasi oleh bias yang signifikan.

Sebagai contoh, terdapat argumen bahwa perempuan kulit hitam dengan kanker payudara lebih mungkin didiagnosis terlambat oleh algoritma yang disetujui FDA di pasaran, justru karena algoritma tersebut diprogram dengan data dari populasi yang kemungkinan

besar tidak memiliki perempuan kulit hitam, atau populasinya sangat sedikit. Hal ini sangat serius dan penting untuk direnungkan, karena pemrograman algoritma dengan data layanan kesehatan dari berbagai populasi dan lokasi geografis sangatlah penting, mengingat variasi ekspresif dalam cara penyakit bermanifestasi pada berbagai ras.

Lebih lanjut, dalam sebuah studi terbaru, ditemukan bahwa antara tahun 2012 dan 2020, hanya 73 dari 161 produk AI yang disetujui oleh Badan Pengawas Obat dan Makanan (FDA) di AS yang telah mengungkapkan secara publik jumlah data yang digunakan untuk memvalidasi produk tersebut, dengan hanya 7 di antaranya yang melaporkan susunan ras untuk populasi studi mereka. Selain itu, di antara 10 produk AI yang disetujui untuk pencitraan payudara, hanya 1 yang mengungkapkan demografi ras dari kumpulan data yang digunakan untuk mendeteksi lesi yang mencurigakan dan menilai risiko kanker.

Dalam studi lain yang dilakukan oleh Universitas Stanford antara Januari 2015 dan Desember 2020, diamati bahwa hampir semua perangkat AI yang disetujui FDA (126 dari 130) hanya menjalani studi retrospektif saat pengajuannya. Tidak satu pun dari 54 perangkat berisiko tinggi dievaluasi oleh studi prospektif dan hanya 17 studi perangkat yang melaporkan bahwa kinerja subkelompok demografis dipertimbangkan dalam evaluasinya.

Disimpulkan dalam studi kedua ini bahwa lebih dari pentingnya mengevaluasi kinerja perangkat AI di beberapa lokasi klinis dan di seluruh populasi yang representatif, juga penting untuk mendorong studi prospektif. Alasan untuk kesimpulan ini adalah bahwa 'studi prospektif dengan perbandingan dengan standar perawatan mengurangi risiko overfitting yang berbahaya dan lebih akurat menangkap hasil klinis yang sebenarnya.' Pengawasan pasca-pemasaran perangkat AI juga diperlukan untuk memahami dan mengukur hasil yang tidak diinginkan dan bias yang tidak terdeteksi dalam uji coba prospektif dan multi-pusat.

Diskusi tentang perlunya regulasi khusus terkait algoritma merupakan tema doktrinal yang berulang di banyak bidang, termasuk sektor kesehatan. Dampaknya menantang pemahaman tentang peran Negara sendiri dalam mengendalikan perkembangan teknologi. Jika, di satu sisi, inovasi diharapkan akan membawa peningkatan kualitas hidup secara keseluruhan, di sisi lain, tidak dapat disangkal bahwa menghadapi masalah ini dari sudut pandang regulasi merupakan tantangan.

Menyusun pendekatan komprehensif untuk menilai kondisi perkembangan teknologi saat ini tampaknya bukan jalan yang masuk akal untuk beberapa tuntutan dan diskusi yang lebih rinci tentang urusan pembuatan undang-undang dalam skenario yang kompleks ini, sementara doktrin hukum perdata telah berupaya membangun model sistematis untuk pembatasan kontur penilaian risiko dalam pengembangan aplikasi yang berpusat pada sistem Kecerdasan Buatan.

Frank Pasquale menyarankan parameterisasi tugas berbasis data untuk pembuatan model standar yang dapat mendukung penilaian akuntabilitas. Menurut penulis, 'standar semacam itu sangat penting mengingat potensi data yang tidak akurat dan tidak tepat untuk mencemari pembelajaran mesin'.

Dalam hal ini, tampaknya proses heuristik berbasis data, jika terkontaminasi sejak awal tahap pemrosesan, dapat menghasilkan hasil yang bias. Dengan kata lain, kurasi data masukan

harus diutamakan dan diamati di seluruh proses algoritmik yang juga harus dapat diaudit jika tidak, substrat akhir yang diperoleh setelah memproses data tersebut (yang disebut 'keluaran') mungkin tidak andal.

Pada dasarnya, parameterisasi model standar tidak lagi bergantung pada upaya regulasi untuk beragam struktur algoritmik, yang bervariasi dalam beberapa aspek, dan menawarkan kebebasan yang lebih besar untuk pengembangan metrik yang diatur sendiri untuk setiap jenis aktivitas.

Dalam konteks ini, dimungkinkan untuk bekerja dengan basis komparatif yang akan menawarkan kondisi yang lebih tepat dan terpetakan dengan baik untuk menentukan kinerja sesuai dengan risiko ekuivalen yang diukur dengan semestinya untuk jenis aktivitas algoritmik yang dimaksud. Stuart Russell dan Peter Norvig telah membahas 'kuantifikasi ketidakpastian' yang rumit dalam konteks prediksi algoritma AI: 'Agen mungkin perlu menangani ketidakpastian, baik karena observabilitas parsial, ketidakpastian nondeterminisme, atau kombinasi keduanya'.

Singkatnya, dugaan yang menjadi dasar tugas berbasis data sejalan dengan pedoman yang sangat penting, yang diusulkan oleh Frank Pasquale sebagai 'hukum robotika keempat' (explainability). Gagasannya memperkuat kebutuhan untuk mengatasi masalah kotak hitam. Sebagaimana telah disebutkan sebelumnya, masalah ini biasanya diidentifikasi dengan penggunaan teknik pembelajaran mesin yang memberikan peningkatan tak terkendali dan tak terawasi pada aplikasi ini, hingga menjadi begitu kompleks sehingga bahkan penciptanya sendiri tidak memahaminya.

Tanggung jawab perdata berkaitan dengan ketidakpastian dan ketidakpastian. Secara tradisional, hal tersebut berasal dari penerapan teori risiko integral sebagai dasar penyelesaian gugatan perdata, khususnya berdasarkan bahaya perwalian dan prinsip kehati-hatian. Logika yang sama, jika diterapkan dalam konteks algoritma AI, akan menghasilkan beberapa konsekuensi yang unik. Mengenai hal ini, Yaniv Benhamou dan Justine Ferland telah menunjukkan lima observasi tentang tugas berbasis data.

1. Pengamatan pertama para penulis adalah bahwa, terkait persyaratan yang dibebankan kepada aktor algoritmik (pemilik, operator, pengecer, dan perancang), perlu untuk mematuhi kewajiban kehati-hatian, yang menyangkut: (a) pemilihan teknologi tertentu, dengan mempertimbangkan tugas-tugas yang perlu dilakukan dan keterampilan serta kemampuan operator itu sendiri; (b) kerangka kerja organisasi yang direncanakan, khususnya terkait tindak lanjut yang memadai; dan (c) pemeliharaan, termasuk rutinitas pemeriksaan keselamatan. Kegagalan untuk mematuhi kewajiban tersebut dapat memicu tanggung jawab mutlak, terlepas dari apakah operator juga bertanggung jawab atas penciptaan atau peningkatan risiko suatu teknologi tertentu. Mempertimbangkan hal ini, tampaknya penting juga bagi dokter atau rumah sakit yang berperan sebagai operator algoritma untuk mematuhi kewajiban-kewajiban ini.
2. Benhamou dan Ferland juga menunjukkan bahwa produsen, termasuk mereka yang bertindak sebagai pengawas algoritmik, harus mematuhi standar perilaku berikut: (b.1) merancang, mendeskripsikan, dan memasarkan produk sedemikian rupa sehingga

memungkinkan mereka memenuhi tugas berdasarkan data, sehingga risiko lebih dapat diprediksi memantau produk dengan tepat setelah diedarkan, mengingat karakteristik teknologi digital yang sedang berkembang, khususnya karena keterbukaan dan ketergantungannya pada lingkungan digital umum, termasuk keusangan, munculnya malware, atau bahkan kerentanannya terhadap kemungkinan serangan eksternal.

3. Apa yang disebut pengawasan, dalam konteks pemantauan tugas-tugas spesifik yang secara hierarkis lebih tinggi bahkan mungkin disebabkan oleh kekuasaan polisi administratif Negara, dalam apa yang disebut Pasquale sebagai 'pengawasan' dalam buku terbarunya. Hal ini dapat dicapai melalui audit dan studi terhadap algoritma spesifik, bahkan setelah peluncurannya di pasaran. Dengan demikian, sebagai hasil dari implementasi sistem pemantauan terawasi, identifikasi anomali dan parameterisasi sistem sebelumnya diharapkan dapat 'memperingatkan' tentang terjadinya perilaku tak terduga, serta pengamatan tren evolusi spesifik dari pembelajaran mesin untuk memprediksi perilaku tersebut. Setelah pemantauan tersebut diimplementasikan, kewajiban untuk memberi tahu calon korban muncul sebagai kewajiban yang melekat pada itikad baik yang objektif.
4. Jika memungkinkan, para penulis berpendapat bahwa produsen harus diwajibkan untuk menyertakan pintu belakang wajib dalam algoritma mereka. Sebutan lain untuk hal ini adalah ungkapan 'rem darurat secara default (atau berdasarkan rancangan)', 'fitur mati', atau fitur yang memungkinkan operator atau pengguna untuk 'mematikan AI' dengan perintah manual, atau membuatnya 'tidak cerdas' hanya dengan menekan tombol panik. Kegagalan untuk menjamin alat dan opsi kontrol tersebut dapat dianggap sebagai cacat desain yang dapat membenarkan pelanggaran tindakan pencegahan umum yang diharapkan darinya, sehingga membuka kemungkinan pengenaan tanggung jawab perdata karena algoritma tersebut dianggap cacat. Bahkan, tergantung pada situasinya, produsen atau operator juga dapat dipaksa untuk 'mematikan' AI sebagai bagian dari tugas pemantauan dan audit algoritmik mereka.
5. Serupa dengan tugas purnajual yang ada, yang terdiri dari peringatan dan instruksi untuk menarik kembali produk cacat, produsen/produsen juga dapat mengemban tugas dukungan dan koreksi yang merupakan konsekuensi dari prinsip auditabilitas dan transparansi sejalan dengan perkembangan terbaru lainnya tentang potensi kewajiban pengembang perangkat lunak untuk memperbarui algoritma yang tidak aman, selama teknologi tersebut masih beredar di pasaran (yaitu, melampaui ketentuan kontrak apa pun tentang masa garansi).

Frank Pasquale menyelidiki potensi liabilitas dalam konteks penggunaan data yang tidak akurat atau tidak tepat (data yang salah) dalam set pelatihan untuk pembelajaran mesin: 'perusahaan yang menggunakan data yang salah dapat diwajibkan untuk memberikan kompensasi kepada mereka yang dirugikan oleh penggunaan data tersebut dan harus dikenakan ganti rugi punitif ketika pengumpulan, analisis, dan penggunaan data yang salah tersebut berulang atau disengaja.' Fungsi punitif dari liabilitas perdata menimbulkan aspek kontroversial yang perlu dipertimbangkan dalam konteks studi singkat ini.

Hal ini karena, khususnya dalam pengalaman hukum umum, manfaat punitif dan dissuasif memiliki penerapan yang lebih luas dan diterima, baik oleh doktrin maupun oleh Pengadilan. Meskipun topik ini kontroversial dan meskipun ganti rugi punitif hanyalah salah satu dari berbagai pilihan untuk mempertimbangkan efek jera dari potensi liabilitas, relevansi pembahasan ini dengan konteks teknologi kompleks di mana algoritma Kecerdasan Buatan dimasukkan tidak dapat dihindari.

Bahasa Indonesia: Menjaga komplementaritas hukum perdata dan regulasi pengumpulan, analisis, dan penggunaan data sangatlah tepat untuk membantunya menghindari kecelakaan yang dapat dicegah dan memperluas peluang bagi mereka yang dirugikan oleh teknologi baru untuk menuntut akuntabilitas.

Saat ini, hukum perdata bergerak ke arah untuk mempromosikan tidak hanya liabilitas tetapi juga akuntabilitas, yang memiliki fungsi prospektif dan lebih kuat dan berdasarkan pada beberapa fungsi, terutama fungsi pencegahan. Skenario yang disajikan oleh Pasquale ini memperkuat, di satu sisi, perhatian penting dengan kepatuhan yang diinginkan, dipertimbangkan dari struktur tata kelola dan kurasi data yang ditujukan pada verifikasi berkelanjutan kualitas pengumpulan yang digunakan dalam algoritma AI.

Di sisi lain, konteks ini ternyata menjadi refleksi tiga kali lipat: (1) jika memungkinkan untuk berasumsi bahwa diagnosis AI akan ditanggung dalam polis asuransi malapraktik penyedia layanan kesehatan saat ini, atau jika pengenalan diagnosis AI ke dalam praktik klinis kemungkinan akan mendorong penyedia asuransi untuk menolak pertanggungan untuk kegiatan tersebut; (2) potensi tanggung jawab perdata dokter sebagai operator algoritmik yang berulang kali mengamati ketidakefektifannya atau menyadari bahwa AI menggunakan pengumpulan data yang bias (data yang salah); (3) pentingnya pendidikan kedokteran dalam AI, kesehatan digital, dan teknologi baru untuk mencegah kejadian tidak diinginkan.

12.4 MASA DEPAN AI DALAM KEDOKTERAN DAN PENDIDIKAN KESEHATAN DIGITAL

Dalam studi ini, diamati bahwa perkembangan berharga P4-Kedokteran dari penggunaan algoritma prediktif tidak dapat dilepaskan dari kebutuhan untuk mempertimbangkan risiko dan ketekunan medis khusus dalam menggunakan teknologi tersebut sebagai alat untuk mendukung pengambilan keputusan. Lebih lanjut, disimpulkan bahwa terdapat tiga dimensi semantik yang berbeda dari ketidakjelasan algoritmik yang relevan dengan Kedokteran:

(1) ketidakjelasan epistemik karena kurangnya pemahaman dokter tentang aturan yang diterapkan sistem AI untuk membuat prediksi dan keputusan; (2) ketidakjelasan atas kurangnya pengungkapan medis tentang penggunaan sistem AI dan ketidaktahuan pasien bahwa pengambilan keputusan otomatis dan aktivitas pembuatan profil tentang mereka sedang dilakukan dengan data pribadi mereka; dan (3) ketidakjelasan penjelasan atas penjelasan yang tidak memuaskan kepada pasien tentang teknologi yang digunakan untuk mendukung pengambilan keputusan profesional. Oleh karena itu, tujuan makalah ini adalah untuk menganalisis setiap jenis ketidakjelasan, dengan mempertimbangkan skenario hipotetis dan dampaknya dalam hal malapraktik medis dan persetujuan pasien.

Mengenai opasitas epistemik, pertanyaan diajukan tentang sejauh mana seorang dokter mungkin bergantung pada AI dan konsekuensi hukum jika dokter mematuhi rekomendasi atau mengesampingkan mesin, yang mengarah pada pertimbangan signifikan tentang penentuan standar uji tuntas medis yang harus menjadi masalah yang selalu terbuka untuk diperdebatkan dalam setiap kasus malapraktik medis tertentu. Hal ini karena, dalam setiap situasi, tingkat akurasi suatu algoritma dan tujuannya berbeda.

Dapat juga disimpulkan dalam makalah ini bahwa ada kemungkinan untuk mengkualifikasi kewajiban dokter sebagai kewajiban hasil ketika ada pelanggaran prinsip etika 'kendali manusia atas teknologi', yaitu, dalam menghadapi ketidakpahaman tentang AI sebagai alat untuk mendukung pengambilan keputusan klinis (AI-sebagai-alat), dengan konsekuensi pelanggaran harapan sah pasien terhadap teknologi sebagai jaminan keberhasilan. Juga diamati bahwa ketidakjelasan atas kurangnya pengungkapan medis dan ketidakjelasan penjelasan menuntut refleksi atas dampak prinsip-prinsip etika penjelasan dan justifikasi, untuk memahami model baru persetujuan pasien dalam AI dan pelanggaran kewajiban medis atas informasi yang berkualitas.

Dalam konteks ini, isu-isu yang disebutkan di atas khususnya, konsekuensi dari tiga dimensi semantik ketidakjelasan algoritmik merupakan tantangan besar bagi para pendidik di bidang ilmu kesehatan. AI dapat membantu para profesional medis dengan menggabungkan sejumlah besar data layanan kesehatan dan mendukung proses pengambilan keputusan mereka tentang diagnosis dan merekomendasikan perawatan. Namun demikian, dokter membutuhkan kemampuan untuk menginterpretasikan hasil dan mengomunikasikan rekomendasi dengan tepat kepada pasien. Dokter perlu mempelajari cara menggunakan dan menginterpretasikan algoritma AI dengan lebih baik, termasuk dalam proses pembelajaran ini pemahaman tentang situasi di mana suatu algoritma harus digunakan secara efektif dalam praktik mereka, dan, yang terpenting, seberapa besar keyakinan yang harus diberikan pada rekomendasi algoritmik, dalam setiap kasus konkret.

Oleh karena itu, keterampilan dan keahlian baru diperlukan seiring kita memasuki era Kecerdasan Buatan di lingkungan layanan kesehatan. Kurangnya pengetahuan mendalam dokter tentang manfaat dan risiko teknologi perawatan kesehatan dapat mengakibatkan hasil yang lebih buruk bagi pasien karena sedikit atau tidak adanya pemahaman tentang alat AI mana yang menambah nilai pada aktivitas mereka atau cara mengintegrasikan AI dengan cara yang sesuai untuk alur kerja klinis. Tugas ini menyerukan model baru untuk mendidik generasi ahli baru dengan pelatihan interdisipliner yang mendalam dalam Kedokteran, etika, dan teknologi. Oleh karena itu, AI perlu diintegrasikan dengan mulus di berbagai aspek kurikulum pendidikan kedokteran.

American Medical Association (AMA) mencatat bahwa dari tahun 2000 hingga 2015 ada 15 laporan nasional yang menyerukan reformasi pendidikan kedokteran. Di AS, ada beberapa inisiatif untuk menggabungkan teknologi baru seperti alat AI dalam pendidikan kedokteran:

(1) Duke Institute for Health Innovation: mahasiswa kedokteran bekerja sama dengan pakar data untuk mengembangkan teknologi yang ditingkatkan perawatannya yang dibuat untuk dokter;
(2) University of Florida: residen radiologi bekerja sama dengan perusahaan berbasis teknologi untuk mengembangkan deteksi berbantuan komputer untuk mamografi;
(3) Carle Illinois College of Medicine: menawarkan kursus bagi ilmuwan dan insinyur klinis untuk mempelajari teknologi baru;
(4) Sharon Lund Medical Intelligence and Innovation Institute: menyelenggarakan kursus musim panas tentang semua teknologi baru dalam perawatan kesehatan, terbuka untuk mahasiswa kedokteran;
(5) Stanford University Center for Artificial Intelligence in Medicine: melibatkan mahasiswa pascasarjana dan pasca sarjana dalam memecahkan masalah perawatan kesehatan dengan menggunakan pembelajaran mesin;
(6) Rocky Vista University College of Osteopathic Medicine: menawarkan kursus untuk melatih mahasiswa kedokteran dalam AI, pemantauan jarak jauh, etika, informatika, telemedicine, analitik, dan kewirausahaan.

Sebagai kesimpulan, isu utama bagi masa depan AI dalam Kedokteran terletak pada seberapa baik pendidikan kedokteran dapat terjamin. Seiring dengan semakin lazimnya AI dan penerapannya di sektor perawatan kesehatan, mahasiswa kedokteran, residen, fellow, dan dokter praktik perlu memiliki pengetahuan yang lebih baik tentang AI. Integrasi kesehatan digital ke dalam pendidikan formal menawarkan cara baru untuk terlibat dalam peluang pendidikan interprofesional.

Menentukan bagaimana membangun pendidikan kesehatan digital dan mengintegrasikannya ke dalam kurikulum formal akan menjadi topik perdebatan di tahun-tahun mendatang. Untuk memastikan bahwa keputusan klinis berbasis AI memenuhi janjinya, perlu adanya peningkatan kesadaran para pengembang, tenaga kesehatan profesional, dan legislator terhadap tantangan dan keterbatasan algoritma yang tidak transparan di sektor kesehatan, serta untuk mendorong pendidikan kedokteran menuju Era Kecerdasan Buatan dalam Kedokteran.

BAGIAN III
HUKUM, TATA KELOLA, DAN REGULASI KECERDASAN BUATAN

Sebagaimana telah disebutkan sebelumnya, Bagian II buku ini mengarahkan perhatiannya pada perdebatan etika dan hukum fundamental, mengevaluasinya berdasarkan transformasi yang dibawa oleh AI. Bagian III menggunakan berbagai pendekatan untuk membahas bagaimana Kecerdasan Buatan diselaraskan oleh hukum, bagaimana penerapannya harus diatur, dan bagaimana regulasi yang ada dan yang akan datang dapat diterapkan untuk mengatasi risikonya.

Bab karya *Ugo Pagallo* mencoba membongkar mitos kedaulatan digital, konstitusionalisme digital, efek Brussel, dan Kecerdasan Buatan yang berpusat pada manusia dalam konteks Hukum Eropa.

Bab karya *Luís Moniz Pereira, Francisco C. Santos, dan António Barata Lopes* mengeksplorasi penerapan pemikiran kontrafaktual oleh agen AI, dan menyarankan bahwa pembelajar kontrafaktual mendorong koordinasi dalam dilema kolektif. Dalam bab karya *Willem Gravett*, penulis secara kritis mengkaji "efek teknologi" kecenderungan manusia terhadap optimisme berlebihan ketika mengambil keputusan yang melibatkan teknologi dan "bias otomatisasi" fenomena di mana hakim menerima rekomendasi dari sistem pengambilan keputusan otomatis, tanpa riset atau konfirmasi tambahan.

Bab karya *Ana Taveira da Fonseca, Elsa Vaz Sequeira, dan Luís Barreto Xavier*, mencoba memastikan apakah terdapat tempat untuk pertanggungjawaban berbasis kesalahan untuk sistem yang digerakkan oleh AI, apakah rezim pertanggungjawaban ketat saat ini sesuai untuk mengatasi kerugian tanpa kesalahan yang disebabkan oleh fungsi sistem AI, dan kapan seorang agen harus dibebaskan dari pertanggungjawaban.

Bab karya *Marcel Lanz dan Stefan Mijic* membahas teknologi pembangkitan bahasa alami (NLG) dan mengeksplorasi berbagai cara pertanggungjawaban perdata atas penerapan NLG dapat dikonstruksi dan bagaimana hal itu harus diselesaikan, dari perspektif hukum Swiss.

Dalam bab karya *Pedro Garcia Marques*, penulis merefleksikan kemungkinan dan batasan penggunaan sistem berbasis AI untuk memprediksi risiko residivisme dalam rangka menjatuhkan hukuman pidana kepada terpidana dan membandingkannya dengan penilaian manusia.

Pembuatan deepfake dan relevansinya dalam konteks administrasi peradilan pidana menjadi topik bab karya *Dalila Durães, Pedro Miguel Freitas, dan Paulo Novais*. Para penulis menjelaskan fondasi teknis deepfake, membahas cara Uni Eropa menanganinya dalam rancangan Undang-Undang AI, dan tantangan yang melekat dalam mengaturnya untuk tujuan peradilan pidana.

Bab karya Maria *José Schmidt-Kessen dan Max Huffman* membahas teknologi penetapan harga berbasis AI dan perannya dalam kolusi algoritmik. Para penulis membahas topik ini dari perspektif perbandingan hukum antimonopoli AS dan Uni Eropa, dan membahas masalah yang masih ada ini secara prospektif. Bab karya *Joana Covelo de Abreu* menekankan

perlunya pendekatan AI yang berpusat pada manusia di bidang peradilan, melalui prinsip-prinsip yang berfokus pada pengguna dan ramah pengguna, serta mengkaji bagaimana rancangan Undang-Undang AI Uni Eropa harus lebih lanjut membahas penggunaan sistem AI oleh peradilan, sebagai mekanisme untuk meningkatkan independensi peradilan, hak prosedural, dan akses terhadap keadilan di Uni Eropa.

Dalam bab karya *Anat Keller, Clara Martins Pereira, dan Martinho Lucas Pires*, para penulis secara kritis mengkaji pengecualian risiko sistemik keuangan dari definisi "risiko tinggi" dalam rancangan Undang-Undang AI Uni Eropa, dan mengadvokasi pendekatan lintas batas yang lebih terintegrasi terhadap AI, dengan mengakui implikasi AI terhadap risiko sistemik keuangan.

Bab terakhir dari Bagian III (*Katerina Yordanova dan Natalie Bertels*) menganalisis potensi kotak pasir regulasi yang dibayangkan dalam rancangan Undang-Undang AI Uni Eropa untuk mengatur AI dan tantangan yang mungkin dihadapinya berdasarkan pengalaman dari kotak pasir regulasi sebelumnya. Penulis kemudian menyarankan solusi yang dibuat khusus untuk mengurangi potensi kerugian dari kotak pasir regulasi AI.

BAB 13
MEMBONGKAR MITOS AI DAN HUKUM UNI EROPA

Abstrak Komisi Eropa baru-baru ini mengusulkan beberapa undang-undang, arahan, dan peraturan yang akan melengkapi undang-undang terkini tentang internet, tata kelola data, dan Kecerdasan Buatan, misalnya, Undang-Undang AI mulai Mei 2021. Beberapa mengusulkan untuk merangkum tren hukum Uni Eropa terkini berdasarkan formula yang menarik, seperti (i) kedaulatan digital; (ii) konstitusionalisme digital; (iii) efek Brussels baru; dan, (iv) pendekatan AI yang berpusat pada manusia. Masing-masing narasi ini memiliki kelebihannya sendiri, tetapi bisa sangat menyesatkan. Narasi-narasi ini harus dipahami dengan skeptis.

Tujuan dari makalah ini adalah untuk membongkar 'mitos' ini melalui informasi hukum 'tentang' realitas, yaitu pengetahuan dan konsep yang membingkai representasi dan fungsi hukum Uni Eropa. Kita perlu mewaspadai hal-hal yang diabaikan oleh mitos-mitos saat ini, seperti isu-isu terbuka tentang keseimbangan kekuasaan antara lembaga-lembaga Uni Eropa dan negara-negara anggota (MS), generasi baru hak-hak digital di tingkat konstitusional Uni Eropa dan MS, hingga interaksi antara model-model baru tata kelola hukum dan potensi fragmentasi sistem, misalnya, antara regulasi teknologi dan hukum lingkungan.

13.1 MITOS REGULASI UE: EMPAT RUMUS AI

Selama beberapa tahun terakhir, Komisi Eropa telah mengusulkan beberapa undang-undang, arahan, dan peraturan yang akan melengkapi legislasi terkini tentang internet, tata kelola data, Kecerdasan Buatan, dan lainnya. Daftar inisiatif dan proposal yang dibahas di tingkat Uni Eropa ('EU') mencakup Undang-Undang Layanan Digital dan Pasar Digital dari Desember 2020, Undang-Undang Tata Kelola Data dari November tahun itu, Undang-Undang Kecerdasan Buatan (AIA) dari Mei 2021, Undang-Undang Keamanan Siber dari Juli 2021, sebagai tambahan atas inisiatif untuk Green Deal, proyek Sains Terbuka, dll.

Dengan mempertimbangkan kompleksitas hukum tersebut, para akademisi telah mengusulkan beberapa rumus menarik yang akan membantu kita menetapkan tingkat abstraksi yang tepat, untuk mengatasi kerumitan regulasi teknologi dan tata kelola data dalam hukum UE. Tujuan dari makalah ini adalah untuk mengkaji empat rumus ini: (i) kedaulatan digital; (ii) konstitusionalisme digital; (iii) efek Brussels baru; dan, (iv) pendekatan yang berpusat pada manusia terhadap AI ('HAI').

Asumsi keseluruhan dari analisis ini adalah bahwa masing-masing tingkat abstraksi ini memiliki kelebihannya sendiri, dan tetap saja, rumus tersebut dapat menyesatkan. Penggunaannya mungkin menyiratkan masalah yang keliru, atau masalah yang dianggap remeh, terkadang mengabaikan inti permasalahan yang sebenarnya. Oleh karena itu, tujuan makalah ini adalah untuk membongkar 'mitos' ini melalui sudut pandang informasi hukum 'tentang' realitas, yaitu pengetahuan dan konsep yang membingkai representasi dan fungsi hukum Uni Eropa. Analisis ini dibagi menjadi lima bagian, yang masing-masing dikhususkan untuk salah satu mitos yang diteliti dalam makalah ini, beserta kesimpulannya. Tujuan

keseluruhannya adalah untuk menawarkan analisis yang lebih serius tentang tren terkini hukum Uni Eropa dan regulasi teknologi.

13.2 KEDAULATAN DIGITAL

Luciano Floridi baru-baru ini menelaah 'perebutan kedaulatan digital' yang terjadi selama beberapa tahun terakhir, mengkaji 'apa itu' (masalah kendali data, perangkat lunak, standar, layanan, infrastruktur, dll.); dan 'mengapa hal itu penting' (perebutan ini menyentuh semua orang) 'terutama bagi Uni Eropa'. Meskipun Floridi merujuk pada 'dunia pasca-Westfalen di mana teritorialitas hukum tidak lagi berlaku secara otomatis dan mungkin tidak relevan', dimensi baru dari konsep lama ini, yaitu 'kedaulatan digital', seharusnya tetap menjelaskan perebutan kendali yang sedang berlangsung antara berbagai sistem regulasi yang bersaing di luar sana: kekuatan pasar, dan norma-norma sosial, kewenangan hukum pemerintah nasional dan organisasi internasional, peran lembaga-lembaga sipil dan sektor keuangan, dan masih banyak lagi.

Namun, dalam hukum Uni Eropa, sejak putusan Pengadilan Eropa dalam kasus Van Gend & Loos tahun 1963, prinsip kedaulatan dan formula terkini tentang 'kedaulatan digital' mengingatkan kita pada simpul hukum tentang siapa yang harus memiliki 'kata terakhir' antara lembaga-lembaga Uni Eropa dan Negara-negara Anggota (MS). Baik atau buruk, 30 tahun yang lalu, kompromi telah dicapai dengan perjanjian Maastricht, dan prinsip subsidiaritas sesuai dengan Pasal 5 Perjanjian Uni Eropa. Sebagian besar inisiatif dan proposal regulasi Komisi, yang disebutkan di atas dalam pendahuluan, memang bergantung pada prinsip subsidiaritas karena skala isu yang dipertaruhkan dengan regulasi aspek-aspek krusial interaksi sosial di internet, tata kelola data, atau AI dan teknologi baru lainnya. Jadi, menyesatkan untuk menyebut tren hukum Uni Eropa saat ini dalam istilah 'kedaulatan digital' karena formula tersebut mungkin menunjukkan bahwa regulasi Uni Eropa terlihat seperti hukum federal. Padahal tidak. Dialihkan oleh MS dan kewenangan konstitusionalnya melalui Perjanjian, kewenangan Uni Eropa bukanlah 'asli' seperti yang terjadi pada kewenangan konstitusional negara bagian federal, misalnya Amerika Serikat.

Detail hukum ini menunjukkan bahwa formula 'kedaulatan digital' tidak mencerminkan keseimbangan kekuasaan antara lembaga Uni Eropa dan MS, atau formula tersebut menunjukkan bahwa beberapa masalah telah diselesaikan atau, setidaknya, ditangani dengan tepat padahal kenyataannya tidak. Para akademisi masih membahas isu yang dijuluki sebagai isu Kompetenz-Kompetenz dalam saga pengadilan konstitusi federal Jerman, kasus Solange, sejak tahun 1970an.

Dalam membahas tata kelola internet, AI, atau menangani aliran data dalam masyarakat informasi saat ini, formula 'kedaulatan digital' tidak membantu kita memecahkan isu yang terus-menerus muncul ini tentang siapa yang berdaulat di Eropa. Lebih lanjut, jika kita tertarik pada arti formula ini 'khususnya bagi Uni Eropa', 'kedaulatan digital' tidak membantu kita menjelaskan jenis tata kelola di balik proposal dan inisiatif Komisi baru-baru ini. Alih-alih mencari satu atau beberapa kedaulatan dalam hukum yang berlaku saat ini, kita seharusnya

lebih teknis dalam memahami tata kelola Uni Eropa saat ini dan yurisprudensinya. Akankah sikap terhadap 'konstitusionalisme digital' menawarkan analisis yang lebih teknis?

13.3 KONSTITUSIONALISME DIGITAL

Mempertimbangkan pendekatan Uni Eropa terhadap tantangan regulasi teknologi dan tata kelolanya saat ini, beberapa pihak berpendapat bahwa "dalam dua puluh tahun terakhir, kebijakan Uni Eropa di bidang teknologi digital telah bergeser dari perspektif ekonomi liberal ke pendekatan yang berorientasi pada konstitusi". Dimensi digital baru konstitusionalisme Uni Eropa ini sering diilustrasikan dengan upaya-upaya terkini untuk menentang kekuatan korporasi transnasional yang beroperasi di dunia maya, dengan serangkaian tanggung jawab dan tugas baru bagi korporasi tersebut, sebagai penyedia layanan di internet, sebagai perancang dan produsen sistem AI berisiko tinggi, sebagai pengendali data pribadi dalam lingkungan digital yang kompleks, dan sebagainya. Serangkaian tugas dan kewajiban baru ini tentu saja sejalan dengan hak-hak baru yang terkait.

Dimulai dengan hak untuk menghapus data yang ditetapkan oleh Pengadilan Luksemburg dalam kasus Google pada tahun 2013, perhatian perlu diarahkan pada hak-hak baru untuk menghapus, untuk dilupakan, portabilitas data, dll. yang diabadikan dalam peraturan perlindungan data umum, atau 'GDPR' dari tahun 2016, atau hak-hak baru untuk tidak diprofilkan, atau diakui oleh sistem AI, yang diusulkan oleh Pasal 5 Undang-Undang AI 2021 Komisi Eropa, hingga kebijakannya saat ini tentang hak akses terbuka, hak sains terbuka, dll. Bukankah seharusnya kita menjuluki semua tren ini sebagai 'konstitusionalisme digital' lembaga-lembaga Uni Eropa?

Menariknya, sikap terhadap konstitusionalisme digital ini merujuk, di satu sisi, pada prinsip sudut pandang kedaulatan digital, seperti perebutan akses, kendali, dan perlindungan atas data dan informasi dalam lingkungan digital, antara pemerintah dan lembaga nasional dan internasional, misalnya Uni Eropa, dan kekuatan perusahaan transnasional. Uni Eropa akan memamerkan kekuatannya, menunjukkan siapa penguasa digital saat ini, dengan menetapkan kewajiban baru bagi para petinggi Silicon Valley, dan hak-hak baru bagi warga negara Uni Eropa. Meskipun penegakan hak dan kewajiban tersebut terkadang tampak bermasalah, misalnya portabilitas data, tampaknya adil untuk mengakui bahwa sikap terhadap konstitusionalisme digital ini, seperti halnya sikap kedaulatan digital yang tumpang tindih, menarik perhatian kita pada sebuah pengubah permainan.

Selama lebih dari 20 tahun terakhir, hukum Uni Eropa memang telah berupaya untuk melengkapi kerangka tradisional hak-hak konstitusional (dan hak asasi manusia) dasar yang terkait dengan tubuh fisik individu dan hak habeas corpus mereka, dengan prinsip baru hak habeas data. Yang terakhir ini dapat ditelusuri kembali ke apa yang telah dibingkai oleh Mahkamah Konstitusi Jerman dalam istilah 'penentuan nasib sendiri informasional' sejak Volkszählungs-Urteil ('keputusan sensus'), tahun 1983. Namun, di sisi lain, formula 'konstitusionalisme digital' dapat menyesatkan, jika diterapkan pada hukum Uni Eropa, karena yang tidak dimiliki Uni Eropa adalah inti dari konstitusionalisme tradisional, yaitu, kekuasaan

atas masalah ketertiban umum, penegakan hukum, dan keamanan nasional di bidang-bidang krusial seperti hukum pidana dan administrasi (termasuk perlindungan prosedural).

Dengan merujuk pada formula konstitusionalisme digital Uni Eropa, risikonya adalah mengabaikan lubang hitam dalam kerangka tersebut, yaitu, hak dan perlindungan bagi tubuh digital individu sehubungan dengan petugas penegak hukum, jaksa penuntut umum, atau dinas rahasia. Untuk memahami bagaimana teknologi berdampak pada prinsip-prinsip supremasi hukum, seperti prinsip habeas corpus dan gagasan 'pengadilan yang adil,' 'kesetaraan senjata,' dll., perhatian harus diarahkan, pertama, ke tingkat hukum nasional. Misalnya, standar ganda perlindungan untuk tubuh fisik dan tubuh digital individu, menurut yurisprudensi Mahkamah Konstitusi dan Pengadilan Kasasi di Italia, dianggap sesuai dengan hukum Uni Eropa dan terlebih lagi, kerangka umum yang disediakan oleh Konvensi Hak Asasi Manusia Eropa 1950 dan Pengadilannya (ECtHR).

Ini berarti bahwa, berurusan dengan tubuh fisik dan perlindungannya dalam hukum ketatanegaraan Italia, sebuah undang-undang dan otorisasi pengadilan memberikan tingkat perlindungan hukum ganda (Pasal 14 Konstitusi), sedangkan, dalam kasus tubuh digital dalam proses pidana, sebagian besar kewenangan hanya sampai pada jaksa penuntut umum (Pasal 2). Apakah sistem AI akan memperkuat asimetri kekuasaan antara jaksa penuntut umum dan tersangka juga, tetapi tidak hanya di Italia tentu saja masih menjadi pertanyaan terbuka. Namun, sejalan dengan klaim konstitusionalisme digital saat ini, pertanyaan terbuka ini, dan secara lebih umum, padanan informasional dari prinsip-prinsip tradisional habeas corpus, peradilan yang adil, kesetaraan senjata, dll., tidak berkisar pada tren hukum Uni Eropa, tetapi sebagian besar pada kewenangan Negara-negara Anggota Uni Eropa dalam kerangka ECtHR.

Ini bukan berarti hukum Uni Eropa tidak punya peran dalam membentuk kerangka hukum untuk perlindungan individu bahkan di hadapan Pengadilan Pidana, misalnya masalah perlindungan data, namun seluruh rangkaian sumber, yang harus disertakan dalam setiap konstitusionalisme digital Eropa seperti kekuasaan dan konstitusi nasional, ECtHR, hukum Uni Eropa dan perjanjiannya, perjanjian internasional, dan lain-lain menimbulkan pertanyaan lebih lanjut.

Saya akui bahwa peran hukum Uni Eropa, meskipun terbatas pada bidang hukum tata negara tertentu, khususnya relevan dalam beberapa bidang baru konstitusionalisme digital, seperti perlindungan data pribadi dan seperangkat hak baru dalam interaksi manusia-AI yang dibentuk, misalnya, oleh AIA Komisi Eropa. Namun, peran hukum Uni Eropa dalam membentuk konstitusionalisme digital di Eropa saat ini dan tata kelola hukumnya yang kompleks, terkadang telah melahirkan mitos-mitos baru. Meskipun dalam hukum Uni Eropa, formula konstitusionalisme digital mengabaikan permasalahan kekuatan nasional dan penggunaan sistem AI yang mengganggu oleh lembaga penegak hukum, rencana konstitusionalisme digital baru (dan bahkan yang diinginkan) di Eropa seringkali melebih-lebihkan peran hukum Uni Eropa. Bagian selanjutnya membahas salah satu pernyataan berlebihan yang populer ini, yang membawa kita kembali pada posisi kedaulatan digital.

13.4 EFEK BRUSSEL

Sepuluh tahun yang lalu, gagasan Anu Bradford tentang 'efek Brussel' menjadi viral. Singkatnya, gagasannya adalah bahwa dalam menangani isu-isu regulasi teknologi, perlindungan data, hukum lingkungan, atau antimonopoli, hukum Uni Eropa secara sepihak telah memberikan efek hukum ekstra-teritorial. Baru-baru ini, Bradford telah menyempurnakan gagasan ini dalam sebuah volume baru, dan beberapa akademisi memperkirakan apakah kita harus mengharapkan efek Brussels baru mengingat inisiatif terbaru Komisi Eropa tentang AI, tata kelola data, layanan dan pasar digital, dll.

Faktanya, demikian pula argumen tentang efek Brussels, yaitu data yang tidak dapat dibagi dan biaya kepatuhan perusahaan multinasional yang berurusan dengan berbagai rezim regulasi, dapat mendorong sebagian besar produsen teknologi dan penyedia layanan untuk mengadopsi dan menyesuaikan diri dengan standar internasional yang paling ketat secara menyeluruh, yaitu kerangka kerja perlindungan data dan lingkungan Uni Eropa, dan sekarang, proposal Komisi Eropa.

Sekali lagi, setelah sikap tentang kedaulatan digital dan konstitusionalisme digital, 'efek Brussels' memiliki kelebihannya. Saya mungkin berani mengatakan bahwa, misalnya, hukum perlindungan data Uni Eropa memang mewakili sebuah model bagi seluruh dunia. Namun, bahkan berdasarkan asumsi umum ini, efek Brussels harus diambil dengan sedikit skeptis. Dengan bersikeras pada kekuatan yang secara sepihak diberikan oleh hukum Uni Eropa, tesis tentang efek Brussels sering mengabaikan berbagai cara di mana peraturan Uni Eropa berkaitan dengan koordinasi dan kerja sama.

Pertama, ketentuan ekstra-teritorial GDPR, yang mengacu pada pengalaman panjang dalam hukum konsumen, dilengkapi dengan perjanjian bilateral tentang pengakuan bersama di tingkat internasional, misalnya Jepang. Kedua, berurusan dengan regulasi teknologi, para pembuat undang-undang Uni Eropa lebih sering memilih solusi ko-regulasi tata kelola hukum, daripada pendekatan top-down. Pasal 5 GDPR tentang prinsip akuntabilitas memberikan ilustrasi model ko-regulasi tersebut. Ketiga, analisis model ko-regulasi yang diadopsi oleh hukum Uni Eropa dengan kebijakan 2017 tentang regulasi yang lebih baik dan cerdas, beberapa perkembangan teknis dari skema Regulasi Lebih Baik Uni Eropa untuk interoperabilitas, hingga 'Undang-Undang Tata Kelola Data' dari November 2020, bertemu dengan tren serupa di sektor hukum lainnya.

Pendekatan ko-regulasi sedang bekerja dengan badan-badan standardisasi, seperti NIST-800-53 dari 2013 dan NIST-800-63C dari 2016, bersama dengan ISO/IEC 27002 dan 27.001 tentang kontrol keamanan dan privasi untuk Sistem dan Organisasi Informasi Federal. Sejalan dengan itu, pendekatan ko-regulasi ini konsisten dengan beberapa model tata kelola di bidang bisnis, seperti kerangka kerja COBIT2019 yang diluncurkan oleh ISACA dan model Arsitektur Perusahaan, yang bertujuan untuk menyelaraskan sistem informasi manajemen dengan kepentingan bisnis.

Dengan menekankan tren tata kelola hukum dan hukum internasional saat ini, tujuannya bukanlah untuk mengabaikan efek Brussel. Saya telah mengakui dampak (unilateral) hukum perlindungan data Uni Eropa terhadap seluruh dunia dan siap mengakui bahwa

ketentuan-ketentuan tertentu dalam AIA tentang pelarangan penggunaan AI tidak hanya akan tetap berlaku, tetapi juga akan menjadi acuan dalam hukum internasional. Namun, setelah kita menerima skenario ini, perhatian harus diarahkan pada isi efeknya, dengan kata lain, yang akan memberikan efek ekstra-teritori unilateral di seluruh yurisdiksi, yang merupakan model bagi seluruh dunia. Perdebatan terkini tentang hukum Uni Eropa dan regulasi teknologi telah memunculkan beberapa mitos dan rumusan populer yang menarik, juga dalam kasus ini. Bagian selanjutnya mengkaji salah satu rumus tersebut: pendekatan 'berpusat pada manusia' terhadap tantangan normatif AI, atau 'HAI'. Sikap ini merangkum narasi dari bagian-bagian sebelumnya, berdasarkan tiga sikap tentang:

(i) HAI Uni Eropa untuk regulasi AI, sebagaimana diilustrasikan oleh proposal AIA dari Komisi, sebagai tindakan kedaulatan digital dalam hukum internasional;
(ii) Hak-hak baru Uni Eropa dalam interaksi manusia-AI yang ditetapkan oleh AIA sebagai penguatan lebih lanjut konstitusionalisme digital Uni Eropa;
(iii) Kemungkinan efek Brussel baru akibat (i) dan (ii).

Tujuan bagian selanjutnya adalah untuk memihak apakah HAI, yaitu pendekatan 'berpusat pada manusia' hukum Uni Eropa untuk regulasi sistem AI, kuat, atau malah menyesatkan.

13.5 'HAI' (KECERDASAN BUATAN YANG BERPUSAT PADA MANUSIA)

'HAI' memiliki kisah yang sudah panjang. Sejak pertengahan 2010-an, Parlemen Eropa menekankan 'nilai-nilai Eropa' yang seharusnya memandu regulasi sistem AI dan teknologi baru lainnya. Pada tahun 2018, Komisi Eropa membentuk Kelompok Pakar Tingkat Tinggi (HLEG) untuk menjelaskan prinsip-prinsip etika AI. HLEG menerbitkan Pedoman Etikanya pada tahun 2019.

Pedoman tersebut mencakup ketahanan lingkungan dan perlindungan kesejahteraan masyarakat dan lingkungan di antara enam persyaratan yang harus dipenuhi oleh sistem AI agar dapat dianggap tepercaya. Namun, dari sudut pandang filosofis, patut dicatat bahwa Pedoman Etika tersebut berulang kali menekankan pendekatan 'berpusat pada manusia': "fondasi bersama yang menyatukan hak-hak ini dapat dipahami berakar pada penghormatan terhadap martabat manusia dengan demikian mencerminkan apa yang kami sebut sebagai 'pendekatan yang berpusat pada manusia' di mana manusia menikmati status moral yang unik dan tak terelakkan sebagai keutamaan dalam bidang sipil, politik, ekonomi, dan sosial."

Dalam pandangan terbaiknya, klaim-klaim tersebut, dan deklarasi serupa, mungkin masuk akal. Pedoman etika HLEG, bagaimanapun, bergantung pada dokumen sebelumnya dari kelompok ahli lain, di mana saya dan rekan-rekan saya menekankan empat risiko AI, yaitu (i) merendahkan keterampilan manusia; (ii) menghilangkan tanggung jawab manusia; (iii) mengurangi kendali manusia; (iv) mengikis hak menentukan nasib sendiri manusia. Terhadap risiko-risiko tersebut, pemahaman yang jelas tentang isu-isu yang sedang diteliti dan inisiatif apa yang dapat diambil untuk melawan penyalahgunaan teknologi secara proaktif sangat disambut baik. Namun, HAI menimbulkan dua masalah besar. Satu bersifat filosofis, yang lainnya praktis. Mengenai sisi filosofis dari kisah ini, batasan setiap pendekatan yang berpusat pada manusia, atau neo-Protagorean, telah ditekankan berulang kali selama beberapa dekade

terakhir, sejak gerakan ekologi pada tahun 1950an dan 1960an, hingga peraturan dan prinsip hukum lingkungan Uni Eropa saat ini. Bioetika dan sikap ontosentrisnya menunjukkan banyak hal tentang tantangan normatif yang ditimbulkan oleh AI dan teknologi baru lainnya: "Perbandingan ini seharusnya tidak mengejutkan.

Dari semua bidang etika terapan, bioetika adalah yang paling mirip dengan etika digital dalam menangani bentuk-bentuk baru agen, pasien, dan lingkungan secara ekologis". Terdapat penelitian yang kuat tentang mengapa sudut pandang ontosentris, alih-alih antroposentris, dapat membantu kita mengatasi apa yang disebut Komisi Eropa, dalam Nota Penjelasan AIA, sebagai 'tantangan kembar', yaitu transformasi hijau dan digital dalam masyarakat kita.

Selain itu, terdapat bukti kekurangan praktis HAI. Dalam AIA, misalnya, Komisi Eropa sepenuhnya mendukung pendekatan yang berpusat pada manusia: semua persyaratan wajib baru untuk sistem AI berisiko tinggi tidak mencakup komitmen apa pun terhadap dampak lingkungan yang merugikan, agar sistem AI tersebut tidak menimbulkan ancaman langsung terhadap "kesehatan dan keselamatan, atau risiko dampak buruk pada hak-hak asasi." Pendekatan Komisi Eropa ini telah dikritik.

Laporan komite khusus Parlemen Eropa tentang Kecerdasan Buatan di Era Digital (AIDA) menganggap bahwa pendekatan tersebut hanya mengabaikan "bahaya apa pun yang terkait dengan lingkungan". Klaimnya adalah bahwa serangkaian aturan yang diusulkan tentang tata kelola AI dan data, transparansi, pengawasan dan keamanan manusia hanya mengabaikan sistem tata kelola yang seharusnya mencegah dampak kritis teknologi terhadap lingkungan. Lagipula, sebagian besar proposal tentang "keberlanjutan lingkungan" teknologi, termasuk AI, diserahkan kepada inisiatif sukarela yang diterapkan oleh penyedia sistem AI non-risiko tinggi, misalnya terkait pembentukan kode etik (AIA Komisi Uni Eropa, sedangkan no. 81 dan pasal 69.2).

Permasalahan hukum Uni Eropa terkait perlindungan lingkungan, harus diakui, lebih tua daripada isu-isu terkini tentang transformasi digital masyarakat kita dan regulasinya. Namun, pemahaman yang berpusat pada manusia tentang tantangan AI justru membuat transformasi hijau masyarakat kita semakin rumit. Hanya pendekatan ontosentris terhadap 'tantangan ganda' masyarakat kita yang tepat untuk tugas ini. Untuk memperkuat asumsi ini, sikap onto-sentris harus mencakup prinsip-prinsip bioetika yaitu, kebaikan, non-maleficence, otonomi, dan keadilan dan melengkapinya dengan prinsip baru, prinsip 'explicability'. Yang terakhir harus menggabungkan kecerdasan AI dan akuntabilitas atas penggunaannya, untuk memahami dan meminta pertanggungjawaban proses pengambilan keputusan AI.

Kita tidak perlu berpusat pada manusia, untuk mengakui risiko penyalahgunaan AI dan dampaknya terhadap keterampilan manusia, tanggung jawab manusia, kendali manusia, atau penentuan nasib sendiri manusia. Namun, kemungkinan besar setiap pendekatan yang berpusat pada manusia terhadap risiko-risiko ini akan gagal dalam menangani bagaimana keterampilan dan tanggung jawab manusia, kendali dan penentuan nasib sendiri tersebut harus dipahami lebih lanjut dalam kaitannya dengan tantangan perlindungan lingkungan dan krisis iklim.

Setidaknya, Komisi Eropa harus melengkapi proposal AIA-nya dengan penilaian dampak lingkungan AI dalam kerangka regulasi Eropa yang ada. Atas dasar ini, kita mungkin bertanya-tanya tentang metrik untuk penilaian dampak lingkungan AI, apakah penilaian jejaknya harus wajib untuk semua sistem AI berisiko tinggi, misalnya, atau diperluas ke aplikasi AI berisiko rendah tertentu. Demikian pula, fokus harus diberikan pada biaya energi dan emisi karbon, limbah elektronik, dan kondisi keberlanjutan lebih lanjut, misalnya, kondisi kerja, hingga metrik yang dioptimalkan untuk sistem AI, atau metrik efisiensi lebih lanjut untuk AI, seperti pelatihan model.

Teknologi AI tingkat lanjut seringkali membutuhkan sumber daya komputasi yang sangat besar yang bergantung pada pusat komputasi besar, dan fasilitas ini memiliki kebutuhan energi dan jejak karbon yang sangat tinggi. Beberapa perkiraan menunjukkan bahwa total permintaan listrik untuk teknologi informasi dan komunikasi (TIK) dapat mencapai 20% dari permintaan listrik global pada tahun 2030, sementara permintaan saat ini berkisar sekitar 1%. AI kemungkinan akan menambah kekhawatiran akan meningkatnya volume limbah elektronik dan tekanan pada unsur tanah jarang yang dihasilkan oleh industri komputasi.

Masalah terakhir dengan postur filosofis dan praktis HAI berkaitan dengan redundansinya. HAI tidak hanya tidak cukup untuk mengatasi tantangan onto-sentris dari transformasi hijau dan digital masyarakat kita dengan tepat, tetapi juga tidak membantu memperjelas teknis bidang kita. Misalnya, ada tradisi gemilang dalam robotika dan AI yang dikhususkan untuk mempelajari interaksi manusia-robot (HRI). Menariknya, para ahli membedakan dua sub-bidang dari disiplin ilmu ini. Beberapa berfokus pada pendekatan HRI yang berpusat pada manusia: penekanan di sini adalah pada apakah dan sejauh mana sistem AI dan robot memenuhi spesifikasi tugas mereka dengan cara yang tampak nyaman dan dapat diterima oleh manusia. Namun, ada juga pendekatan HRI yang berpusat pada robot: ini tidak berarti bahwa para ahli dan akademisi mengabdikan diri untuk mengurangi keterampilan manusia, atau merendahkan tanggung jawab manusia. Sebaliknya, yang ingin dipahami oleh para ilmuwan dan insinyur komputer adalah entitas, seperti robot pintar, yang mengejar tujuan 'sendiri', berdasarkan isyarat seperti motivasi, pendorong, atau emosinya.

Bidang-bidang pengembangan dan inovasi teknologi yang dinamis ini, misalnya, pembentukan 'mesin moral', telah didanai oleh program-program penelitian Uni Eropa (dan itu hal yang baik). Haruskah kita menyimpulkan bahwa, dalam semua proyek penelitian HRI yang berpusat pada robot, para akademisi harus mematuhi pendekatan yang berpusat pada manusia?

Pertanyaan ini redundan atau sangat bisa diperdebatkan. Hal ini redundan, karena peneliti AI harus mematuhi hukum; hal ini sangat bisa diperdebatkan, karena beberapa hukum, seperti hukum lingkungan Uni Eropa, bergantung pada basis ontosentris, misalnya Pasal 37 Piagam Hak Asasi Uni Eropa (CFR), dan Pasal 11 Perjanjian tentang Fungsi Uni Eropa (TFEU). Oleh karena itu, sebagaimana rumus-rumus menarik sebelumnya tentang kedaulatan digital, konstitusionalisme digital, dan efek Brussel, HAI juga harus ditanggapi dengan skeptis.

Kecurigaan yang sama kita terapkan pada diri kita sendiri ketika bertanya kapan matahari terbenam, atau akan terbit besok, meskipun kita bukanlah orang-orang yang tinggal

di Bumi, melainkan orang-orang Copernicus. AI menimbulkan tantangan unik bagi keterampilan dan tanggung jawab manusia, kendali manusia, dan penentuan nasib sendiri. Namun, keunikan ini tidak menyiratkan pandangan neo-Protagorean, melainkan harus dipahami sesuai dengan sikap ontosentris etika digital yang melengkapi keempat prinsip bioetika.

13.6 PELAJARAN MITOS UE: KEDAULATAN HINGGA HAI

Bab ini membahas tren hukum Uni Eropa terkini dan serangkaian proposal lanjutan dari Komisi, yang membongkar empat narasi atau 'mitos' populer tentang kedaulatan digital, konstitusionalisme digital, efek Brussel baru, dan HAI. Empat pelajaran dipetik dari sikap terhadap informasi hukum ‘tentang’ realitas ini, yaitu, tentang pengetahuan dan konsep yang membingkai representasi dan fungsi hukum Uni Eropa:

(a) Bertentangan dengan prinsip kedaulatan digital, perhatian diarahkan pada prinsip subsidiaritas berdasarkan Pasal 5 Perjanjian Uni Eropa dan tata kelola lembaga-lembaga Uni Eropa yang kompleks;
(b) Bertentangan dengan pandangan tentang konstitusionalisme digital Uni Eropa, batasan hukum Uni Eropa dalam hukum pidana, keamanan nasional, ketertiban umum, dan penegakan hukum ditekankan, untuk menawarkan gambaran yang lebih realistis tentang perdebatan dan tren terkini tentang bagaimana hukum seharusnya melindungi tubuh digital individu (juga, tetapi tidak hanya, dalam hukum pidana dan hukum administrasi);
(c) Bertentangan dengan para pendukung efek Brussel baru, pandangan tentang penerapan sepihak efek hukum ekstrateritorial ini dilengkapi dengan inisiatif bilateral pengakuan bersama di tingkat internasional dan model-model baru ko-regulasi, koordinasi, dan kerja sama di dalam Uni Eropa;
(d) Bertentangan dengan asumsi HAI, fokus diarahkan pada kelemahan filosofis dan praktisnya, serta bagaimana pendekatan ontosentris etika digital memberikan perspektif yang lebih baik terhadap tantangan ganda transformasi hijau dan digital dalam masyarakat kita.

Sikap terhadap tren hukum Uni Eropa saat ini menyoroti hal yang masih krusial: keseimbangan antara kekuatan Uni Eropa dan negara-negara anggota, generasi baru hak digital di tingkat konstitusional Uni Eropa dan MS, hingga interaksi antara model tata kelola hukum baru dan potensi fragmentasi sistem, misalnya antara regulasi teknologi dan hukum lingkungan. Mitos-mitos terkini tentang kedaulatan digital, konstitusionalisme digital, efek Brussel baru, dan HAI tidak membantu kita mengatasi permasalahan yang belum terselesaikan ini. Sebaliknya, mitos-mitos tersebut justru dapat mendorong kita untuk mengabaikannya. Meskipun beberapa permasalahan ini tidak bergantung pada hukum Uni Eropa dan lembaga-lembaganya, permasalahan tersebut berkontribusi dalam membentuk tren hukum Uni Eropa saat ini.

BAB 14
AI UNTUK PENALARAN KONTRAFAKTUAL DAN HUKUM

Abstrak Ketika berbicara tentang penilaian moral, kita mengacu pada fungsi mengenali tindakan yang pantas atau tercela dan kemungkinan pilihan di antara tindakan tersebut oleh agen. Kemampuan mereka untuk membangun kemungkinan urutan kausal memungkinkan mereka untuk merancang alternatif di mana memilih satu menyiratkan mengesampingkan yang lain. Musyawarah internal ini membutuhkan kemampuan kognitif, yaitu membangun argumen kontrafaktual. Ini berfungsi tidak hanya untuk menganalisis kemungkinan masa depan, karena bersifat prospektif, tetapi juga untuk menganalisis situasi masa lalu, dengan membayangkan keuntungan atau kerugian yang dihasilkan dari alternatif tindakan yang sebenarnya dilakukan, berdasarkan informasi evaluatif yang kemudian diketahui.

Pemikiran kontrafaktual dengan demikian merupakan prasyarat bagi agen AI yang terlibat dalam kasus Hukum, untuk memberikan penilaian dan, sebagai tambahan, untuk mengevaluasi tata kelola agen AI tersebut yang sedang berlangsung. Lebih lanjut, mengingat pemberdayaan kognitif yang luas dari penalaran kontrafaktual pada individu manusia, yaitu dalam membuat penilaian, muncul pertanyaan tentang bagaimana kehadiran individu dengan kemampuan ini dapat meningkatkan kerja sama dan konsensus dalam populasi yang biasanya hanya mementingkan diri sendiri.

Hasil penelitian kami, menggunakan Teori Permainan Evolusioner (EGT), menunjukkan bahwa pemikiran kontrafaktual mendorong koordinasi dalam masalah tindakan kolektif yang terjadi pada populasi besar dan memiliki dampak yang terbatas pada dilema kerja sama yang tidak memerlukan koordinasi tersebut.

14.1 KONTRAFAKTUAL HUKUM: KAUSALITAS DAN RASA BERSALAH

Hukum dengan jelas menyatakan bahwa teori kausalitasnya adalah ketergantungan kontrafaktual. Fokus teori kontrafaktual terletak pada moralitas. Minimum sosialnya adalah kita tidak melakukan hal yang merugikan. Tanggung jawab moral kita secara alami tercermin dalam semacam uji kontrafaktual tertentu, yang membandingkan bagaimana dunia setelah tindakan kita dengan bagaimana dunia akan terjadi jika, bertentangan dengan fakta, kita tidak melakukan tindakan tersebut. Demikian pula, seseorang dapat bernalar tentang tindakan alternatif yang akan memperbaiki dunia atau menghasilkan kebaikan yang lebih besar.

Kelompok pernyataan yang kami anggap kontrafaktual adalah pernyataan kondisional yang dihubungkan dengan kepalsuan anteseden dan klausa konsekuennya. Kontrafaktual, kemungkinan, dan hipotesis merupakan bagian dari asal mula keberadaan, dan keberadaan itu sendiri karena sebelumnya tidak demikian. Dalam kami juga mempertimbangkan kondisional yang antesedennya tidak diketahui dan mengevaluasi kondisional tersebut dengan menerapkan revisi dan abduksi untuk memenuhinya. Hukum fisika, misalnya, dapat diinterpretasikan sebagai pernyataan kontrafaktual, seperti "Seandainya berat pegas ini berlipat ganda, panjangnya juga akan berlipat ganda" (Hukum Hooke).

Kausalitas sebagai prasyarat tanggung jawab hukum berkaitan erat dengan kausalitas sebagai hubungan alamiah yang merupakan inti dari penjelasan ilmiah. Tanggung jawab moral bergantung pada sifat-sifat alamiah seperti kausalitas, intensi, dan sejenisnya. Teori kontrafaktual tentang hubungan kausal dominan dalam Hukum dan Filsafat modern. Kita lebih tercela ketika kita menyebabkan suatu kejahatan, daripada sekadar mencoba menyebabkannya. Kita mengalami penyesalan ketika kita telah menyebabkan kerugian meskipun kita sama sekali tidak bersalah. Bukan penyesalan, melainkan rasa bersalah yang mengganggu kita; dalam kasus seperti itu, kita menilai diri sendiri patut disalahkan. Rasa bersalah, bukan penyesalan, yang konsisten dengan penilaian diri semacam itu.

Ketika orang tergerak untuk berpikir kontrafaktual, mereka umumnya berpikir tentang bagaimana segala sesuatunya mungkin akan menjadi lebih baik ('kontrafaktual ke atas' dalam istilah psikologi eksperimental). Ketika memikirkan bagaimana suatu peristiwa mungkin menjadi lebih buruk, seseorang berbicara tentang 'kontrafaktual ke bawah'. Lysias dari Yunani, setelah Perang Peloponnesos (382 SM), mengatakan, "Jika kita tetap bersatu dan setiap orang melakukan apa yang saya lakukan, oligarki dan perang saudara tidak akan terjadi." Sejarawan dan jenderal Thucydides dari Yunani lainnya, tidak hanya menekankan betapa mengerikannya perang itu, tetapi juga menggarisbawahi momen-momen ketika perang itu mungkin lebih buruk bagi Athena dan warganya.

Menurut teori disosiasi penilaian, kontrafaktual ke atas cenderung berfokus pada tujuan fungsional untuk mengidentifikasi cara-cara mencegah hasil negatif. Pikiran-pikiran ini dapat membatalkan hasil tidak hanya dengan meniadakan penyebab langsung, tetapi juga dengan meniadakan kondisi yang memungkinkan atau menambahkan kondisi yang melemahkan. Hal ini menunjukkan bahwa ada lebih banyak cara seorang aktor dapat mencegah suatu hasil daripada cara-cara yang dapat menyebabkannya.

Oleh karena itu, kontrafaktual ke atas yang melibatkan diri sendiri cenderung menarik perhatian pada tindakan-tindakan yang melibatkan kesalahan. Penelitian menunjukkan bahwa program-program penjara yang dirancang untuk merangsang dan mengeksplorasi perilaku ke atas narapidana Pemikiran kontrafaktual tentang kejahatan, penangkapan, hukuman, dan vonis mereka dapat meningkatkan atribusi menyalahkan diri sendiri oleh narapidana, dan meningkatkan rasa bersalah mereka.

Studi teoretis kami tentang rasa bersalah, yang didasarkan pada Teori Permainan Evolusioner (EGT), memberikan bukti bahwa, dalam populasi yang sejak awal terdapat sedikit agen yang merasa bersalah, kerja sama yang lebih baik cenderung muncul seiring dengan penyebaran rasa bersalah. Selama beberapa dekade atau bahkan berabad-abad, para pengacara telah menggunakan uji yang relatif sederhana untuk menilai kesalahan terdakwa yang disebut 'tetapi karena sebab-akibat': "Kerugian tidak akan terjadi jika bukan karena tindakan terdakwa." Mengingat kondisional "Jika terdakwa melakukan tindakan A, maka kerugian I akan terjadi," kontrafaktualnya dapat memajukan anteseden menjadi penyebab dari konsekuensinya: "Jika terdakwa tidak melakukan tindakan A, maka kerugian I tidak akan terjadi." Klausul "but-for" juga bisa bersifat tidak langsung.

Jika Joe menghalangi pintu keluar kebakaran gedung dengan furnitur, dan Judy meninggal dunia setelah tidak dapat mencapai pintu keluar, maka Joe bertanggung jawab secara hukum atas kematiannya meskipun ia tidak menyalakan api. Demikian pula, pertanyaan utama dalam setiap kasus diskriminasi pekerjaan adalah apakah pemberi kerja akan mengambil tindakan yang sama seandainya karyawan tersebut berasal dari ras yang berbeda (usia, jenis kelamin, agama, asal usul, dll.).

Teori-teori psikologi sosial dan kognitif terkini mengusulkan arsitektur mental 'pemrosesan ganda'. Sebagian besar aktivitas pikiran dicapai melalui pemrosesan yang cepat, otomatis, dan sarat heuristik, contohnya adalah sistem visual kita. Sistem kognitif pertama ini sering disebut sistem otomatis, atau sistem intuitif, atau singkatnya 'sistem 1'. Namun, terkadang, kita perlu memikirkan suatu masalah, mempertimbangkan situasi kontrafaktual, mempertimbangkan asumsi, mempertimbangkan kemungkinan, dan secara sadar memutuskan solusi.

Pemikiran semacam ini lambat, melelahkan, dan mudah terganggu oleh tugas-tugas lain; terkadang disebut sistem penalaran, atau pemrosesan terkendali, atau singkatnya 'sistem 2'. Namun, tidak ada hal dalam sistem 2 yang menghalangi penggunaan heuristik secara sadar dan disengaja, yang secara umum disebut sebagai aturan praktis, sebuah domain utama studi AI. Hukum dan prosedur legislatif dapat mendorong orang untuk menggunakan kedua sistem tersebut. Kami telah mengkaji bagaimana penalaran kontrafaktual dapat digunakan untuk membahas tanggung jawab moral dan, lebih lanjut, menunjukkan bagaimana penalaran tersebut dapat dimanfaatkan untuk selanjutnya menghasilkan kebaikan yang lebih besar dan menghindari kerugian, setelah mengetahui hasil bersama dari tindakan seseorang dan orang lain dalam permainan sosial abstrak.

Dalam bab ini, kami berfokus pada penggunaan EGT untuk membuktikan mengapa dan bagaimana penalaran kontrafaktual yang diatur oleh AI dapat menjadi pendorong kerja sama dalam suatu populasi, dan pada kejadiannya dalam ranah tata kelola Hukum dan penerapan Hukum. Kami tidak akan membahas secara rinci isu tata kelola inovasi AI oleh Hukum, karena kami telah membahasnya di tempat lain, tetapi kami memberikan, di bagian 5, garis besar isu-isu regulasi AI tersebut. Sisanya disusun sebagai berikut. Pertama, kami mengingat kembali beberapa latar belakang sosial dan sejarah terkait masa lalu alternatif dan masa depan yang prospektif.

Selanjutnya, kami memberikan gagasan dasar tentang penalaran kontrafaktual. Hal ini diikuti dengan penggunaannya dalam model teori permainan evolusioner, yang secara intuitif diilustrasikan dengan contoh Stag-Hunt yang terkenal. Selanjutnya, kami mengajukan argumen untuk penggunaan penalaran kontrafaktual dalam hukum, khususnya dalam hal peningkatan Plea Bargaining bersama, dengan analogi dengan permainan Stag-Hunt, dan menguraikan konsekuensi yuridis positifnya. Setelah itu, kami akan membahas lebih detail penggunaan pemikiran kontrafaktual dalam pemodelan permainan evolusioner, dan akhirnya menyimpulkan dengan beberapa catatan.

14.2 BEBERAPA LATAR BELAKANG SOSIAL DAN SEJARAH

Hidup dalam masyarakat yang lebih baik pertama-tama membutuhkan perenungan tentang seperti apa masyarakat yang lebih baik itu. Tugas ini sama sekali tidak mudah. Sepanjang sejarah, manusia selalu membayangkan utopia. Ketika kita memikirkan Republik ideal Plato, atau Kota Tuhan St. Augustinus, atau Utopia Thomas Moro, atau Masyarakat Tanpa Kelas Karl Marx, kita selalu jauh dari masyarakat konkret.

Sepanjang sejarah kita, kita telah menghuni dunia sebagaimana adanya, tetapi membayangkan alternatif yang akan membuatnya lebih baik. Permainan dialektika antara ranah deskriptif dan ranah preskriptif ini sangat kaya dan bermanfaat. Tentu saja, kita belum pernah mencapai utopia apa pun sejauh ini; lebih lanjut, kita tidak yakin apakah, seandainya kita berhasil, utopia itu akan baik bagi umat manusia. Namun, baik atau buruk, utopia telah memainkan peran kunci dalam keputusan individu dan kolektif kita.

Dari sudut pandang kolektif, utopia telah memberikan model untuk menggambarkan apa yang kita bayangkan sebagai tujuan ideal. Kita terbiasa berpikir bahwa memiliki tujuan, atau maksud yang komprehensif, sangatlah positif. Namun, tujuan ini juga telah menimbulkan banyak kekerasan antarkelompok dengan kepentingan yang berseberangan.

Cukuplah untuk memikirkan berbagai Kediktatoran Proletar yang telah menjamur di seluruh planet ini, dan bagaimana, dengan dalih menciptakan masyarakat yang egaliter dan adil, mereka telah mengesahkan tindakan kekerasan ekstrem, dengan pembunuhan massal terhadap manusia. Di sisi lain, tanpa beragam kemungkinan utopia, kita akan relatif tersesat, karena kita tidak akan memiliki cukup keragaman dalam jawaban atas pertanyaan kolektif tentang ke mana kita ingin pergi. Kita membutuhkan keragaman ini agar tidak bergantung hanya pada satu kemungkinan. Bayangkan satu jawaban yang bersifat religius, misalnya untuk pertanyaan ini.

Jawaban itu tidak akan diterima oleh semua orang beriman, apalagi oleh orang yang tidak beriman. Bahkan tanpa mencapai konsensus tentang apa itu masyarakat ideal dan menerima gagasan bahwa berbagai dugaan tentangnya dapat hidup berdampingan, kita akan dengan tegas sepakat bahwa masyarakat manusia tidak boleh digunakan sebagai dalih untuk memperkaya 10% populasi dunia yang sedikit jumlahnya. Mengonsumsi semuanya, masing-masing, juga tidak mungkin memberikan makna yang kredibel bagi kehidupan individu dan kolektif kita. Namun, inilah yang semakin sering kita saksikan. Artinya, kita sedang menapaki jalan yang berbahaya, baik dalam hal kapasitas kita untuk mengidealkan (atau ketiadaannya), maupun dalam hal apa secara konkret yang kita lakukan untuk mencoba dan memperbaiki masa kini.

Merefleksikan isu-isu ini membutuhkan penerapan berpikir kritis, sebuah kapasitas yang kita akui langka. Memang, data dari Psikologi Sosial cukup simbolis dalam bidang ini; kita tahu dari eksperimen Salomon Asch bahwa persentase konformis dalam populasi tertentu jauh lebih tinggi daripada persentase nonkonformis. Kita juga tahu setidaknya sejak eksperimen Stanley Milgram bahwa kecenderungan untuk patuh kepada figur yang tampak berwibawa sangat kuat di antara manusia.

Jika pemberi perintah kredibel, jika ia menjalin hubungan dekat dengan pengikut perintah, pengikut tersebut akan melakukan apa pun yang diperintahkan kepadanya, tanpa melawan. Dalam konteks ini, kita harus mengangkat isu berpikir kritis dan konsepsi dunia alternatif. Mengharapkan setiap orang untuk bersikap nonkonformis, kritis, dan terinformasi akan menyiratkan keyakinan akan perubahan sosial yang sangat tidak mungkin, dengan konsekuensi yang sangat sulit diprediksi.

Di sisi lain, dalam ranah moralitas individu, salah satu persyaratan struktural untuk dapat menegaskan bahwa suatu tindakan tertentu bermoral adalah kemungkinan tindakan tersebut tidak dilakukan. Kewajiban bukanlah tentang kewajiban yang membatasi. Bahkan mengetahui apa itu kebaikan, seperti yang diakui Santo Paulus, kita dapat melakukan kejahatan: dalam ketegangan inilah martabat semua tindakan didirikan. Bahkan dalam teologi Kristen, masalah kehendak bebas menemukan jawaban yang selaras dengan pertanyaan tentang kejahatan. Artinya, Tuhan mengizinkannya atas nama kebaikan yang lebih besar, yaitu kebebasan. Jika kita hanya diberi satu pilihan, tidak akan ada martabat dalam memilihnya. Dalam ranah emosi pun, imajinasi skenario alternatif menempati tempat yang menonjol. Perhatikan situasi tokoh Camus dalam The Stranger: Jika cuaca tidak sepanas ini, jika tidak ada keputusasaan yang diakibatkannya, akankah ia membunuh orang Arab itu? Akankah ia tetap menjalani hukuman mati yang tidak perlu? Kemungkinan besar tidak.

Permainan antara apa yang ada dan apa yang mungkin terjadi ini, bukti adanya fungsi kognitif yang lebih tinggi, mendasari setiap spekulasi tentang kemungkinan dunia, dan memungkinkan kita mengantisipasi skenario respons. Nah, kemungkinan pra-adaptasi, evaluasi hasil, dan spekulasi tentang revisi strategis ini, merupakan inti dari penalaran hipotetis kontrafaktual. Bagaimana pendekatan ilmiah terhadap isu ini dapat membantu kita lebih memahami peran tersebut, dan bagaimana pendekatan tersebut berkaitan dengan isu moralitas?

14.3 TENTANG PENALARAN KONTRAFAKTUAL

Pemikiran Kontrafaktual (KT) adalah kemampuan kognitif manusia yang dipelajari dalam berbagai domain, yaitu Psikologi, Kausalitas, Keadilan, Moralitas, Sejarah Politik, Sastra, Filsafat, Logika, dan AI. Khususnya dalam AI, terdapat upaya berkelanjutan dalam pengembangan solusi algoritmik yang mampu mengidentifikasi penjelasan kontrafaktual terhadap keputusan yang dihasilkan oleh sistem otomatis. KT menangkap proses penalaran tentang peristiwa masa lalu yang tidak terjadi, yaitu, apa yang akan terjadi seandainya peristiwa tersebut terjadi, yang mungkin memperhitungkan apa yang kita ketahui saat ini. KT juga digunakan untuk bernalar tentang suatu peristiwa yang memang terjadi, mengenai apa yang akan terjadi jika peristiwa itu tidak terjadi; atau apakah peristiwa lain mungkin terjadi menggantikannya.

Contoh situasi: Petir menyambar hutan, dan kebakaran hutan yang dahsyat pun terjadi. Hutan itu kering setelah musim panas yang panjang dan terik, dan banyak hektar lahan hancur. Sebuah pemikiran kontrafaktual adalah: Seandainya tidak ada petir, kebakaran hutan tidak akan terjadi. Saat ini, terdapat penemuan kembali dan apresiasi terhadap peran kontrafaktual

dalam bidang Sastra, penelitian Sejarah, Psikologi Kognitif, Psikologi Moral, dan Kecerdasan Buatan, hanya untuk menyebutkan beberapa bidang yang lebih relevan.

Secara spesifik, dalam contoh ini, penalaran kontrafaktual terdiri dari membayangkan skenario alternatif yang berkaitan dengan skenario yang benar-benar terjadi, dan mengeksplorasi konsekuensinya: "Seandainya lantai hutan tidak tertutup daun kering setelah musim panas yang panjang dan terik, maka petir tidak akan menyebabkan kebakaran yang begitu dahsyat." Diterapkan pada moralitas kelompok, relevansinya berkaitan erat dengan konstruksi skenario hipotetis dan kredibel alternatif tentang masa lalu, pilihan yang dibuat, atau tentang peristiwa yang terjadi, dan, bersamaan dengan itu, penilaian berbagai konsekuensi yang akan mengikutinya.

Penalaran kontrafaktual yang dilakukan dengan tepat dapat memberikan wawasan yang sangat relevan tentang langkah-langkah ke depan dalam domain penerapannya. Dengan demikian, penalaran ini merupakan alat yang sangat baik untuk memahami dan menjelaskan mutabilitas perilaku tertentu, didukung oleh tinjauan strategi, dan peninjauan ulang masa lalu berdasarkan apa yang kita ketahui secara a posteriori saat ini. Kita dapat mengidentifikasi beberapa alasan yang mendorong individu membangun kontrafaktual: Kebutuhan untuk meningkatkan kinerja di masa depan, atau untuk memperbaiki suatu peristiwa faktual agar lebih dapat diterima oleh diri mereka sendiri, atau dapat dibenarkan oleh orang lain, baik karena alasan kita tidak mengejar alternatif tersebut, atau karena pengalaman mengajarkan kita tentang apa yang sebenarnya bisa kita lakukan secara berbeda dari apa yang telah kita lakukan.

Cara penalaran ini juga dapat diterapkan pada peristiwa yang tidak terjadi tetapi bisa saja terjadi. Misalnya, untuk menduga seperti apa rupa wilayah perkotaan Amerika Serikat jika, alih-alih membangun rel kereta api besar, investasi justru lebih banyak mengandalkan sungai sebagai sarana komunikasi. Atau tentang peristiwa-peristiwa yang terjadi, dengan demikian menalar apa yang akan terjadi seandainya peristiwa itu tidak terjadi; misalnya, membayangkan Revolusi Portugis 25 April 1974 tidak terjadi, dan bagaimana evolusi dari apa yang disebut "Musim Semi Marcellis" sebelumnya. Atau jika suatu peristiwa tertentu tidak terjadi, tetapi peristiwa lain akan terjadi, misalnya, jika eksploitasi besar-besaran bahan bakar fosil tidak terjadi, dan jika kita sudah beralih ke eksploitasi energi surya dan angin. Dan bahkan untuk memverifikasi apakah alternatif-alternatifnya akan acuh tak acuh terhadap konsekuensi yang relevan.

Dalam arti tertentu, kita dapat menganggap bahwa semua laboratorium ilmiah adalah tempat kontra-faktualitas, karena mereka menciptakan skenario alternatif, yang merupakan penyederhana realitas, di mana variabel tertentu dapat diuji. Artinya, realitas terlalu kaya dan kompleks untuk dijadikan tempat yang tepat untuk uji ilmiah tertentu.

Jika kita ingin tahu apakah "x" adalah penyebab "y", kita harus menciptakan skenario kontrafaktual yang memungkinkan hal ini dibuktikan. Faktanya, kita mungkin meramalkan kemunculan "y" dalam urutan waktu di mana "x" telah terjadi, dan ini terjadi secara berurutan karena "x" dikaitkan dengan "z" dan "z"-lah yang sebenarnya menyebabkan "y" dan juga "x". Menemukan hal ini dengan mengamati realitas mungkin sama sekali mustahil jumlah item

yang hadir bersamaan terlalu tinggi dan dapat menyebabkan kesalahpahaman yang tidak perlu dan keyakinan yang tidak berdasar.

Dengan demikian, di laboratorium, memiliki dugaan yang baik dan menguji satu variabel pada satu waktu memungkinkan kita untuk mengamati jaringan kausal yang tak terduga dan tidak ambigu. Ketika Galileo menduga bahwa dalam ruang hampa semua benda jatuh dengan kecepatan yang sama, memperoleh kecepatan yang sama pada waktu yang sama, terlepas dari massanya, ia tidak memiliki sarana teknis untuk menguji teori tersebut. Dari pengalaman mentalnya, ia merancang sistem saluran yang sangat halus tempat bola-bola dengan massa berbeda menggelinding.

Pemolesan saluran dan penyempurnaan bola dapat meminimalkan ketiadaan ruang vakum pada saat itu, karena gesekannya diminimalkan. Galileo kemudian membangun skenario yang memungkinkan pada masanya untuk menguji teori yang hanya sedikit orang yang mau menerimanya. Meskipun demikian, vakum sempurna, seperti yang kita ketahui saat ini, mustahil, karena ia niscaya terdiri dari fluktuasi vakum, yang tanpanya Prinsip Ketidakpastian Heisenberg akan dilanggar.

14.4 PENALARAN KONTRAFAKTUAL DAN KONFLIK KEPENTINGAN POPULASI BESAR

Secara khusus, mengenai penerapan penalaran kontrafaktual dalam ranah AI, pendekatan ilmiah terhadap pertanyaan moralitas dan penilaian dapat dikaji melalui pertimbangannya sebagai salah satu kasus model teori permainan yang diimplementasikan komputer. Teori permainan saat ini merupakan bahasa umum untuk mengkodekan setiap konflik kepentingan, dengan penerapan yang mencakup mulai dari teologi hingga ekonomi, mencakup ilmu komputer, matematika, fisika, antropologi, psikologi, dan banyak disiplin ilmu lainnya. Permainan juga diakui sebagai salah satu landasan uji coba utama yang mendasari kemajuan dalam kecerdasan buatan (AI), yang secara tepat disebut sebagai "Drosophila AI".

Secara umum, teori permainan mempelajari bagaimana, dalam hubungan strategis, pemain yang bertindak rasional menghasilkan hasil terbaik bagi diri mereka sendiri. Untuk melakukan ini, setiap pemain harus menganalisis permainan, dan mengidentifikasi strategi yang tersedia untuk mencapai tujuannya. Biasanya, pendekatan teori permainan klasik mengabaikan proses dinamis skala besar yang muncul dalam banyak skenario sosial dan sistem ekonomi serta politik modern. Sebaliknya, di sini kita akan berfokus pada analisis penalaran kontrafaktual yang terjadi dalam populasi besar, dengan mengadopsi varian dinamis dari teori permainan yang disebut Teori Permainan Evolusioner (EGT).

EGT mempertimbangkan populasi pemain yang berinteraksi melalui permainan, sebuah metafora dari sebuah konflik. Hasil yang diperoleh dari serangkaian interaksi tertentu dijumlahkan dan dikaitkan dengan kesuksesan sosial atau kebugaran individu. Dalam konteks alami, kita dapat mengatakan bahwa strategi yang berhasil bereproduksi lebih cepat. Dalam sistem sosial, strategi yang berhasil cenderung lebih sering ditiru dan dengan demikian akan menyebar dalam populasi. Hal ini menghasilkan kesamaan (formal dan dinamis) yang praktis antara pembelajaran sosial dan evolusi Darwin.

Dalam konteks sistem manusia, EGT memungkinkan penemuan pola perilaku yang paling mungkin ditemukan dalam populasi manusia, beserta mekanisme yang memungkinkan seseorang mencapai kondisi tersebut. EGT juga memungkinkan deskripsi kuantitatif baru tentang dinamika pengaruh teman sebaya, termasuk rasionalitas terbatas dan bias kognitif yang berkaitan dengan sebagian besar proses sosial.

Di sini kita akan mengilustrasikan ide-ide ini dalam konteks konflik kepentingan sederhana, yang dijelaskan oleh permainan non-kooperatif. Pertanyaan terkait kolaborasi atau tidak relevan dalam berbagai bidang seperti Psikologi Evolusioner, Biologi Evolusioner, Ekonomi, atau Hukum, dan lain-lain. Oleh karena itu, penting untuk mengetahui apakah penalaran kontrafaktual merupakan alat esensial untuk memahami dinamika perilaku, dan untuk meningkatkan pencapaian individu maupun kolektif dalam konteks di mana kolaborasi yang telah berevolusi memberikan manfaat terbesar.

Mengingat spektrumnya yang luas dan nilai kognitifnya, pertanyaan ilmiah yang relevan, dan bermanfaat dalam penelitian, adalah apa peran efektif, jika memadai, dari sekelompok kecil individu yang diberkahi rasionalitas kontrafaktual ini dalam suatu populasi tertentu. Lebih spesifik lagi, untuk memahami apakah minoritas ini misalnya 10% individu dapat memengaruhi seluruh kelompok, mendorong perilaku kooperatif berdasarkan kemampuan mereka untuk berpikir kontrafaktual demi kebaikan bersama.

Sangatlah relevan untuk menentukan apakah penalaran kontrafaktual, bahkan ketika diadopsi oleh minoritas, dapat memengaruhi pola perilaku kolektif. Yang penting, minoritas ini dapat mewakili sekelompok individu yang berbeda yang ingin mengadopsi penalaran yang lebih rinci dibandingkan dengan individu yang hanya belajar dari orang lain. Minoritas ini juga dapat dilihat sebagai agen atau algoritma buatan, meniru tantangan saat ini untuk memahami dunia hibrida yang akan segera kita hadapi, yang terdiri dari manusia dan mesin. Memang, selain bertujuan untuk memahami keputusan manusia dengan lebih baik, penelitian AI akan terus menyelidiki bagaimana kita dapat mendorong perilaku prososial dalam situasi di mana kerja sama tetap tidak ada atau berpotensi tidak muncul. Hal ini dapat dicapai dengan cara yang berbeda namun halus dengan mentransformasi sifat-sifat dilema yang dihadapi manusia, seperti yang diilustrasikan di bawah ini.

Kami juga menyinggung masalah yang sangat kompleks yang muncul dari moralitas yang bergantung pada sistem agama atau filsafat. Untuk menghindari masalah yang dihasilkan, pendekatan ilmiah akan memilih aspek-aspek yang fundamental bagi moralitas kelompok, yang dapat dikaitkan dengan semua konteks, terlepas dari budaya asli masing-masing kelompok, atau fakta bahwa agen otonom tersebut bersifat biologis, atau berbasis silikon. Ia akan membahas secara abstrak unsur-unsur katakanlah, unsur-unsur atom dari semua sistem moral, seperti: berkolaborasi atau tidak berkolaborasi, mengakui kesalahan dan meminta maaf, mengakui atau mengungkapkan niat, dan sebagainya; dan cara bagaimana aspek-aspek ini dapat atau tidak, secara individu atau saling terkait satu sama lain, memupuk kohesi kelompok.

14.5 PERBURUAN RUSA JANTAN: HUKUM DAN REGULASI AI

Berbekal dua peringatan dini yang telah disebutkan, mari kita telaah pendekatan kita terhadap peran kontrafaktual. Dalam kasus permainan Perburuan Rusa Jantan yang terkenal, dilema kerja sama direnungkan, yang membantu kita menetapkan pentingnya membangun kontrafaktual. Ini adalah permainan yang dimainkan oleh dua agen mana pun dalam suatu populasi, dan misi mereka yang terlibat adalah memburu rusa jantan, sebuah tugas yang harus dilakukan bersama-sama untuk memaksimalkan kemungkinan keberhasilan dan dengan imbalan yang besar.

Dengan demikian, kita juga dapat melihatnya sebagai metafora dari masalah koordinasi. Setiap pemain dapat memutuskan untuk tidak berkolaborasi dan memilih untuk mencoba memburu kelinci betina sendiri. Meskipun merupakan alternatif yang kurang menguntungkan, keputusan tersebut dapat diartikan lebih aman, karena pemburu hanya bergantung pada dirinya sendiri, dan kelinci betina lebih mudah diburu daripada rusa jantan.

Dilema ini muncul karena setiap pemburu tidak mengetahui apa yang akan dilakukan oleh pemburu lainnya; Artinya, apakah ia akan berkolaborasi dan berburu rusa jantan, atau akan bertindak sendiri, memutuskan untuk membelot dan berburu kelinci. Jadi, masing-masing pemain dapat tergoda untuk melindungi diri dengan berburu kelinci. Dengan kata lain, skenario yang paling kooperatif (kedua pemain memilih rusa jantan) tidak tercapai karena takut yang lain tidak akan mengikuti jalan yang sama. Imbalannya berbeda sesuai dengan setiap pilihan yang diambil. Seseorang dapat, misalnya, mempertimbangkan imbalan R sebesar 4 unit untuk keputusan berburu rusa jantan, jika diambil secara bersamaan oleh kedua pemain; imbalan sebesar 3 unit untuk keputusan berburu kelinci sendirian; dan 0 unit untuk pemain yang memutuskan berburu rusa jantan tanpa yang lain melakukannya.

Dengan demikian, kita menghadapi dilema kerja sama di mana maksimalisasi hasil bergantung pada keputusan efektif untuk bekerja sama oleh kedua pemain. Dalam konteks EGT, pemain meninjau strategi mereka, mengamati tindakan satu sama lain, dan meniru tindakan yang paling berhasil. Dalam ranah Hukum, contoh dilema koordinasi semacam itu berlimpah. Strategi yang dikenal sebagai *Plea Bargain* (PB) dapat meningkatkan hasilnya secara substansial jika didasari oleh kesimpulan abstrak dari permainan Stag Hunt. Bayangkan situasi whistleblowing ganda, di mana masing-masing dari dua pelaku dalam konteks imbalan seperti yang dialami para pemain Stag Hunt mengakui kesalahan yang lebih besar dari keduanya, sehingga memperoleh peningkatan PB yang menguntungkan, yang juga menguntungkan pihak Hukum.

Kemudian, kesimpulan kami yang dihasilkan memvalidasi peningkatan substansial dalam pandangan dan penggunaan PB saat ini, termasuk tata kelola Hukum yang lebih baik. Whistleblowing ganda saat ini tidak lagi dianggap sebagai tindakan yang lebih menguntungkan secara individual, karena apa pun yang divalidasinya hanya divalidasi oleh salah satu whistleblower, sehingga tidak menambah nilai pada bukti.

Hal ini mungkin masuk akal dalam konteks penetapan kesalahan dalam proses pidana; namun, whistleblowing ganda dapat memberikan Hukum bukti kesaksian yang lebih luas dan konfirmatif. Selain itu, jika dianalisis dari sudut pandang moralitas kelompok, sikap tentang PB

ini dapat dilihat sebagai upaya mendorong terjadinya beberapa PB. Hal ini tidak hanya mendorong pengakuan bersalah dalam populasi asal para terdakwa kejahatan, sesuatu yang kita tahu diinginkan, tetapi juga relevan untuk menyusun bukti forensik yang lebih kuat. Dengan demikian, bidang penelitian terbuka bagi para filsuf hukum yang tertarik pada moralitas evolusioner dan topik-topik yang bersifat menghakimi dengan menggunakan perangkat EGT.

Dari perspektif yang lebih umum, permainan Stag-Hunt juga merupakan contoh prototipe kontrak sosial, sebuah kesepakatan kolektif antara yang diperintah dan penguasa mereka, yang mendefinisikan tugas dan hak masing-masing. Dalam ranah ini, contoh permainan Stag-Hunt dapat ditemukan dalam tulisan-tulisan Rousseau, Hobbes, dan Hume. Smith dan Szathmary juga telah membahas analogi kontrak sosial yang tersirat dalam berbagai latar alami, yang dapat dipahami melalui lensa dinamika adaptif, evolusi budaya, dan pembelajaran sosial.

Dengan memasukkan dinamika kelompok yang terkait dengan jenis masalah ini, Stag-Hunt dapat dengan mudah digeneralisasikan ke situasi N-pemain di mana jumlah minimum kooperator diperlukan untuk berburu rusa jantan. Menetapkan ambang batas seperti itu meniru situasi yang umum terjadi pada sebagian besar upaya publik, di mana upaya gabungan minimum diperlukan untuk mencapai tujuan kolektif. Hal ini juga berlaku dalam perjanjian internasional, yang sering kali menuntut jumlah ratifikasi minimum untuk dapat dipraktikkan.

Adopsi undang-undang baru, baik di tingkat nasional maupun internasional, seperti yang terkait dengan aksi dan regulasi iklim, menawarkan contoh-contoh utama upaya kolektif yang dapat dibingkai sebagai N-player Stag-Hunt dari permainan koordinasi. Penyalahgunaan antibiotik, keraguan vaksinasi, dan bahkan mengoordinasikan populasi untuk mematuhi peraturan SARS-CoV-2, memberikan contoh lebih lanjut dari kelas dilema ini.

Dalam semua kasus, sifat non-linier dari pengembalian yang terkait dengan sistem adaptif yang kompleks ini (misalnya, seperti dalam kasus langkah-langkah kesehatan masyarakat), secara alami mengarah pada ambang batas dan tingkat adopsi yang kritis untuk menghasilkan dampak yang terukur. "Permainan" iklim dan kesehatan masyarakat memang memiliki kompleksitas tambahan karena sifat pengembalian yang tertunda dan tidak pasti, sebuah kompleksitas yang tidak akan kami uraikan lebih lanjut di sini.

Dilema lain dari kelas ini secara alami muncul dari diskusi yang sedang berlangsung tentang regulasi AI. Kemajuan teknologi yang pesat dalam AI, serta semakin meluasnya penerapan teknologi cerdas dalam domain aplikasi baru, telah menimbulkan kecemasan dan ketakutan akan ketinggalan di antara berbagai pemangku kepentingan, sehingga mendorong narasi perlombaan. Nyata atau tidak, keyakinan akan perlombaan untuk supremasi domain melalui AI dapat mewujudkannya, hanya dari konsekuensinya.

Konsekuensi ini dapat bersifat negatif, karena perlombaan untuk supremasi teknologi menciptakan ekologi pilihan yang kompleks yang dapat mendorong para pemangku kepentingan untuk meremehkan atau bahkan mengabaikan prosedur etika dan keselamatan. Akibatnya, berbagai aktor didesak untuk mempertimbangkan dampak normatif dan sosial dari kemajuan teknologi ini, dengan mempertimbangkan penggunaan prinsip kehati-hatian dalam

inovasi dan penelitian AI. Namun, hal ini menciptakan dilema regulasi baru, di mana non-linearitas dan ambang batas seperti yang dijelaskan di atas niscaya akan memainkan peran penting.

Menyetujui atau tidak menerapkan langkah-langkah ini melibatkan permainan koordinasi N-pemain lainnya, ditambah dengan dinamika inovasi yang terkait dengan sistem AI. Model teori permainan juga dapat digunakan dalam konteks ini. Kami menunjukkan bagaimana langkah-langkah regulasi ini dapat memberikan solusi untuk skenario tertentu, bergantung pada jangka waktu pengembangan produk AI dan risiko eksternalitas negatif. Namun, langkah-langkah ini juga dapat melampaui target, sehingga menghambat inovasi, dan menghambat investasi dalam mengembangkan inovasi baru karena menjadi upaya yang terlalu berisiko.

Kini, terlepas dari konflik atau contoh yang kita minati, jika kita ingin memiliki mesin yang diberkahi kapasitas moral, yang mampu memilih keputusan moral yang mengoptimalkan hasil yang diharapkan dan, setidaknya, memaksimalkan utilitas yang diharapkan (menggunakan paradigma utilitarian di sini, dengan beberapa pertimbangan), sangatlah penting bagi kita untuk belajar memprogramnya dengan kapasitas untuk mengembangkan skenario kontrafaktual. Skenario-skenario ini terbukti menjadi alat yang sangat baik untuk memilih alternatif yang tidak tersedia dalam portofolio perilaku untuk sekadar ditiru dan dapat menghasilkan peningkatan kohesi dan kerja sama dalam kelompok. Pernyataan ini memiliki relevansi eksperimental yang signifikan; menurut sebuah studi yang dilakukan oleh Departemen Kehakiman Inggris, dalam konteks rehabilitasi pelaku tindak pidana yang dihukum dalam kasus pengadilan, kasus residivisme sangat dimitigasi oleh strategi yang melibatkan penggunaan penalaran kontrafaktual.

Faktanya, dalam kegiatan mentoring yang bertujuan agar pelaku tindak pidana menjadikan alternatif lain sebagai alternatif dalam hidup, terbukti bahwa mereka yang memproses stimulus hingga menginginkan alternatif eksistensial lain adalah mereka yang paling kecil kemungkinannya untuk kembali ke ranah kriminal. Dalam semua contoh ini, penalaran kontrafaktual juga dapat digunakan untuk menilai, secara moral, niat dari tindakan suatu agen. Seseorang secara kontrafaktual berasumsi bahwa efek samping berbahaya tertentu yang terjadi mungkin tidak terjadi. Meskipun demikian, akankah tujuan agen yang bertindak tercapai? Jika tidak, maka efek samping ini sangat diperlukan dan, oleh karena itu, mungkin disengaja. Jika demikian, maka tidak perlu untuk mencapai tujuan agen, dan oleh karena itu, efek berbahaya tersebut tidak perlu disengaja.

14.6 PERMAINAN EVOLUSIONER DENGAN PEMIKIRAN KONTRAFAKTUAL (CT)

Di bagian ini, kami mengilustrasikan bagaimana penerapan pemikiran kontrafaktual pada Perburuan Rusa Jantan bertentangan dengan apa yang terjadi dengan proses mimetik yang diusulkan oleh teori pembelajaran sosial individu dapat menduga apa yang akan terjadi jika ia menggunakan strategi lain sebagai strateginya sendiri (seperti berkolaborasi) alih-alih strategi yang sebenarnya ia gunakan (seperti membelot). Kami berangkat dari pemodelan komputer agen buatan yang umum untuk mengilustrasikan bahwa penalaran kontrafaktual

jauh lebih efisien dan efektif dalam merevisi strategi daripada sekadar meniru strategi paling sukses yang digunakan oleh musuh.

Perhatikan bahwa permainan ini juga menunjukkan bahwa penciptaan kontrafaktual hanyalah aktivitas mental instrumental yang sepenuhnya bergantung pada diri sendiri. Artinya, ini juga merupakan sumber daya yang tersedia bagi mereka yang secara sistematis memilih strategi egois. Terdapat kontrafaktualitas untuk kebaikan, dan untuk kejahatan ... (misalnya, Mafia).

Mengingat pemberdayaan kognitif CT yang luas pada individu manusia, muncul pertanyaan tentang bagaimana kehadiran individu dengan strategi yang didukung CT memengaruhi evolusi kerja sama dalam populasi yang terdiri dari individu-individu dengan beragam strategi interaksi. Yang penting, bergantung pada permainan dan strategi terkait, individu dapat merevisi strategi mereka dengan cara yang berbeda. Asumsi umum teori permainan klasik adalah bahwa pemain bersifat rasional, dan bahwa Ekuilibrium Nash merupakan prediksi yang wajar tentang apa yang diadopsi oleh agen rasional yang mementingkan diri sendiri. Namun, seringkali pemain memiliki kemampuan kognitif yang terbatas atau menggunakan heuristik yang lebih sederhana untuk merevisi pilihan mereka. Teori permainan evolusioner (EGT) menawarkan solusi untuk situasi ini, dengan mengadopsi deskripsi populasi interaksi permainan di mana individu menggunakan pembelajaran sosial dan imitasi. Akibatnya, strategi yang berhasil menyebar dalam populasi.

Namun, berbeda dengan pembelajaran sosial, agen yang lebih canggih (seperti manusia) justru mungkin membayangkan bagaimana hasil yang lebih baik bisa diperoleh, jika mereka memutuskan secara berbeda, dan kemudian belajar sendiri dengan merevisi strategi mereka. Di sinilah Pemikiran Kontrafaktual (CT) berperan. Di sini, kami sebelumnya telah mengusulkan model matematika untuk mempelajari dampak pada kerja sama dari populasi agen yang menggunakan penalaran kontrafaktual semacam itu, jika dibandingkan dengan populasi yang hanya terdiri dari pembelajar sosial. Secara khusus, kami menjawab positif untuk tiga pertanyaan utama:

1. Dapatkah kita memformalkan revisi perilaku kontrafaktual dalam populasi besar (dengan mengambil dinamika kerja sama sebagai studi kasus aplikasi)?
2. Akankah kerja sama muncul dalam dilema kolektif jika, alih-alih dinamika evolusi dan pembelajaran sosial, individu merevisi pilihan mereka melalui pemikiran kontrafaktual?
3. Apa dampak pada tingkat kerja sama secara keseluruhan dari adanya sebagian kecil pemikir kontrafaktual dalam populasi pembelajar sosial? Apakah kerja sama diuntungkan oleh keragaman metode pembelajaran tersebut?

CT dapat dilakukan setelah mengetahui hasil yang diperoleh setelah satu langkah bermain dengan rekan pemain. CT menggunakan pemikiran kontrafaktual: Seandainya saya bermain berbeda, apakah saya akan mendapatkan hasil yang lebih baik daripada yang saya dapatkan? Informasi ini dapat diperoleh dengan mudah dengan melihat matriks hasil permainan, dengan asumsi rekan pemain akan melakukan permainan yang sama, yaitu, jika faktor-faktor lain tetap sama. Dalam kasus positif, pemain CT akan belajar untuk selanjutnya mengadopsi strategi

permainan alternatif. Dalam EGT, bentuk pembelajaran standar yang sering digunakan adalah apa yang disebut Pembelajaran Sosial (SL).

Pembelajaran ini pada dasarnya terdiri dari mengubah strategi seseorang dengan meniru strategi individu yang lebih sukses dalam populasi, dibandingkan dengan kesuksesannya sendiri. CT, sebaliknya, dapat dibayangkan sebagai bentuk pembelajaran pembaruan strategi yang mirip dengan debugging, dalam arti: jika langkah permainan saya yang sebenarnya tidak menghasilkan hasil yang baik dan terakumulasi, maka, setelah mengetahui langkah rekan pemain, saya dapat membayangkan bagaimana saya akan bermain lebih baik jika saya membuat pilihan strategi yang berbeda.

Dibandingkan dengan SL, jenis penalaran ini kemungkinan memiliki dampak yang kecil dalam permainan kerja sama dengan satu ekuilibrium Nash (atau satu strategi stabil evolusioner, dalam konteks EGT) seperti Dilema Tahanan atau permainan Barang Publik, di mana dominasi pembelotan berlaku. Namun, seperti yang diilustrasikan di bawah ini, pemikiran kontrafaktual berpotensi memiliki dampak yang kuat dalam permainan koordinasi, yang dicirikan oleh beberapa Ekuilibria Nash: CT akan memungkinkan penalaran meta yang ekuilibrianya memberikan hasil yang lebih tinggi.

Mari kita pertimbangkan populasi berukuran Z di mana individu-individu terlibat dalam dilema N-Stag-Hunt (lihat di atas) yang dicirikan oleh serangkaian perilaku terbatas: bekerja sama atau membelot. Para kooperator (Cs) menyumbangkan biaya c untuk barang publik, sementara pembelot (D) menolak untuk melakukannya. Kontribusi yang terakumulasi dikalikan dengan faktor peningkatan F, dan hasil yang dihasilkan didistribusikan secara merata di antara semua individu dalam kelompok, terlepas dari apakah mereka berkontribusi atau tidak. Persyaratan koordinasi diperkenalkan dengan memperhatikan bahwa kita sering menemukan situasi di mana jumlah minimum M dari Cs diperlukan dalam suatu kelompok untuk menciptakan manfaat kolektif.

Apa dampak keberadaan sebagian kecil pemikir kontrafaktual dalam populasi pembelajar sosial terhadap tingkat kerja sama secara keseluruhan? Apakah kerja sama diuntungkan oleh keberagaman metode pembelajaran tersebut? Untuk menjawab pertanyaan-pertanyaan ini, kami mengembangkan model dinamika populasi baru berdasarkan permainan evolusi, yang memungkinkan perbandingan langsung antara dinamika perilaku yang diciptakan oleh individu yang merevisi perilaku mereka melalui pembelajaran sosial dan melalui pemikiran kontrafaktual.

Pada Gambar 1a, kami mengilustrasikan dinamika perilaku baik di bawah CT maupun SL untuk parameter yang sama dari permainan Stag-Hunt N-orang. Untuk setiap fraksi kooperator (C), jika gradien G (untuk SL maupun CT) positif (negatif), maka kemungkinan fraksi C akan meningkat (menurun). Sebagaimana ditunjukkan, dalam kedua kasus, dinamika dicirikan oleh dua cekungan tarik-menarik dan dua titik tetap interior: satu tidak stabil (juga dikenal sebagai titik koordinasi), dan keadaan koeksistensi yang stabil antara C dan D.

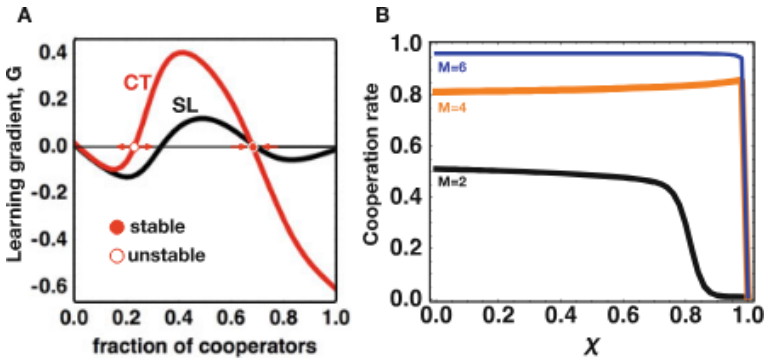
Untuk mencapai tingkat kerja sama yang stabil (dalam keadaan koeksistensi), individu harus berkoordinasi agar dapat mencapai cekungan tarik-menarik kooperatif di sisi kanan plot, sebuah fitur umum dalam banyak dilema barang publik non-linier. Gambar 14.1 juga

menunjukkan bahwa CT memungkinkan terciptanya strategi bermain baru, yang sebelumnya tidak ada dalam populasi, karena strategi baru dapat muncul secara spontan berdasarkan penalaran individu.

Dengan demikian, CT secara menarik menghasilkan hasil yang berbeda jika dibandingkan dengan SL. Dalam skenario khusus ini, terlihat jelas bagaimana CT dapat memfasilitasi koordinasi tindakan, karena individu dapat bernalar tentang hasil suboptimal yang terkait dengan tidak tercapainya ambang batas koordinasi, dan secara individual bereaksi terhadapnya.

Pada Gambar 14.1a, individu dapat merevisi strategi mereka melalui pembelajaran sosial atau penalaran kontrafaktual. Namun, seseorang juga dapat membayangkan situasi di mana setiap agen dapat menggunakan CT dan SL dalam keadaan yang berbeda, situasi yang cenderung terjadi pada populasi manusia. Untuk mencakup heterogenitas tersebut pada tingkat agen, mari kita pertimbangkan model sederhana di mana agen menggunakan SL dengan probabilitas χ, dan menggunakan CT dengan probabilitas $(1-\chi)$.

Pada Gambar 1b, kami menunjukkan dampak χ terhadap tingkat kerja sama rata-rata dalam dilema Stag-Hunt N-orang di mana, tanpa adanya CT, kerja sama tidak mungkin bertahan. Hebatnya, hasil kami menunjukkan bahwa sedikit prevalensi individu yang menggunakan CT sudah cukup untuk mendorong seluruh populasi pembelajar sosial menuju standar yang sangat kooperatif, memberikan indikasi lebih lanjut tentang kekokohan kerja sama yang didorong oleh penalaran kontrafaktual. Hasil ini menjadi lebih jelas ketika koordinasi lebih sulit dicapai (yaitu, ambang batas koordinasi yang lebih besar, M).



Gambar 14.1 (a) Panel kiri: Gradien pembelajaran untuk pembelajar sosial (SL, garis hitam) dan pembelajar kontrafaktual (CT, garis merah) untuk permainan SH N-orang. Jika gradien pembelajaran positif (negatif), fraksi kooperator akan cenderung meningkat (menurun). Lingkaran kosong dan lingkaran penuh masing-masing merepresentasikan analogi populasi terhingga dari titik tetap tidak stabil dan stabil. Panel kanan: Distribusi stasioner proses Markov yang dihasilkan oleh probabilitas transisi yang digambarkan pada panel kiri; distribusi ini mengkarakterisasi prevalensi setiap fraksi kooperator dalam populasi terhingga. (b) Panel kanan: Kerja sama secara keseluruhan sebagai fungsi dari prevalensi individu yang menggunakan pembelajaran sosial (SL, χ) dan penalaran kontrafaktual (CT, 1-χ). Hal ini menunjukkan bahwa hanya prevalensi pemikiran kontrafaktual yang relatif kecil yang diperlukan untuk mendorong kerja sama pada seluruh populasi agen yang mementingkan diri sendiri. Parameter lain: Z = 50, N = 6, F = 5,5. $M = N/2$ (panel A), c = 1,0, μ = 0,01, $\beta_{SL} = \beta_{CT}$ = 5,0

Hasil ini mungkin memiliki berbagai implikasi menarik jika populasi heterogen dipertimbangkan. Misalnya, kita dapat membayangkan masa depan yang dekat dengan masyarakat hibrida yang terdiri dari manusia dan mesin. Dalam skenario seperti itu, penting untuk tidak hanya memahami bagaimana perilaku manusia berubah di hadapan entitas buatan, tetapi juga untuk memahami sifat apa yang harus disertakan dalam agen buatan yang mampu memanfaatkan kerja sama antarmanusia. Hasil kami menunjukkan bahwa sebagian kecil agen CT buatan dalam populasi pembelajar sosial Manusia dapat secara signifikan memengaruhi dinamika kerja sama menuju keadaan kooperatif.

14.7 KONTRAFAKTUAL AI: KOORDINASI STAG-HUNT

Kami berpendapat bahwa penalaran atau pemikiran kontrafaktual merupakan perangkat kognitif dengan sejarah panjang manusia, yang menyediakan dasar bagi penjelasan kausal, dan karenanya bagi atribusi kesalahan dalam penilaian moral dan yudisial. Kami mengilustrasikan dampak potensial penalaran kontrafaktual dalam konteks dilema barang publik non-linier, yang juga dikenal sebagai permainan Stag-Hunt N-pemain, suatu kelas dilema yang relevan dalam berbagai domain, mulai dari hukum dan kesehatan masyarakat hingga perjanjian internasional dan regulasi AI. Hasil kami menunjukkan bahwa pembelajar kontrafaktual mendorong koordinasi dalam dilema kolektif semacam ini, yang mentransformasi dinamika perilaku yang biasanya terkait dengan permainan ini.

Kami juga menunjukkan bagaimana pembelajar kontrafaktual ini dapat memengaruhi orang lain. Khususnya, di era yang semakin dibentuk oleh sistem dan artefak cerdas yang memperkuat kemampuan manusia untuk memanipulasi informasi, hal ini mendorong pemahaman tentang bagaimana instrumen tersebut dapat mengubah perilaku manusia dan meningkatkan kapasitas kita untuk bekerja sama. Dalam ranah ini, hasil kami menunjukkan bahwa sebagian kecil agen buatan yang menggunakan CT mampu mengarahkan kerja sama manusia ketika ditempatkan dalam populasi hibrida yang terdiri dari manusia dan mesin. Efek serupa telah terbukti hadir dalam konteks dilema lain.

Jelas, proses pengambilan keputusan nyata di antara manusia melibatkan kompleksitas yang melampaui batas yang kita gunakan untuk menggambarkan ide-ide ini. Di sisi lain, kesederhanaan konseptual model-model ini membuatnya dapat diterapkan secara umum pada berbagai masalah yang melibatkan tindakan kooperatif kolektif, yang muncul dalam berbagai situasi yang saling bertentangan di alam dan masyarakat, sehingga memberikan wawasan tentang kekayaan, keindahan, keragaman, dan kompleksitas interaksi sosial kolektif. Akhirnya, penelitian kami sebelumnya tentang etika mesin mendorong kami untuk mempertimbangkan dan berargumen tentang dampak positif terhadap tata kelola Hukum dan penerapannya terkait keuntungan dari promosi Tawar-Menawar Pembelaan bersama, berdasarkan analogi situasionalnya dengan permainan evolusi Stag-Hunt dan hasil-hasilnya. Lebih lanjut, penelitian kami sebelumnya tentang rasa bersalah, menunjukkan keuntungan melatih tahanan dalam berpikir kontrafaktual tentang tindakan dan pilihan alternatif mereka, dengan tujuan mengasah rasa moral mereka, mempercepat pembebasan bersyarat mereka, dan meningkatkan perilaku mereka di masa mendatang.

Memang, terdapat intuisi yang kuat bahwa antisipasi penyesalan (atas hasil yang tidak diinginkan) merupakan faktor signifikan dalam pengambilan keputusan. Secara umum, teori penyesalan menyiratkan bahwa daya tarik suatu pilihan tidak dapat dievaluasi tanpa mengacu pada konteks pilihan lain yang tersedia. Karena penyesalan merupakan respons terhadap hasil kontrafaktual dari pilihan yang berbeda, pengetahuan yang diharapkan dimiliki oleh pembuat keputusan tentang hasil tersebut seharusnya memengaruhi antisipasi penyesalan. Pengetahuan tentang matriks hasil permainan memungkinkan evaluasi kemungkinan hasil alternatif.

BAB 15
PENGAMBILAN KEPUTUSAN PERADILAN DI ERA AI

Abstrak Kecerdasan buatan (AI) telah menjadi bagian yang merambah hampir di setiap aspek masyarakat dan bisnis: mulai dari pemberian skor kredit kepada orang-orang, mengidentifikasi kandidat terbaik untuk suatu posisi pekerjaan, hingga pemeringkatan pelamar untuk penerimaan di universitas. Salah satu inovasi paling mencolok dalam sistem peradilan pidana Amerika Serikat dalam tiga dekade terakhir adalah pengenalan perangkat lunak penilaian risiko, yang didukung oleh algoritma canggih, untuk memprediksi apakah pelaku individu kemungkinan akan mengulangi tindak pidananya.

Fokus kontribusi ini adalah pada penggunaan alat penilaian risiko ini dalam penjatuhan hukuman pidana. Terlepas dari pertimbangan sosial, etika, dan hukum yang lebih luas, hingga saat ini, belum banyak yang diketahui tentang bagaimana persepsi teknologi memengaruhi kognisi dalam pengambilan keputusan, terutama dalam konteks hukum. Penelitian menunjukkan bahwa manusia cenderung mempercayai algoritma sebagai sesuatu yang objektif, dan, dengan demikian, tidak dapat dibantah. Kontribusi ini meneliti dua fenomena dalam hal ini: (i) “efek teknologi” kecenderungan manusia terhadap optimisme berlebihan ketika membuat keputusan yang melibatkan teknologi; dan (ii) “bias otomatisasi” fenomena di mana hakim menerima rekomendasi dari sistem pengambilan keputusan otomatis, dan berhenti mencari bukti konfirmasi, bahkan mungkin mengalihkan tanggung jawab pengambilan keputusan ke mesin.

15.1 AI DAN ALGORITMA: PENGAMBILAN KEPUTUSAN OTOMATIS

Semakin berkembang, sistem Kecerdasan Buatan (AI) dan algoritma yang menggerakkannya ditugaskan untuk membuat keputusan krusial yang sebelumnya dibuat oleh manusia. Pengambilan keputusan algoritmik berdasarkan big data telah menjadi alat penting dan merambah ke semua aspek kehidupan kita sehari-hari: artikel berita yang kita baca, film yang kita tonton, orang-orang yang kita habiskan waktu bersama, apakah kita diperiksa di antrean keamanan bandara, apakah lebih banyak petugas polisi dikerahkan di lingkungan kita, dan apakah kita memenuhi syarat untuk mendapatkan kredit, layanan kesehatan, perumahan, pendidikan, dan kesempatan kerja, di antara segudang keputusan komersial dan pemerintah lainnya.

Karena "keajaiban" teknologi telah menjadi begitu umum, karena pengaruhnya terhadap kehidupan kita begitu mendalam, dan karena kebanyakan dari kita kurang memahami cara kerjanya, makna "teknologi" yang dikonstruksi secara sosial telah secara implisit dikaitkan dengan optimisme terhadap apa yang akan dihadirkan teknologi di masa depan. Hingga saat ini, belum banyak yang diketahui tentang bagaimana persepsi teknologi memengaruhi kognisi dalam pengambilan keputusan, terutama dalam konteks hukum. Kehadiran teknologi yang merajalela di hampir setiap aspek masyarakat dan bisnis dan penyebarannya yang semakin pesat dalam hukum menjadikan hal ini isu yang krusial.

Deskripsi klasik tentang proses pengadilan biasanya menekankan martabat, kelambatan, dan keahlian hukum yang telah lama dihormati para hakim dan jaksa dalam sistem peradilan pidana. Namun, kini, pengadilan telah menjadi tempat berkembangnya analitik data dan algoritma. Salah satu inovasi paling mencolok dalam sistem peradilan pidana Amerika Serikat selama tiga dekade terakhir adalah pengenalan perangkat lunak penilaian risiko, yang didukung oleh algoritma canggih dan seringkali bersifat hak milik, untuk memprediksi apakah pelaku individu cenderung mengulangi tindak pidana (yang disebut "risiko residivisme"). Fokus bab ini adalah pada penggunaan terbaru, dan mungkin yang paling meresahkan, dari alat penilaian risiko ini: penggabungannya ke dalam proses penjatuhan hukuman pidana.

Secara umum, otomatisasi dapat meningkatkan konsistensi dan prediktabilitas pengambilan keputusan dengan mengurangi kesewenang-wenangan yang umum terjadi dalam keputusan manusia. Dengan banyaknya pertanyaan serupa, algoritma akan memberikan jawaban yang dapat diprediksi dan konsisten. Simon Chesterman menyatakan: "Sementara banyak keputusan evaluatif yang dibuat oleh manusia didasarkan pada bias bawah sadar dan reaksi intuitif, algoritma mengikuti parameter yang ditetapkan untuk mereka."

Banyak akademisi dan praktisi memandang sistem penilaian risiko otomatis sebagai jalur yang menjanjikan menuju penjatuhan hukuman yang lebih efisien, tidak bias, dan berbasis empiris. Mengganti pengambilan keputusan diskresioner hakim dengan pengambilan keputusan otomatis yang terstruktur dan kuantitatif, demikian argumennya, akan mencegah hakim dari "menjatuhkan hukuman secara membabi buta", yaitu "menghukum berlebihan" (memenjarakan pelaku yang risikonya kecil terhadap keselamatan publik) atau "menghukum kurang" (melepaskan penjahat berbahaya ke masyarakat untuk mengulangi kejahatannya).

Sistem penilaian risiko otomatis sering dikatakan meminimalkan tingkat dan lamanya hukuman bagi pelaku berisiko rendah, sehingga menghasilkan biaya anggaran yang lebih rendah dan mengurangi dampak sosial. Selain itu, algoritma prediktif dapat menghemat waktu berharga bagi hakim, jaksa, dan staf pengadilan yang bekerja terlalu keras.

Masih sangat terbatas penelitian empiris tentang apakah algoritma penilaian risiko otomatis benar-benar mencapai tujuan-tujuan ini. Namun, terdapat penelitian signifikan yang menunjukkan bahwa alat-alat otomatis ini menghasilkan hasil yang bias karena variabel sosioekonomi, bias dalam data, dan prediksi yang tidak akurat. Sejumlah faktor telah diidentifikasi yang menjadikan alat penilaian risiko otomatis lebih mirip kotak Pandora daripada obat mujarab. Misalnya, algoritma hanya sebaik data yang diberikan dan pertanyaan yang diajukan. Dalam praktiknya, algoritma dapat memperkuat disparitas yang ada. Data dan pertanyaan ini sama sekali bukan latihan statistik yang netral; praktisi diharuskan untuk mengajukan serangkaian pertanyaan terarah tentang riwayat kriminal, aktivitas rekreasi, pendidikan, hukuman pidana sebelumnya, rekan kerja, keluarga dan hubungan, kesejahteraan emosional, perumahan, penyalahgunaan zat, pengasuhan anak dalam keluarga, sikap, bantuan sosial, keuangan, pekerjaan, dan berbagai isu lainnya.

Oleh karena itu, tidak mengherankan bahwa, secara umum, di Amerika Serikat, status sebagai orang Afrika-Amerika kemungkinan besar akan menghasilkan klasifikasi berisiko tinggi,

karena banyak orang Afrika-Amerika hidup dalam kondisi kemiskinan, dan memiliki karakteristik "berisiko tinggi" (misalnya masalah sosial, pendidikan, pekerjaan, dan keluarga, penyalahgunaan zat terlarang, serta riwayat trauma dan kekerasan). Penelitian telah menunjukkan bahwa ras dan gender merupakan konstruksi sosial yang kompleks yang tidak dapat begitu saja direduksi menjadi variabel biner dalam sistem penilaian risiko otomatis.

Dengan demikian, sistem penilaian risiko otomatis gagal mengendalikan secara memadai disparitas gender atau ras dan potensi hasil diskriminatif. Singkatnya, sistem otomatis memutus hubungan antara hukuman dan tindakan individu, kurang bernuansa, dan mengabaikan pentingnya berbagai faktor motivasi, kontekstual, dan struktural yang berkontribusi pada tindakan manusia. Namun, fokus bab ini adalah pada potensi dampak yang mungkin ditimbulkan oleh alat-alat ini terhadap bagaimana hakim menggunakan diskresi mereka sendiri dalam menjatuhkan hukuman, meskipun mereka sendiri tidak merasakan adanya perbedaan. Meskipun algoritma penilaian risiko ini belum mengambil keputusan untuk menggantikan hakim (belum), belum jelas bagaimana hakim harus mengintegrasikannya ke dalam proses pengambilan keputusan mereka, atau bagaimana algoritma tersebut dapat memengaruhi keputusan mereka. Kuantifikasi konon membantu menjaga akuntabilitas hakim dan membuat putusan lebih konsisten dan efisien. Namun, masalahnya adalah masih sedikit yang diketahui tentang kemanjuran intervensi semacam itu.

Sebagaimana dibahas di bawah, penelitian menunjukkan bahwa manusia cenderung mempercayai algoritma sebagai sesuatu yang objektif, dan, dengan demikian, tidak dapat dibantah. Istilah "bias otomatisasi" menggambarkan fenomena di mana hakim menerima panduan atau rekomendasi dari sistem pengambilan keputusan otomatis, dan berhenti mencari bukti konfirmasi, bahkan mungkin mengalihkan tanggung jawab pengambilan keputusan kepada mesin.

15.2 PROSES PENJATUHAN HUKUMAN

Hakim memang memandang penjatuhan hukuman sebagai tanggung jawab yang berat. Dalam konteks praperadilan, keputusan hakim pada dasarnya bersifat biner: haruskah terdakwa tetap di penjara selama masa praperadilan, atau tidak? Namun, setiap perkara pidana mengandung segudang fakta, faktor, dan fitur yang dapat memengaruhi hukuman yang akan dijatuhkan. Kesulitan umum ini diperparah oleh banyaknya keputusan yang harus dibuat hakim untuk mencapai hukuman yang tepat.

Sebagai pertimbangan awal, hakim harus menentukan fakta, faktor, dan fitur mana yang relevan dengan hukuman, dan bobot yang tepat untuk masing-masingnya. Kemudian, hakim harus mencapai keputusan mendasar tentang apakah akan mengeluarkan pelaku dari masyarakat atau memilih dari sederet kemungkinan tidak dipenjara. Keputusan selanjutnya, antara lain, melibatkan lamanya hukuman, dan apakah sebagian darinya harus ditangguhkan, dan, jika ya, untuk jangka waktu berapa lama dan dalam kondisi apa.

Lebih lanjut, hakim tidak hanya harus mempertimbangkan hukuman yang tepat untuk pelanggaran tersebut, tetapi juga risiko yang ditimbulkan oleh pelaku, yaitu memprediksi kemungkinan residivisme pelaku. Secara historis, menilai risiko residivisme terdakwa

membutuhkan ketergantungan pada "intuisi, insting, dan rasa keadilan" hakim, yang dapat menghasilkan "hukuman yang lebih berat" berdasarkan "prediksi klinis yang tak terucapkan".

Karena alasan-alasan ini, daya tarik perangkat lunak penilaian risiko untuk sistem peradilan pidana yang terbebani hampir tak tertahankan. Daya tarik sistem penilaian risiko otomatis adalah bahwa sistem tersebut mengusulkan untuk menyuntikkan objektivitas ke dalam sistem peradilan pidana yang telah dikompromikan, terlalu lama dan terlalu sering, oleh kegagalan manusia. Para pendukung sistem penilaian risiko otomatis juga mengklaim bahwa sistem tersebut membuat penjatuhan hukuman lebih transparan dan rasional.

Metode otomatis diyakini memberikan substansi pada keputusan dan membuatnya lebih ilmiah, dapat diaudit, dan, akibatnya, memberikan kesan legitimasi. Dukungan untuk pengenalan alat penilaian risiko algoritmik kemudian didasarkan pada premis bahwa alat tersebut meningkatkan profesionalisme dengan meningkatkan pembelaan dan akuntabilitas keputusan, menghasilkan keseragaman di seluruh wilayah dan yurisdiksi, dan mempertahankan persepsi validitas ilmiah yang objektif.

Tidak ada bukti empiris yang menunjukkan bahwa hukuman pidana yang lebih lama memiliki dampak signifikan terhadap residivisme seseorang. Dengan demikian, tidak serta merta berarti bahwa hukuman penjara yang lebih lama akan menurunkan risiko residivisme terdakwa. Oleh karena itu, seorang hakim menghadapi pertanyaan yang lebih rumit tentang bagaimana menggunakan skor penilaian risiko otomatis dalam menjatuhkan hukuman, dan keputusan akhir mungkin bergantung pada teori penologi hakim itu sendiri. Atau, hakim mungkin hanya mengambil pendekatan yang menghindari risiko dan menjatuhkan hukuman yang lebih berat kepada terdakwa yang dilabeli "berisiko tinggi" oleh perangkat lunak, alih-alih menanggung risiko pribadi dan sosial dari seorang residivis yang melakukan kejahatan lagi. Jelas, hakim menghadapi tantangan unik dalam proses pengambilan keputusan mereka di era AI dan big data. Konsekuensinya diilustrasikan dengan baik oleh putusan Mahkamah Agung Wisconsin dalam kasus S v. Loomis.

15.3 S V LOOMIS

Pada tahun 2013, Eric Loomis, seorang pria kulit hitam berusia 31 tahun, ditangkap di La Crosse, Wisconsin, atas tuduhan terkait penembakan saat berkendara. Loomis membantah terlibat dalam penembakan tersebut, tetapi ia tetap melepaskan haknya untuk diadili dan mengaku bersalah atas dua dakwaan yang lebih ringan melarikan diri dari petugas lalu lintas dan mengemudikan kendaraan tanpa izin pemiliknya. Semua ini merupakan pelanggaran berulang.

Loomis juga menjalani masa percobaan karena mengedarkan obat resep, dan ia terdaftar sebagai pelaku kejahatan seksual karena pernah dihukum atas penyerangan seksual tingkat tiga. Sebagai mitigasi, pengacaranya menekankan masa kecilnya yang dihabiskan di panti asuhan tempat ia mengalami pelecehan. Dengan seorang putra yang masih bayi, Loomis juga sedang berlatih menjadi seniman tato. Menyusul pembelaan tersebut, pengadilan sirkuit (pengadilan tingkat pertama) memerintahkan laporan investigasi pra-putusan hukuman, yang

mencakup penilaian risiko oleh sistem otomatis, COMPAS, untuk membantu pengadilan dalam menentukan hukuman Loomis.

Penilaian COMPAS memperkirakan risiko residivisme berdasarkan wawancara dengan terdakwa dan informasi dari riwayat kriminal terdakwa. COMPAS menilai variabel dalam lima bidang utama: keterlibatan kriminal, hubungan/gaya hidup, kepribadian/sikap, keluarga, dan pengucilan sosial. Penilaian risiko COMPAS menetapkan Loomis berisiko tinggi untuk ketiga jenis residivisme yang diukur oleh sistem: residivisme praperadilan, residivisme umum, dan residivisme kekerasan. Dalam menjatuhkan hukuman maksimum 6 tahun penjara dan 5 tahun pengawasan lanjutan, hakim secara khusus menyebutkan skor COMPAS:

> Anda diidentifikasi melalui penilaian COMPAS sebagai individu yang berisiko tinggi bagi masyarakat... Saya menolak masa percobaan karena keseriusan kejahatan dan karena riwayat Anda... dan alat penilaian risiko yang telah digunakan, menunjukkan bahwa Anda berisiko sangat tinggi untuk mengulangi tindak pidana S v. Loomis.

Loomis menentang hukumannya, dengan alasan bahwa penggunaan skor COMPAS oleh pengadilan tingkat pertama melanggar haknya atas proses hukum yang semestinya, karena, di antara argumen lainnya, hal itu melanggar haknya atas hukuman individual karena COMPAS mengandalkan informasi tentang karakteristik kelompok yang lebih besar untuk menghitung kesimpulan tentang kemungkinan pribadinya untuk melakukan kejahatan di masa mendatang. Mahkamah Agung Wisconsin akhirnya menolak semua tuntutan Loomis. Menanggapi argumen Loomis tentang haknya atas hukuman individual, pengadilan membedakan kasus ini dari kasus hipotetis di mana skor penilaian risiko merupakan satu-satunya faktor atau faktor penentu dalam putusan hukuman. Dalam kasus Loomis, skor risiko otomatis hanyalah salah satu informasi di antara banyak informasi lain yang dipertimbangkan hakim dalam menjatuhkan hukuman. Pengadilan berpendapat bahwa argumen peradilan yang adil mungkin berhasil jika skor risiko merupakan faktor penentu atau satu-satunya yang dipertimbangkan hakim.

Masalah yang jelas adalah, tanpa pernyataan yang jelas dari hakim mengenai hal ini, mustahil untuk menentukan sejauh mana hakim benar-benar mengandalkan skor risiko otomatis untuk menentukan hukuman terdakwa. Lebih parah lagi, karena adanya bias implisit, hakim sendiri mungkin tidak mengetahuinya.

Untuk memastikan bahwa hakim yang menjatuhkan hukuman mempertimbangkan hasil penilaian risiko otomatis dengan tepat, Mahkamah Agung Wisconsin dalam kasus Loomis menetapkan bagaimana penilaian ini harus disajikan ke pengadilan negeri, dan sejauh mana hakim dapat menggunakannya. Meskipun skor risiko mungkin berguna untuk memahami pertimbangan keselamatan publik terkait pengurangan dan manajemen risiko pelaku, skor tersebut tidak boleh digunakan untuk menentukan tingkat keparahan atau lamanya hukuman, dan tentu saja tidak boleh menjadi faktor resmi yang memberatkan atau meringankan dalam putusan hukuman.

Dalam upaya untuk memastikan bahwa batasan-batasan ini dipatuhi, pengadilan mengamanatkan hakim untuk menjelaskan pada saat putusan “faktor-faktor selain penilaian risiko COMPAS yang secara independen mendukung hukuman yang dijatuhkan” dalam kasus S v. Loomis. Pengadilan Loomis berusaha memberikan perlindungan prosedural untuk

mengingatkan hakim tentang bahaya penilaian ini. Pengadilan menetapkan bahwa "nasihat tertulis" harus disertakan dalam setiap laporan investigasi pra-putusan hukuman yang berisi skor penilaian risiko COMPAS. "Nasihat tertulis tentang batasan-batasannya" ini harus menjelaskan bahwa:

1. COMPAS adalah alat yang bersifat kepemilikan, yang telah mencegah pengungkapan informasi spesifik tentang bobot faktor-faktor atau bagaimana skor risiko dihitung;
2. Skor COMPAS didasarkan pada data kelompok, sehingga mengidentifikasi kelompok dengan karakteristik yang menjadikan mereka pelaku tindak pidana berisiko tinggi, bukan individu berisiko tinggi tertentu;
3. Beberapa penelitian menunjukkan bahwa algoritma COMPAS mungkin bias dalam mengklasifikasikan pelaku tindak pidana minoritas;
4. COMPAS membandingkan terdakwa dengan sampel nasional, tetapi belum menyelesaikan studi validasi silang untuk populasi Wisconsin, dan alat seperti ini harus terus dipantau dan diperbarui keakuratannya seiring perubahan populasi; dan
5. COMPAS awalnya tidak dikembangkan untuk digunakan dalam putusan S v. Loomis.

"Peringatan tertulis tentang pembatasan" ini memberikan peringatan kepada hakim tentang penggunaan skor COMPAS secara bermakna, yang ditegaskan kembali dalam pendapat yang sependapat. Ketua Mahkamah Agung Roggensack menulis pendapat yang sependapat terpisah untuk mengklarifikasi bahwa:

> Meskipun keputusan kami hari ini mengizinkan pengadilan yang menjatuhkan putusan untuk mempertimbangkan COMPAS, kami tidak menyimpulkan bahwa pengadilan yang menjatuhkan putusan dapat mengandalkan COMPAS untuk hukuman yang dijatuhkannya . . . [Karena] pendapat mayoritas secara bergantian menggunakan istilah "mempertimbangkan" dan "mengandalkan" ketika membahas kewajiban pengadilan yang menjatuhkan hukuman dan alat penilaian risiko COMPAS, keputusan kami dapat secara keliru diartikan sebagai mengizinkan penggunaan COMPAS dalam kasus S v. Loomis.

Sebagai cara untuk mengatasi kekhawatiran tentang penggunaan alat penilaian risiko algoritmik, Hakim Abrahamson juga menulis secara terpisah untuk menekankan bahwa, dalam mempertimbangkan alat-alat ini dalam menjatuhkan hukuman, seorang hakim "harus mencatat proses penalaran yang bermakna yang membahas relevansi, kekuatan, dan kelemahan alat penilaian risiko" dalam kasus S v. Loomis.

Meskipun menyatakan kekhawatiran tentang potensi ketidakadilan dan diskriminasi yang melekat dalam alat penilaian risiko algoritmik, Mahkamah Agung Wisconsin tetap menyetujui penggunaannya dalam kasus Loomis dengan suara bulat. Pengadilan hanya secara dangkal membahas risiko yang melekat pada alat penilaian risiko algoritmik ini dengan menambahkan peringatan dan mewajibkan pengungkapan tertentu untuk menyertai skor COMPAS dalam laporan investigasi pra-putusan. Namun, pengadilan tidak membahas pertanyaan mendasar tentang mengapa skor tersebut harus dimasukkan dalam laporan penilaian risiko jika skor tersebut tidak akan memengaruhi lamanya hukuman. Pengadilan tidak menjelaskan bagaimana hakim dapat menggunakan skor risiko terdakwa jika ia tidak dapat mengubah lamanya hukuman berdasarkan skor tersebut.

Kesimpulan yang dapat diterima adalah bahwa pengadilan pada akhirnya gagal membatasi penggunaan sistem ini secara signifikan, sebagian besar karena gagal mempertimbangkan tekanan eksternal dan internal terhadap hakim untuk menggunakan alat penilaian risiko otomatis ini, ketidakmampuan hakim untuk mengevaluasi alat penilaian risiko, dan pengaruh bias kognitif terhadap proses pengambilan keputusan hakim. Pendekatan "label peringatan" pengadilan bukannya menerapkan pembatasan yang berarti merupakan cara yang tidak efektif untuk mengubah cara hakim mengevaluasi penilaian risiko otomatis, dan hal ini telah membuka peluang bagi hakim untuk sangat dipengaruhi oleh penilaian risiko tersebut.

Tidak realistis mengharapkan hakim yang menjatuhkan hukuman, setelah meninjau skor risiko otomatis, untuk menggunakan diskresi tanpa pandangan yang telah ditentukan sebelumnya, atau bahkan bias terhadap, terdakwa. Jika semua faktor sama, skor risiko yang tinggi akan memperkecil kemungkinan seorang pelaku menerima hukuman minimum atau menghindari hukuman penjara. Terlepas dari kenyataan bahwa tidak ada hakim yang ingin berada dalam posisi harus membela hukuman ringan yang dijatuhkan kepada terdakwa "berisiko tinggi", terutama jika terdakwa tersebut benar-benar melakukan kejahatan di masa mendatang, pengadilan sama sekali mengabaikan "efek teknologi" dan peran bias kognitif yang mendukung ketergantungan data (khususnya, yang disebut "bias otomatisasi" dan "penahan") pada proses pengambilan keputusan hakim.

15.4 “EFEK TEKNOLOGI”

Para peneliti telah mengidentifikasi kecenderungan optimisme yang berlebihan ketika membuat keputusan yang melibatkan teknologi. Karena terobosan teknologi seringkali menghasilkan hasil yang dramatis dan berkesan, seperti merevolusi industri dan meningkatkan kualitas hidup kita misalnya ponsel pintar, jam tangan pintar, mobil tanpa pengemudi, pencetakan tiga dimensi, dan layanan streaming hiburan peristiwa semacam itu sangat menonjol. Sebaliknya, kegagalan teknologi kurang menonjol, karena tidak cenderung mengubah status quo, dan juga tidak mungkin dibahas di depan umum. Akibatnya, dalam konteks pengambilan keputusan, orang mengembangkan asosiasi yang tidak disadari atau “implisit” antara teknologi dan kesuksesan melalui akumulasi pengalaman yang menghubungkan keduanya. Hal ini terjadi karena kemajuan teknologi yang luar biasa telah mengondisikan kita untuk berharap bahwa teknologi akan menjadi pendorong kesuksesan dan kemajuan. Bias terhadap optimisme dalam teknologi ini telah diberi label "efek teknologi".

Setelah berkembang secara tidak sadar, asosiasi implisit beroperasi dengan cepat dan otomatis terkait kognisi dan perilaku. Model heuristik-sistematis Chaiken menunjukkan bahwa pemrosesan informasi dapat terjadi melalui dua jalur: (i) jalur sistematis yang lebih membutuhkan usaha; atau (ii) jalur yang lebih otomatis (heuristik) yang tidak melibatkan pemrosesan informasi yang kompleks. Seseorang menggunakan jalur heuristik ketika terdapat isyarat kuat tentang keandalan suatu pesan, yang menurunkan motivasi orang tersebut untuk terlibat dalam pemrosesan yang lebih membutuhkan usaha dan sistematis. Para peneliti berpendapat bahwa:

> Seorang . . . Gagasan tentang teknologi telah begitu erat kaitannya dengan kemajuan dan pencapaian, atau, "kesuksesan", sehingga penggunaan teknologi dalam konteks pengambilan keputusan dapat memicu asumsi otomatis bahwa pilihan keputusan yang melibatkan teknologi akan berhasil.

Implikasi yang meresahkan dalam konteks peradilan adalah adanya karakteristik situasional utama yang mungkin memicu efek teknologi. Misalnya, individu dalam konteks di mana mereka mengalami beban kognitif tinggi seperti yang dialami hakim setiap hari mungkin lebih rentan terhadap efek teknologi dan pemrosesan heuristik.

15.5 "BIAS OTOMATISASI" DAN EFEK PENJANGKARAN

Selain tekanan eksternal, hakim rentan terhadap bias psikologis yang mendorong penggunaan alat penilaian risiko otomatis. Sejumlah penelitian telah menunjukkan bahwa di pengadilan yang mengandalkan perangkat ilmiah dan teknologi, hakim (dan individu lainnya) cenderung tunduk pada angka dan hasil yang dihasilkan komputer, yang dapat membingkai dan mengondisikan pandangan hakim. Individu cenderung lebih mempertimbangkan penilaian empiris yang dianggap ahli daripada bukti non-empiris yang dapat menciptakan bias yang mendukung penilaian risiko otomatis daripada narasi pelaku sendiri. Penelitian menunjukkan bahwa sulit dan tidak umum bagi individu untuk menentang rekomendasi algoritmik. Misalnya, dalam sebuah eksperimen baru-baru ini di Institut Teknologi Georgia, seorang mahasiswa ditempatkan di sebuah kantor kecil bersama robot untuk menyelesaikan survei akademik. Tiba-tiba alarm berbunyi dan asap memenuhi lorong di luar pintu. Robot yang dilengkapi tanda bertuliskan "Robot Pemandu Darurat" itu mulai bergerak. Hal ini memaksa siswa tersebut untuk membuat keputusan dalam sekejap, antara melarikan diri melalui pintu keluar yang ditandai dengan jelas, atau mengikuti robot tersebut melalui jalur yang tidak diketahui dan melalui pintu yang tersembunyi.

Dua puluh enam dari 30 peserta memilih untuk mengikuti robot tersebut, meskipun robot tersebut membimbing mereka menjauh dari pintu keluar yang sebenarnya. "Kami terkejut," ujar peneliti utama, Paul Robinette, dalam sebuah wawancara: "Kami pikir tidak akan ada cukup kepercayaan, dan bahwa kami harus melakukan sesuatu untuk membuktikan bahwa robot tersebut dapat dipercaya."

Hasil penelitian menunjukkan bahwa ketika orang-orang diberi tahu bahwa sebuah robot (atau mesin lain) dirancang untuk melakukan tugas tertentu, seperti dalam kasus percobaan ini, mereka kemungkinan besar akan secara otomatis memercayainya untuk melakukan tugas tersebut dengan benar. Faktanya, para peserta dalam penelitian ini memberikan robot tersebut keuntungan dari keraguan, meskipun instruksi robot tersebut agak berlawanan dengan intuisi. Dalam versi lain dari studi tersebut, mayoritas partisipan bahkan terus mengikuti robot setelah robot tersebut tampak "rusak" atau membeku di tempatnya, mendorong seorang peneliti untuk muncul dan meminta maaf atas "kinerja buruk" robot tersebut.

Para pendukung hukuman berbasis risiko otomatis berpendapat bahwa algoritma hanya memberikan prediksi "indikatif" risiko. Sebagian besar hakim juga berpendapat bahwa mereka tidak secara membabi buta mengikuti hasil yang diberikan oleh algoritma ketika

memutuskan nasib seorang pelaku tindak pidana. Sebaliknya, mereka mengklaim mengandalkan keahlian dan pengalaman klinis mereka untuk menilai kepribadian pelaku tindak pidana, situasi sosial-ekonomi, dan risiko residivisme. Namun, penelitian dalam ekonomi perilaku dan psikologi kognitif telah menunjukkan bahwa secara psikologis sulit dan jarang bagi manusia untuk "mengesampingkan" rekomendasi suatu algoritma. Hakim cenderung mengikuti prediksi algoritma penilaian risiko otomatis. Dalam survei terhadap lebih dari 100 hakim dan praktisi hukum Kanada, persepsi umum adalah bahwa sistem pengambilan keputusan otomatis “lebih baik daripada penilaian klinis atau bentuk penilaian subjektif lainnya . . . ”.

Dari perspektif hakim, penilaian kuantitatif oleh program perangkat lunak umumnya tampak lebih andal, ilmiah, dan sah daripada hampir semua sumber informasi lainnya, termasuk perasaannya sendiri tentang seorang pelaku. Hal ini berlaku, tidak hanya bagi orang awam, tetapi juga bagi para profesional yang sangat terampil. Sulit untuk mempertanyakan angka dan persamaan jika Anda belum terlatih dalam statistik. Dengan demikian, bahayanya adalah ketika algoritma memprediksi risiko residivisme yang "tinggi", hakim cenderung memenjarakannya, terlepas dari faktor-faktor lain.

Permasalahan khusus dengan penalaran pengadilan dalam kasus Loomis adalah bahwa pengadilan tersebut menaruh kepercayaannya pada hakim untuk mempertimbangkan "nasihat tertulis" dan mengevaluasi skor penilaian risiko otomatis yang sesuai dan ini terjadi di era di mana masyarakat secara keseluruhan sangat dipengaruhi oleh "efek teknologi". Dengan atau tanpa nasihat tertulis, hakim secara konsisten memberikan bobot yang lebih besar kepada teknologi dan bukti berbasis forensik daripada faktor-faktor lain, baik disadari maupun tidak.

Dorongan untuk mengikuti rekomendasi komputer bersumber dari "bias otomatisasi". Studi telah menunjukkan bahwa "bias otomatisasi" terjadi karena kecenderungan kebanyakan orang untuk lebih memercayai kemampuan analitis sistem otomatis daripada kemampuan mereka sendiri, bahkan ketika dihadapkan pada bukti ketidakakuratan sistem. Sebagaimana dicatat oleh Danielle Citron, "sibia otomatisasi secara efektif mengubah jawaban yang disarankan program komputer menjadi keputusan akhir yang tepercaya". Meskipun sistem pengambilan keputusan otomatis berpotensi menghilangkan kesalahan tertentu yang terkait dengan pengambilan keputusan manusia, pada kenyataannya, sistem ini tampaknya hanya mengganti kesalahan tersebut dengan kesalahan baru. Menurut profesor psikologi Linda Sitka:

> Kebanyakan orang akan mengambil jalan dengan upaya kognitif paling sedikit, dan alih-alih menganalisis setiap keputusan secara sistematis, mereka akan menggunakan aturan praktis pengambilan keputusan atau heuristik... Alat bantu keputusan otomatis dapat bertindak sebagai salah satu heuristik pengambilan keputusan ini, dan digunakan sebagai pengganti sistem pemantauan atau pengambilan keputusan yang lebih cermat.

Dengan demikian, Sitka memandang bias otomatisasi sebagai akibat dari seseorang yang menggunakan sistem pengambilan keputusan otomatis sebagai pengganti heuristik untuk pencarian dan pemrosesan informasi yang cermat. Definisi ini memperlakukan bias otomatisasi serupa dengan bias dan heuristik lain dalam pengambilan keputusan manusia

(seperti, misalnya, bias konfirmasi), kecuali bahwa bias otomatisasi secara khusus berasal dari interaksi dengan sistem otomatis.

Tiga faktor utama telah diasumsikan berkontribusi terhadap bias otomatisasi. Pertama, ada kecenderungan manusia untuk memilih jalan dengan upaya kognitif paling sedikit (yang disebut "hipotesis kikir kognitif"). Dengan demikian, manusia cenderung menggunakan arahan atau rekomendasi sistem otomatis sebagai heuristik pengambilan keputusan yang kuat, alih-alih proses kognitif analisis dan evaluasi informasi yang membutuhkan usaha.

Faktor kedua adalah kepercayaan manusia terhadap sistem otomatis sebagai agen yang kuat dengan kemampuan analitis yang superior. Akibatnya, manusia mungkin melebih-lebihkan kinerja sistem pengambilan keputusan otomatis dan mungkin menganggap sistem otomatis memiliki kemampuan dan otoritas yang lebih besar daripada diri mereka sendiri atau manusia lain.

Faktor ketiga yang berkontribusi terhadap bias otomatisasi adalah fenomena "difusi tanggung jawab". Ketika manusia berbagi tugas pengambilan keputusan dengan mesin, efek psikologis yang sama terjadi ketika manusia berbagi tugas dengan manusia lain, yaitu apa yang disebut "kemalasan sosial", yang tercermin dalam kecenderungan manusia untuk mengurangi upaya mereka sendiri ketika bekerja dalam kelompok, dibandingkan ketika mereka bekerja secara individu.

Sejauh pengguna manusia memandang sistem pengambilan keputusan otomatis sebagai anggota tim lainnya, manusia tersebut mungkin merasa kurang bertanggung jawab atas hasilnya, dan, akibatnya, mengurangi upaya mereka sendiri dalam memantau dan menganalisis data lain yang tersedia. Selain itu, karena individu dan lembaga sering beralih ke algoritma dengan tujuan yang jelas untuk mengurangi bias dan kesalahan manusia, algoritma ini dapat dianggap sebagai sumber yang otoritatif, dengan pengetahuan yang lebih luas daripada manusia yang menafsirkannya. Dengan demikian, pengguna manusia, seperti hakim, cenderung mematuhi apa yang diputuskan oleh algoritma, meskipun kepatuhan tersebut dapat merugikan orang lain. Hal ini disebabkan oleh kekuasaan umum yang dimiliki figur otoritas dan "kesediaan masyarakat untuk menyesuaikan diri dengan tuntutan ... otoritas". Bias otomatisasi merupakan fenomena kuat yang menjadikan "nasihat tertulis" yang diamanatkan oleh Mahkamah Agung Wisconsin menjadi tidak relevan.

Mahkamah Agung Wisconsin dalam kasus Loomis menyatakan ekspektasinya bahwa "pengadilan wilayah akan menggunakan kebijaksanaan ketika menilai skor risiko COMPAS terhadap terdakwa individu". Sebagaimana dijelaskan di atas, ekspektasi pengadilan tidak berdasar. Manusia jauh lebih memercayai keputusan yang dihasilkan komputer daripada yang seharusnya. Bahkan, mereka mengandalkan keputusan otomatis bahkan ketika mereka mencurigai adanya malfungsi.

Masalah terkait adalah bahwa sistem pengambilan keputusan otomatis juga dapat "memberikan perlindungan bagi agen manusia". Misalnya, sebuah survei terhadap hakim dan pengacara di Kanada menemukan bahwa banyak yang menganggap perangkat lunak, seperti COMPAS, sebagai peningkatan dibandingkan penilaian subjektif manusia.

Meskipun para praktisi ini tidak menganggap perangkat lunak penilaian risiko sebagai prediktor perilaku masa depan yang sangat andal, mereka tetap lebih menyukai sistem ini karena penggunaannya meminimalkan risiko hakim dan pengacara sendiri disalahkan atas konsekuensi keputusan mereka. Sebagaimana dicatat Chesterman dengan tepat, sistem penilaian risiko otomatis:

> Seharusnya tidak menjadi dasar untuk menghindari akuntabilitas dalam arti sempit, yaitu kewajiban untuk memberikan pertanggungjawaban atas suatu keputusan, meskipun setelah kejadian, atau untuk menghindari tanggung jawab atas kerugian akibat keputusan tersebut.

Terlepas dari bias otomatisasi, pengadilan di Amerika Serikat telah berulang kali mengakui bahwa pernyataan peringatan tidak banyak membantu mencegah pencari fakta mempertimbangkan faktor-faktor tertentu setelah kesadaran pencari fakta terpapar pada faktor-faktor tersebut. Pengadilan dalam kasus Amerika Serikat v Rodriguez menjelaskan mengapa instruksi juri untuk mengabaikan pendapat pribadi yang diungkapkan oleh jaksa tidak efektif: "seseorang tidak dapat menghilangkan sesuatu yang tidak diinginkan"; "setelah tusukan pedang, sulit untuk mengatakan lupakan lukanya"; dan "jika Anda melempar sigung ke dalam kotak juri, Anda tidak dapat memerintahkan juri untuk tidak menciumnya".

Kekhawatirannya adalah hakim mungkin mengadaptasi praktik penjatuhan hukuman mereka agar sesuai dengan prediksi algoritma penilaian risiko. Ekonom perilaku menyebut "penjangkaran" untuk menggambarkan fenomena umum di mana individu memanfaatkan setiap bukti yang tersedia terlepas dari seberapa lemahnya untuk membuat keputusan selanjutnya. Dalam eksperimen klasik mereka tentang "efek penjangkaran", Amos Tversky dan Daniel Kahneman memanipulasi roda keberuntungan yang ditandai dari 0 hingga 100 agar berhenti di angka 10 atau 65.

Mereka akan berdiri di depan sekelompok mahasiswa Universitas Oregon yang mereka rekrut sebagai partisipan, memutar roda, dan meminta para mahasiswa untuk menuliskan angka di mana roda tersebut berhenti (yang tentu saja adalah 10 atau 65). Kemudian, para peneliti mengajukan pertanyaan berikut kepada partisipan: "Apakah persentase negara-negara Afrika di antara anggota PBB lebih besar atau lebih kecil daripada angka yang baru saja Anda tulis?"

Seperti yang dijelaskan Kahneman:

> Putaran roda keberuntungan bahkan yang tidak dicurangi tidak mungkin menghasilkan informasi yang berguna tentang apa pun, dan para partisipan... seharusnya mengabaikannya begitu saja. Namun, mereka tidak mengabaikannya.

Ketika roda berhenti di angka 10, para partisipan memberikan perkiraan rata-rata sebesar 25%; Ketika roda berhenti di angka 65, para partisipan memberikan estimasi rata-rata sebesar 45%. Dengan demikian, hanya dengan diberikan sebuah angka bahkan angka yang mereka tahu sepenuhnya acak dan sama sekali tidak berpengaruh pada kuantitas yang diminta untuk mereka estimasi memberikan dampak yang nyata terhadap respons partisipan.

Sejak studi penting Tversky dan Kahneman, efek penjangkaran telah terbukti sebagai "fenomena yang benar-benar ada di mana-mana yang telah diamati dalam beragam domain penilaian yang berbeda". Efek penjangkaran telah terbukti sebagai bias kognitif yang kuat,

andal, dan persisten. Banyak temuan menunjukkan bahwa angka-angka yang jelas-jelas tidak relevan meskipun ditentukan secara acak dapat memandu penilaian numerik yang dihasilkan dalam kondisi ketidakpastian.

Tidak mengherankan bahwa hakim tidak kebal terhadap efek jangkar. Keputusan pengadilan sering kali melibatkan kuantifikasi. Dan kuantifikasi pengadilan umumnya terjadi dalam keadaan yang secara inheren tidak pasti. Pertama, hakim harus membuat keputusan mereka setidaknya sebagian berdasarkan bukti yang kontroversial dan kontradiktif. Kedua, hakim sering diminta untuk mengkuantifikasi hal-hal yang tidak dapat dikuantifikasi kesalahan kualitatif dari pihak yang bersalah yang harus dinyatakan sebagai pemberian ganti rugi moneter atau penentuan denda pidana atau lamanya hukuman penjara. Tanpa adanya pedoman algoritmik yang ketat atau spesifikasi kelembagaan lainnya, proses ini dapat menjadi ambigu dan sangat subjektif.

Dengan demikian, jika prediksi algoritma penilaian risiko tentang risiko residivisme pelaku lebih tinggi daripada yang ditetapkan hakim dalam pikirannya sendiri, ia dapat meningkatkan hukuman, bahkan tanpa menyadari bahwa ia mengikuti algoritma tersebut. Sebagaimana dinyatakan di atas, mustahil untuk menentukan sejauh mana seorang hakim sebenarnya mengandalkan skor risiko otomatis untuk menentukan hukuman terdakwa. Seorang hakim yang dihadapkan dengan penilaian yang mengungkapkan risiko residivisme yang lebih tinggi daripada yang diprediksi, dapat meningkatkan hukuman terdakwa tanpa menyadari bahwa penjangkaran mungkin berperan dalam putusan tersebut.

15.6 BIAS ALGORITMA HUKUMAN: KASUS LOOMIS

Ilmuwan data, Cathy O'Neil, menggambarkan era saat ini sebagai era optimisme teknologi yang tak terbantahkan. Keyakinan yang cenderung kita berikan pada kekuatan teknologi melindungi sistem algoritmik dari interogasi kritis secara umum. Ironisnya, kata jurnalis teknologi investigasi, Julia Angwin, bahwa kita sebagai sistem peradilan pidana, badan politik, dan budaya mengambil pendekatan yang terlalu manusiawi terhadap infrastruktur algoritmik:

> Kita terlalu mempercayainya. Kita belum berpikir seketat atau sestrategis yang kita perlukan tentang dampaknya. Kita belum sepenuhnya mempertimbangkan apakah, dan bahkan bagaimana, mengatur algoritma yang... mengatur hidup kita....

Loomis penting karena menunjukkan, tidak hanya tantangan yang dihadapi pengadilan dalam memahami cara kerja sistem penilaian risiko otomatis, tetapi juga fakta bahwa hampir tidak ada preseden yang memandu pengambilan keputusan hakim dalam menggunakan alat-alat otomatis ini dan alat-alat otomatis lainnya.

Peringatan tertulis saja tampaknya tidak dapat memberikan informasi yang memuaskan kepada hakim sebagai gatekeeper yang efektif, terutama ketika mereka mungkin tidak cukup dibekali dengan pengetahuan dan pemahaman tentang cara kerja alat-alat otomatis ini. Meskipun peringatan mungkin mengingatkan hakim akan kekurangan alat-alat ini, nasihat tersebut mungkin gagal meniadakan tekanan eksternal dan internal yang cukup

besar dari sistem peradilan pidana yang mendukung penggunaan penilaian risiko kuantitatif otomatis.

Mahkamah Agung Wisconsin dalam kasus Loomis tampaknya tidak menyadari kenyataan ini. Mahkamah Agung tersebut menerima begitu saja klaim pengadilan sirkuit dalam proses pasca-putusan bahwa pengadilan "akan menjatuhkan hukuman yang sama persis" bahkan tanpa skor risiko otomatis. Nasihat yang diwajibkan pengadilan tersebut menunjukkan bahwa hakim harus menjadi pemeriksa bias terhadap alat yang dirancang untuk mengoreksi bias hakim.

Titik awal yang bermanfaat mungkin adalah merenungkan kembali pertanyaan: "Siapakah pembuat keputusan?" Seperti yang dicatat Liu dkk.: "Dalam dunia yang seringkali secara membabi buta menggambarkan angka sebagai sesuatu yang ilmiah, netral, dan objektif, para pembuat keputusan manusia cenderung menyerahkan kekuasaan mereka kepada data." Sebagaimana terlihat dalam kasus Loomis, penghormatan yang kurang berdasar terhadap algoritma meminggirkan peran otoritas publik dan pengawasan dalam tata kelola.

Karena "kata-kata kalah oleh angka" dalam putusan pidana, lembaga peradilan harus sangat berhati-hati dalam menilai nilai kualitatif teknologi-teknologi baru ini. Meskipun pemerintah mungkin mewajibkan manusia untuk membuat keputusan akhir, mengatasi efek jangkar tetaplah bermasalah, selama pendekatan berbasis data terhadap proses pengambilan keputusan di sektor publik ditoleransi.

Hakim harus dilatih tentang fenomena bias otomatisasi. Studi telah menunjukkan bahwa individu yang menerima pelatihan semacam itu lebih cenderung untuk mencermati saran-saran dari sistem otomatis. Selain pelatihan, cara terbaik untuk memastikan keadilan sistem penilaian otomatis adalah melalui perlindungan prosedural. Dalam pengembangan algoritma untuk diterapkan dalam sistem peradilan pidana, akuntabilitas dan pengawasan adalah kuncinya. Para pembuat kebijakan harus memastikan bahwa sistem ini telah dirancang sesuai tujuan penggunaannya, dan terus dipantau serta dinilai keakuratan dan keandalannya.

Memfasilitasi penelitian dan audit eksternal untuk mengevaluasi dan menguji algoritma untuk bias juga sangat penting. Desain, implementasi, dan evaluasi sistem otomatis yang akan digunakan dalam sistem peradilan pidana harus konsisten dengan nilai-nilai inti sistem tersebut, termasuk perlindungan yang setara dan proses hukum yang wajar. Pertanyaan normatif dan etika yang penting muncul di setiap kesempatan ketika alat penilaian risiko algoritmik ini diintegrasikan ke dalam sistem yang ada dan keputusan tersebut tidak boleh dibuat dengan mudah atau tanpa informasi yang memadai.

Setidaknya untuk saat ini, manusia tetap memegang kendali atas pemerintahan, dan mereka dapat menuntut penjelasan atas keputusan dalam bahasa alami, bukan kode komputer. Kegagalan dalam konteks peradilan pidana berisiko menyerahkan fungsi-fungsi pemerintahan dan hukum yang inheren kepada "elit komputasional yang tidak akuntabel". Kebijakan peradilan pidana seharusnya didasarkan pada data, tetapi kita tidak boleh membiarkan bahasa sains yang steril mengaburkan pertanyaan-pertanyaan tentang keadilan, akuntabilitas, dan keadilan.

Angwin berpendapat bahwa "akuntabilitas algoritmik" memerlukan pendekatan yang lebih skeptis terhadap algoritma secara umum. Kita hidup di era optimisme teknologi secara umum, era di mana teknologi-teknologi baru menjanjikan untuk membuat hidup kita lebih efisien dan menyenangkan. Teknologi-teknologi tersebut mungkin membantu menjadikan sistem peradilan lebih adil; mungkin juga tidak. Intinya adalah kita berutang kepada diri kita sendiri dan kepada Eric Loomis serta setiap orang lain yang hidupnya mungkin diubah oleh sebuah algoritma untuk mencari tahu.

Pada akhirnya, manusia harus mengevaluasi setiap proses pengambilan keputusan dan mempertimbangkan bentuk-bentuk otomatisasi apa yang bermanfaat, tepat, dan konsisten dengan supremasi hukum. Pada akhirnya, ada sesuatu yang perlu dikatakan tentang hukuman yang dijatuhkan oleh hakim manusia tanpa bantuan algoritma. Hakim, sebagai manusia, tidak diselimuti oleh nuansa mistik dan kesempurnaan yang menyelimuti teknologi. Dalam beberapa hal, lebih mudah untuk memeriksa dan menggugat keputusan hakim ketika terdakwa mencurigai adanya bias yang memengaruhi keputusan hakim, karena hakim, sebagian besar, harus memberikan alasan atas tindakan mereka. Sedangkan Eric Loomis sendiri, ia dibebaskan dari Lembaga Pemasyarakatan Jackson pada Agustus 2019, setelah menjalani masa hukuman penuh enam tahun. Menurut COMPAS, setidaknya, ia berisiko tinggi untuk kembali.

BAB 16
TANGGUNG JAWAB UNTUK SISTEM BERBASIS AI

Abstrak Artikel ini mencoba mengkaji apakah rezim tanggung jawab perdata saat ini menyediakan kerangka kerja yang kuat untuk menangani kerugian ketika sistem AI—terutama yang berbasis pembelajaran mesin terlibat. Kami mencoba menemukan jawaban untuk tiga pertanyaan: adakah tempat untuk tanggung jawab berbasis kesalahan, ketika mustahil untuk memastikan, di antara banyak aktor, tindakan siapa yang menyebabkan kerugian? Apakah rezim tanggung jawab ketat saat ini tepat untuk menangani kerugian tanpa kesalahan yang disebabkan oleh fungsi sistem AI atau diperlukan sistem baru?

Kapan seorang agen harus dibebaskan dari tanggung jawab? Analisis ini mempertimbangkan karya penting yang dihasilkan di Uni Eropa, khususnya Laporan 2019 tentang "Tanggung Jawab atas AI dan Teknologi Digital Lain yang Berkembang" (oleh Kelompok Pakar yang dibentuk oleh Komisi Eropa), Resolusi Parlemen Eropa 2020 tentang Tanggung Jawab Perdata atas AI, Rancangan Undang-Undang AI 2021, Rancangan Arahan Tanggung Jawab AI 2022, dan Rancangan Arahan Tanggung Jawab Produk 2022.

16.1 PRESENTASI PERMASALAHAN

Teknologi digital berkembang pesat dan kecerdasan buatan (AI) memengaruhi semua sektor ekonomi dan kehidupan kontemporer. Pengoperasian sistem AI modern yang mandiri atau berbasis perangkat lunak kemungkinan besar terkait dengan kerugian. Dalam artikel ini, kami membahas pertanyaan apakah tanggapan tradisional terhadap masalah kompensasi melalui tanggung jawab perdata memadai untuk mengatasi kerugian ketika sistem AI diterapkan. Meskipun kami beranjak dari sudut pandang yurisdiksi Hukum Perdata, diskusi kami mencoba melampaui batas-batas tradisi hukum kami sendiri.

Tantangan yang dihadapi rezim pertanggungjawaban perdata tradisional karena penyebaran sistem AI terkait dengan fitur-fitur khusus dari pengoperasian sistem tersebut: kemampuan sistem AI untuk membuat keputusan dengan cara yang semakin otonom; ketidakjelasan teknologi berbasis pembelajaran mesin; keterlibatan berbagai agen dalam membangun, merakit, memperkenalkan ke pasar, menyesuaikan, memilih dan mengawasi data, melatih, memperbarui dan menggunakan sistem (Kelompok Pakar tentang Pertanggungjawaban dan Teknologi Baru Pembentukan Teknologi Baru, Pertanggungjawaban atas kecerdasan buatan dan teknologi baru lainnya, Direktorat Jenderal Kehakiman dan Konsumen; kerentanan terhadap serangan siber. Semua faktor ini berkontribusi pada kesulitan dalam memutuskan siapa jika ada yang harus bertanggung jawab atas kerugian atau kerugian. Oleh karena itu, kecuali kita menyempurnakan atau memikirkan kembali pendekatan tradisional, mereka yang menderita kerugian kemungkinan besar akan kehilangan kompensasi yang adil.

Kasus yang diajukan berdasarkan tanggung jawab berbasis kesalahan tradisional terhadap operator atau pengguna sistem AI sangat sulit untuk dimenangkan. Proses

kausalitasnya biasanya tidak diketahui oleh korban. Efek kotak hitam dari algoritma pembelajaran mesin menghambat transparansi dan keterjelasan proses pengambilan keputusan. Jumlah agen yang berpotensi terlibat menambah kompleksitas tugas. Beban penggugat untuk membuktikan kesalahan, seringkali, mustahil untuk dipenuhi.

Dalam artikel ini, kami mempertimbangkan tiga pertanyaan.

Pertanyaan pertama berkaitan dengan kemungkinan menetapkan tanggung jawab berbasis kesalahan ketika berbagai aktor yang terlibat dalam proses tersebut mengabaikan aturan yang berlaku, tetapi mustahil untuk menentukan tindakan mana yang merupakan penyebab sebenarnya dari kerugian tersebut. Pertanyaan kedua: ketika tidak ada kesalahan yang dilakukan dan kerugian disebabkan oleh fungsi sistem AI, haruskah kita menerapkan rezim pertanggungjawaban ketat yang berlaku? Haruskah kita, sebagai gantinya, merancang rezim khusus untuk kerugian yang terkait dengan sistem AI? Pertanyaan ketiga dan terakhir: kapan pertanggungjawaban agen harus dikecualikan? Dengan kata lain, apa saja pembelaan yang dapat diajukan agen untuk menghindari pertanggungjawaban?

Uni Eropa (UE) telah menerbitkan dokumen-dokumen penting yang membahas AI dan pertanggungjawaban perdata secara umum. Laporan tentang “Tanggung Jawab atas AI dan Teknologi Digital Baru Lainnya”, yang dipresentasikan oleh Kelompok Pakar tentang Pertanggungjawaban dan Teknologi Baru Pembentukan Teknologi Baru (Kelompok Pakar), yang dibentuk oleh Komisi Eropa, membahas penerapan rezim pertanggungjawaban yang ada pada teknologi digital baru, dengan fokus pada AI, dan perlunya reformasi rezim tersebut. Parlemen Eropa (EP) telah mengadopsi pada tanggal 20 Oktober 2020 sebuah “Resolusi dengan Rekomendasi kepada Komisi tentang Rezim Tanggung Jawab Perdata untuk AI”, termasuk Proposal untuk Regulasi Parlemen Eropa dan Dewan tentang Tanggung Jawab atas Operasi Sistem AI.

Rancangan Undang-Undang AI 2021 tidak membahas masalah tanggung jawab perdata. Sebaliknya, Komisi Eropa pada tahun 2022 telah mengusulkan sebuah Arahan tentang Adaptasi Aturan Tanggung Jawab Perdata Non-Kontraktual terhadap Kecerdasan Buatan (‘Arahan Tanggung Jawab AI’) dan, pada saat yang sama, sebuah Arahan baru tentang Tanggung Jawab atas Produk Cacat, menggantikan Arahan Tanggung Jawab Produk (PLD) yang lama.

Laporan tahun 2019 membedakan dua jenis pelaku potensial. Operator front-end adalah "orang atau badan hukum yang menjalankan kendali atas risiko yang terkait dengan pengoperasian dan fungsi sistem AI dan mendapatkan manfaat dari pengoperasiannya". Operator backend adalah "orang atau badan hukum yang, secara berkelanjutan, mendefinisikan fitur-fitur teknologi, menyediakan data dan layanan dukungan backend yang penting dan karena itu juga menjalankan sejumlah kendali atas risiko yang terkait dengan pengoperasian dan fungsi Sistem AI". Dasar pertanggungjawaban agen-agen ini tampaknya terletak pada posisi kendali yang mereka jalankan atas sistem AI dan, dalam kaitannya dengan operator frontend, juga pada manfaat yang diperolehnya darinya. Meskipun tidak jelas, produsen, programmer, dan orang yang bertanggung jawab untuk memberi makan sistem akan bertindak sebagai operator backend. Tampaknya lebih sulit untuk mengidentifikasi siapa

yang mungkin mengambil peran sebagai operator frontend, terutama karena laporan tersebut ambigu tentang topik ini. Jika, di satu sisi, pengguna sistem ini tampaknya termasuk di sini, di sisi lain pertimbangan 11 dari peraturan yang diusulkan menyatakan bahwa pengguna hanya boleh (secara objektif) menanggapi jika ia adalah seorang operator. Hal ini menunjukkan bahwa terdapat pengguna yang merupakan operator mereka yang, selain mendapatkan manfaat dari penggunaan, juga memiliki kendali atas prosesnya dan yang lainnya bukan, karena mereka tidak memiliki atribut serupa.

Perlu ditekankan bahwa konsep operator frontend dan backend digunakan di sini dengan cara yang tidak sepenuhnya sejalan dengan usulan Parlemen Eropa. Dalam pendekatan ini, operator frontend adalah "orang yang terutama memutuskan dan mendapatkan manfaat dari penggunaan teknologi yang relevan" dan operator backend adalah "orang yang secara terus-menerus mendefinisikan fitur-fitur teknologi yang relevan dan menyediakan dukungan backend yang esensial dan berkelanjutan".

Tampaknya konsep operator frontend yang diusulkan dalam studi Kelompok Pakar lebih luas dan sekaligus lebih jelas daripada yang digunakan dalam usulan Parlemen Eropa. Keduanya mengharuskan pengguna sistem AI untuk mendapatkan manfaat dari penggunaannya. Namun, sementara operator frontend juga mengharuskan kendali atas sumber bahaya yang dibentuk oleh sistem tersebut, operator frontend merasa puas dengan keputusan untuk menggunakan sumber bahaya tersebut.

16.2 TANGGUNG JAWAB SUBJEKTIF DALAM KASUS KAUSALITAS ALTERNATIF

Keterlibatan sistem AI dalam menghasilkan kerugian meningkatkan kesulitan dalam menetapkan kesalahan dan kausalitas. Tidak selalu mudah untuk menunjukkan adanya imputasi subjektif atas kerugian yang dialami pelaku, karena kurangnya tujuan atau kelalaian karena kebaruan situasi seperti ini belum memungkinkan pengembangan kewajiban kehati-hatian sama seperti sulitnya mengevaluasi kesalahan pelaku. Otonomi sistem ini membuat mustahil, pada tingkat yang lebih besar atau lebih kecil, untuk meramalkan bagaimana mereka akan bertindak dalam kasus tertentu. Kurangnya prediktabilitas menghambat kemampuan untuk membuat prognosis mengenai kemungkinan akibat dari tindakan tersebut. Hal ini, pada gilirannya, dapat menghambat penilaian kesalahan pelaku karena telah menciptakan atau menggunakan AI dalam keadaan konkret tersebut.

Dengan cara yang sama, akan sangat sulit untuk melanjutkan imputasi objektif atas kerugian tersebut terhadap perilaku salah satu peserta dalam proses tersebut, karena ketidakmungkinan untuk memastikan penyebab sebenarnya dari kerugian tersebut. Salah satu kemungkinan hasil pada tingkat ini adalah kesimpulan bahwa setiap partisipasi dalam proses penciptaan atau penggunaan sistem AI dapat menyebabkan kerugian tersebut. Dengan kata lain, salah satu dari mereka dapat menjadi sumber kerugian, tetapi tidak diketahui siapa di antara mereka yang sebenarnya bertanggung jawab atas kerugian tersebut.

Solusi untuk masalah kausalitas alternatif banyak dibahas dalam teori hukum, dan masih diperdebatkan apakah seseorang harus:

(a) mengesampingkan tanggung jawab para pelaku, karena kurangnya kausalitas;

(b) mengecualikan dalam kasus-kasus tersebut persyaratan kausalitas;
(c) mengganti kausalitas aktual dengan kausalitas yang mungkin. Alih-alih menunjukkan bahwa tindakan masing-masing pelaku merupakan penyebab aktual kerugian, cukuplah membuktikan bahwa tindakan tersebut merupakan penyebab potensial atau yang mungkin dari kerugian tersebut. Pada saat yang sama, dapat ditetapkan pembalikan beban pembuktian terkait asumsi ini, dengan mengasumsikan adanya kausalitas potensial.

Mayoritas penulis condong ke opsi terakhir, menolak opsi pertama karena, dalam menyeimbangkan kepentingan calon pelaku kerugian dan orang yang terdampak, opsi pertama lebih diutamakan. Tampaknya tidak dapat dipertahankan jika kita menganggap bahwa salah satu pelaku melakukan tindakan yang dapat menyebabkan kerugian dan mungkin benar-benar menyebabkannya. Opsi kedua ditolak karena menimbulkan penyimpangan yang tidak perlu dan tidak terpuji terhadap aturan sistem tanggung jawab kesalahan.

Ketika persyaratan kausalitas diabaikan, baik kausalitas aktual dari tindakan maupun kesesuaiannya untuk menimbulkan kerugian tidak dievaluasi. Hal ini dapat mengakibatkan tanggung jawab seseorang yang tidak melakukan tindakan yang dapat menimbulkan kerugian. Di sisi lain, pilihan ketiga, selain didasarkan pada kepentingan orang yang terdampak yang lebih dominan daripada kepentingan calon pelaku cedera yang bertentangan, tidak melibatkan risiko meminta pertanggungjawaban seseorang yang tidak dapat berkontribusi pada peristiwa tersebut.

Praduga kausalitas potensial merupakan cara untuk meringankan persyaratan pembuktian dalam hal ini, melindungi orang yang terdampak, karena sulit untuk membuktikan kecukupan partisipasi setiap individu dalam menghasilkan keseluruhan hasil. Perlu ditegaskan bahwa keringanan beban pembuktian kausalitas ini tidak mengecualikan pemenuhan unsur-unsur pertanggungjawaban perdata lainnya bagi setiap pelaku yang terlibat dalam proses kausalitas. Bahkan dalam kaitannya dengan kausalitas, sebagaimana telah disebutkan, kecukupan perilaku individu terhadap produksi seluruh kerugian harus dibuktikan, kecuali jika legislator telah menetapkan praduga kecukupan untuk melindungi posisi orang yang terdampak.

Dalam kasus seperti itu, tanggung jawab dapat dikecualikan jika calon pelaku membuktikan bahwa perilakunya tidak menyebabkan kerugian, bahwa perilaku pelaku lain menyebabkan kerugian, atau bahwa dalam kasus ini terdapat alasan pembenar, pembebasan, atau bahkan impunitas, yang menguntungkan dirinya atau salah satu pelaku lain yang terlibat.

Terkadang dipertanyakan apakah penerapan rezim tanggung jawab bersama dan terpisah dari semua pelaku ini harus bergantung pada verifikasi tiga persyaratan: adanya hubungan kronologis antara tindakan individu, adanya hubungan spasial antara tindakan yang sama, dan/atau kesamaan sifat tindakan yang dilakukan oleh masing-masing pelaku. Selain hubungan objektif ini, juga dipertanyakan apakah hubungan subjektif harus diberlakukan, yaitu kebutuhan agar semua pelaku saling menyadari atau, dalam versi yang lebih longgar, agar mereka saling menyadari.

Ketiadaan partisipasi bersama dalam kasus-kasus ini yang dicirikan oleh kesadaran bilateral akan kerja sama tampaknya membuktikan ketidakbenaran formulasi pertama dari hubungan subjektif yang telah disebutkan sebelumnya. Memang benar bahwa formulasi kedua tidak tercakup dalam rezim partisipasi bersama, karena didasarkan justru pada ketidaktahuan (meskipun bersalah) akan kerja sama tersebut. Secara teoretis, akan ada ruang untuk persyaratan semacam itu. Namun, kami yakin bahwa persyaratan ini tidak pada tempatnya, karena jarang menghormati kepentingan orang yang terdampak. Kecuali untuk kasus-kasus di mana orang-orang bertindak berdampingan, hampir mustahil bagi seorang agen untuk menyadari pihak lain.

Demikian pula, tidak ada alasan untuk membatasi tanggung jawab agen berdasarkan kedekatan fisik mereka, kedekatan temporal perilaku mereka, atau kesamaan di antara mereka. Memang benar bahwa dalam beberapa kasus hal ini akan terjadi secara alami. Pertimbangkan, misalnya, kasus-kasus di mana dua orang yang tidak saling mengenal menembak satu sama lain, tanpa mampu menunjukkan tembakan mana yang fatal, sejauh salah satu di antaranya dapat menjadi penyebab kematian. Kontur situasi menunjukkan bahwa kedua subjek memang dekat secara fisik, bahwa tindakan yang sesuai relatif sinkron, dan bahwa mereka memiliki karakteristik yang identik. Namun, terkadang hal ini tidak terjadi dan tidak ada alasan yang relevan secara material untuk memperlakukan masalah secara berbeda, yaitu untuk menolak klaim perlindungan korban. Hal ini terjadi, misalnya, dalam kasus-kasus di mana seseorang terinfeksi AIDS dan tidak mungkin untuk menentukan, pada saat penyakit tersebut terdeteksi, apakah asal-usulnya ditemukan dalam transfusi darah yang terkontaminasi yang telah diterimanya di masa lalu atau dalam hubungan intim yang juga ia miliki di masa lalu dengan orang yang terinfeksi. Jika dalam kedua kasus tersebut terdapat kesalahan atau kelalaian dari calon pelaku, apa dasar untuk menolak klaim kompensasi dari orang yang terdampak?

Mempertanggungjawabkan pelaku yang berpotensi merugikan menyiratkan penekanan pada bahaya tindakan yang dilakukan oleh masing-masing pelaku. Bukan kerugian yang membenarkan pertanggungjawaban pelaku karena ia mungkin bukan pelakunya melainkan kemampuan tindakan untuk menghasilkan kerugian tersebut. Logika yang mendasarinya tampaknya lebih konsisten dengan gagasan bahwa, dalam kasus-kasus ini, pelanggaran harus dilihat sebagai pelanggaran dengan bahaya konkret dan bukan sebagai pelanggaran dengan akibat.

Berdasarkan hal ini, masuk akal untuk berpendapat bahwa, setiap kali semua orang yang terlibat dalam sistem AI secara individual melakukan tindakan melawan hukum dan patut disalahkan, yang secara abstrak mampu menghasilkan kerugian, solusinya pada prinsipnya adalah pertanggungjawaban bersama dari mereka yang terlibat, meskipun tidak mungkin untuk mengidentifikasi tindakan konkret di balik kerugian tersebut. Ini berarti bahwa setiap pelaku bertanggung jawab atas keseluruhan kerugian dan kemudian dapat mengklaim kembali bagian yang menjadi tanggung jawab masing-masing dalam hubungan internal.

Jika, secara umum, kebutuhan akan hubungan objektif dan/atau subjektif sangat diragukan, dalam situasi ini kami yakin bahwa persyaratan tersebut tampaknya tidak masuk

akal. Penyebaran, baik dalam hal subjek yang mengintervensi proses penciptaan dan penggunaan AI maupun dalam hal waktu, mengingat waktu yang mungkin memediasi antara intervensi tersebut, hampir tidak memungkinkan perlindungan kepentingan orang yang terdampak.

Di sisi lain, mengingat perbedaan sifat keterlibatan pembuatan, pemrograman, penyisipan data, pembaruan, penggunaan sistem AI orang yang terdampak hampir tidak akan mendapatkan kompensasi atas kerugian yang diderita. Meskipun dari sudut pandang konseptual ini bisa menjadi solusi, dari sudut pandang nilai, ini bukanlah hasil yang paling tepat. Dalam kasus seperti itu, yang penting adalah memutuskan kepentingan mana yang layak dilindungi: kepentingan orang yang terdampak atau kepentingan pelaku. Mengingat bahwa pelaku telah melakukan kesalahan yang dapat menyebabkan kerugian, tidak ada alasan untuk mengutamakan kepentingan mereka di atas kepentingan orang yang terdampak.

Salah satu masalah yang dibahas dalam Laporan 2019 justru adalah masalah kausalitas alternatif. Disarankan agar rezimnya serupa dengan kausalitas berganda, dengan setiap peserta bertanggung jawab secara tanggung renteng atas semua kerugian yang diderita. Meskipun penyebab sebenarnya dari kerugian tersebut tidak diketahui, dimungkinkan untuk menetapkan tingkat probabilitas di antara tindakan-tindakan berbagai pelaku. Dalam skenario seperti itu, disarankan agar beban pembuktian diletakkan di pihak orang yang tindakannya memiliki probabilitas lebih tinggi untuk menyebabkan kerugian.

Perlu ditegaskan bahwa Laporan 2019 mengusulkan sistem pertanggungjawaban berbasis kesalahan sebagai aturan pertanggungjawaban perdata, meskipun laporan tersebut mengakui keringanan aturan beban pembuktian dalam hal kausalitas, dengan mempertimbangkan:

1. "kemungkinan bahwa teknologi setidaknya berkontribusi terhadap kerugian";
2. "kemungkinan bahwa kerugian disebabkan oleh teknologi atau oleh sebab lain dalam lingkup yang sama";
3. "risiko adanya cacat yang diketahui dalam teknologi, meskipun dampak kausal aktualnya tidak terbukti dengan sendirinya";
4. "tingkat ketertelusuran dan kejelasan proses dalam teknologi yang mungkin berkontribusi terhadap penyebab (asimetri informasi)";
5. "tingkat aksesibilitas dan pemahaman data yang dikumpulkan dan dihasilkan oleh teknologi secara ex-post";
6. "jenis dan tingkat kerugian yang berpotensi dan benar-benar ditimbulkan".

"Apabila kerusakan tersebut merupakan jenis yang seharusnya dihindari oleh aturan keselamatan, kegagalan untuk mematuhi aturan keselamatan tersebut, termasuk aturan tentang keamanan siber, harus mengarah pada pembalikan beban pembuktian:

(a) kausalitas, dan/atau.
(b) kesalahan, dan/atau.
(c) keberadaan cacat".

Resolusi Parlemen Eropa 2020 tidak membahas masalah kausalitas alternatif. Meskipun menetapkan tanggung jawab bersama dari berbagai operator yang dapat dimintai

pertanggungjawaban, tidak jelas apakah aturan tersebut ditujukan untuk kasus-kasus partisipasi bersama, kepengarangan paralel, kausalitas alternatif, atau untuk semua. Ini berarti bahwa kemungkinan tanggung jawab para peserta dalam proses kausalitas tidak dapat dipastikan ketika seseorang tidak dapat menentukan tindakan mana yang secara efektif menyebabkan kerusakan. Selain itu, meskipun tanggung jawab mereka diterima, tidak ada posisi yang diambil mengenai kemungkinan perlunya membuktikan kesesuaian tindakan tersebut untuk menghasilkan kerusakan.

Menurut usulan ini, sistem pertanggungjawaban dasar haruslah pertanggungjawaban berbasis kesalahan, meskipun sistem ini juga menyediakan praduga kesalahan yang dapat dibantah dari pihak operator.

Secara keseluruhan, dalam kasus kausalitas alternatif, solusinya tentu akan berupa salah satu dari tiga hal berikut:

(a) pengecualian pertanggungjawaban perdata;
(b) pertanggungjawaban sebagian dari masing-masing pihak atas sebagian dari total kerugian;
(c) pertanggungjawaban bersama dan tanggungan semua pihak atas seluruh kerugian.

Dari sudut pandang teknis, semua solusi ini layak. Solusi pertama didasarkan pada ketiadaan hubungan sebab akibat yang konkret. Solusi kedua dan ketiga, secara berbeda, menekankan kesalahan masing-masing pelaku atau kerugian yang diderita oleh orang yang terdampak. Satu-satunya perbedaan adalah cara menilai hubungan sebab akibat. Alih-alih memerlukan bukti hubungan sebab akibat in concreto, hubungan sebab akibat in abstracto sudah cukup. Perbedaannya hanya terletak pada rezim kepatuhan terhadap kewajiban yang dibebankan kepada mereka.

Dalam solusi kedua, sistem tanggung jawab bersama diterapkan, artinya, setiap pelaku hanya bertanggung jawab atas sebagian dari kompensasi. Pihak yang terdampak tidak dapat menuntut kompensasi penuh dari calon pelaku kerugian, sebagaimana tidak ada satu pun calon pelaku kerugian yang terikat oleh keseluruhan kompensasi. Dalam solusi ketiga yang diusulkan, rezim tanggung jawab bersama dan masing-masing berlaku, di mana setiap pelaku bertanggung jawab atas kompensasi total, dengan kemungkinan menuntut pembayaran kembali dalam hubungan internal.

Solusi pertama lebih menekankan kepentingan calon pelaku kerugian dibandingkan kepentingan orang yang terdampak. Hal ini tampaknya bukan pilihan yang paling tepat. Banyaknya orang yang terlibat dalam pembuatan dan penggunaan sistem AI, yang berlokasi atau berasal dari berbagai wilayah dunia dan bidang kegiatan yang berbeda, membuat identifikasi dan penentuan lokasi mereka menjadi sangat sulit.

Oleh karena itu, terlalu memberatkan untuk memaksakan kepada orang yang terdampak seringkali orang perseorangan yang tidak mengetahui semua detail ini kebutuhan untuk menuntut masing-masing peserta guna mendapatkan kompensasi atas semua kerugian yang diderita. Bahkan, akan lebih mudah bagi salah satu peserta untuk menemukan peserta lainnya dan menjalankan haknya untuk menuntut kembali pembayaran. Oleh karena itu,

sistem tanggung jawab bersama dan tanggung jawab bersama tampaknya lebih tepat untuk situasi ini.

Anggapan kecukupan kausal juga tampaknya tepat dalam konteks ini, mengingat, di satu sisi, sifat keseluruhan sistem yang sangat teknis, spesifik, dan kompleks, dan, di sisi lain, (bukan karena kesalahan) kurangnya pengetahuan calon korban tentang cara kerja sistem. Catatan akhir untuk menyebutkan bahwa 'Arahan Tanggung Jawab AI', meskipun tidak membahas masalah kausalitas alternatif, mengusulkan dua langkah yang sangat penting terkait tanggung jawab berbasis kesalahan: pemberian wewenang kepada pengadilan nasional untuk memerintahkan penyedia atau pengguna sistem AI untuk mengungkapkan bukti relevan yang dimilikinya mengenai sistem AI berisiko tinggi tertentu (pasal 3) dan penetapan praduga yang dapat dibantah tentang hubungan sebab akibat antara kesalahan tergugat dan keluaran yang dihasilkan oleh sistem AI atau kegagalan sistem AI untuk menghasilkan keluaran tersebut (pasal 4).

16.3 TANGGUNG JAWAB MUTLAK

Quid iuris ketika tidak ada kesalahan yang dilakukan dan kerugian disebabkan oleh berfungsinya sistem AI? Satu-satunya jalan yang mungkin adalah tanggung jawab mutlak. Dalam konteks ini, dua pertanyaan telah diajukan:

1. Apakah ada rezim pertanggungjawaban ketat yang berlaku saat ini yang dapat diterapkan secara langsung atau analogi terhadap masalah ini?
2. Apakah perlu atau disarankan untuk merancang rezim khusus untuk kerugian yang disebabkan oleh sistem AI?

Rezim pertanggungjawaban ketat saat ini yang disajikan sebagai solusi yang memungkinkan untuk masalah ini terutama: pertanggungjawaban produk, pertanggungjawaban atas kerusakan yang disebabkan oleh hewan, dan pertanggungjawaban atas kerusakan yang disebabkan oleh kendaraan bermotor.

Menurut Arahan Dewan 85/374/EEC (PLD), produsen yang dipahami sebagai produsen atau importir barang ke UE untuk didistribusikan sebagai bagian dari kegiatan komersialnya bertanggung jawab atas cacat pada produknya. Produk berarti "semua barang bergerak, kecuali produk pertanian primer dan hewan buruan, meskipun digabungkan ke dalam barang bergerak lain atau ke dalam barang tidak bergerak". Listrik juga dianggap sebagai "produk". Produsen hanya bertanggung jawab atas cacat produk pada saat dipasarkan, bukan cacat yang muncul setelahnya. Korban bertanggung jawab untuk membuktikan kerusakan, cacat, dan penyebab kerusakan akibat cacat tersebut.

Beberapa kesulitan telah diidentifikasi dalam penerapan rezim ini terhadap kerusakan yang disebabkan oleh sistem AI. Pertama, rezim ini tidak akan sepenuhnya menyelesaikan masalah karena tidak membahas kemungkinan tanggung jawab pemilik, pemegang, atau pengguna sistem AI. Artinya, rezim ini hanya dapat menawarkan solusi parsial. Bahkan dalam bidang pembuatan sistem AI, terdapat kendala dalam penerapannya.

Hal ini juga berlaku untuk definisi produsen dan produk. Meskipun tidak diragukan lagi bahwa pembuatan perangkat keras dapat dilihat sebagai aktivitas produksi dan hasilnya

sebagai produk, hal yang sama tidak berlaku dalam hal pembuatan algoritma yang menjadi dasar AI, atau untuk memberi makan sistem tersebut.

Definisi produk dalam PLD mungkin memberi kesan bahwa hanya benda bergerak dan berwujud yaitu, benda yang dapat dirasakan oleh indra yang layak mendapatkan kualifikasi tersebut. Algoritma atau data yang memasukkannya hampir tidak mungkin termasuk dalam kategori tersebut. Oleh karena itu, sulit juga untuk menganggap pemrogram atau orang yang memasukkan data sebagai produsen dalam pengertian PLD. Tentu saja, seseorang selalu dapat mencoba melihat norma sebagai instrumen hidup yang tunduk pada interpretasi evolusioner, menyesuaikannya dengan realitas saat ini (dan bukan dengan standar tahun 1985), atau, jika ini tidak memungkinkan, menggunakan analogi. Namun, tidak pasti apakah ini akan membuahkan hasil, karena produsen hanya bertanggung jawab atas cacat pada produk yang ada pada saat produk tersebut dipasarkan.

Masalah besar dengan kerusakan yang disebabkan oleh sistem AI terletak pada fakta bahwa risiko cedera lebih berkaitan dengan otonomi sistem tersebut daripada kemungkinan cacat dalam desainnya. Dalam kebanyakan kasus, tidak ada cacat. Evolusi sistem tidak dapat dikontrol oleh perancang, pemrogram, atau orang lain yang terlibat dalam penyediaan dan pembaruannya. Selain itu, biasanya, kesalahan terjadi lama setelah sistem dipasarkan dan tidak diketahui atau tidak dapat diidentifikasi pada saat itu.

Menghadapi kesulitan-kesulitan ini, usulan PLD baru menetapkan: perluasan pengertian produk untuk secara eksplisit mencakup berkas dan perangkat lunak manufaktur digital (pasal 4), sehingga menghilangkan ketidakpastian tentang kualifikasi sistem AI sebagai produk; praduga cacat (pasal 6); dan praduga hubungan sebab akibat antara cacat dan kerusakan (pasal 9). Wewenang pengadilan nasional untuk memerintahkan tergugat mengungkapkan bukti relevan yang dimilikinya juga ditetapkan (pasal 8).

Tanggung jawab mutlak biasanya didasarkan pada salah satu dari dua pilar: posisi kendali atas sumber bahaya atau pemanfaatan sumber bahaya tersebut. Tanggung jawab atas kerusakan yang disebabkan oleh hewan tampaknya justru mencari dukungan dalam pemanfaatan hewan tersebut oleh pemiliknya. Kemungkinan penerapan rezim ini terhadap kerusakan akibat sistem AI hanya dapat dilakukan melalui analogi. Kesamaan antara kedua kasus ini terletak pada ketidakpastiannya. Layaknya hewan, kinerja sistem AI juga tidak dapat diprediksi. Ketidakpastian inilah yang menciptakan risiko yang mendasari kedua realitas ini. Dari perspektif ini, tidak ada yang dapat menghalangi penerapan aturan pertanggungjawaban atas kerusakan yang disebabkan oleh hewan dengan kerusakan yang disebabkan oleh sistem AI, dengan analoginya.

Pertanyaannya adalah apakah solusi yang diusulkan paling tepat untuk masalah ini. Jika dicermati, pilihan yang diambil adalah meminta pertanggungjawaban mereka yang memanfaatkan sumber bahaya atau mereka yang memanfaatkan sumber bahaya untuk kepentingan mereka sendiri. Ini berarti bahwa mereka yang menciptakan sumber bahaya atau mereka yang memanfaatkan sumber bahaya tersebut untuk kepentingan orang lain tidak akan bertanggung jawab atas kerugian yang timbul darinya. Dengan menerapkan hal ini ke dunia digital, hal ini sama saja dengan mengecualikan pertanggungjawaban pencipta sistem AI dan,

jika demikian, pengguna yang menggunakannya untuk kepentingan orang lain. Dalam banyak kasus, ini tampaknya bukan solusi yang paling tepat untuk masalah kita. Bahkan, jangan dilupakan bahwa mereka yang merancang, memprogram, memberi makan, dan memperbarui sistem ini menentukan fungsinya. Dalam sistem perangkat lunak tertutup, tidak ada yang memiliki akses selain entitas-entitas ini. Tingkat informasi dan pemahaman sistem juga sama sekali tidak sama dengan tingkat pemahaman penggunanya. Dengan mempertimbangkan hal ini, apakah masuk akal untuk mendasarkan tanggung jawab sepenuhnya dan semata-mata pada pemanfaatan risiko, dengan mengecualikan mereka yang menciptakannya atau yang dapat membatasinya, pada tingkat yang lebih besar atau lebih kecil? Dapat dikatakan bahwa keputusan akhir untuk menggunakan sistem AI terletak pada penggunanya. Namun, keputusan itu tidak berarti mengendalikan risiko sistem. Keputusan ini tidak memengaruhi desain model AI. Penting untuk membedakan antara bahaya intrinsik, yang diakibatkan oleh konfigurasi sistem, dan bahaya yang diakibatkan oleh keputusan untuk menggunakan sistem tersebut dalam keadaan yang tidak tepat. Dalam kasus pertama, bahaya berasal dari sistem itu sendiri, sedangkan dalam kasus kedua, bahaya diakibatkan oleh keputusan buruk penggunanya.

Di sini, penting untuk memulai dengan mempertanyakan apakah keputusan buruk tersebut cukup menjadi dasar pertanggungjawaban berdasarkan kesalahan, khususnya pelanggaran aturan lalu lintas. Jika tidak demikian, kita harus mengandalkan pertanggungjawaban mutlak, yang didasarkan pada pemanfaatan sumber bahaya oleh pelaku. Sebaliknya, dalam kasus ini, sumber pertanggungjawaban mutlak harus ditemukan dalam bahaya sistem, dan dapat dibahas apakah tanpa skenario seperti itu akan lebih tepat untuk meminta pertanggungjawaban mereka yang memiliki posisi kendali (relatif) atas sumber bahaya atau mereka yang memanfaatkan sumber tersebut, atau keduanya.

Kami yakin bahwa solusi terakhir adalah yang paling tepat. Tidak masuk akal untuk membebaskan perancang algoritma, pemrogram, atau orang yang memasukkan atau memperbarui data dari pertanggungjawaban. Mereka berada dalam posisi terbaik untuk mengendalikan sumber bahaya ini, dan mereka juga mendapatkan manfaat darinya, meskipun secara tidak langsung. Meskipun, pada umumnya, mereka tidak mendapatkan keuntungan dari keuntungan yang diciptakan oleh sistem, mereka memanfaatkan nilainya dengan memperdagangkannya. Demikian pula, tidaklah masuk akal untuk membebaskan pengguna dari segala kemungkinan tanggung jawab. Selain memanfaatkan sumber bahaya, mereka sendiri memiliki wewenang untuk memutuskan apakah akan menggunakan sistem AI dalam keadaan tertentu tersebut. Keputusan mereka untuk menggunakannya, meskipun bukan satu-satunya penyebab bahaya, berkontribusi pada pemeliharaan atau peningkatannya. Yang harus ditentukan adalah apakah kerusakan tersebut sesuai dengan perwujudan salah satu bahaya yang dihasilkan atau diperparah oleh penciptaan dan/atau penggunaan sistem AI. Dengan kata lain, pertanyaannya adalah apakah kerusakan tersebut merupakan akibat dari perwujudan bahaya-bahaya tersebut dari sistem itu sendiri dan keputusan untuk menggunakannya dalam konteks dan tujuan tertentu tersebut atau hanya salah satunya.

Jika kerusakan berasal dari keduanya, seharusnya ada tanggung jawab bersama dan masing-masing dari semua peserta dalam proses kausal. Jika kerusakan berasal dari hanya salah satu dari mereka, hanya dia yang bertanggung jawab. Namun, kita harus memperhatikan fakta bahwa dalam setiap kelompok pencipta atau pengguna mungkin terdapat beberapa orang yang berpotensi membahayakan. Karena mustahil untuk menentukan tingkat bahaya dari setiap partisipasi individu dan kontribusi masing-masing terhadap terjadinya kerusakan, setiap orang yang berpotensi membahayakan harus bertanggung jawab secara bersama-sama atas kerusakan tersebut.

Rezim pertanggungjawaban atas kerusakan yang disebabkan oleh hewan tampaknya tidak mencakup semua situasi ini. Rezim pertanggungjawaban atas kerusakan yang disebabkan oleh kendaraan bermotor juga tampaknya bukan solusi untuk masalah kita, karena akan kembali menghukum pengguna sistem AI. Faktanya, pertanggungjawaban atas kerusakan yang disebabkan oleh kendaraan bermotor selalu berada di tangan pemilik atau pengguna kendaraan, karena ia adalah orang yang diuntungkan. Karena alasan-alasan yang telah disebutkan, visi semacam itu bukanlah jawaban yang paling tepat untuk permasalahan kita. Namun, beberapa sistem hukum tidak hanya menuntut orang yang bertanggung jawab menggunakan kendaraan untuk kepentingannya sendiri, tetapi juga bahwa ia benar-benar mengemudi, yang menunjukkan perlunya memiliki posisi kendali atas sumber bahaya. Rezim semacam itu mengesampingkan tanggung jawab pemilik atau pengguna, ketika mereka menggunakan kendaraan untuk kepentingan pihak ketiga, dan potensi tanggung jawab perancang sistem. Namun, penting untuk dipahami bahwa posisi kendali ini hanya menyangkut kemungkinan untuk menentukan apakah kendaraan digunakan dan bagaimana penggunaannya. Tidak diperlukan kendali atas konstruksi dan kinerja kendaraan yang tepat. Dari sudut pandang subjektif, hal ini penting karena mengecualikan produsen kendaraan tersebut dari cakupan penerapan rezim ini. Artinya, perancang algoritma, pemrogram, dan orang-orang yang memberi makan atau memperbarui sistem juga dikecualikan dari tanggung jawab di sini. Hanya pemilik, pemegang, atau pengguna yang tetap bertanggung jawab.

Ketidaksempurnaan mesin membenarkan sistem tanggung jawab yang ketat. Ketidaksempurnaan ini juga terdapat dalam sistem AI. Oleh karena itu, analogi juga dapat diterapkan pada kerusakan yang disebabkan oleh sistem AI. Dalam beberapa kasus, analogi mungkin tidak diperlukan dan aturan yang dimaksud bahkan dapat diterapkan secara langsung pada situasi tersebut. Hal ini terjadi, misalnya, dalam kecelakaan yang melibatkan kendaraan otonom, meskipun masih dipertanyakan sejauh mana akan ada pengarahan kendaraan yang efektif dalam kasus otomatisasi penuh. Namun demikian, kami meragukan kecukupan rezim ini untuk menyelesaikan masalah dasar kami, karena rezim ini sekali lagi menghukum pengguna sistem AI demi kepentingannya sendiri, mengabaikan tanggung jawabnya dalam hipotesis penggunaan demi kepentingan pihak ketiga, dan, yang lebih penting, mengesampingkan potensi tanggung jawab perancang sistem. VI. Laporan 2019 mendukung penerapan tanggung jawab ketat bagi operator yang mendapatkan manfaat dari atau mengendalikan sistem. Namun, laporan tersebut membatasi tanggung jawab tersebut pada kasus-kasus di mana sistem AI digunakan "di lingkungan non-pribadi" dan "biasanya dapat

menyebabkan kerugian yang signifikan". Oleh karena itu, hal ini mengecualikan kasus-kasus di mana sistem digunakan di lingkungan tertutup, yang memaparkan sejumlah kecil orang pada risiko cedera, yang dapat terjadi, misalnya, dalam penggunaan AI dalam pelaksanaan prosedur medis.

Kemungkinan tingkat keparahan cedera tampaknya lebih besar daripada tingkat keparahannya. Jika ada operator yang diuntungkan dari risiko dan operator lain yang mengendalikannya, "tanggung jawab mutlak harus berada di tangan operator yang memiliki kendali lebih besar atas risiko operasi", dengan demikian menunjukkan keutamaan kendali dibandingkan posisi yang mengambil keuntungan dari sumber bahaya. Laporan tersebut juga menganjurkan perluasan tanggung jawab produk untuk mencakup cacat perangkat lunak yang terjadi setelah dipasarkan.

Resolusi EP 2020 membatasi tanggung jawab mutlak operator terhadap kerusakan yang disebabkan oleh sistem AI berisiko tinggi, yang tercantum dalam lampiran proposal. Menurut rezim yang diusulkan, sistem AI berisiko tinggi harus dipahami sebagai sistem yang memiliki "potensi signifikan dalam sistem AI yang beroperasi secara otonom untuk menyebabkan kerugian atau kerusakan pada satu orang atau lebih secara acak dan melampaui apa yang dapat diperkirakan secara wajar; signifikansi potensi tersebut bergantung pada interaksi antara tingkat keparahan potensi kerugian atau kerusakan, tingkat otonomi pengambilan keputusan, kemungkinan risiko tersebut terwujud, serta cara dan konteks penggunaan sistem AI".

Kesamaan dalam proposal-proposal ini adalah gagasan untuk mencoba membatasi tanggung jawab mutlak pada kasus-kasus tertentu. Pembenaran untuk hal ini tidak terletak pada pertimbangan hukum melainkan pada kebijakan. Tujuannya adalah untuk membangun rezim yang tidak menghalangi para ilmuwan dan pengembang sistem AI untuk melanjutkan penelitian dan aktivitas mereka. Pilihan ini dapat dimengerti, meskipun sulit untuk menerima hasil yang dihasilkannya. Mustahil, misalnya, seorang pasien yang menderita kerusakan serius pada jiwa atau integritas fisiknya akibat prosedur medis menggunakan sistem AI tidak akan mendapatkan kompensasi.

Pertimbangan proporsionalitas dan kewajaran mencegah pelanggaran terhadap kebaikan apa pun yang setara atau lebih unggul dari kebaikan yang dilindungi. Pengembangan teknologi kecil kemungkinannya akan menjadi kebaikan yang lebih unggul daripada kehidupan atau bahkan, dalam kasus tertentu, daripada integritas fisik. Lebih lanjut, fakta bahwa pengembang tidak bertanggung jawab tidak mendorong mereka untuk menginvestasikan sumber daya mereka dalam meningkatkan sistem.

Dana kompensasi atau asuransi wajib hanya akan mencegah ketidaknyamanan ini jika terdapat tanggung jawab atas kerugian yang diderita seseorang akibat tindakan Sistem AI, terlepas dari apakah terdapat kesalahan atau tanggung jawab mutlak. Bukti sederhana atas kerugian tersebut meskipun mungkin masuk akal untuk membatasi kerugian yang dapat dikompensasi oleh dana ini sesuai dengan sifat dan tingkat keparahannya dan penyebabnya masing-masing akan cukup untuk membenarkan kompensasi yang didukung oleh dana atau perusahaan asuransi. Jika tidak, dana atau perusahaan asuransi hanya akan menggantikan

pihak yang dirugikan dalam memenuhi kewajibannya untuk memberikan kompensasi, tanpa memperluas perlindungan kepentingan calon pihak yang dirugikan.

Berdasarkan semua pemikiran ini, kami menganjurkan rezim tanggung jawab yang menangani masalah-masalah ini secara tepat. Jika tidak, di satu sisi, banyak situasi akan tetap tidak terlindungi dan, di sisi lain, pengguna sistem jenis ini akan dianggap bertanggung jawab, membebaskan pengembang dari segala tanggung jawab. Sebagaimana telah disebutkan, hal ini tampaknya tidak tepat. Sayangnya, usulan awal Rancangan Arahan Tanggung Jawab AI tidak menerima pendekatan tanggung jawab mutlak. Sebaliknya, ia menerima tanggung jawab berdasarkan kesalahan, dengan beberapa instrumen khusus: praduga kausalitas yang dapat dibantah dan rezim pengungkapan bukti. Berdasarkan alasan-alasan yang telah dijelaskan sebelumnya, kami berpendapat bahwa rezim yang diusulkan tidak memadai untuk menangani kerugian yang disebabkan oleh sistem AI. Selain itu, patut dipertanyakan apakah Rancangan Arahan tersebut konsisten dengan tingkat perlindungan yang dibayangkan oleh Rancangan Undang-Undang AI.

16.4 PEMBEBASAN TANGGUNG JAWAB ATAS KERUSAKAN AI

Tergantung pada sifat tanggung jawabnya, pertanyaan lain perlu diajukan: dalam situasi apa agen dapat terhindar dari tanggung jawab? Di sini, kami akan membahas kasus-kasus di mana tanggung jawab agen harus dikecualikan. Kami tidak memasukkan faktor-faktor yang dapat membebaskan produsen dari tanggung jawab dan apakah faktor-faktor yang tercantum dalam Pasal 7 PLD perlu ditinjau kembali, karena tidak memadai untuk mengatasi kekhususan yang timbul dari kerusakan yang disebabkan oleh sistem AI. Kami mengakui bahwa sudah terdapat doktrin yang relevan dan masuk akal yang menggarisbawahi bahwa arahan tersebut memungkinkan produsen AI untuk menghindari tanggung jawab dengan menggunakan apa yang disebut pembelaan risiko pengembangan.

Kekhawatiran yang tampaknya telah diatasi oleh proposal PLD baru adalah bahwa pembelaan risiko pengembangan tidak dapat diterapkan ketika evolusi pengetahuan ilmiah dan teknis terjadi pada periode ketika produk masih berada dalam kendali produsen. Amandemen ini menunjukkan bahwa produsen tetap bertanggung jawab jika, misalnya melalui pembaruan, ia dapat menghilangkan cacat yang terungkap oleh evolusi pengetahuan dan teknik. Namun, perlu dipertanyakan apakah perubahan tersebut cukup untuk menjamin keamanan produk berbasis AI yang sudah beredar.

Oleh karena itu, kami akan berfokus pada pengecualian yang berlaku ketika tanggung jawab tidak didasarkan pada cacat sistem AI. Tentu saja, tidak ada jawaban tunggal untuk pertanyaan ini, karena alasan untuk mengecualikan agen bervariasi karena perbedaan sifat tanggung jawab. Jika agen bertanggung jawab berdasarkan tanggung jawab mutlak, alasan untuk menghindari tanggung jawab akan berbeda dari alasan yang seharusnya diterima ketika kita memiliki rezim tanggung jawab berbasis kesalahan, bahkan dengan pembalikan beban pembuktian. Untuk tujuan ini, menentukan apakah agen tersebut merupakan pemrogram atau pengguna, operator backend atau operator frontend, sebagaimana didefinisikan di atas,

tidaklah sepenting memahami apakah ia bertanggung jawab berdasarkan sistem tanggung jawab mutlak atau rezim tanggung jawab berbasis kesalahan.

Sebenarnya, baik dalam undang-undang perdata negara anggota maupun dalam proposal Uni Eropa untuk menyelaraskan rezim hukum perdata atas kerugian yang disebabkan oleh sistem AI, terdapat kecenderungan untuk mengecualikan "solusi satu untuk semua", yang berarti bahwa kewajiban untuk mengganti kerugian yang disebabkan oleh sistem AI dapat didasarkan pada tanggung jawab mutlak atau tanggung jawab berbasis kesalahan. Meskipun demikian, proposal Komisi Eropa untuk Arahan Tanggung Jawab AI hanya membahas harmonisasi aturan untuk praduga kausalitas dan pengungkapan bukti, sehingga menyerahkan kepada masing-masing Negara Anggota untuk menentukan apakah tanggung jawab agen harus didasarkan pada tanggung jawab mutlak atau sistem tanggung jawab berbasis kesalahan.

Jika agen akan dimintai pertanggungjawaban atas kerugian yang disebabkan oleh sistem AI, ia tidak dapat mengabaikan pertanggungjawaban hanya karena kerugian tersebut disebabkan oleh sistem tersebut atau merupakan konsekuensi dari otonominya. Namun, secara teori, beberapa faktor lain dapat membebaskan agen dari pertanggungjawaban:

(i) bukti bahwa agen telah mematuhi tugas-tugas tertentu, seperti, misalnya, uji tuntas, penjagaan, dan pengawasan, serta bertindak dengan kehati-hatian yang semestinya;
(ii) bukti bahwa kerugian atau kerusakan disebabkan oleh keadaan kahar;
(iii) bukti bahwa kerugian atau kerusakan tersebut dapat dikaitkan dengan pihak ketiga;
(iv) bukti bahwa korban atau orang yang terdampak telah menyebabkan kerugian atau kerusakan;

Faktor pertama hanya dapat diterima dalam sistem berbasis kesalahan, bahkan dengan pembalikan pembuktian, karena, dalam kasus pertanggungjawaban mutlak, pertanggungjawaban agen tidak didasarkan pada adanya tindakan melawan hukum atau pelanggaran kewajiban. Dalam hukum perdata negara-negara anggota, rezim pertanggungjawaban berdasarkan praduga kesalahan memungkinkan agen untuk lepas dari tanggung jawab, ketika mereka terbukti telah bertindak dengan kehati-hatian yang semestinya untuk menghindari kerugian. Proposal yang disusun oleh Kelompok Pakar dan EP juga tampaknya menerima solusi ini, dengan berupaya menyesuaikan bukti yang harus diajukan untuk mengatasi kekhasan yang timbul dari kerusakan yang disebabkan oleh sistem AI. Dalam hal ini, Kelompok Pakar mengusulkan: “Operator teknologi digital yang sedang berkembang harus mematuhi serangkaian kewajiban kehati-hatian yang telah disesuaikan”.

Dalam sistem pertanggungjawaban ketat, agen akan bertanggung jawab terlepas dari apakah telah melanggar kewajiban yang berlaku. Tanggung jawabnya akan dibenarkan oleh risiko yang ditimbulkan oleh agen dengan pengembangan aktivitasnya atau oleh keuntungan yang diperolehnya. Mengenai kerugian yang disebabkan oleh keadaan kahar atau suatu peristiwa kebetulan, tidak diragukan lagi bahwa pembuktian keberadaannya pada prinsipnya harus mengarah pada pengecualian tanggung jawab agen, baik dalam sistem praduga bersalah maupun sistem tanggung jawab mutlak. Meskipun demikian, beberapa klarifikasi perlu dilakukan.

Pertama, bahkan dalam sistem kesalahan, jika terbukti bahwa kerugian tersebut secara langsung disebabkan oleh keadaan kahar, tetapi pada saat yang sama, ditunjukkan bahwa, jika agen telah bertindak dengan sungguh-sungguh, ia dapat menghindari kerugian tersebut, tanggung jawab agen harus tetap berlaku.

Kedua, dalam rezim tanggung jawab mutlak, kita cenderung menganggap bahwa pembuktian keberadaan keadaan kahar yang menyebabkan kerugian sudah cukup untuk menghapus tanggung jawab agen. Rumusan komprehensif yang terkadang digunakan perlu dipertimbangkan kembali, dari sudut pandang kami. Dalam Pasal 4 dan 8 Resolusi EP 2020, diusulkan bahwa “operator tidak bertanggung jawab jika kerugian atau kerusakan disebabkan oleh keadaan kahar”.

Untuk menghindari tanggung jawab hukum, pembuktian keberadaan force majeure saja tidak cukup. Pembuktian bahwa force majeure tersebut "asing" terhadap pengoperasian sistem AI juga diperlukan. Pertimbangkan dua contoh: haruskah agen bertanggung jawab atas kerusakan yang disebabkan oleh robot bedah yang, setelah gempa bumi, jatuh menimpa pasien dan melukainya? Haruskah agen bertanggung jawab atas kerusakan yang disebabkan oleh robot bedah yang menyebabkan cedera pada pasien akibat kegagalan koneksi akibat badai besar?

Dalam kasus pertama, kerusakan yang disebabkan oleh robot tersebut dapat berasal dari instrumen lain yang ada di ruang operasi. Hal yang sama tidak berlaku untuk kerusakan yang disebabkan oleh kurangnya koneksi, bahkan karena fenomena atmosfer yang luar biasa. Sangat diragukan bahwa, ketika agen bertanggung jawab penuh, tanggung jawabnya dapat dihapuskan ketika peristiwa force majeure tersebut tidak asing bagi pengoperasian sistem AI. Faktanya, salah satu risiko khas sistem AI adalah dapat menyebabkan kerugian bagi pihak ketiga ketika terjadi kegagalan konektivitas. Oleh karena itu, meskipun kegagalan tersebut disebabkan oleh situasi yang luar biasa atau tidak lazim, agen tidak boleh lepas dari tanggung jawab.

Kami sepenuhnya menyadari bahwa menggunakan konsep yang tidak pasti atau samar seperti kerugian yang disebabkan oleh keadaan kahar yang asing bagi penggunaan sistem AI akan membutuhkan upaya yang lebih besar dari pengadilan. Meskipun situasi di mana terjadi kegagalan koneksi mungkin lebih mudah dipahami dan dirumuskan, praktik dan kejadian sehari-hari tentu akan menghadirkan contoh lain yang tentu akan menimbulkan lebih banyak pertanyaan.

Pertanyaan lain yang sering muncul adalah bagaimana menangani kasus di mana agen dapat membuktikan bahwa pihak ketiga menyebabkan kerugian atau kerusakan. Otonomi pertanyaan ini dan pada akhirnya pengecualian tanggung jawab bergantung pada apakah kerugian tersebut semata-mata disebabkan oleh pihak ketiga yang bukan produsen atau operator sistem AI. Sebelumnya, kami telah membahas masalah yang timbul dari hipotesis di mana beberapa operator dapat bertanggung jawab berdasarkan sistem tanggung jawab mutlak atau rezim tanggung jawab berbasis kesalahan.

Permasalahannya di sini adalah mengidentifikasi dan memisahkan kasus-kasus di mana pihak ketiga mengganggu sistem AI, memodifikasi, atau memengaruhi operasinya. Kita dapat

mengidentifikasi beberapa contoh. Peretas yang dengan sengaja mengganggu sistem AI. Subjek yang secara lalai mengganggu fungsi robot dengan memutus catu dayanya. Anak-anak yang membajak drone pengiriman barang, dll.

Dalam situasi apa pun yang dijelaskan, tidak diragukan lagi bahwa pihak ketiga harus bertanggung jawab atas kerugian yang diderita oleh pihak yang dirugikan. Permasalahannya adalah memastikan apakah tanggung jawab pihak ketiga ini dapat mengecualikan tanggung jawab produsen atau operator, sebagaimana didefinisikan di atas.

Sebagaimana telah disebutkan sebelumnya, kita akan mengecualikan dari analisis kita tanggung jawab produsen atau pabrikan atas produk cacat. Dalam sistem tanggung jawab berbasis kelalaian dengan dugaan kesalahan, bukti bahwa pihak ketiga secara eksklusif menyebabkan cedera pada umumnya memungkinkan agen untuk lepas dari tanggung jawab, kecuali jika tindakan pihak ketiga dimungkinkan oleh pelanggaran uji tuntas agen. Dengan kata lain, pembuktian bahwa pihak ketiga secara eksklusif menyebabkan kerugian seharusnya mengecualikan tanggung jawab agen, kecuali dalam kasus di mana agen dapat mencegah tindakan pihak ketiga jika ia bertindak dengan tekun.

Mengenai tanggung jawab mutlak, tidak ada solusi seragam untuk kerugian yang secara eksklusif disebabkan oleh pihak ketiga. Aturannya adalah bahwa tanggung jawab mutlak harus dikecualikan dalam kasus-kasus ini, meskipun tanggung jawab bersama dan tanggung renteng diakui dalam rezim tertentu. Contoh paling paradigmatis adalah tanggung jawab perwakilan, ketika ditetapkan bahwa prinsipal bertanggung jawab secara perwakilan dan secara mutlak atas perbuatan melawan hukum yang dilakukan agennya.

Tidak adanya sistem yang seragam untuk menangani kerugian yang secara eksklusif disebabkan oleh pihak ketiga merupakan argumen yang mendukung rezim otonom untuk kerugian yang disebabkan oleh sistem AI. Resolusi EP 2020 sangat inovatif dalam hal ini. Menurut Pasal 8.3, "apabila kerugian atau kerusakan disebabkan oleh pihak ketiga yang mengganggu sistem AI dengan memodifikasi fungsi atau dampaknya, operator tetap bertanggung jawab atas pembayaran kompensasi jika pihak ketiga tersebut tidak dapat dilacak atau tidak memiliki aset". Dengan asumsi bahwa seringkali sulit untuk mengidentifikasi orang dari pihak ketiga dan/atau bahwa pihak ketiga tersebut mungkin tidak memiliki aset yang cukup untuk membayar kompensasi, diusulkan agar tanggung jawab agen tetap dipertahankan, bahkan dalam kasus-kasus yang dibingkai dalam rezim tanggung jawab berbasis kesalahan. Di sisi lain, opsi ini menyiratkan bahwa, a fortiori, solusi tersebut harus diterapkan pada hipotesis tanggung jawab mutlak.

Meskipun kami memahami kekhawatiran di balik usulan tersebut, sulit untuk tidak mempertanyakan apakah kami tidak menghadapi tanggung jawab mutlak, terlepas dari kualifikasi yang diusulkan oleh Parlemen Eropa. Ketika terbukti bahwa agen bertindak dengan tekun dan tidak dapat menghindari tindakan pihak ketiga, dan meskipun demikian, ia tetap bertanggung jawab karena pihak tersebut tidak dapat dilacak atau tidak memiliki uang, kesimpulannya adalah Resolusi EP 2020 mendukung pendekatan tanggung jawab mutlak.

Ketika sistem AI menyebabkan kerusakan, seseorang tidak dapat mengesampingkan kemungkinan bahwa orang yang terdampak, melalui tindakan atau kelalaiannya, telah

berkontribusi terhadap kerusakan yang diderita atau sejauh mana kerusakan tersebut terjadi. Menurut pemahaman yang lebih modern, tanggung jawab agen hanya dapat dikecualikan ketika perilaku orang yang terdampak merupakan satu-satunya penyebab kerusakan. Dalam kasus lain, perilaku lalai dari pihak yang dirugikan seharusnya hanya menjadi dasar untuk mengurangi tanggung jawab. Solusi ini dianut oleh Laporan 2019 dan Resolusi EP 2020. Namun, terdapat sistem hukum, seperti Portugal, di mana Kitab Undang-Undang Hukum Perdata masih mengatur pengecualian tanggung jawab sepenuhnya dalam beberapa kasus yang dijelaskan. Namun, solusi ini sering dikritik dan tidak memadai untuk situasi di mana kerugian disebabkan oleh sistem AI dan pihak yang dirugikan secara bersamaan. Ini hanyalah contoh lain dari perlunya rezim tanggung jawab perdata yang otonom untuk kerugian AI.

BAB 17
RISIKO HUKUM BAHASA ALAMI: PERSPEKTIF SWISS

Abstrak Penggunaan dan peningkatan Pembangkitan Bahasa Alami (NLG) merupakan perkembangan terkini yang berkembang pesat. Manfaatnya berkisar dari kemudahan penerapan alat otomatisasi tambahan untuk tugas-tugas sederhana yang berulang hingga bot penasihat yang berfungsi penuh yang dapat menawarkan bantuan untuk masalah kompleks dan solusi yang bermakna di berbagai bidang. Dengan sistem otonom yang terintegrasi penuh, pertanyaan tentang kesalahan dan tanggung jawab menjadi area perhatian yang kritis. Meskipun berbagai cara untuk memitigasi dan meminimalkan kesalahan telah tersedia dan sedang ditingkatkan dengan memanfaatkan berbagai set data pengujian kesalahan, hal ini tidak menutup kemungkinan adanya cacat signifikan dalam keluaran yang dihasilkan.

Dari perspektif hukum, harus ditentukan siapa yang bertanggung jawab atas hasil yang tidak diinginkan dari algoritma NLG: Apakah pembuat kode yang memikul tanggung jawab utama atau operator yang tidak mengambil langkah-langkah yang wajar untuk meminimalkan risiko keluaran yang tidak akurat atau tidak diinginkan? Jawaban atas pertanyaan ini menjadi semakin kompleks ketika pihak ketiga berinteraksi dengan algoritma NLG yang dapat mengubah hasil. Meskipun teori gugatan tradisional mengaitkan tanggung jawab dengan kemungkinan adanya kendali, NLG mungkin merupakan penerapan yang mengabaikan gagasan ini karena algoritma NLG tidak dirancang untuk dikendalikan oleh operator manusia.

17.1 DASAR-DASAR TEKNIS PEMBANGKITAN BAHASA ALAMI

Pendahuluan Aspek Teknis

Pembangkitan Bahasa Alami (NLG) merupakan subbidang utama dari Pemrosesan Bahasa Alami dan Pembelajaran Mendalam secara keseluruhan. Terobosan terbaru dalam model autoregresif seperti GPT-2 OpenAI, GPT-3, InstructGPT, atau Primer Google telah menghasilkan demonstrasi teks yang dihasilkan mesin yang terbukti sulit atau bahkan mustahil dibedakan dari teks tertulis biasa. Saat menggunakan NLG, klaim pertanggungjawaban dapat muncul di area mana pun yang menggunakan komunikasi verbal.

Aplikasi pertama yang mengomersialkan teknologi ini sudah mulai tersedia dengan bot obrolan yang saling memperkuat, pembangkitan kode otomatis, dan lainnya yang mulai memasuki pasar. Dalam karya ini, kami akan mengeksplorasi area aplikasi baru tersebut dan berfokus terutama pada area yang kemungkinan dianggap cocok untuk penggunaan dengan niat baik, tetapi dapat menyebabkan hasil negatif yang tidak diinginkan dalam kondisi tertentu.

Kami selanjutnya mengkaji kapasitas deteksi manusia dan algoritmik terhadap teks yang dihasilkan mesin untuk memitigasi penyebaran cepat konten yang dihasilkan dalam bentuk artikel berita dan lainnya. Berdasarkan karya sebelumnya yang berfokus pada pembuatan teks hukum dengan perangkat NLG generasi sebelumnya, kami melatih model transformator auto-regresif yang lebih canggih untuk mengilustrasikan cara kerja model

tersebut dan pada titik mana operator model memiliki pengaruh langsung atau tidak langsung terhadap kemungkinan keluaran yang dihasilkan.

Di bagian kedua artikel ini, kami mengkaji isu-isu pertanggungjawaban perdata yang mungkin timbul ketika menggunakan NLG, khususnya berfokus pada Arahan tentang produk cacat dan pertanggungjawaban berbasis kesalahan berdasarkan Hukum Swiss. Mengenai hal tersebut, kami membahas dasar hukum spesifik yang dapat menimbulkan pertanggungjawaban ketika NLG digunakan.

Risiko Pembelajaran Penguatan

Pembangkitan Bahasa yang Tidak Diinginkan

Salah satu cara yang memungkinkan untuk mengadaptasi model ini dan model serupa dengan masukan penggunanya adalah dengan menerapkan pembelajaran penguatan pada model tersebut. Salah satu cara tersebut, menggunakan model berbasis transformator yang kami perkenalkan sebelumnya, adalah dengan menambahkan masukan pengguna yang relevan ke dalam set data fine-tuning. Hal ini memungkinkan operator untuk menyesuaikan model dengan perilaku pengguna dan secara teori meningkatkan keterbacaan dan pemahaman secara keseluruhan.

Potensi bahaya pembelajaran penguatan yang tidak terkontrol dengan memanfaatkan masukan pengguna yang tidak difilter serta menggunakan sumber data yang tidak diverifikasi secara cermat untuk set data fine-tuning utama, ditunjukkan oleh keluaran yang tidak diinginkan dari NLG. Dua contoh terbaru yang lebih menonjol adalah Bot Twitter Microsoft Tay pada tahun 2016 dan Watson IBM pada tahun 2013.

Dalam kasus Tay, Bot tersebut melatih dirinya sendiri melalui interaksi tanpa filter yang dilakukannya dengan pengguna Twitter yang menggunakan bahasa yang provokatif dan menyinggung. Berdasarkan interaksi tersebut, hal ini akan menghasilkan bahasa yang provokatif dan ofensif, bahkan ketika merespons pengguna yang tidak menggunakan bahasa tersebut. Salah satu pendekatan terbaru yang diadopsi oleh OpenAI adalah pemfilteran konten pada tingkat prompt input. Dalam lingkungan yang membutuhkan tingkat moderasi yang tinggi, pemfilteran konten dapat diterapkan di sumber untuk menghindari prompt yang kemungkinan akan menghasilkan respons yang penuh kebencian, kekerasan, atau hal-hal yang tidak diinginkan. Meskipun hal ini menangani prompt input yang paling ekstrem dan dapat dideteksi, hal ini tidak menjanjikan respons bebas bias input, yang harus ditangani pada tingkat model.

Pembuatan Kode dan Data Kode yang Rentan

Dengan kemajuan pembuatan teks berbasis transformator, pelatihan model pada tugas-tugas yang sangat spesifik dan menantang secara teknis mulai menjadi mungkin. Salah satu bidang yang sedang berkembang adalah pembuatan kode otomatis berdasarkan input bahasa alami. Yang mungkin paling terkenal dan banyak digunakan adalah Github Copilot yang dirilis pada Juni 2021. Ini menghasilkan urutan kode dalam berbagai bahasa dengan beberapa komentar, nama fungsi, atau kode di sekitarnya. Copilot sebagian besar didasarkan pada model GPT-3 dasar yang kemudian disempurnakan di Github menggunakan kode sumber terbuka.

Karena tidak ada peninjauan manual untuk setiap entri, Github menyarankan bahwa kumpulan data yang mendasarinya dapat berisi pola pengkodean yang tidak aman yang kemudian disintesis menjadi kode yang dihasilkan di tingkat pengguna akhir. Evaluasi awal terhadap kode yang dihasilkan telah menemukan sekitar 40% dari cuplikan kode yang dihasilkan mengandung kerentanan keamanan dalam skenario yang relevan dengan Common Weakness Enumeration berisiko tinggi.

Pembuatan kode yang berpotensi rentan tanpa peninjauan akan menimbulkan risiko serius bagi pemilik kode tersebut, sehingga aplikasi alat tersebut tanpa peninjauan atau hampir otonom tidak mungkin diterapkan secara otonom menggunakan model yang mendasarinya. Namun, terdapat solusi peninjauan kode otomatis yang tersedia untuk memeriksa bagian kode tertentu untuk potensi masalah kualitas dan keamanan (misalnya Sonar). Memungkinkan operator non-teknis menggunakan bahasa alami untuk menghasilkan cuplikan kode sederhana yang dapat dipindai secara otomatis untuk menemukan kerentanan dan diterapkan dalam lingkungan pengujian yang akan digunakan untuk pembuatan prototipe cepat tampaknya merupakan pemanfaatan pembuatan kode semi-otonom yang paling masuk akal.

Deteksi Teks yang Dihasilkan Mesin

Dengan tersedianya komputasi awan yang luas yang memungkinkan produksi konten yang dihasilkan mesin dan media sosial yang memungkinkan distribusi massal, hambatan terakhir tetaplah kualitas teks yang dihasilkan dan kemampuan konsumen konten biasa untuk membedakannya dari teks yang diproduksi biasa. Berdasarkan penelitian terkait sebelumnya, kemampuan evaluator non-terlatih bergantung pada tingkat tertentu pada domain subjek serta kualitas model itu sendiri.

Untuk kutipan yang berfokus pada bahasa hukum yang bersumber dari opini hukum AS selama beberapa dekade, kemampuan membedakan berkisar antara 49% (dihasilkan oleh GPT-2) hingga 53% (dihasilkan oleh Transformer-XL), keduanya mendekati acak. Penelitian terkait juga menunjukkan bahwa akurasi meningkat untuk domain cerita yang ditulis manusia yang lebih kreatif dengan prompt "Once upon a time" di mana GPT-2 mencapai hasil 62% dan GPT-3 kembali mencapai nilai tebakan acak sebesar 49%.

Meskipun kecil kemungkinan kita akan melihat literatur yang dihasilkan mesin siap untuk konsumsi massal dalam waktu dekat, satu faktor yang mengkhawatirkan adalah bahwa nilai akurasi untuk mendeteksi artikel berita juga sebesar 57% untuk GPT-2 dan nilai tebakan acak dasar sebesar 51% untuk GPT-3. Meskipun sebagian besar konsumen kesulitan membedakan teks yang dihasilkan mesin dan manusia, model yang sama dapat dilatih untuk membedakan keduanya. Jika model yang diterapkan (GPT-2/Transformer-XL) diketahui sebelumnya, tingkat deteksinya tinggi, yaitu 94% hingga 97%. Namun, jika model tersebut tidak diketahui sebelumnya, tingkat deteksinya mencapai 74% hingga 76%. Oleh karena itu, kemungkinan deteksi akan sangat sulit terutama untuk model yang tidak bersumber terbuka dan tidak mudah direplikasi. Hal ini kemungkinan akan semakin sering terjadi, seperti halnya dengan GPT-3 yang telah dilisensikan kepada Microsoft, dengan kode sumber model yang tidak tersedia untuk umum.

Pengaruh Operator terhadap Keluaran

Keterangan Umum

Saat menyiapkan alat yang menghasilkan konten teks, hanya ada beberapa opsi untuk memengaruhi keluaran. Lapisan pertama dan paling dasar dari sebagian besar algoritma NLG adalah kumpulan data dasar yang digunakan untuk melatih model dasar. Dalam kasus GPT-2 dengan 1,5 miliar datanya, GPT-2 dilatih menggunakan data dari 8 juta Halaman Web atau 40 GB teks internet. Seleksi dilakukan sebagian dengan memilih halaman web keluar dari Reddit yang menerima "3 Karma" sebagai cara seleksi kualitas yang relatif manusiawi. Lapisan ini tidak dapat direplikasi dalam banyak kasus oleh pengguna akhir alat dan harus diterima apa adanya (sementara opsi untuk melatih model dari awal tersedia tetapi sulit diimplementasikan pada tingkat yang cukup canggih bagi sebagian besar pengguna akhir).

Lapisan kedua dan yang lebih mudah dipengaruhi adalah kumpulan data yang digunakan untuk "menyempurnakan" model. Langkah ini memungkinkan pengguna akhir untuk mengkhususkan keluaran mereka ke domain tertentu, gaya bahasa tertentu, atau yang serupa. Di sini, pengguna akhir memiliki pengaruh tertinggi terhadap keluaran NLG aktual yang akan dihasilkan. Area domain tertentu, seperti "bahasa hukum", misalnya, memungkinkan pengguna untuk mengkhususkan keluaran bahasa yang dihasilkan agar terdengar sangat mirip, bahkan identik, dengan bahasa hukum yang memenuhi syarat.

Tren terkini untuk LLM serta keterbatasan teknis sebagian besar operator akan membuat lapisan ini semakin sulit diakses, dengan sebagian besar model hanya mengizinkan interaksi operator melalui API komersial. Pendekatan ini membatasi pengaruh operator, tetapi juga meninggalkan jejak pemanfaatan yang dapat diaudit yang kemudian dapat selalu dilacak kembali ke penyedia model yang digunakan.

Perubahan ketiga dan paling langsung pada kualitas bahasa keluaran dapat diatur pada pengaturan parameter dasar model. Biasanya, dan khususnya dalam kasus model contoh kita, hal tersebut adalah panjang teks keluaran yang diinginkan, prompt teks awal, dan "suhu". Suhu di sini memungkinkan pengguna akhir untuk meningkatkan kemungkinan kata-kata berprobabilitas tinggi dan mengurangi kemungkinan kata-kata berprobabilitas rendah, yang seringkali menghasilkan teks yang biasanya lebih koheren dengan suhu yang sedikit lebih tinggi. Lapisan ini adalah yang paling mudah dimodifikasi dan paling banyak berinteraksi dengan pengguna akhir. Parameter-parameter tersebut kemungkinan akan menjadi kurang tersedia untuk sebagian besar aplikasi LLM yang dikomersialkan, sehingga penyedia model memiliki pengaruh yang lebih besar untuk mengoptimalkan parameter dan keluaran berdasarkan panjang hasil optimal dan skor probabilitas yang ada.

Parameter tambahan yang terkadang juga diterapkan adalah untuk mencegah munculnya bahasa kasar. Ini dapat berarti bahwa meskipun perintah teks yang diberikan atau dataset pelatihan atau fine-tuning yang mendasarinya mengandung bahasa kasar, model tersebut tetap tidak akan pernah mengeluarkan kata-kata yang dianggap menyinggung berdasarkan daftar kata kunci yang telah ditetapkan.

Data dan Metode

Untuk mengilustrasikan lebih lanjut dampak langsung pengaturan parameter operator terhadap keluaran, kami melatih model GPT-Neo pada kumpulan data teks hukum, yang menerapkan beberapa metode dan data dari penelitian sebelumnya, dengan menggunakan model yang lebih baru dan lebih canggih.

Lokasi empiris kami adalah Pengadilan Sirkuit AS, pengadilan banding tingkat menengah dalam sistem peradilan federal. Hakim Pengadilan Sirkuit meninjau putusan Pengadilan Distrik, memutuskan apakah akan menguatkan atau membatalkan putusan. Para hakim menjelaskan putusan mereka dengan memberikan opini tertulis. Korpus kami terdiri dari 50.000 opini Pengadilan Sirkuit AS ini, yang diambil sampelnya secara seragam dari keseluruhan opini untuk tahun 1890 hingga 2010. 1 Sampel tersebut mencakup opini utama (mayoritas) dan opini tambahan (konsensus dan perbedaan pendapat).

Kami melakukan pra-pemrosesan minimal agar generator kami dapat mereplikasi gaya teks asli. Kami menghapus beberapa metadata dan markup XML, tetapi tetap mempertahankan kapitalisasi, tanda baca, dll. Kami mempertahankan notasi sitasi hukum khusus yang digunakan oleh pengadilan AS. Pendapat-pendapat tersebut umumnya cukup panjang, rata-rata berisi 2024 token (kata) per artikel. Panjang rata-rata tersebut secara bertahap menurun dari tahun 1890an hingga mencapai titik minimum pada tahun 1970an.

Setelah itu, panjang rata-rata pendapat-pendapat ini terus meningkat hingga saat ini. Perlu dicatat, sekitar tahun 1970an basis data penelitian hukum digital mulai digunakan. Pendekatan kami dalam merepresentasikan dokumen hukum adalah model bahasa auto-regresif. Kami diberikan korpus tanpa pengawasan. Kami menyempurnakan titik pemeriksaan GPT-Neo yang ada pada korpus hukum kami.

Contoh Pengaruh Operator

Kami mendemonstrasikan bagaimana pengguna akhir dapat berinteraksi dengan model berdasarkan fine-tuning model dan pemilihan parameter dengan contoh-contoh berikut menggunakan model referensi kami. Dalam contoh pertama (c.), kami menggunakan dataset berukuran penuh dengan nilai suhu default yang diterapkan, yang menghasilkan beberapa hasil yang bermakna.

Dalam dua contoh lainnya, kami menggunakan dataset berukuran penuh standar, tetapi kemudian mengubah parameter suhu menjadi 0,9 (d.) dan 0,1 (e.). Kami mengamati bahwa contoh yang ditetapkan untuk suhu yang lebih tinggi juga memiliki keterbacaan yang lebih baik, tidak terlalu banyak mengandung preseden hukum yang tercantum, melainkan lebih banyak menggunakan bahasa yang alami.

(a.) Petunjuk: Pengadilan Banding Sirkuit Kesepuluh melanggar prinsip-prinsip yang telah ditetapkan di sini. Tidak satu pun keputusan yang diandalkan (...)

(b.) Putusan Akhir yang Sejati: (...) oleh Pengadilan Banding—Estate of Ceballos v. Husk, 919 F. 3d 1204, Hastings v. Barnes, 252 Fed. Appx. 197, Allen, 119 F. 3d 837, dan Sevier v. Lawrence, 60 F. 3d 695—hampir membuktikan bahwa tindakan para petugas tersebut melanggar hukum. Pengadilan sangat bergantung pada Allen. Namun, fakta-fakta Allen sangat berbeda dari fakta-fakta di sini.

(c.) Dataset lengkap, suhu default 0,7 (...) oleh mayoritas menyatakan bahwa terdakwa yang bukan merupakan pihak dalam konspirasi tidak dapat dimintai pertanggungjawaban pribadi atas tindakan nyata yang mendukung konspirasi tersebut. Sebaliknya, Pengadilan Banding Sirkuit Kesepuluh menyatakan bahwa terdakwa bertanggung jawab secara pribadi atas tindakan nyata jika ia "merupakan anggota konspirasi pada saat tindakan nyata tersebut dan berpartisipasi di dalamnya dengan cara tertentu."

(d.) Dataset lengkap, suhu tinggi 0,9 (...) oleh Pengadilan Banding Sirkuit Kesepuluh bahkan sedikit mirip dengan kasus ini. Dalam kasus Amerika Serikat v. Smith, 7 F.3d 1433, pengadilan memutuskan bahwa mosi terdakwa untuk menekan bukti yang diperoleh dengan melanggar Amandemen Keempat ditolak karena pemerintah tidak memiliki alasan yang cukup untuk menggeledah rumah terdakwa.

(e.) Dataset lengkap, suhu sangat rendah 0,1 (...) oleh Pengadilan Banding Kesepuluh dalam kasus ini tidak konsisten dengan keputusan Mahkamah Agung dalam kasus Amerika Serikat v. Booker, 543 U.S. 220, 125 S. Ct. 738, 160 L. Ed. 2d 621. Dalam kasus Booker, Mahkamah Agung memutuskan bahwa sifat wajib Pedoman Penjatuhan Hukuman menjadikannya inkonstitusional. 543 U.S. di hlm. 244. Mahkamah Agung memutuskan bahwa Pedoman tersebut inkonstitusional karena tidak "cukup andal untuk mendukung temuan 'kewajaran'."

17.2 ASPEK HUKUM

Pendahuluan Analisis Hukum

Penggunaan kecerdasan buatan (AI) seperti algoritma NLG menimbulkan berbagai tantangan hukum, termasuk masalah pertanggungjawaban. Sebagian besar aplikasi AI dirancang untuk berkembang secara mandiri guna mengatasi masalah yang tidak atau belum dapat dipertimbangkan oleh pengembangnya saat memprogramnya. Akibatnya, AI yang belajar mandiri dapat berevolusi secara tak terduga. Dalam kasus terburuk, suatu algoritma dapat membahayakan orang lain melalui perilaku berbahaya yang dipelajari sendiri.

Saat menggunakan NLG, klaim pertanggungjawaban dapat muncul di area mana pun yang menggunakan komunikasi lisan, baik lisan maupun tertulis. Oleh karena itu, rumah sakit atau perusahaan asuransi yang menggunakan bot berbasis NLG untuk berkomunikasi dengan pasien, pengacara yang menggunakan NLG untuk menyusun ringkasan, atau media berita yang menggunakan NLG untuk menyunting artikel menghadapi klaim pertanggungjawaban jika keluaran algoritma tersebut menyebabkan kerugian bagi orang lain.

Literatur hukum yang membahas implikasi AI terhadap gugatan gugatan di masa mendatang berfokus pada arahan Dewan Eropa tentang tanggung jawab atas produk cacat (Arahan) dan apakah ketentuan tanggung jawab baru diperlukan atau tidak. Artikel ini menganalisis lebih lanjut bagaimana komunikasi verbal yang dihasilkan oleh algoritma NLG dapat melanggar hak pribadi, melanggar hak kekayaan intelektual, atau menjadi penyebab gugatan persaingan tidak sehat.

Tanggung Jawab atas Tindakan Otonom AI Secara Umum

Tindakan Tak Terduga AI yang Belajar Mandiri sebagai Tantangan bagi Hukum Gugatan

Kemampuan belajar mandiri AI menimbulkan tantangan besar bagi pengembang dan operator. Di satu sisi, AI secara otonom mengembangkan solusi baru untuk masalah. Di sisi lain, karakteristik yang sama ini menimbulkan tantangan yang luar biasa, karena pengembang dan operator tidak selalu mampu mengantisipasi risiko yang mungkin ditimbulkan oleh AI yang belajar mandiri terhadap orang lain.

Seseorang mungkin menyimpulkan bahwa tindakan AI yang diadopsi dari mekanisme pembelajaran mandiri tidak dapat diramalkan oleh pengembang atau operator AI, sehingga mencegah mereka menerapkan tindakan penanggulangan yang memadai. Persepsi ini akan mengguncang fondasi hukum perdata, karena hukum tersebut menetapkan ketidakmampuan pengembang untuk mengendalikan risiko yang berasal dari AI.

Para ahli telah mencoba membahas aspek otonomi AI dan telah mengusulkan berbagai gagasan berdasarkan hukum pertanggungjawaban yang ada, seperti analogi untuk semua jenis pertanggungjawaban perwakilan. Sebagaimana teknologi baru lainnya, beberapa pihak berpendapat perlunya landasan hukum baru untuk mengatur risiko secara memadai dan menetapkan tanggung jawab kepada produsen dan operator. Lebih lanjut, beberapa pihak mendukung konsep hukum e-person yang sepenuhnya baru, yang menjadikan AI itu sendiri sebagai tergugat dalam gugatan perbuatan melawan hukum.

Sebagaimana kebanyakan teknologi baru, harus dianalisis secara cermat apakah AI menciptakan risiko dengan kualitas baru atau hanya mengubah kuantitasnya, karena hanya risiko pertama yang memerlukan penerapan aturan tanggung jawab baru. Apakah AI memenuhi syarat untuk hal tersebut masih harus ditentukan.

Tergugat dalam Gugatan Perbuatan Melawan Hukum

Dengan AI yang menyebabkan kerugian bagi orang lain, produsen atau pengembang perangkat lunak akan memiliki peran yang lebih signifikan sebagai tergugat dalam gugatan perbuatan melawan hukum karena merekalah yang menjadi sumber tindakan yang tidak diinginkan. Dalam kasus algoritma NLG, keluaran yang dihasilkan didasarkan pada kode program yang dikembangkan oleh produsen AI, pihak yang dapat disalahkan jika keluaran tersebut berdampak negatif.

Untuk algoritma AI, fase pembelajaran mandiri, ketika produk sudah beredar, menjadi semakin penting. Karena pergeseran pengembangan produk ke fase pasca pemasaran ini, para ahli hukum berpendapat bahwa tidak hanya produsen AI yang menanggung risiko liabilitas, tetapi juga operator. Untuk perangkat lunak individual, operator juga dapat bertanggung jawab atas tindakan produsen jika produsen tersebut dapat dianggap sebagai perwakilan operator dan produsen tersebut tidak dapat membuktikan bahwa mereka telah mengambil semua langkah yang wajar dan diperlukan untuk menginstruksikan dan mengawasi perwakilan tersebut guna mencegah terjadinya kerusakan. Misalnya, untuk media berita yang memanfaatkan NLG, pemimpin redaksi atau staf pengawas lainnya dapat bertanggung jawab atas berfungsinya perangkat lunak dengan baik dan bertanggung jawab jika perangkat lunak tersebut menyebabkan kerugian bagi orang lain.

Kausalitas sebagai Faktor Pembatas Tanggung Jawab

Fakta bahwa algoritma pembelajaran mandiri berkembang secara independen setelah pengembang mengedarkannya menyulitkan untuk membatasi kontribusi kausal setiap aktor terhadap kerusakan. Dalam kebanyakan kasus, agen buatan pembelajaran mandiri (seperti NLG) bukanlah produk standar, melainkan dirancang khusus sesuai kebutuhan operator. Oleh karena itu, produsen dan operator bertindak bersama-sama ketika mengembangkan dan melatih AI untuk digunakan oleh operator. Berdasarkan hukum Swiss, jika dua terdakwa bertindak bersama-sama, keduanya bertanggung jawab secara tanggung renteng atas semua kerugian yang ditimbulkan (Pasal 50 Kitab Undang-Undang Hukum Kewajiban Swiss ("KUHP")).

Terutama dalam aplikasi AI dan algoritma NLG, interaksi dengan pihak ketiga, seperti pelanggan operator, menjadi semakin penting bagi algoritma untuk berkembang lebih lanjut. Sebagaimana ditunjukkan oleh contoh-contoh nyata baru-baru ini, masukan yang dihasilkan oleh pelanggan dapat menimbulkan efek yang tidak diinginkan pada perilaku AI. Secara umum, produsen harus mengambil langkah-langkah yang wajar untuk mencegah algoritma menggunakan data masukan yang tidak memenuhi syarat (seperti ujaran kebencian) untuk mengadaptasi perilakunya. Namun, produsen tidak dapat diharapkan untuk mengantisipasi setiap kemungkinan penyalahgunaan produknya. Berdasarkan hukum Swiss, produsen dapat terhindar dari tanggung jawab jika ia membuktikan bahwa tindakan tak terduga dari pihak ketiga secara signifikan lebih relevan dalam menyebabkan kerugian daripada tindakannya sendiri, sehingga memutus rantai kausalitas.

Demikian pula, Arahan tersebut menetapkan bahwa produsen tidak bertanggung jawab jika ia membuktikan bahwa kemungkinan besar cacat yang menyebabkan kerusakan tidak ada pada saat produsen mengedarkan produk atau bahwa cacat tersebut muncul setelahnya. Beberapa penulis berpendapat bahwa interaksi pengguna dengan AI dapat menjadi akar penyebab kerugian dan oleh karena itu, produsen terhindar dari tanggung jawab.

Terlepas dari tantangan-tantangan spesifik ini, membuktikan kausalitas dalam setiap klaim ganti rugi merupakan hal yang menantang dan, dalam banyak kasus, membutuhkan sumber daya yang signifikan untuk membuktikannya. Dalam banyak kasus gugatan perdata, tidak akan ada satu pun penyebab yang diidentifikasi sebagai penyebab terjadinya kerugian, tetapi berbagai penyebab akan berkontribusi sebagian terhadap kerugian penggugat. Bagi penggugat, pembuktian sebab akibat akan tetap menjadi rintangan signifikan untuk mendapatkan kompensasi atas kerusakan.

Arahan tentang Produk Cacat

Keterangan Umum

Sebagian besar algoritma NLG akan menyebabkan kerugian ekonomi yang tidak tercakup dalam Arahan (Pasal 9) atau hukum pertanggungjawaban produk Swiss yang sejalan dengan Arahan tersebut. Namun demikian, dapat dibayangkan bahwa algoritma NLG juga akan menyebabkan cedera pribadi atau kerusakan properti. Hal ini terjadi ketika algoritma NLG memberikan informasi yang salah yang menyebabkan cedera fisik pada orang lain (misalnya, seorang dokter menerima diagnosis dari perangkat yang menggunakan NLG yang cacat untuk

berkomunikasi, bot komunikasi dari fasilitas panggilan darurat swasta memberikan nasihat medis palsu).

Para ahli telah membahas secara ekstensif apakah Arahan tersebut berlaku untuk perangkat lunak atau tidak. Meskipun ambigu, sebagian besar berpendapat bahwa perangkat lunak termasuk dalam Arahan tersebut. Untuk mengatasi keraguan yang masih ada, Komisi Uni Eropa telah menerbitkan amandemen terhadap Arahan yang menyebut perangkat lunak sebagai produk. Oleh karena itu, analisis berikut mengasumsikan bahwa perangkat lunak memenuhi syarat sebagai produk berdasarkan Arahan tersebut.

Berbagai aspek Arahan dibahas dalam literatur hukum, dengan dua hal yang menonjol: Pertama, harus ditentukan apakah tindakan sistem AI dianggap cacat dalam arti Arahan tersebut. Kedua, produsen sistem AI dapat dibebaskan dari tanggung jawab berdasarkan pembelaan mutakhir jika mereka membuktikan bahwa, pada saat produk tersebut diedarkan, tindakan tertentu dari sistem AI, terutama yang dikembangkan sistem melalui mekanisme pembelajaran mandiri, tidak dapat diramalkan dengan sarana teknis dan pengetahuan ilmiah yang tersedia pada saat itu.

Defektivitas Sistem AI

Uji Ekspektansi Konsumen

Banyak akademisi kesulitan menentukan apakah suatu sistem AI cacat atau tidak. Arahan tersebut menganggap suatu produk cacat jika tidak memberikan keamanan yang diharapkan seseorang (Pasal 6 (1) Arahan). Oleh karena itu, suatu produk cacat jika konsumen yang wajar akan menganggapnya cacat dengan mempertimbangkan penyajian produk, penggunaan yang secara wajar dapat diharapkan dari produk tersebut, dan waktu produk tersebut diedarkan.

Uji yang didasarkan pada ekspektasi konsumen ini mungkin tidak memadai untuk menentukan defektivitas teknologi mutakhir, karena sulit untuk ditentukan dan tidak memiliki titik acuan. Pendekatan risiko-manfaat yang menentukan apakah desain alternatif yang wajar akan secara signifikan mengurangi terjadinya kerugian mungkin lebih tepat.

AI Menentang Konsep Cacat

Berbagai penyebab dapat menyebabkan kesalahan perangkat lunak. Beberapa di antaranya lebih mudah dibuktikan dan tidak menentang definisi cacat sebagaimana tercantum dalam Arahan. Di antara kasus-kasus angka tersebut di mana produsen menyebabkan kesalahan dalam kode algoritma, melatih algoritma (sebelum mengedarkannya) dengan data yang tidak sesuai, atau tidak menerapkan langkah-langkah yang memadai untuk mencegah pihak ketiga merusak kode (misalnya peretasan). Namun, aspek-aspek lain yang mempersulit pembuktian cacat atau menantang pemahaman konsep cacat muncul dengan AI. Dari sudut pandang teknis, sulit untuk menganalisis tindakan AI yang menyebabkan kerusakan karena proses yang berlangsung dengan cara yang belum dapat dirasakan dari luar (masalah kotak hitam). Khususnya dalam kasus NLG, mungkin sudah sulit bagi penggugat untuk membuktikan bahwa keluaran yang menyebabkan kerusakan dihasilkan secara artifisial, sehingga Arahan tersebut berlaku.

Dari sudut pandang normatif, fakta bahwa suatu algoritma, melalui mekanisme pembelajaran mandiri, dapat mengadopsi perilaku yang tidak diinginkan oleh pengembangnya, menantang persepsi cacat: Para ahli membahas berbagai cara untuk menentukan ekspektasi konsumen yang wajar terhadap sistem AI. Agen AI mengungguli keterampilan manusia untuk tugas-tugas tertentu.

Membandingkan hasil algoritma AI dengan hasil manusia tidak cukup mempertimbangkan kinerja AI yang superior dibandingkan dengan manusia yang terbatas tugasnya. Membandingkan hasil dua algoritma untuk menentukan ekspektasi wajar pelanggan tidaklah tepat karena konsekuensinya adalah hanya algoritma dengan kinerja terbaik yang dianggap aman, sementara algoritma lainnya cacat.

Menentukan cacatnya proses pembelajaran suatu algoritma mungkin lebih sulit karena proses tersebut terutama berkembang setelah produk diedarkan dan terjadi di luar kendali produsen, khususnya dengan NLG. Fase di mana algoritma NLG berinteraksi dengan pengguna secara khusus menantang pemahaman tentang cacatnya: Saat AI menyediakan layanannya kepada pengguna, ia secara bersamaan meningkatkan kemampuannya, sehingga menimbulkan pertanyaan apakah algoritma tersebut dapat dianggap cacat saat digunakan atau tidak.

Seperti yang telah ditunjukkan oleh contoh-contoh sebelumnya, interaksi dengan pengguna dapat menyebabkan algoritma mengembangkan perilaku tertentu yang tidak diinginkan oleh produsen. Harus ditentukan apakah produsen harus menyediakan langkah-langkah yang wajar untuk mencegah algoritma berevolusi secara tidak diinginkan.

Para ahli sepakat bahwa produsen harus menerapkan perlindungan untuk mencegah algoritma memasukkan perilaku pengguna yang tidak pantas atau ilegal ke dalam kodenya. Hal ini mungkin terbukti lebih mudah secara teori daripada praktik karena sangat sulit untuk memprediksi perilaku pengguna apa yang dapat menyebabkan algoritma pembelajaran mandiri berevolusi dengan cara yang tidak diinginkan oleh produsen. Jika pengguna berinteraksi dengan AI dengan cara yang tidak terduga dan menyebabkan kerugian, suatu produk tidak dapat dianggap cacat.

Pembelaan Mutakhir

Produsen dapat terhindar dari tanggung jawab jika dapat membuktikan bahwa suatu cacat tidak dapat dideteksi ketika produk diedarkan dengan pengetahuan teknis dan ilmiah yang tersedia. Teknologi baru dengan efek negatif yang belum diketahui seperti AI memenuhi syarat untuk pembelaan mutakhir. Oleh karena itu, para ahli mengusulkan pengecualian aplikasi tertentu dari pembelaan mutakhir, sebagaimana yang telah dilakukan oleh badan legislatif di berbagai yurisdiksi untuk fitur teknologi lain seperti GMO dan xenotransplantasi.

Membedakan antara kondisi yang memenuhi syarat sebagai cacat produk dan kondisi yang termasuk dalam pembelaan mutakhir dalam hal AI sulit dilakukan. Algoritma pembelajaran mandiri dapat mengembangkan perilaku yang tidak diinginkan yang tidak dapat diramalkan oleh produsen yang teliti. Namun, fakta bahwa produsen tidak dapat meramalkan potensi perilaku berbahaya dari perangkat lunak AI-nya tidak serta merta memicu pembelaan mutakhir. Contoh-contoh sebelumnya menunjukkan bahwa tidaklah cukup jika produsen tidak

dapat meramalkan risiko spesifik dari produknya. Pembelaan tersebut hanya dapat diajukan jika produsen juga tidak dapat mengantisipasi risiko umum bahaya yang ditimbulkan oleh produknya.

Para penyusun Arahan ini bermaksud agar pembelaan ini berlaku untuk kasus-kasus yang sangat terbatas. Oleh karena itu, produsen diharuskan telah menerapkan kehati-hatian dan ketekunan yang maksimal untuk mengantisipasi dampak negatif dari produk mereka agar dapat mengajukan pembelaan tersebut. Beberapa penulis berpendapat bahwa risiko AI yang belajar mandiri sudah cukup diketahui untuk mencegah produsen berhasil mengajukan pembelaan tersebut.

Dari perspektif praktis, rintangan untuk menggunakan pembelaan tersebut juga signifikan. Produsen yang menggunakannya kemungkinan besar harus mengungkapkan rahasia bisnis (seperti kode pemrograman) kepada pihak yang dirugikan, sehingga sangat kecil kemungkinan pembelaan tersebut akan digunakan secara luas untuk membela klaim tanggung jawab produk.

Terakhir, terdapat beragam kemungkinan penerapan AI, meskipun tidak setiap kategori produk menimbulkan bahaya yang sama bagi konsumen. Dalam kebanyakan kasus, bahaya yang mengancam dari produk konvensional merupakan risiko bahaya terbesar; peningkatan aplikasi AI pada produk-produk ini tidak meningkatkan risiko tersebut secara signifikan. Oleh karena itu, pengecualian umum AI dari pembelaan mutakhir tidak akan mempertimbangkan risiko bahaya individual dari setiap kategori produk.

Kesimpulannya, pengecualian seperti yang diusulkan oleh beberapa penulis memerlukan analisis yang lebih mendalam tentang risiko spesifik AI dan presedennya. Seruan umum untuk mengecualikan teknologi baru dari pembelaan bersifat kontraproduktif dan dapat menghalangi produsen untuk berinvestasi dalam produk yang menggunakan AI.

Tanggung Jawab atas Kelalaian

Jika tidak ada ketentuan khusus yang memungkinkan tergugat untuk menuntut kompensasi atas kerugian, dalam hukum Swiss, tanggung jawab perdata berbasis kesalahan umum berlaku (Pasal 41 Kitab Undang-Undang Hukum Pertanggungjawaban Swiss). Bagi NLG, tanggung jawab berbasis kesalahan akan menjadi relevan jika keluaran yang dihasilkan melanggar hak pribadi, melanggar hak kekayaan intelektual, atau memicu ketentuan undang-undang persaingan tidak sehat.

Pelanggaran Hak Kekayaan Intelektual

Keluaran yang dihasilkan oleh algoritma NLG tanpa campur tangan manusia tidak dilindungi oleh hak cipta karena kurangnya masukan kreatif. Oleh karena itu, keluaran yang dihasilkan oleh algoritma NLG dapat digunakan oleh pihak lain tanpa melanggar undang-undang hak cipta atau membayar royalti atas penggunaannya. Di sisi lain, algoritma NLG dapat mengandalkan sumber yang tersedia di internet. Risiko penggunaan karya berhak cipta atau paten harus dipertimbangkan oleh pengembangnya.

Pelanggaran Hak Pribadi

Beberapa contoh menunjukkan bahwa algoritma kecerdasan buatan untuk NLG dapat menghasilkan keluaran yang melanggar hak pribadi orang lain (pencemaran nama baik,

pencemaran nama baik, dll.). Hukum Swiss memberikan serangkaian upaya hukum kepada korban pelanggaran hak pribadi, mulai dari putusan pengadilan hingga tuntutan ganti rugi. Otonomi algoritma NLG tidak mengecualikan tanggung jawab perdata operator jika keluaran yang dihasilkan oleh algoritma NLG melanggar hak pribadi orang lain (Pasal 28 (1) Kitab Undang-Undang Hukum Perdata Swiss). Jika penggugat membuktikan bahwa operator algoritma NLG bersalah, ia dapat menuntut kompensasi moneter (Pasal 28a (3) Kitab Undang-Undang Hukum Perdata Swiss dan Pasal 41 Kitab Undang-Undang Kewajiban Swiss) (Meili 2018, Pasal 28a n16).

Media berita rentan terhadap tuntutan hukum jika mereka secara luas menggunakan algoritma NLG tanpa pengawasan yang memadai. Seiring dengan semakin pendeknya lingkaran pemberitaan dan semakin pentingnya pemain baru, risiko bahwa keluaran yang dihasilkan oleh NLG melanggar hak pribadi meningkat. Media berita bukanlah satu-satunya operator yang mungkin terlibat dalam gugatan pencemaran nama baik ketika menggunakan NLG yang tidak berfungsi dengan baik.

Khususnya, portal pemeringkatan yang menggunakan NLG untuk membuat komentar tentang bisnis (misalnya, yang dikumpulkan dari pelanggan formulir umpan balik individual) dapat melanggar hak pribadi jika umpan balik (agregat) tersebut salah atau melanggar hak pribadi orang lain. Apakah mesin pencari atau pemilik situs web yang menyediakan tautan ke konten yang melanggar hak pribadi orang lain juga bertanggung jawab, hal ini belum ditentukan berdasarkan hukum Swiss.

Persaingan tidak Sehat

Hasil algoritma NLG dapat rentan terhadap klaim persaingan tidak sehat jika melanggar persyaratan persaingan sehat. Kasus-kasus di mana isu persaingan tidak sehat yang melibatkan NLG menjadi relevan adalah semua jenis aktivitas penjualan di mana NLG digunakan untuk mengiklankan produk. Ini dapat melibatkan iklan umum yang meluas atau perbandingan otomatis dengan produk serupa milik pesaing, atau deskripsi yang disesuaikan untuk setiap pelanggan guna membujuk mereka untuk membeli produk tertentu. Dengan munculnya situs web pemeringkatan (seperti Google Maps, Yelp, dll.), bisnis memanfaatkan peringkat yang baik. Algoritma NLG dapat membantu bisnis dengan mudah membuat ulasan palsu. Pembuatan atau pemesanan ulasan palsu untuk meningkatkan atau melemahkan peringkat bisnis pesaing secara tidak wajar, memenuhi syarat sebagai persaingan tidak sehat, dan dapat memberikan klaim ganti rugi kepada pesaing.

Undang-Undang Persaingan Tidak Sehat Swiss (UWG) memberikan sanksi terhadap berbagai bentuk perilaku persaingan tidak sehat. Secara khusus, undang-undang ini memberikan sanksi atas tindakan yang menyesatkan pelanggan tentang produk operator NLG sendiri atau produk pesaing (Pasal 3 (1) UWG). Undang-undang ini mengatur berbagai upaya hukum, seperti ganti rugi berupa perintah pengadilan bagi korban, bagi negara atau asosiasi profesi (Pasal 9 dan 10 UWG). Korban dapat menuntut ganti rugi lebih lanjut berdasarkan tanggung jawab berdasarkan kesalahan dalam Pasal 41 CO.

Kewajiban Kehati-hatian

Pemilik hak cipta yang dilindungi atau orang yang hak pribadinya telah dilanggar oleh hasil karya NLG memiliki berbagai upaya hukum untuk melawan pelanggaran hak-hak mereka. Selain perintah pengadilan, korban dapat menuntut ganti rugi. Ganti rugi ini didasarkan pada ketentuan umum tanggung jawab berdasarkan kesalahan dalam Kitab Undang-Undang Kewajiban Swiss (Pasal 41 CO).

Oleh karena itu, penggugat harus membuktikan, di antara unsur ganti rugi dan kausalitas, bahwa pelaku perbuatan melawan hukum telah melanggar kewajiban kehati-hatian yang berlaku. Kewajiban kehati-hatian berasal dari standar hukum atau privat, yang untuk teknologi baru belum ditetapkan. Jika standar khusus tidak ada, prinsip umum untuk semua jenis aktivitas berbahaya berlaku. Dengan demikian, seseorang yang menimbulkan risiko bahaya bagi orang lain harus mengambil semua tindakan pencegahan yang diperlukan dan wajar untuk mencegahnya.

Mahkamah Agung Swiss telah menangani kasus-kasus di mana tautan dari blog daring mengarah ke halaman web yang melanggar hak pribadi. Tanpa penilaian mendalam, Mahkamah menyimpulkan bahwa operator blog tidak dapat terus-menerus memantau konten semua halaman web yang ditautkan. Demikian pula, Mahkamah Agung Jerman menyimpulkan bahwa operator mesin pencari tidak dapat dimintai pertanggungjawaban atas pelanggaran hak pribadi apa pun atas saran pelengkapan otomatis yang dihasilkan oleh perangkat lunaknya. Operator mesin pencari hanya boleh mengambil langkah-langkah yang wajar untuk mencegah pelanggaran hak pribadi. Performa perangkat lunak yang lancar dan efisien tidak boleh dihambat oleh sistem penyaringan yang ketat.

Di sisi lain, keahlian khusus produsen atau operator yang memungkinkan mereka menilai risiko pelanggaran hak pihak ketiga oleh agen AI harus dipertimbangkan untuk menetapkan standar perawatan yang berlaku. Lebih lanjut, operator juga bertanggung jawab untuk secara berkala mengontrol kumpulan data algoritma yang digunakannya guna meningkatkan kemampuannya. Namun, meskipun telah direncanakan dengan matang, pengembang atau operator algoritma NLG mungkin tidak selalu dapat memprediksi siapa yang mungkin dirugikan oleh algoritma yang digunakan, sehingga mencegahnya mengambil tindakan terhadapnya. Terakhir, khususnya dengan algoritma NLG yang belajar mandiri, pengembang dan operator harus mencegah algoritma tersebut mengambil perilaku berbahaya dari interaksi dengan penggunanya.

17.3 NLG AI: TANTANGAN HUKUM PERDATA

NLG menawarkan beragam kemungkinan aplikasi. Algoritma mutakhir memungkinkan terciptanya keluaran verbal yang tidak dapat dibedakan dari ucapan manusia. Permainan kucing-kucingan sedang berlangsung antara mereka yang memprogram NLG dan mereka yang mengembangkan algoritma yang mampu menentukan apakah keluaran tertentu bersifat manusiawi atau buatan. Sebagaimana ditunjukkan, verifikasi teks yang dihasilkan komputer sangat penting dari perspektif hukum, karena landasan hukumnya hanya berlaku untuk salah satu atau yang lainnya.

Fungsi pembelajaran mandiri dari kecerdasan buatan menantang hukum perdata. Interaksi dengan pengguna dapat mengakibatkan perilaku yang tidak diinginkan, dan dalam kasus terburuk, bahkan menyebabkan kerugian. Hal ini menimbulkan pertanyaan sensitif tentang sejauh mana seorang programmer atau operator AI harus bertanggung jawab atas tindakannya karena mereka mungkin tidak selalu dapat mengantisipasi perilaku AI di masa mendatang yang berasal dari interaksi dengan pihak ketiga. Penelitian hukum perlu bergulat untuk beberapa waktu dengan cara menangani tantangan spesifik AI sebelum menyerah pada godaan undang-undang baru.

BAB 18
AI DAN RISIKO RESIDIVISME: CATATAN AWAM

Abstrak Berikut ini adalah kisah seorang pengacara pidana yang prihatin, seorang awam, saat ia mengamati perubahan yang diramalkan oleh AI dalam bidang penilaian risiko residivisme individu untuk tujuan penjatuhan hukuman pidana kepada narapidana. Oleh karena itu, teks ini akan merefleksikan sifat penilaian tersebut ketika dipromosikan oleh program AI baru berdasarkan informasi yang diturunkan secara statistik dengan makna aktuaria. Kemudian, teks ini membandingkan penilaian risiko residivisme tersebut dengan penilaian yang dilakukan dalam paradigma manusia tradisional saat ini.

Mengidentifikasi tantangan yang ditimbulkan oleh alternatif teknologi terhadap kelangsungan hidup landasan prinsip-prinsip hukum pidana. Catatan akhir akan membahas apa yang kurang dalam proposal teknologi ini, terlepas dari semua kelebihan teknis dan keuntungan yang dirasakan. Pendekatannya di sini berubah. Dari literatur, kita akan mengemukakan catatan kemanusiaan yang menjadi inti dari segala sesuatu yang menyerupai penghakiman.

Baik penghakiman individu yang dinilai maupun penghakiman pengadilan yang melakukan penilaian. Meskipun keduanya manusiawi, kita perlu memperhatikan jenis kemanusiaan yang harus diakui oleh seluruh ilmu pengetahuan yaitu hukum dan pendekatan spesifiknya. Kemanusiaan itulah yang tampaknya kurang dalam proposal-proposal AI teknologi.

18.1 AI PIDANA: PREDIKSI RESIDIVISME TANPA MORAL

Catatan-catatan singkat, tidak lebih, akan ditemukan di bawah ini. Refleksi seorang pengacara pidana, yang pasti merasakan ketidaknyamanan akibat redundansi. Seorang awam yang didorong oleh rasa ingin tahu dan minat ilmiah. Seseorang yang, terlepas dari semua investasinya, tidak ragu bahwa ia hanya dapat mengikuti revolusi teknologi yang tak henti-hentinya, yang di sini maupun di tempat lain, menciptakan kemungkinan-kemungkinan baru dan tantangan yang menyeluruh. Tantangan-tantangan semacam itu, agar dapat diatasi dengan baik, akan membutuhkan satu hal yang tidak dapat disediakan: waktu untuk berhenti sejenak.

Baris-baris berikut bukanlah tulisan seorang ahli. Jika keahlianlah yang Anda, pembaca yang budiman, cari, silakan cari di tempat lain. Di sisi lain, hal-hal tersebut merupakan hasil dari (sedikit) waktu untuk berhenti sejenak dan merenung. Jika, pembaca yang budiman, Anda bersedia menerima hal itu, silakan, bergabunglah dengan saya. Apa yang saya tawarkan hanyalah sekadar refleksi. Siapa tahu, mungkin waktu Anda pada akhirnya akan termanfaatkan dengan baik.

Objek perhatian utama kita akan dipusatkan pada mesin dan program komputer yang diberkahi kecerdasan buatan, dengan penekanan khusus pada mesin dan program yang

tingkat otonominya ditentukan oleh kapasitas pengambilan keputusan independen dengan kemampuan untuk belajar, pada tingkat pembelajaran mendalam.

Pertimbangan alat prediktif untuk tingkat risiko residivisme akan menjadi sangat penting, yaitu alat komputer untuk mengantisipasi risiko residivisme pada narapidana. Dibekali otonomi dalam mempertimbangkan sifat dan ukuran hukuman individual, COMPAS, LSI-R, VRAG, dan ORAS akan dijadikan acuan, karena merupakan akronim dari program komputer yang tersedia saat ini untuk "prediksi" risiko residivisme, yang masing-masing bergantung pada informasi yang diperoleh secara statistik dan memiliki makna aktuaria.

Semuanya berkaitan dengan program komputer yang, saat ini, menghitung, melalui atribusi poin dan definisi peringkat, risiko residivisme dari setiap pelaku kejahatan tertentu yang telah dihukum dalam melakukan kejahatan di masa mendatang. Ditawarkan saat ini sebagai alat bantu untuk putusan pengadilan, program-program ini belum memiliki atribut pengambilan keputusan yang otonom. Meskipun kapasitas teknologi tersebut mungkin sudah dapat dicapai.

Oleh karena itu, yang dipertaruhkan dalam instrumen prediksi ini saat ini adalah klaim kapasitasnya untuk memprediksi perilaku individu di masa mendatang. Dan bukan hanya itu, tetapi juga untuk melakukannya secara akurat. Dan selalu, tak terelakkan, klaim lebih lanjut bahwa penilaian probabilitas tersebut, justru karena akurasinya yang tak terbantahkan, harus dapat berfungsi sebagai kriteria dan dasar, tidak hanya untuk membatasi kebebasan fisik individu, tetapi juga untuk menentukan tingkat keparahan dan sifatnya.

Kemungkinan putusan pidana oleh mesin dan program, yaitu, oleh artefak yang dilengkapi kecerdasan buatan, kepada setiap individu berdasarkan serangkaian fakta tertentu tentang hukuman tertentu, yang diajukan dalam bentuk tertentu dan pada tingkat keparahan tertentu, berdasarkan prediksi perilaku kriminalnya di masa mendatang, kini telah hadir. Dan bersamanya, untuk pertama kalinya, peran sentral dimainkan dalam proses pengambilan keputusan putusan bersalah pidana dan penetapan kesalahan oleh mesin. Artinya, oleh objek yang sama sekali tidak memiliki kemampuan untuk penilaian dan refleksi moral. Pertanyaan yang tak terelakkan akan muncul mengenai kemungkinan dan justifikasi hukum atas putusan pidana individual tanpa pengakuan moral.

Ketiadaan pengakuan moral tersebut menjadi fakta ketika, dalam proses atribusi tersebut, mesin kalkula bergeser dari peran instrumental belaka (meskipun sudah signifikan) dan mencapai otonomi penuh hingga menjadi penengah tunggal dalam proses pengambilan keputusan putusan pidana individual. Sebuah fakta yang diiklankan dengan baik dalam waktu dekat oleh penyedia sistem masing-masing, karena pengembangan instrumen-instrumen ini secara berkelanjutan dan peningkatan kapasitas pembelajarannya diduga akan memberi mereka jenis algoritma yang mampu melakukan kalkulasi prediksi dengan margin kesalahan yang semakin kecil. Momennya mungkin akan tiba ketika instrumen-instrumen untuk menghitung residivisme ini akan sepenuhnya mampu menggantikan pengadilan, Kejaksaan Umum, dan pada umumnya aparat kesejahteraan sosial dalam kapasitas pembantu pengadilannya, yang memutuskan sendiri, dengan sudut pandang mereka sendiri, tanpa didampingi dan tanpa pengawasan, sifat hukuman, menentukan ukurannya, durasinya, dan

syarat-syarat penyelesaiannya. Pada saat itu, proses pengambilan keputusan putusan pidana akan menjadi, bukan berbasis mesin, bahkan bukan digerakkan oleh mesin, melainkan proses pengambilan keputusan oleh mesin.

Pertimbangan gabungan instrumen-instrumen ini dan adopsinya (dalam statusnya saat ini dan dalam waktu dekat) akan menantang keberadaan serangkaian prinsip-prinsip tradisional pertanggungjawaban pidana, seperti, sifat personal dari setiap hukuman pidana individu, in dubio pro reo, atau penerimaan kehendak bebas sebagai dasar pembenaran filosofis kontemporer atas putusan pidana individu.

18.2 PREDIKTABILITAS PERILAKU MASA DEPAN

Penerimaan Penilaian Probabilitas sebagai Kriteria dan Dasar untuk Membatasi Kebebasan Fisik: Implikasi dan Konsekuensinya

Mari kita bahas setiap tantangan satu per satu. Dan mari kita melakukannya melalui sudut pandang jenis penilaian yang dapat ditegaskan oleh program-program ini: penilaian yang tidak diragukan lagi didasarkan, bukan pada kepastian, tetapi hanya pada probabilitas.

Probabilitas residivisme tidak berarti kepastian residivisme dan perhitungan penilaian risiko tidak sama dengan kepastian bahaya. Oleh karena itu, penilaian probabilitas risiko, meskipun ditentukan secara ketat terkait dengan suatu agen, tidak menyiratkan bahwa ia sebenarnya berbahaya.

Dengan demikian, pengenaan penalti berdasarkan perhitungan tersebut dapat menyiratkan:

a. Penerapannya pada agen yang, terlepas dari probabilitas statistik yang telah dihitung, sebenarnya tidak menimbulkan risiko apa pun yang dihitung sebagai probable;
b. Margin negatif inefisiensi ini diterima, artinya diakui sebagai biaya sistem bahwa sejumlah narapidana yang tidak dapat ditentukan (karena hukuman diterapkan kepada semua orang berdasarkan perhitungan risiko yang telah divalidasi oleh sistem) akan dikenakan jenis hukuman dan tingkat keparahan yang, dalam kasus individual, tidak dibenarkan, karena ketidaksesuaian antara perhitungan probabilitas risiko dan risiko individu yang sebenarnya diwakili oleh terpidana.

Hukuman atau hukuman-hukuman yang dipilih, ukuran keparahannya, yaitu hukuman yang dijatuhkan kepada terdakwa, tunduk pada putusan yang didasarkan pada perhitungan risiko berbasis probabilitas belaka, akan berlaku bagi individu tersebut terlepas dari risiko yang, sebagaimana disebutkan di atas, sebenarnya diwakili oleh terpidana tersebut.

Dengan demikian, masalahnya: jika hukuman yang berlaku didasarkan pada pengejaran kebutuhan preventif, maka risiko pembenaran hukuman berdasarkan penilaian probabilitas, pasti akan menyiratkan bahwa perampasan kebebasan individu dari pelaku individu tersebut mungkin tidak memiliki tujuan yang dapat diramalkan, sah atau tidak. Oleh karena itu, sesuai dengan hukuman yang pengenaannya tidak dapat dibenarkan oleh tujuan yang sah.

Dengan demikian, pengabaian tersebut didikte, meskipun tanpa disadari, dari salah satu prinsip dasar peradilan pidana kontemporer: prinsip tentang sifat pribadi yang inheren dan tidak dapat dibatalkan dari setiap hukuman individu. Probabilitas, risiko, yaitu keraguan

bahkan jika didukung oleh perhitungan akan cukup. Hal itu akan cukup untuk perampasan kebebasan dan terutama untuk pengenaan hukuman pidana dengan jenis ketajaman tertentu yang tersirat (dan seharusnya tersirat) dalam setiap atribusi pidana. Penjahat yang dihukum disalahkan, dipilih, dan dicap berbahaya, bukan karena ia sebenarnya berbahaya, tetapi karena segala sesuatu yang bersifat teknologi mesin, spreadsheet, algoritma menunjukkan bahwa ia berbahaya. Bahkan jika dan bahkan ketika ia tidak berbahaya.

Komitmen tersebut mungkin terletak pada penyempurnaan terus-menerus mekanisme yang terlibat dalam perhitungan probabilitas risiko sebagai dasar untuk usulan atau keputusan tentang sifat dan kuantifikasi tingkat keparahan hukuman yang akan diterapkan pada setiap penjahat yang dihukum. Namun, penjahat yang dihukum itu akan selalu dianggap oleh sistem sebagai objek perhitungan. Objek yang akan ditargetkan dengan ukuran-ukuran tersebut.

Jika disederhanakan menjadi kondisi objek perhitungan, kalkula pada rei extensae yang terlibat dalam penilaian risiko yang diperlukan untuk penentuan hukuman pasti akan gagal memahami ketiadaan orang itu sendiri, individu itu sendiri, yang membentuk objek pengukuran tersebut, dalam persamaan tersebut.

Sekali lagi, seseorang dapat bersumpah demi komitmen untuk memungkinkan otonomi dalam perhitungan hanya ketika kondisi teknis terbaik tercapai. Namun, dengan keraguan yang membayangi setiap penilaian risiko dan dengan objek, alih-alih orang, yang menjadi fokus, perspektif yang diadopsi tentu saja adalah pembatasan kerugian jika terjadi kesalahan perhitungan oleh pihak pelaksana yang akan mengabaikan pertimbangan apa pun terhadap orang tersebut, menggantikannya dengan sekadar objek yang sedang diukur. Namun, jika ini benar, seseorang selalu dapat berargumen bahwa probabilitas justru merupakan inti dari setiap penilaian kemungkinan tindakan residivisme di masa mendatang oleh para penjahat terpidana. Baik itu evaluasi yudisial tradisional yang tunduk pada diskusi pengadilan terbuka, yang berkaitan dengan bukti-bukti relevan mengenai kepribadian terdakwa, serta keluarga, status sosial dan ekonomi, serta perilakunya sebelum dan sesudah melakukan kejahatan; baik itu aktivasi sibernetik teknologi dari parameter-parameter yang terorganisir secara algoritmik yang, sejujurnya, akan memasukkan faktor dan parameter yang sama ke dalam kalkulasi yang akan diteliti oleh hakim dan juri untuk mencapai keputusan yang beralasan tentang hukuman. Semua itu tidak memungkinkan lebih dari sekadar perkiraan. Probabilitas pada akhirnya mendasari setiap penilaian prognostik tentang residivisme dan mendukung setiap keputusan tentang pengenaan dan penegakan hukuman pidana dalam setiap kasus tertentu, pada masing-masing penjahat terpidana. Maka dapat dikatakan bahwa, setidaknya dalam bentuk teknologinya, metode-metode baru yang ditemukan untuk penilaian residivisme kriminal ini, masalah ketelitian atau kurangnya perhitungan probabilitas menjadi tematik secara tegas. Seiring dengan janji untuk mencapainya secara sukses dengan cara yang terukur dan terverifikasi.

Sedangkan bentuk penilaian risiko peradilan tradisional tidak akan menumbuhkan ambisi semacam itu karena tidak akan memungkinkan kapabilitas yang setara. Sebab, pengawasan ketat semacam itu akan membutuhkan jenis pengolahan data massal dan pengendalian parameter terdiferensiasi secara simultan yang akan jauh di bawah cakupan

kapabilitas yang diberikan oleh metode peradilan tradisional, yang manusiawi sebagaimana adanya dalam sifat dan bentuknya.

Lebih lanjut, pengolahan data dan kehati-hatian yang diambil dalam pemilihannya untuk mendukung penilaian risiko residivisme dengan tepat, dapat dikatakan lebih lanjut, tidak kalah ketatnya, karena metode manusia tradisional dan alternatif teknologi baru disandingkan. Dengan demikian, dapat dikatakan bahwa kemungkinan-kemungkinan teknologi baru ini menyediakan platform bagi kritik dan kecaman yang sangat dibutuhkan terhadap sistem penilaian dan atribusi risiko tradisional dengan atribusi kesalahan dan putusan pertanggungjawaban pidana individual yang inheren karena tidak mengakui fakta bahwa sistem tersebut, sama seperti sistem sibernetik, semata-mata bergantung dan bergantung pada probabilitas. Probabilitas, di satu sisi dan sisi lainnya, tetap merupakan paradigma.

Dan dengan tidak menerima realisasi tersebut, seseorang dapat berpendapat bahwa metode manusia tradisional secara efektif akan menyembunyikan persyaratan ketelitian tertinggi dalam perhitungan penilaian risiko individualnya. Hal ini menghilangkan instrumen yang memungkinkan peninjauan dan pengawasan menyeluruh serta meyakinkan terhadap setiap keputusan individual tentang hukuman. Dengan demikian, tanpa disadari, membuka pintu bagi kesewenang-wenangan yudisial.

Jika kritik tersebut benar. Dan memang benar. Dan jika kita dapat mengaitkan dorongan khususnya dengan keberadaan alternatif teknologi saat ini dengan kemampuannya yang tak tertandingi; proposal-proposal teknologi ini tetaplah kurang. Dan mereka kekurangan hal yang tidak dimiliki pendekatan tradisional. Pada inti pilihan epistemik fundamentalnya, mereka memang gagal. Dan mereka memang gagal di mana pendekatan tradisional yang dikritik, terlepas dari relevansi kritiknya, tidak. Model-model teknologi gagal dalam dan karena kemandirian epistemiknya.

Kemandirian sebagai fitur sentral epistemiknya terletak, bukan pada fakta bahwa seluruh model kalkulasi yang disediakan oleh algoritma yang selalu adaptif merupakan epilog dari sebuah latihan probabilitas. Sebaliknya, kemandirian epistemiknya yang kokoh terletak pada asumsi bahwa probabilitas sudah mencukupi. Tidak ada lagi yang dibutuhkan.

Probabilitas adalah latihan yang menjadi dasar kalkulator sibernetik untuk seluruh penilaian risikonya. Keraguan oleh karena itu menjadi hal yang mendasari setiap upaya kalkulasi. Oleh karena itu, keraguan, yaitu keraguan yang di bawah keyakinan kemampuan teknis luar biasa telah direduksi menjadi ketidakrelevanan statistik, tidak akan cukup untuk membenarkan pembalikan, apalagi mempertanyakan tingkat penilaian risiko yang telah ditetapkan terkait individu terpidana tersebut. Terlepas dari keraguan tersebut, pernyataan probabilitas statistik akan cukup.

Seberapa pun keraguan mungkin ada di mana-mana dalam setiap perhitungan, tidak satu pun dari hal tersebut akan menghalangi penjatuhan hukuman kepada terpidana sesuai dengan ukuran risiko yang, terlepas dari keraguan tersebut, secara statistik dianggap mewakili dirinya. Seseorang dihukum, dirampas kebebasannya dan sebagaimana disebutkan di atas, disalahkan, dikucilkan, dan dicap berbahaya dihadapkan dengan pembenaran yang mungkin

sangat lemah. Karena, terlepas dari penilaian tersebut, dalam kasus tertentu, orang tersebut mungkin sebenarnya tidak menimbulkan risiko sama sekali.

Tentu saja, seseorang selalu dapat beralih ke cara manusia tradisional dalam melakukan penilaian risiko tersebut. Berdasarkan probabilitas, sama seperti padanan teknologinya, dan keinginan, seolah-olah, akan kepastian yang akan memberikan validitas yuridis yang memadai. Namun, perubahan dalam analisis tersebut tidak akan memberikan jawaban yang masuk akal atas pertanyaan yang diajukan. Dua kesalahan tidak akan menghasilkan kebenaran, dan keraguan dalam pendekatan manusia tradisional akan gagal memberikan perbaikan yang sangat dibutuhkan oleh pendekatan teknologi.

Namun, dan yang terpenting, meskipun mendasari sebuah fakta, pendekatan tersebut tidak menggambarkan realitas. Dan realitas dari upaya manusia tersebut adalah akarnya yang kuat dalam pencarian kepastian. Kepastian yang memberikan penjelasan atas hukuman yang didasarkan pada keyakinan hakim tersebut dan melaluinya, Negara (dan Persemakmuran bagi mereka yang berada di sana) mengenai risiko spesifik yang sebenarnya ditimbulkan oleh individu tersebut. Dan pada tanggung jawab Negara atas penilaian risiko yang keliru tersebut. Dalam pendekatan tradisional manusia ini, keraguan dihargai, dapat diapresiasi, ditantang di pengadilan terbuka, dapat diperdebatkan oleh para pihak, diperdebatkan, ditinjau berulang kali, dan ditantang lagi. Keraguan, bahkan di ambang ketidakrelevanan statistik, tetaplah keraguan. Bahkan, dapat dikatakan bahwa dalam pendekatan manusia ini tidak ada keraguan yang tidak relevan, baik secara statistik maupun lainnya.

Perbedaan dalam cara penanganan keraguan di setiap sistem sangat mencolok. Dan mungkin di situlah letak jenis kecukupan lain yang dibanggakan dan diemban oleh pendekatan teknologi: kemandirian prosedural. Dengan mengabaikan diskusi dalam bentuk apa pun, di pengadilan terbuka, tertutup, atau apa pun, pendekatan teknologi menghitung, menilai, dan mengusulkan hasil akhirnya. Suatu hari nanti, pendekatan teknologi dapat memutuskan sendiri. Sementara itu, ketika membantu pengadilan dalam menyampaikan putusannya, apa, mungkin seseorang bertanya, yang akan dimiliki pengadilan untuk melawan kemandirian keputusan siap pakai yang memutuskan berdasarkan keraguan yang dapat diabaikan, yang menyamakannya dengan kepastian yang memadai, tanpa pernah mengambil bagian dalam rasionalnya, atau memungkinkan diskusi yang kontradiktif tentang apa yang telah menjadi satu-satunya domain aktuaria mesin (menggunakan istilah yang menjadi dasar mesin memutuskan melawan dirinya sendiri)?

Keraguan dan probabilitas dengan demikian memenuhi, dalam pendekatan manusia tradisional, bagian dari biaya sistem. Keduanya merupakan bukti kekeliruannya, sumber ketidakadilannya. Kesadaran akan apa yang terkandung dalam ancaman ketidakadilan yang nyata menunjukkan komitmen pendekatan tradisional terhadap prinsip praduga tak bersalah dan berfungsi sebagai tulang punggung klaimnya atas validitas konstitusional.

Klaim ini mensyaratkan penerimaan, sebagai penegasan moral yang tak terbantahkan tentang kejahatan yang lebih kecil, bahwa lebih baik membebaskan orang yang bersalah daripada menghukum orang yang tidak bersalah. Dan untuk mengungkapnya secara menyeluruh dalam pelaksanaan kalkulasi dan penilaian risiko residivisme setiap terpidana.

Oleh karena itu, keraguan akan dipertimbangkan demi kepentingan orang yang dievaluasi, dengan menerima risiko kesalahan sebagai wujudnya. Penilaian risiko yang tidak memadai, dalam menghadapi keraguan, menjadi biaya sistem tradisional yang, dengan menyatakan prinsip kejahatan yang lebih kecil yang telah disebutkan sebelumnya, melakukannya sebagai bagian dari tugasnya untuk menghormati dan mempertimbangkan secara setara setiap individu, mengakui kemanusiaan yang melekat pada setiap orang. Pernyataan semacam itu diberkahi dengan konsekuensi praktis langsung, mengedepankan prinsip yang persis sama yang merupakan terjemahan praktis manusiawi dari pernyataan tersebut dan metode tradisional: prinsip in dubio pro reo. Prinsip inilah yang gagal disucikan dan diintegrasikan oleh alternatif teknologis. Prinsip inilah yang mengakui validitas konstitusional prinsip pertama, sehingga merugikan prinsip kedua.

18.3 RISIKO BIAS TEKNOLOGI

Lebih lanjut, pemahaman linear tentang perilaku prospektif yang menjadi inti dari setiap penilaian risiko yang otonom secara teknologi akan memungkinkan pemahaman perilaku manusia secara linear. Kelayakan penilaian risiko teknologi hanya dimungkinkan sejauh perilaku manusia di masa depan dapat ditentukan. Sebuah penentuan yang, mari kita ingatkan diri kita sendiri, hanya dimungkinkan melalui transformasi keraguan atas risiko aktual, ketika secara statistik tidak mungkin, menjadi kepastian risiko.

Pada tingkat ini, tidak akan ada ruang bagi momen-momen langka penyesalan, penghentian, pertobatan, introspeksi, atau pencerahan. Penelusuran kembali, perubahan pikiran, karena secara statistik sangat tidak mungkin, akan sama dengan ketiadaan di dunia baru yang berani dengan keputusan teknologi yang terlalu pasti tentang risiko. Keputusan sibernetik yang tepat yang akan menjalankan fungsinya, jika dan hanya jika, unsur manusia dilepaskan dari persamaan, sepenuhnya diabaikan dari perhitungan. Momen-momen itu, meskipun jarang, sama saja dengan ketiadaan. Namun, kodrat dan kondisi manusia, pada momen-momen langka tersebut, dengan keras kepala menantang kepastian yang seperti mesin, meskipun tampak menenangkan dalam keakuratannya yang diduga.

Jika seseorang menerima keadaan teknologi ini, beserta teka-teki metodologis yang telah disebutkan, akan terjadi perubahan pada premis filosofis yang menopang dan membentuk dasar bagi setiap pelaksanaan peradilan pidana dan penetapan kesalahan. Premis filosofis tersebut, yaitu ketergantungannya pada pemahaman setiap orang sebagai agen bebas. Premis filosofis tersebut, pada akhirnya, komitmen terhadap indeterminisme. Yang dipertaruhkan adalah pengabaian konsepsi indeterministik tentang tindakan manusia. Dan di pinggiran statistik, di mana ketidakmungkinan mengintai, ketidakrelevanan kehendak bebas diterima untuk semua maksud dan tujuan. Digantikan oleh semacam determinisme statistik yang masuk akal, meskipun tidak tertantang dan diterima tanpa kritik.

Lalu ada risiko pelestarian dan promosi model realitas sosial yang ada dengan prasangka dan asimetrinya. Dan, bersamaan dengan itu, serangkaian masalah konjungtural yang terus-menerus menguji kredibilitas semua upaya teknologi dalam penilaian risiko. Karena setiap upaya perlu mempertimbangkan data tentang lingkungan sosial dan asal-usul

terpidana, upaya semacam itu dapat menyiratkan risiko mempertimbangkan terdakwa sebagai sebuah model, dan, dengan demikian, sebagai anggota komunitas paradigmatik untuk tujuan statistik. Dan risiko bias pun muncul.

Selain risiko inheren berupa salah perhitungan, model semacam itu cenderung melanggengkan persepsi mayoritas dominan tentang realitas sosial dengan asimetri yang menyuburkan prasangka yang mendasarinya. Ketika tidak memiliki instrumen kritis yang mampu mengungkapnya, model algoritmik penilaian risiko yang diajukan oleh terdakwa, alih-alih mampu memastikan bahaya aktual yang diwakilinya sebagai individu, justru berisiko membebani orang-orang yang sama persis yang, sebagai anggota komunitas yang kurang beruntung, telah menderita dampak dari asimetri dan prasangka yang sama persis.

Lebih buruk lagi. Dengan menyediakan gudang data dan metode kalkulasi yang tampaknya mengurutkan kesimpulan yang diperlukan secara logis, masuk akal, dan linear, sistem tersebut akan cenderung memvalidasi dan, dengan demikian, melegitimasi cara pengumpulan dan pemrosesan data serta usulan hukuman yang dihasilkan, yang di balik gudang data tersebut, mereplikasi dengan cara yang memungkinkan untuk disembunyikan dan disamarkan bias yang tepat yang akan melabeli individu-individu yang risikonya dinilai dengan faktor-faktor kriminogenik yang sama yang memengaruhi secara nyata anggota komunitas asal mereka. Dan tidak ada alasan lain selain fakta bahwa mereka berasal dari komunitas tersebut. Oleh karena itu, mereka yang, pada kenyataannya, paling sering bergulat dengan keadilan mungkin, dalam penilaian linear yang logis dan masuk akal terhadap mesin, mendapati diri mereka dilabeli sebagai personifikasi bahaya karena logika mekanis menuntutnya. Terlepas dari siapa dirinya secara individu dan terlepas dari apa yang mungkin dapat dilakukan oleh pertimbangan manusia, di bawah cahaya penalaran (yang tidak kalah) manusiawi, dalam menantang logika (yang dihitung seperti mesin) tersebut.

18.4 KESIMPULAN

Tentu saja pertimbangan tentang perubahan pikiran yang mustahil, meskipun selalu mungkin.Kami akan mencoba memberikan jawaban dengan cara yang sama seperti pada kesempatan sebelumnya. Jalannya akan tidak ortodoks. Kami mulai dengan sastra.

Jenis perubahan yang penulis Brasil Jorge Amado, dalam novelnya Terras do Sem Fim, masukkan ke dalam pikiran Damião, seorang pembunuh bayaran (jagunço), saat ia bersiap untuk menyergap dan membunuh korban lainnya. Sebuah perubahan pikiran, bahkan sebuah pencerahan, yang membuatnya, untuk pertama kalinya, gagal menembak. Damião tersiksa oleh pertanyaan-pertanyaan yang pernah didengarnya dari Sinhô Badaró yang ditujukan kepada pembunuh lain (jagunço), “Apakah menurutmu membunuh orang itu baik? Tidakkah kau merasakan apa pun? Tidak ada apa-apa di dalam?”. Dan kini dalam penyergapan itu, ketika ia bersiap untuk membunuh Firmo, sebuah ide muncul. Awalnya, hanya sebagai dugaan “Dan tiba-tiba, ide mengerikan itu memotong kepalanya: bagaimana jika Dona Teresa hamil, seorang anak dalam perutnya? (...) Ia akan lahir tanpa ayah, sang ayah akan menjadi sasaran” Damião.

«Dan ia menggigil di sekujur tubuhnya, tubuhnya yang besar dan raksasa.

> Anda dapat melihat wajah Dona Teresa (...) sebelum ada cahaya bulan, seputih susu, yang tumpah ke tanah. (...) Ia memintanya untuk tidak membunuh Firmo, demi Tuhan ia tidak membunuh... Di lantai yang diterangi cahaya bulan, pria berkulit hitam itu melihat wajah Teresa dengan sempurna.»
>
> Bagaimana jika aku tidak membunuh Firmo?
>
> Seandainya pun tak terduga, di situlah letak kemanusiaan.

Dan kata-kata Arendt muncul di benak: "fakta bahwa manusia mampu bertindak berarti bahwa hal yang tak terduga dapat diharapkan darinya, bahwa ia mampu melakukan apa yang mustahil tak terhingga". Maka, karena manusia "unik", karena, dengan setiap kelahiran, "sesuatu yang unik dan baru datang ke dunia". Setiap awal, demikian Arendt, setiap asal membawa serta unsur ketidakpastian. Oleh karena itu, setiap momen perbedaan merupakan momen kebaruan yang tak terduga dan tak terduga, momen kejutan, yang menegaskan bahwa "yang baru" "selalu muncul dalam kedok keajaiban".

Di sini, pemikiran lahir sebagai kondisi politis untuk tindakan, sebagai kemungkinan bahasa dan wacana, dan oleh karena itu, kehidupan yang didasarkan pada, karena ditandai oleh, kemanusiaan. Dan, dengan itu, kata, pemikiran yang merefleksikannya, dan penilaian yang kemudian, melalui keduanya, menjadi mungkin. Nah, yang dipertaruhkan di sini justru, menurut ARENDT, kembalinya pemikiran ke tindakan. Tindakan yang dianggap sebagai kebebasan. Sebab, ketika tindakan, yang melahirkan kreativitas manusia, dikesampingkan, pemikiran itu sendiri, sebagai "aktivitas berbagi dalam perbedaan yang di dalamnya kesetaraan diakui", akan dilupakan. Dan, pada saat itu, penilaian, yang menjadi sandaran kemungkinan kebebasan, akan lenyap dari eksistensi manusia.

Oleh karena itu, apa yang secara definitif ditinggalkan? Sekali lagi, sastra. Kita beralih ke PRIMO LEVI dalam bukunya Se Questo è un Uomo. Saat ditawan di Auschwitz, ia menceritakan bahwa suatu hari ketika dikawal oleh seorang penjaga, bernama Alex, sebuah pipa baja yang diminyaki melintasi jalan mereka. Saat penjaga itu lewat, ia bersandar pada pipa dan secara tidak sengaja mengotori tangannya dengan minyak. "Tanpa kebencian dan tanpa ejekan", Levi memberi tahu kita, "Alex menyeka tangannya di bahu saya, telapak tangan dan punggung tangan Anda, untuk membersihkannya". Dan Levi menyimpulkan, "(...) ia akan sangat terkejut, Alex yang malang dan kejam, jika seseorang mengatakan kepadanya bahwa hari ini saya menghakiminya atas tindakan ini, ia dan Pannwitz, dan banyak orang yang seperti dia, besar dan kecil, di Auschwitz dan di tempat lain. "

Hal ini penilaian manusia sebagai hakikat pemikiran manusia dan tanda martabat manusia menjadi terlupakan dan, seiring dengan munculnya kelupaan, ratapan P. F. TRAWSON menjadi ratapan kita sendiri. Sebuah ratapan yang ia ungkapkan sebagai berikut: "sangat disayangkan bahwa pembicaraan tentang sentimen moral telah kehilangan dukungan". Yang tidak disukai bukan hanya keraguan mengenai pembatasan konsep moral ke dalam satu asal rasional yang tunggal, tetapi juga persepsi terhadap ketidakcukupan akal budi semata sebagai dorongan motivasi utama untuk bertindak.

Sebuah ratapan yang menjadi titik tolak untuk menguraikan sikap reaktif partisipasi yang diusulkannya. Sebagaimana sikap-sikap yang dalam menghadapi setiap perilaku kita masing-masing memicu sikap reaktif umum (sikap reaktif biasa). Sikap-sikap semacam itu yang

bersifat partisipatif yang disebut oleh Penulis sebagai "reaksi yang pada hakikatnya alami dan manusiawi", yang tanpanya "diragukan kita memiliki sesuatu yang dapat kita anggap dapat dipahami sebagai sebuah sistem hubungan antarmanusia, sebagai masyarakat manusia".

Yang dipertaruhkan adalah perlunya mempertimbangkan jaringan sikap dan perasaan yang membentuk bagian esensial dari kehidupan moral sebagaimana kita mengenalnya, sebagai sumber makna yang tersirat dalam "bahasa moral", dan, dengan demikian, dari segala sesuatu yang ingin kita katakan ketika kita berbicara tentang kelayakan, tanggung jawab, rasa bersalah, kutukan, dan keadilan. Masing-masing beresonansi dalam sikap dan perasaan yang membentuk perekat jenis hubungan spesifik yang membentuk sebuah komunitas. Jenis hubungan yang setara dengan kehidupan itu sendiri, sekali lagi dengan ARENDT, dalam kebersamaan yang secara ontologis mencirikannya. Sebuah hubungan yang, pada akhirnya, terletak pada fondasi tanggung jawab moral.

BAB 19
RELEVANSI DEEPFAKE DALAM ADMINISTRASI PERADILAN PIDANA

Abstrak Saat ini, sulit untuk membedakan antara konten asli yang dibuat oleh manusia atau deepfake yang dibuat oleh algoritma deepfake. Oleh karena itu, masyarakat dan negara-negara perlu memiliki sistem yang dapat mendeteksi dan mengevaluasi konten tersebut tanpa campur tangan manusia. Makalah ini menyajikan tantangan kecerdasan buatan, khususnya pembelajaran mesin dan pembelajaran mendalam, dalam melawan deepfake. Selain itu, makalah ini juga menyajikan relevansi deepfake dalam administrasi peradilan pidana.

19.1 DEEPFAKE: AI MEDIA PALSU BERBAHAYA

Deepfake muncul pada akhir tahun 2017 ketika seorang pengguna anonim di jejaring sosial Reddit mengunggah video yang menampilkan wajah aktris porno yang diganti dengan wajah selebritas. Potensi teknik ini dengan cepat menyebar di internet, sehingga dapat diakses oleh semua orang. Deepfake adalah salah satu penggunaan kecerdasan buatan (AI) yang paling berbahaya dan paling terlihat. Namanya berasal dari gabungan "pembelajaran mendalam" dan "media palsu". Untuk membuat Deepfake, teknik AI, khususnya pembelajaran mesin (ML), digunakan untuk membuat atau memanipulasi konten, baik berupa audio maupun video. Konten yang mungkin sangat sulit dibedakan dari konten asli oleh manusia maupun solusi teknologi.

19.2 DEEPFAKE: DEFINISI DAN KATEGORI

Deepfake adalah konten (video, audio, atau lainnya) yang sepenuhnya atau sebagian direkayasa atau dimanipulasi dari konten yang sudah ada (video, audio, atau lainnya). Namun, kasus deepfake yang paling signifikan muncul dalam format video, yang kesulitan autentikasinya memungkinkan informasi apa pun karena konten audiovisual apa pun dapat direkayasa. Deepfake muncul dari aplikasi AI yang menggabungkan, mencampur, mengganti, atau melapisi gambar dan video, menciptakan video palsu sehingga terlihat autentik. Deepfake adalah contoh nyata dari kerumitan dan kompleksitas teknologi yang digunakan. Namun, aplikasi yang tersedia memungkinkan siapa pun dengan daya komputasi rendah dan pengetahuan terbatas untuk membuat video palsu.

Deepfake dapat dibuat secara legal untuk melibatkan atau memicu pengamatan kritis. Namun demikian, deepfake juga digunakan untuk melakukan penipuan, menipu, atau mengintimidasi orang dengan merilis gambar atau video tanpa persetujuan mereka. Terdapat beberapa kategori deepfake: penggantian wajah atau pertukaran tubuh, rekonstruksi wajah, pembangkitan wajah, sintesis ucapan, dan shallowfake. Penggantian wajah atau pertukaran tubuh menggantikan bagian tubuh seseorang dengan bagian tubuh lain. Rekonstruksi wajah memanipulasi bagian tubuh seseorang untuk menyusunnya sehingga tampak seolah-olah mereka mengekspresikan sesuatu yang bukan dirinya. Pembangkitan wajah memungkinkan terciptanya citra sintetis dari orang-orang yang meyakinkan tetapi sepenuhnya fiktif. Sintesis

ucapan menggunakan algoritma pelatihan untuk menghasilkan suara deepfake atau berkas audio buatan. Shallowfake memungkinkan terciptanya penipuan audiovisual dengan menggunakan metode penyuntingan dasar. Tabel 19.1 menyajikan contoh-contoh deepfake yang berbeda.

19.3 AI DAN DEEPFAKE

Seperti yang telah disebutkan sebelumnya, deepfake dibuat menggunakan AI. Tapi bagaimana caranya? Salah satu keunggulan AI adalah ia menyerap banyak pengetahuan dari lingkungan tempat ia ditempatkan, mempelajari, dan meningkatkan responsnya setiap hari.

Untuk lebih memahami konsep AI, AI dapat dibagi menjadi dua kategori: Kecerdasan Buatan Kuat dan Kecerdasan Buatan Sempit. Kecerdasan Buatan Kuat mencakup sistem yang menunjukkan kecerdasan manusia atau bahkan lebih unggul di semua bidang. Lebih lanjut, jenis kecerdasan ini dapat berbagi pengalaman dari berbagai domain. Beberapa pengujian digunakan untuk menunjukkan apakah suatu sistem menunjukkan AI Kuat, tetapi hal ini belum terjadi, yang menyebabkan para ahli berpendapat bahwa ini hanyalah sebuah aspirasi.

AI Sempit atau AI Lemah mencakup semua sistem yang memiliki jawaban benar atau salah yang terdefinisi dengan baik, yang memiliki pola dan struktur dasar yang jelas, dan yang memiliki kecepatan riset dan komputasi yang menawarkan keunggulan dibandingkan manusia. Sistem AI yang ada saat ini belum dirancang untuk menerapkan penalaran abstrak, memahami konsep, atau keterampilan pemecahan masalah spektrum luas secara umum. Jadi, sistem dalam kategori ini memiliki aplikasi yang sempit dan terbatas untuk memecahkan masalah spesifik.

Ada tiga pendekatan utama yang berbeda untuk mengembangkan sistem AI: Metode berbasis aturan, Pembelajaran Mesin (ML), dan Pembelajaran Mesin (DL). Dari ketiga pendekatan ini, yang paling banyak digunakan adalah dua pendekatan terakhir, yang akan dibahas di bagian selanjutnya.

Tabel 19.1 Jenis dan contoh deepfake

Jenis	Format	Deskripsi	Aplikasi bisnis
Penggantian wajah atau pertukaran tubuh	Deepfake Foto	Tukar Wajah dan Tubuh—mengganti wajah atau tubuh orang lain.	Kosmetik, kacamata, gaya rambut, atau pakaian virtual.
	Deepfake Video	Tukar Wajah—mengganti wajah orang lain.	Film, di mana wajah aktor ditampilkan pada tubuh ganda.
Rekonstruksi wajah (tubuh)	Deepfake Audio dan Video	Puppetry seluruh tubuh—memindahkan gerakan tubuh dari satu orang ke orang lain.	Direktur perusahaan dan atlet dapat menyamarkan kondisi fisik selama pameran video.
		Sinkronisasi Bibir—menyesuaikan aktivitas mouse dan frasa yang	Video institusional dapat dikonversi menjadi pidato lain dengan menggunakan juru

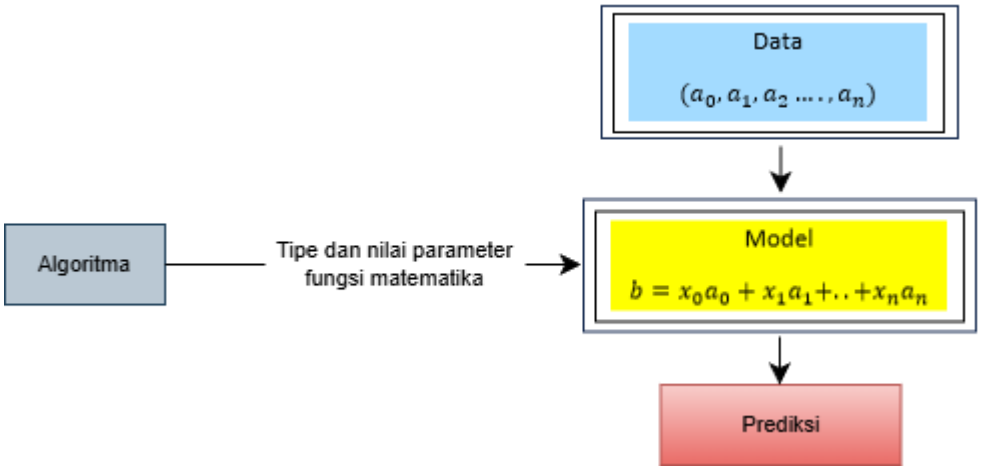
		diartikulasikan dalam video percakapan.	bicara yang sama dalam rekaman aslinya.
Pemalsuan dangkal	Deepfake Video	Morphing Wajah—wajah berubah menjadi wajah lain melalui transisi yang mulus.	Peserta gim video dapat memperkenalkan penampilan mereka kepada tokoh favorit mereka.
Sintesis ucapan	Deepfake Audio	Tukar Suara—mengganti juru bicara atau meniru orang lain.	Juru bicara narasi dapat berupa orang muda, tua, pria, atau wanita.
		Text-to-Speech - memodifikasi audio dengan menulis konten baru.	Mengganti kata-kata tanpa perlu membuat berkas audio baru.
Pembuatan wajah	Deepfake Foto	Pembuatan Wajah—menghasilkan wajah yang tampak realistis.	Ini dapat berfungsi sebagai solusi untuk menghapus foto tanpa melanggar privasi atau hak gambar siapa pun, karena semuanya telah dihasilkan secara artifisial, sehingga tidak ada yang difoto.

19.4 PEMBELAJARAN MESIN

ML adalah jenis AI yang memungkinkan aplikasi perangkat lunak menjadi lebih tepercaya dengan meningkatkan kapasitasnya untuk mengantisipasi hasil yang akurat. Dalam ML, algoritma disajikan dengan kumpulan data berisi beberapa data dan sebuah tag untuk data tersebut. Dengan cara ini, sistem akan menemukan pola umum dalam kumpulan data tersebut. Gambar 19.1 menyajikan posisi ML terhadap AI. Semakin banyak contoh yang terdapat dalam kumpulan data, semakin efektif algoritma merespons. Selain itu, sistem belajar dari kesalahannya.

Berdasarkan konsep teoretis, ML dapat dijelaskan oleh sebuah model, yang memiliki data sebagai masukan dan prediksi sebagai keluaran. Namun, model tidak lebih dari sekadar rumus matematika. Rumus matematika adalah hasil implementasi algoritma ML. Rumus matematika akan mengukur parameter yang dapat digunakan untuk prediksi. Model dapat dilatih dan dipelajari dari data pelatihan.

Gambar 19.1 Posisi ML dalam konteks AI



Gambar 19.2 Visualisasi diagram Model ML

Gambar 19.2 menyajikan visualisasi diagram Model ML. Hal ini sangat umum di masyarakat, memungkinkan perusahaan untuk menyelidiki tren perilaku individu dan pola-polanya, tidak hanya untuk meningkatkan produk mereka tetapi juga untuk memberikan kualitas layanan yang lebih baik kepada pengguna.

Beberapa perusahaan di dunia menggunakan ML, dan ML berperan sebagai faktor pembeda yang penting antar perusahaan. Saat ini, ML digunakan di banyak bidang dan aplikasi. Salah satu aplikasi ML yang paling menonjol adalah dalam sistem rekomendasi, yang memungkinkan pemberian saran kepada klien tentang topik-topik tertentu. Contohnya adalah berita, permainan, dan film.

Sebagaimana telah disebutkan, terdapat beberapa area penerapan ML, dan terdapat beberapa kategori yang mengelompokkan berbagai algoritma berdasarkan tujuannya. Sebagaimana disajikan pada Gambar 19.3, tiga kategori utama adalah pembelajaran terawasi,

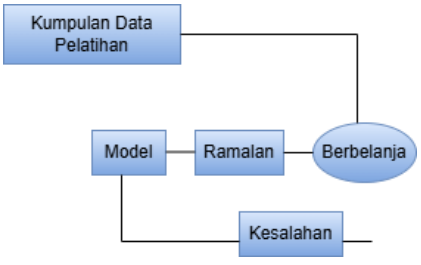
tanpa pengawasan, semi-supervised, dan pembelajaran penguatan. Gambar ini juga menyajikan beberapa aplikasi dari setiap jenis kategori ML.

Pembelajaran Terawasi

Pembelajaran terawasi dapat digambarkan sebagai proses yang menggunakan algoritma yang mampu menghasilkan pola dan hipotesis dari contoh yang diberikan, dan menerapkannya untuk memprediksi contoh yang tidak diketahui. Jenis pembelajaran ini mencoba memprediksi variabel dependen dari daftar variabel independen. Semua algoritma ML mengikuti proses yang serupa: set data, fitur, algoritma, evaluasi, dan pelatihan.

Untuk menggunakan model ML, selalu dibutuhkan set data masukan yang dibagi menjadi pelatihan dan pengujian, yang harus berisi jutaan data. Semakin besar set data, semakin baik respons algoritma. Pelatihan mencakup variabel yang akan diprediksi/diklasifikasikan, dan dengan data inilah algoritma akan belajar sehingga dapat menerapkan pengetahuan ini dalam set data pengujian dan memprediksi/mengklasifikasikan variabel yang sama.

Gambar 19.3 Metode Pembelajaran Mesin



Gambar 19.4 Model pembelajaran terbimbing

Fitur adalah data yang akan diubah menjadi representasi numerik agar dapat dipahami oleh algoritma. Algoritma tidak menggunakan semua data, karena akan mempelajari data mana yang relevan.

Pada pembelajaran terbimbing, kumpulan data memiliki beberapa sampel data, yang terdiri dari pasangan contoh input-output yang melatih model. Ketika model disiapkan, model akan memprediksi hasil label yang diharapkan. Kemudian prediksi tersebut dibandingkan dengan label. Jika tidak ada kecocokan, kita memiliki apa yang disebut kesalahan. Kesalahan tersebut hanyalah umpan balik dalam model, yang akan diperbarui. Gambar 19.4 merepresentasikan model pembelajaran terbimbing.

Berbagai algoritma pembelajaran mesin dapat digunakan, yaitu algoritma regresi atau klasifikasi (Tabel 19.2). Algoritma regresi bertujuan untuk mengetahui bagaimana satu variabel berevolusi terhadap variabel lainnya. Jenis algoritma ini memprediksi nilai kontinu. Contoh penerapan algoritma regresi adalah prediksi, peramalan, atau estimasi.

Algoritma klasifikasi berusaha menjelaskan variabel kategoris dengan dua kategori atau lebih, membagi data ke dalam kelas-kelas, menggunakan fitur-fitur umum. Beberapa contoh penerapan algoritma klasifikasi adalah deteksi penipuan, klasifikasi citra, retensi pelanggan, dan diagnostik personal.

Tabel 19.2 Jenis aplikasi algoritma

Jenis	**Nama Algoritma**	**Deskripsi**
Regresi	Regresi Linier	Membandingkan setiap fitur dengan hasil untuk membantu prakiraan di masa mendatang.
Klasifikasi Regresi	Pohon Keputusan	Nilai sumber daya data dipisahkan menjadi cabang-cabang pada simpul keputusan oleh model klasifikasi atau regresi.
	Bayes Naif	Ini adalah sekelompok pengklasifikasi probabilistik dasar yang bergantung pada penerapan teori Bayes dengan tata kelola mandiri yang kuat dari fitur-fitur Bayes Naif.
	Hutan Acak	Menerapkan keputusan sederhana berdasarkan pendekatan yang paling banyak dipilih. Dalam regresi, prediksi dihitung berdasarkan rata-rata.
	AdaBoost	Menggunakan berbagai model untuk memutuskan, tetapi skala merupakan ukuran penting untuk mendapatkan presisi dalam hasil prakiraan.
	Pohon Penguat Gradien	Ini berfokus pada kesalahan yang dibuat oleh pohon sebelumnya dan mencoba memperbaikinya.
	Mesin Vektor Pendukung	Digunakan untuk tugas klasifikasi dan menemukan hiperbidang yang memisahkan kelas-kelas.
Klasifikasi	Regresi Logika	Mengukur hubungan antara variabel absolut dan setidaknya satu faktor bebas dengan mengevaluasi probabilitas menggunakan kapabilitas logistik.

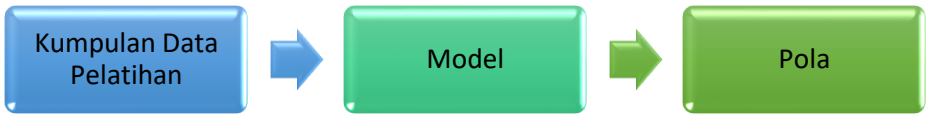
Pembelajaran Tanpa Supervisi

Pembelajaran tanpa supervisi berbeda dari pembelajaran tersupervisi karena tidak ada data yang mendefinisikan label. Cara algoritma menciptakan pengetahuan dari data dilakukan

secara berbeda, dengan menganalisis afinitas antar objek yang dianalisis untuk mendeteksi persamaan/perbedaan karakteristiknya; dari sana, label akan dibuat dan ditetapkan. Pembelajaran ini menghasilkan temuan pola dalam data yang jika tidak demikian akan dianggap sebagai noise, karena tidak mengandung informasi bermanfaat.

Keuntungan pembelajaran tanpa supervisi adalah data tidak perlu dikategorikan, sehingga sejumlah besar data tidak terstruktur dapat diakses untuk dianalisis. Algoritma mencoba menarik asumsi dari data yang tidak berlabel, dan menemukan pola data baru. Gambar 19.5 menyajikan model pembelajaran tanpa supervisi.

Aplikasi penting dari pembelajaran tanpa supervisi adalah deteksi anomali. Dalam metode ini, jaringan dilatih untuk menemukan komposisi dan tampilan keseluruhan aliran data, menentukan apakah satu titik data terlihat berbeda dari yang lain. Penerapan metode ini memungkinkan, misalnya, deteksi upaya penipuan siber dalam transaksi yang kompleks. Model pembelajaran tanpa pengawasan dapat dibagi menjadi pengelompokan dan pengurangan dimensionalitas data.



Gambar 19.5 Model pembelajaran tanpa pengawasan

Tabel 19.3 Algoritma yang dapat diterapkan untuk pengelompokan dan reduksi dimensionalitas

Tipe	**Nama Algoritma**	**Deskripsi**
Pengelompokan	Model Campuran Gaussian	Lebih fleksibel dalam rentang dan struktur klaster k-means.
	Pengelompokan K-Means	Menempatkan data ke dalam beberapa klaster (k), yang masing-masing berisi data dengan atribut yang sama.
	Pengelompokan Hirarkis	Ini adalah perhitungan yang menciptakan hierarki kelompok. Perhitungan ini dimulai dengan mendistribusikan data berdasarkan kelompok berdasarkan informasi masing-masing. Di sini, dua kelompok yang berdekatan akan berada dalam kelompok yang serupa. Perhitungan ini ditutup ketika hanya tersisa satu kelompok.
	Sistem Rekomendasi	Membantu menentukan data penting untuk menyusun saran.
	K-Tetangga Terdekat	Ini adalah pengukuran langsung yang mempertahankan semua topik yang tersedia dan mendeskripsikan contoh baru berdasarkan estimasi kemiripan.

Pengurangan Dimensi	PCA/T-SNE	Proses ini mengurangi jumlah fitur menjadi 3 atau 4 lintasan dengan varians tertinggi.

Pengelompokan adalah teknik yang membagi dan mengelompokkan sampel data yang serupa. Pengelompokan ini disebut klaster. Contoh pengelompokan adalah sistem yang direkomendasikan, pemasaran yang ditargetkan, dan segmentasi pelanggan.

Reduksi dimensionalitas adalah metode untuk meringkas fitur menjadi apa yang disebut nilai inti yang secara ringkas menyampaikan informasi serupa. Dengan memilih beberapa komponen saja, jumlah sumber daya berkurang, dan sebagian kecil data hilang. Tabel 19.3 merangkum algoritma yang dapat diterapkan untuk pengelompokan dan reduksi dimensionalitas.

Pembelajaran Semi-Supervisi

Pembelajaran semi-supervisi memiliki banyak kegunaan di dunia digital, mendeteksi penipuan, baik dalam berita, email, dll. Algoritma yang dilatih dalam kumpulan data kecil dapat belajar memberi label pada data dan digunakan dalam penerjemahan, memungkinkan algoritma untuk menerjemahkan bahasa menggunakan kamus yang tidak lengkap.

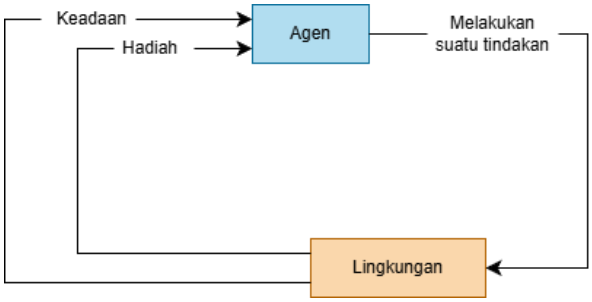
Model ML dilatih dan diuji dengan data yang hadir dalam proporsi yang tidak sama. Proporsi data pelatihan lebih rendah daripada rasio data uji. Algoritma pembelajaran mesin semi-supervised berada di antara pembelajaran tanpa supervisi dan tersupervisi, dan data tanpa label dapat meningkatkan akurasi pembelajaran secara signifikan.

Pembelajaran Penguatan

Terakhir, pembelajaran penguatan terdiri dari pengajaran model pembelajaran mesin, mendefinisikan aturan spesifik yang harus diikuti, memberikan penghargaan ketika algoritma menyelesaikan tugas dengan baik, atau memberikan hukuman ketika berperilaku salah.

Pembelajaran penguatan berbeda secara signifikan dari pembelajaran seperti yang telah disebutkan sebelumnya. Dalam jenis pembelajaran ini, pertukaran antara agen dan lingkungan tempat ia bekerja sangat penting. Dengan cara ini, agen berinteraksi dengan lingkungan yang menghasilkan tindakan yang akan mengubah lingkungan tersebut, sehingga mesin menerima penghargaan atau penalti.

Perlu dicatat bahwa agen tidak menyadari sebelumnya tindakan yang perlu diambil. Keputusan yang diambilnya akan memengaruhi tindakan selanjutnya. Gambar 19.6 merinci model pembelajaran penguatan. Tujuan mesin adalah belajar berperilaku sedemikian rupa sehingga menghasilkan imbalan (atau mengurangi penalti) selama masa pakainya. Pembelajaran ini hanya bergantung pada dua kriteria: hasil yang tertunda dan penelitian coba-coba.



Gambar 19.6 Model Pembelajaran Penguatan

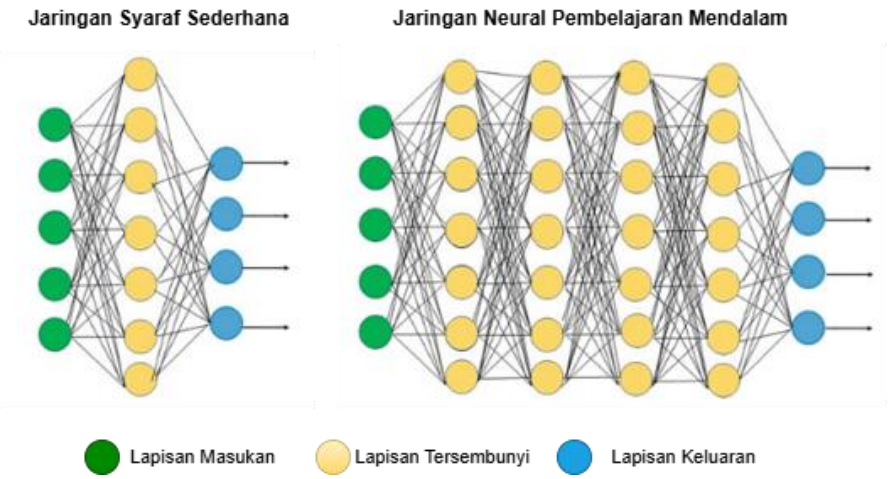
Agen bermaksud membangun kebijakan yang optimal; namun, ia memecahkan masalah mempelajari keadaan baru sambil meningkatkan manfaatnya secara keseluruhan. Dilema trade-off ini disebut Eksplorasi versus Eksplorasi. Agen harus menganalisis kedua sisi dilema dan memilih strategi berdasarkan hasil keseluruhan. Oleh karena itu, untuk membuat keputusan umum terbaik di masa mendatang, agen harus menyimpan informasi secara memadai. Contoh pembelajaran penguatan adalah keputusan yang dibuat secara waktu nyata, visi komputer, dan mengemudi otonom.

Terdapat dua pendekatan model yang berbeda untuk pembelajaran penguatan, yaitu Proses Keputusan Markov (MDP) dan model Q-learning. MDP terdiri dari klaster status domain terbatas S, sekelompok kemungkinan tindakan $A(s)$ dalam setiap kondisi, fungsi imbalan nyata $R(s)$, dan model transisi $P(s', s|a)$. Q-learning adalah metode gratis yang diterapkan untuk membuat agen PacMan yang dapat bermain sendiri. Metode ini berputar di sekitar nilai Q yang merevisi yang merepresentasikan nilai a yang berperilaku dalam status s.

19.5 PEMBELAJARAN MENDALAM

Pembelajaran Mendalam (Pembelajaran Mendalam/*DeepLearning*/DL) adalah subkelas teknik Pembelajaran Mesin (ML) di mana sistem terdiri dari beberapa lapisan untuk mempelajari representasi data dengan berbagai tingkat konsepsi. Sebagian besar metode DL menggunakan arsitektur jaringan saraf tiruan, sehingga disebut jaringan saraf tiruan dalam (*deep neural network*/DNN). Gambar 19.7 menggambarkan jumlah lapisan tersembunyi dalam jaringan saraf tiruan. Jaringan saraf standar terdiri dari 2-3 lapisan tersembunyi, tetapi jaringan dalam terdiri dari sekitar 150 lapisan.

Pembelajaran tugas klasifikasi dari gambar, teks, atau suara dilakukan oleh model pembelajaran jarak jauh (DL). Pembelajaran ini, dari sisi model, agar berhasil, membutuhkan data berlabel dalam jumlah besar dan daya komputasi yang besar, yang seringkali membutuhkan GPU berkinerja tinggi untuk menjalankan pelatihan model.



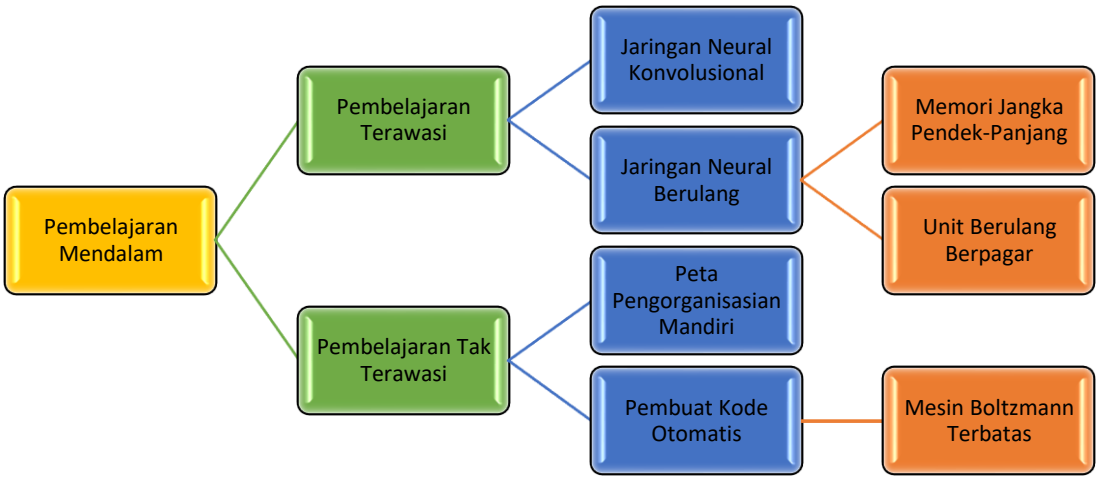
Gambar 19.7 Perbandingan antara jaringan saraf tiruan sederhana dan jaringan pembelajaran mendalam.

Beberapa bidang menggunakan pembelajaran mendalam (DL), tetapi telah mencapai keberhasilan tertentu dalam mengemudi otonom. DL digunakan untuk mendeteksi objek dan orang secara otomatis dalam sistem ini. Bidang lain di mana DL banyak digunakan adalah bidang Dirgantara dan Pertahanan, misalnya satelit mendeteksi objek atau mengidentifikasi zona aman bagi pasukan. Teknik-teknik ini juga relevan dalam penerjemahan ucapan otomatis dan asisten rumah tangga.

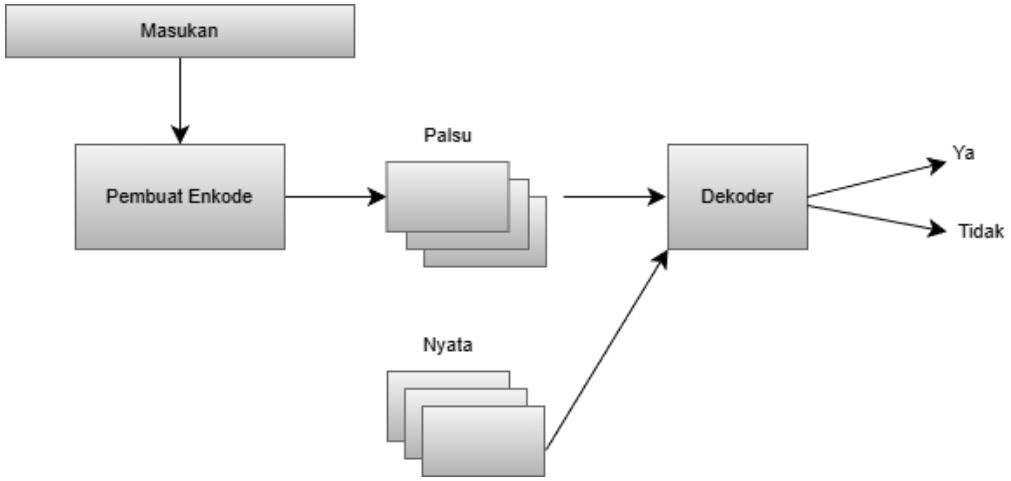
Sebagian besar model DL menggunakan apa yang disebut jaringan saraf tiruan, yang terinspirasi oleh koneksi neuron di otak manusia, dan dapat mengekstrak/mempelajari fitur secara otomatis. Misalnya, Anda dapat mengirimkan gambar ke jaringan, dan jaringan ini dapat mengekstrak fitur dari gambar tersebut tanpa campur tangan manusia. Saat jaringan ini dilatih, mereka secara otomatis belajar mengklasifikasikan/memecahkan masalah. Salah satu keuntungan utama yang mereka miliki adalah kemampuan untuk meningkatkan kapasitas klasifikasi mereka seiring bertambahnya data dan berjalannya waktu. Banyak aplikasi pembelajaran jarak jauh (DL) menggunakan teknik *Transfer Learning* (TF).

TF adalah prosedur yang memerlukan penyesuaian prototipe yang telah dilatih sebelumnya. Pelatihan jaringan dimulai dengan dataset yang sudah ada. Kemudian, jaringan ini diberikan data atau dataset baru yang berisi kelas-kelas yang tidak diketahui. Dari sana, jaringan menyesuaikan diri, mentransposisi pengetahuan sebelumnya ke situasi baru ini. Waktu komputasi lebih singkat dibandingkan dengan jaringan yang tidak terlatih.

Seperti ditunjukkan pada Gambar 19.8, model pembelajaran jarak jauh (DL) dapat diklasifikasikan sebagai pembelajaran terawasi (*supervised learning*) atau pembelajaran tak terawasi (*unsupervised learning*). Dalam pembelajaran jarak jauh (DL) terawasi, jaringan saraf tiruan konvolusional dan jaringan saraf tiruan rekuren (RNN) digunakan. RNN juga dapat dibagi menjadi *gated recurrent unit* (GRU) dan memori jangka pendek. Dalam pembelajaran jarak jauh (DL) tanpa pengawasan, terdapat jaringan *self-organizing map* (SOM) dan *autoencoder* (AE). AE adalah tipe mesin Boltzmann terbatas (RBM).



Gambar 19.8 Model Pembelajaran Mendalam



Gambar 19.9 Arsitektur GAN

19.6 PEMBUATAN DEEPFAKE

Jenis lain dari jaringan saraf tiruan dalam disebut Jaringan Adversarial Generatif (GAN). Jaringan ini dapat digunakan untuk membuat deepfake. GAN belajar dari set data pelatihan dan menghasilkan sampel data dengan fitur serupa. GAN memiliki arsitektur yang terdiri dari tidak lebih dari dua komponen jaringan saraf tiruan: enkoder dan dekoder. Model menggunakan enkoder untuk melatih set data yang luas guna menghasilkan data palsu. Dekoder pada dasarnya adalah pengklasifikasi biner yang menerima input (konten asli atau wajah) dan menggunakan fungsi SoftMax untuk mengidentifikasi data autentik (Gambar 19.9). Contoh aplikasi deepfake adalah VGGFace, FakeApp, Faceswap, dan CycleGAN.

19.7 DETEKSI DEEPFAKE

Deteksi deepfake merupakan teknologi yang memungkinkan, terutama dalam dua domain: gambar dan video.

Model Deteksi Citra

Literatur menyajikan beberapa proses untuk membedakan citra yang dihasilkan oleh GAN menggunakan jaringan dalam. Salah satu metode didasarkan pada prosedur

prapemrosesan untuk menganalisis karakteristik statistik citra dan meningkatkan pengenalan citra palsu yang dibuat oleh manusia. Metode lain berpusat pada jaringan saraf konvolusional dalam yang mengidentifikasi citra palsu yang dihasilkan oleh GAN. Xuan dkk. mengungkapkan jaringan saraf konvolusional (CNN) berbasis Gaussian Blur dan Gaussian Noise untuk mengidentifikasi citra manusia palsu. Sebuah metodologi hibrida dikembangkan untuk mengidentifikasi citra palsu secara efektif.

Model Deteksi Video

Model deteksi video menggunakan salah satu dari dua pendekatan: analisis fitur biologis atau spasial temporal. Pendekatan pertama dapat diterapkan pada tiga metode berbeda: menangkap wajah palsu, pengatur waktu untuk deepfake, dan hubungan antara audio dan video. Metode pertama didasarkan pada aspek fisik kedipan mata. Dengan demikian, metode ini memantau kedipan mata untuk mendeteksi wajah palsu. CNN dengan RNN dan pengklasifikasi biner digunakan untuk mengawasi kedipan mata. Metode kedua didasarkan pada aspek fisik pengatur waktu (pulsasi) untuk mendeteksi video palsu. Metode ini menggunakan GAN untuk membandingkan video palsu dengan video asli. Terakhir, metode ketiga didasarkan pada hubungan antara audio dan video. Metode ini menggunakan model DL dengan fungsi triple loss yang mendeteksi video palsu dari video asli.

Pendekatan kedua adalah menggunakan jaringan saraf tiruan konvolusional (CNN) untuk mengekstrak fitur dari sebuah frame. Kemudian, fitur-fitur ini dapat dilewatkan melalui LSTM yang menganalisis urutan temporal dalam frame. Akhirnya, video diklasifikasikan sebagai asli atau palsu dengan fungsi Softmax. Pendekatan lain adalah Recycle-GAN, yang menggunakan jaringan adversarial generatif dependen untuk menggabungkan data spasial dan temporal. Hasil evaluasi menunjukkan bahwa informasi spasial dan temporal dapat memberikan hasil yang baik dalam mendeteksi deepfake.

19.8 DEEPFAKE DAN ADMINISTRASI PERADILAN PIDANA

Deepfake dapat memiliki dampak negatif yang mendalam dengan menargetkan reputasi individu, menciptakan peristiwa atau konten palsu yang dapat mengakibatkan hukuman yang salah atau tuntutan hukum, mengurangi kepercayaan terhadap lembaga; dan mengancam keamanan nasional atau merusak hubungan internasional jika disalahgunakan oleh pemerintah.

Dari perspektif administrasi peradilan pidana, deepfake dapat sangat berbahaya karena menciptakan kekaburan antara yang benar dan yang salah. Video, gambar, audio, atau dokumen yang dimanipulasi dapat disajikan sebagai bukti dan menipu hakim, pengacara, dan petugas polisi, sehingga "menimbulkan keraguan pada bukti audio-visual sebagai satu kategori bukti utuh". Selain itu, seiring meningkatnya daya pemrosesan sistem informasi dan perkembangan teknologi deepfake yang lebih jauh, terutama dalam aplikasi waktu nyata, konferensi video atau telekonferensi saksi, ahli, atau pihak-pihak dapat dimanipulasi dengan membuat mereka terdengar, melakukan, dan bertindak dengan cara yang tidak pernah terjadi.

Dua tahun lalu, kasus pertama deepfake yang diajukan sebagai bukti di pengadilan Inggris menjadi berita utama. Dalam kasus hak asuh anak, ibu dari anak tersebut menunjukkan

rekaman audio yang berisi ancaman dari ayah anak tersebut. Keaslian rekaman tersebut kemudian dipertanyakan dan akhirnya dibatalkan oleh pengadilan. Menurut para ahli yang memeriksa berkas audio tersebut, rekaman tersebut telah dimanipulasi untuk memasukkan kata-kata yang tidak pernah diucapkan oleh sang ayah.

Sikap tradisional pengadilan yang menerima begitu saja jenis bukti tertentu video atau audio, misalnya mungkin telah mencapai titik akhirnya. Pertanyaan utamanya adalah bagaimana pengadilan dan masyarakat secara umum dapat melindungi diri mereka sendiri secara memadai dari dampak negatif teknologi deepfake.

Usulan untuk Peraturan Parlemen Eropa dan Dewan yang menetapkan regulasi terkoordinasi tentang kecerdasan buatan (Undang-Undang Kecerdasan Buatan) memberi kita gambaran tentang apa yang mungkin menjadi pandangan Uni Eropa dalam masalah ini.

Yang termasuk dalam pokok bahasannya (pasal 1(d)) adalah aturan untuk sistem AI yang digunakan untuk "menghasilkan atau memanipulasi konten gambar, audio, atau video". Sistem AI yang didefinisikan dalam pasal 3(1) "sebagai perangkat lunak yang dikembangkan dengan satu atau lebih teknik dan pendekatan yang tercantum dalam Lampiran I dan dapat, untuk serangkaian tujuan yang ditentukan manusia, menghasilkan keluaran seperti konten, prediksi, rekomendasi, atau keputusan yang memengaruhi lingkungan tempat mereka berinteraksi".

Di antara teknik dan pendekatan yang dirujuk, kita menemukan pembelajaran mesin dan pembelajaran mendalam (lampiran I(a)). Kerangka kerja yang diusulkan ini mengadopsi pendekatan berbasis risiko terhadap kecerdasan buatan yang membedakan empat tingkat risiko: tidak dapat diterima, tinggi, terbatas, dan minimal. Untuk lebih memahami tingkat risiko yang ditimbulkan oleh deep fake, perlu dijelaskan secara singkat masing-masing tingkat risiko tersebut.

Pertama, risiko yang tidak dapat diterima mencakup penggunaan AI yang dilarang oleh Undang-Undang Kecerdasan Buatan, karena tidak mematuhi nilai-nilai Uni Eropa (misalnya hak asasi manusia). Penggunaan AI untuk memanipulasi perilaku manusia yang menyebabkan kerugian fisik atau psikologis (Pasal 5, 1(a)(b)), penilaian sosial (Pasal 5, 1(c)), dan identifikasi biometrik jarak jauh secara real-time di ruang publik untuk tujuan penegakan hukum (Pasal 5, 1(d)) dilarang. Identifikasi biometrik jarak jauh secara real-time hanya dapat dilakukan dalam kasus-kasus tertentu, terutama jika proporsional dan sangat diperlukan untuk mencapai salah satu dari tiga tujuan: pencarian korban yang terarah, menghentikan serangan teroris atau ancaman langsung terhadap nyawa dan keselamatan fisik, atau melacak tersangka atau pelaku kejahatan berat (Pasal 5, 1(d) i,ii,iii).

Pengkategorian sistem AI sebagai berisiko tinggi berarti sistem tersebut diizinkan untuk digunakan, sejauh persyaratan yang tercantum dalam pasal 8 hingga 15 terpenuhi (misalnya dokumentasi teknis terkini, kemampuan pencatatan, transparansi yang memadai, pengawasan manusia, tingkat akurasi, ketahanan, dan keamanan siber yang memadai) dan penyedia, pengguna, serta pihak lain mematuhi kewajiban yang tercantum dalam pasal 16 hingga 29, khususnya penilaian kesesuaian. Legislator Eropa memasukkan sistem AI dalam domain

identifikasi biometrik, perangkat rekrutmen, skor kredit, manajemen layanan darurat, dan lain-lain ke dalam kategori risiko ini.

Meskipun proposal Eropa menyebutkan pendekatan berbasis risiko tiga tingkat (tidak dapat diterima, tinggi, dan rendah atau minimal), tingkat keempat umumnya diakui: risiko terbatas. Sistem AI "tertentu" mengingat ungkapan yang digunakan dalam proposal tunduk pada kewajiban transparansi (pasal 52). Artinya, penggunaannya diizinkan, tetapi penyedia dan pengguna harus memberi tahu pengguna akhir bahwa mereka berinteraksi dengan sistem AI atau konten yang dihasilkan secara artifisial. Cakupan sistem AI yang tercakup dalam tingkat risiko ini cukup luas karena mencakup semua sistem AI yang berinteraksi dengan orang perseorangan, mendeteksi emosi, mengkategorikan berdasarkan data biometrik, atau menghasilkan deepfake.

Sistem AI yang tidak menimbulkan risiko atau berisiko rendah termasuk dalam kategori risiko terakhir. Tidak ada kewajiban yang dibebankan kepada penyedia atau pengguna sistem AI tersebut. Namun demikian, penyedia dapat memutuskan, secara sukarela, untuk membuat kode etik dan memastikan kepatuhan terhadap persyaratan yang ditetapkan untuk sistem berisiko tinggi (pasal 69). Namun, penerapan persyaratan ini merupakan keputusan penyedia sistem AI non-berisiko tinggi, karena persyaratan tersebut hanya wajib untuk sistem berisiko tinggi.

Berdasarkan klasifikasi risiko yang telah disebutkan sebelumnya, kita dapat dengan yakin menegaskan bahwa penggunaan deepfake secara tegas termasuk dalam kategori ketiga: risiko terbatas. Dengan demikian, deepfake tidak dilarang atau dianggap berisiko tinggi, tetapi juga tidak diklasifikasikan sebagai risiko rendah atau minimum. Oleh karena itu, pengguna "yang menggunakan sistem AI untuk menghasilkan atau memanipulasi konten gambar, audio, atau video yang sangat mirip dengan orang, tempat, atau peristiwa yang ada dan secara keliru tampak autentik bagi seseorang, harus mengungkapkan bahwa konten tersebut telah dibuat atau dimanipulasi secara artifisial dengan memberi label pada keluaran kecerdasan buatan tersebut dan mengungkapkan asal buatannya" (pertimbangan 70 proposal). Oleh karena itu, pengguna sistem AI yang menghasilkan deep fake harus mengungkapkan asal buatan dan sifat konten yang dihasilkan (pasal 52(3)). Namun, kewajiban transparansi memiliki beberapa pengecualian.

Anggota parlemen Eropa tersebut berupaya mencapai keseimbangan antara perlindungan nilai dan tujuan yang sangat berbeda: pengembangan, pemasaran, dan penggunaan sistem AI (pertimbangan 1); pertumbuhan ekonomi dan pembangunan sosial (pertimbangan 2 dan 3); martabat manusia, kesehatan, keselamatan, kebebasan, kesetaraan, demokrasi, supremasi hukum, hak atas non-diskriminasi, perlindungan data pribadi dan privasi, serta hak-hak anak (pertimbangan 1 dan 15).

Dengan demikian, proposal tersebut memungkinkan penghapusan kewajiban transparansi jika deep fake digunakan untuk alasan yang sah. Pasal 52(3) secara khusus menyatakan bahwa asal-usul buatan deep fake tidak boleh diungkapkan jika digunakan untuk tujuan mendeteksi, mencegah, menyelidiki, dan menuntut tindak pidana atau merupakan

pelaksanaan yang sah atas hak atas kebebasan berekspresi dan hak atas kebebasan seni dan sains.

Memahami apakah deep fake harus diumumkan atau ditandai demikian, atau bahkan apakah konten tersebut legal, memerlukan analisis yang dilakukan kasus per kasus. Mungkin ada kasus di mana jenis konten buatan ini merupakan pelanggaran hak asasi pihak ketiga yang ditetapkan sebagai tindak pidana (misalnya pencemaran nama baik, pemerasan, pornografi anak). Namun, dalam kasus lain, konten ini mungkin hanya merupakan ekspresi kreativitas. Batasan kebebasan berekspresi, seni, dan sains agak sulit didefinisikan, dan mengekang risiko manipulasi massal dengan deep fake mungkin terbukti menjadi tantangan yang terlalu sulit diatasi oleh hukum. Alat kendali atau identifikasi konten teknologi dapat menjadi alat yang berharga bagi pemerintah, bisnis, dan individu.

19.9 TANTANGAN PENEGAKAN HUKUM TERHADAP DEEPFAKE DAN AI SEBAGAI SOLUSI

Masalah besar yang harus dipecahkan oleh administrasi peradilan pidana adalah penegakan hukum. Kerangka hukum saat ini tidak mampu menangani deep fake. Masalah penerapan aturan hukum yang ada dalam kasus deep fake dapat diringkas sebagai berikut. Pertama, evolusi teknologi yang pesat membuat norma hukum apa pun cepat usang. Kedua, ada kebutuhan mendesak untuk mendefinisikan apa dan bagaimana teknologi harus digunakan. Ketiga, karena sifat lintas batas teknologi, mungkin sangat rumit untuk mengidentifikasi aturan yang harus dipatuhi oleh teknologi ini; oleh karena itu, teknologi ini biasanya terdaftar di negara-negara dengan aturan yang lebih longgar. Keempat, sulit untuk menegakkan undang-undang ketika pengembangan dan penggunaan teknologi tidak terbatas pada satu negara saja. Kelima, tugas pihak-pihak yang berkepentingan dengan deepfake seringkali bersifat parsial. Keenam, aturan yurisdiksi tertentu dapat dengan mudah dielakkan.

AI harus membantu mengidentifikasi dan menghapus deepfake yang bermasalah. Sebagaimana ML dan DL menimbulkan masalah, keduanya juga harus menjadi bagian dari solusi. Namun demikian, literasi digital tidak boleh diabaikan, karena kecepatan solusi teknologi yang digunakan dalam pembuatan deepfake seringkali lebih cepat daripada solusi deteksinya. Tidak ada pengganti yang nyata untuk sikap kritis terhadap konten digital.

BAB 20
HUKUM ANTIMONOPOLI DAN AI DALAM PENETAPAN HARGA

Abstrak Harga merupakan elemen inti dari transaksi komersial dan parameter penting persaingan. Salah satu tujuan hukum antimonopoli adalah memastikan bahwa harga pasar terbentuk berdasarkan hukum penawaran dan permintaan, dan bukan berdasarkan keinginan pelaku monopoli atau kartel. Inovasi dalam ilmu komputer dan data telah menghasilkan teknologi penetapan harga yang mengandalkan analitik canggih atau teknik pembelajaran mesin (ML), yang dapat memperkuat disparitas daya tawar yang ada, sebagian dengan mendukung koordinasi harga antar pesaing.

Penelitian yang ada menetapkan kerangka kerja teoretis untuk kerugian persaingan melalui koordinasi, menunjukkan bahwa teknologi penetapan harga dapat mengarah pada tingkat harga yang mendekati kartel sekaligus menghindari larangan anti-kartel. Kontribusi ini dibangun berdasarkan kerangka kerja tersebut, dengan mempertimbangkan literatur empiris, teori permainan, dan ilmu komputer terkini tentang teknologi penetapan harga untuk menghasilkan taksonomi teknologi-teknologi tersebut.

Kami kemudian menggunakan pendekatan komparatif untuk mengidentifikasi dampak hukum dari berbagai teknologi penetapan harga pada tingkat yang lebih terperinci berdasarkan hukum antimonopoli Uni Eropa dan AS. Kontribusi ini mendukung pemahaman yang lebih baik antara para ekonom dan pembuat kebijakan mengenai analisis dan penanganan teknologi penetapan harga berbasis AI.

20.1 KOORDINASI HARGA ALGORITMIK DAN TANTANGAN HUKUM ANTIMONOPOLI

Harga merupakan elemen inti dari transaksi komersial dan parameter penting persaingan. Salah satu tujuan hukum antimonopoli adalah memastikan terbentuknya harga pasar dalam persaingan bebas, sehingga hukum penawaran dan permintaan berlaku untuk memaksimalkan kesejahteraan konsumen. Salah satu contoh di mana harga tidak dihasilkan dari persaingan adalah ketika terdapat koordinasi anti-persaingan di pasar. Koordinasi anti-persaingan dilarang berdasarkan berbagai ketentuan hukum antimonopoli, terutama berdasarkan larangan kartel (misalnya Pasal 1 Undang-Undang Sherman di AS dan Pasal 101 Perjanjian tentang Fungsi Uni Eropa) dan berdasarkan aturan merger (di mana merger yang menghasilkan dampak terkoordinasi harus dilarang).

Pendapat yang berlaku dalam kajian antimonopoli dan praktik penegakan hukum adalah bahwa, kecuali terdapat koordinasi yang eksplisit antar pesaing, koordinasi kecil kemungkinannya terjadi di pasar dengan empat pemain atau lebih. Ketika koordinasi tidak eksplisit melainkan diam-diam, hal tersebut jarang tercakup dalam aturan antimonopoli. Berdasarkan kondisi hukum saat ini di AS dan Uni Eropa, koordinasi diam-diam menurut definisi tidak tercakup dalam larangan kartel, karena aturan yang ada mensyaratkan adanya bukti adanya kesepakatan antara peserta kartel untuk memicu penerapan larangan tersebut. Untuk menunjukkan adanya kesepakatan, secara tradisional, diperlukan bukti tentang

komunikasi antara peserta kartel, catatan pertemuan, atau pertukaran informasi melalui pihak ketiga.

Meskipun terdapat alasan untuk meyakini bahwa pembatasan penegakan hukum terhadap hasil terkoordinasi ini memungkinkan terjadinya dampak harga dan non-harga yang merugikan konsumen, hingga saat ini belum ada pendekatan yang diterima secara luas untuk mencegah hasil ini sebagai bentuk penegakan perilaku. Meningkatnya kemungkinan terjadinya koordinasi diam-diam akibat merger berpotensi tercakup dalam hukum merger, tetapi penegak persaingan jarang melarang merger karena potensi dampak terkoordinasinya. Dalam kasus-kasus di mana mereka mencoba memblokir merger semacam itu, mereka seringkali gagal karena standar hukum yang tinggi untuk membuktikan koordinasi diam-diam yang diberlakukan oleh pengadilan.

Keyakinan konvensional bahwa koordinasi anti-persaingan dalam bentuk kolusi diam-diam tidak mungkin terjadi sedang ditantang oleh studi terbaru dalam teori permainan dan perkembangan teknologi yang pesat. Inovasi dalam ilmu komputer dan data telah menghasilkan teknologi penetapan harga yang mengandalkan analitik canggih atau teknik pembelajaran mesin (ML), yang dianggap mampu belajar mandiri dan mempertahankan koordinasi harga di antara para pesaing. Belum pasti apakah undang-undang antimonopoli akan menangkap koordinasi harga algoritmik semacam itu.

Hal itu bergantung pada (a) kemampuan teknologi yang digunakan dan potensinya untuk menyebabkan kerugian kompetitif, dan (b) interpretasi hukum antimonopoli. Bab ini bertujuan untuk memberikan kejelasan lebih lanjut pada kedua dimensi ini untuk berkontribusi pada perdebatan yang sedang berlangsung tentang hukum antimonopoli dan teknologi penetapan harga berbasis AI.

Kami mulai dengan memaparkan perkembangan terkini teknologi penetapan harga dan kolusi diam-diam algoritmik dalam literatur hukum, ekonomi mikro, dan ilmu komputer. Selanjutnya, kami menggunakan beragam teknologi berbasis AI yang dapat diterapkan untuk penetapan harga dinamis. Kami kemudian mengkaji berbagai opsi teknologi dari perspektif antimonopoli Uni Eropa dan AS sebelum menguraikan pendekatan kebijakan dan prospek masa depan.

20.2 PENETAPAN HARGA ALGORITMIK DAN KOLUSI

Penetapan harga merupakan bagian inti dari strategi bersaing perusahaan. Dalam situasi di mana harga dapat disesuaikan dengan mudah dan sering, seperti misalnya dalam kasus ritel daring atau ritel alat tulis dengan label harga elektronik, perusahaan dapat mengoptimalkan penetapan harga mereka dengan menggunakan teknik penetapan harga dinamis. Penetapan harga dinamis memungkinkan perusahaan untuk bereaksi cepat terhadap perubahan permintaan atau tingkat persediaan, dan untuk merespons perilaku pesaing dan tren pasar.

Semakin banyak data yang tersedia bersama dengan algoritma dapat mendukung dan meningkatkan penetapan harga dinamis. Algoritma sederhana dapat menerapkan aturan penetapan harga "analog" yang sudah ada sebelumnya, seperti menyamai harga terendah

pesaing. Dalam hal ini, penggunaan algoritma memungkinkan otomatisasi penetapan harga dan reaksi yang lebih cepat terhadap perubahan harga pesaing. Sebagai alternatif, algoritma pembelajaran mesin yang lebih canggih dapat secara mandiri memilih data mana yang akan dijadikan dasar keputusan penetapan harga dan dapat mempelajari strategi penetapan harga yang optimal.

Dampak algoritma penetapan harga terhadap persaingan pasar masih ambigu. Di satu sisi, algoritma penetapan harga dapat memiliki efek pro-persaingan, seperti memungkinkan manajer membuat keputusan yang lebih cepat dan lebih baik atau secara langsung mengurangi biaya dengan mengotomatiskan pengambilan keputusan sepenuhnya. Selain itu, algoritma penetapan harga dapat berkontribusi pada penyelesaian pasar yang lebih cepat dan lebih efisien.

Di sisi lain, muncul kekhawatiran bahwa algoritma penetapan harga dapat memfasilitasi koordinasi perilaku pasar yang eksplisit, implisit, atau bahkan tidak disengaja, yang mengarah pada hasil pasar yang kolusif dan kerugian konsumen. Literatur akademis dalam hukum dan kebijakan antimonopoli terbagi mengenai risiko persaingan dari penetapan harga algoritmik perusahaan. Beberapa penulis berpendapat bahwa pasar sedang ditransformasikan secara fundamental oleh penggunaan algoritma, dan bahwa risiko kolusi algoritmik yang meluas adalah nyata.

Ezrachi dan Stucke mengusulkan empat skenario di mana penggunaan algoritma dalam penetapan harga dapat menyebabkan hasil yang kolusif. Pertama, dalam situasi di mana algoritma digunakan untuk mendukung perjanjian penetapan harga eksplisit dengan menerapkan perilaku yang disepakati atau mendeteksi penyimpangan apa pun darinya. Contohnya adalah kartel penjual poster Amazon yang dituntut di AS dan Inggris. Dalam kasus ini, para pesaing di Amazon mencapai kesepakatan tegas mengenai fitur algoritma yang akan digunakan dalam menetapkan harga di Amazon, dengan efek koordinasi harga yang tidak akan terjadi tanpa adanya perjanjian.

Sebagaimana dijelaskan oleh Otoritas Persaingan dan Pasar Inggris, perjanjian penetapan harga lisan awal kemudian diimplementasikan menggunakan perangkat lunak penetapan harga. Hal ini bertentangan dengan hukum antimonopoli. Kedua, dalam situasi hub-and-spoke di mana banyak pelaku pasar mendelegasikan keputusan penetapan harga mereka kepada algoritma yang sama dari pengembang pihak ketiga, Ezrachi dan Stucke berpendapat bahwa terdapat risiko penyelarasan harga yang diarahkan oleh hub. Ketiga, dalam situasi di mana perusahaan secara sepihak menerapkan algoritma sederhana dengan tindakan yang dapat diprediksi, misalnya algoritma yang menerapkan aturan 'menang-lanjutkan-kalah-balik' atau algoritma yang menyesuaikan harga pesaing, penyelarasan harga dapat terjadi.

Terakhir, ketika algoritma pembelajaran mesin canggih dirancang oleh perusahaan untuk mencapai suatu tujuan, seperti memaksimalkan keuntungan, dan dibiarkan sendiri untuk menentukan strategi penetapan harga yang optimal (skema agen otonom), mereka secara otonom dapat belajar untuk berkolusi. Khususnya, dalam kasus skema agen otonom,

literatur ini menyatakan bahwa undang-undang antimonopoli yang berlaku saat ini tidak cocok untuk memberikan sanksi atas perilaku semacam ini.

Pandangan sebaliknya berpendapat bahwa klaim tentang risiko penetapan harga algoritmik dalam perdebatan tentang AI dan hukum antimonopoli adalah "fiksi ilmiah". Aliran literatur ini khususnya menganggap kolusi algoritmik otonom sebagai sesuatu yang jauh dari kemampuan teknologi saat ini dan menunjukkan fakta bahwa belum ada bukti empiris kolusi otonom yang disajikan sejauh ini. Selain itu, aliran literatur ini berpendapat bahwa bentuk-bentuk kolusi algoritmik yang memungkinkan saat ini tidak menimbulkan tantangan mendasar apa pun terhadap aturan antimonopoli yang ada.

Beberapa kritik mungkin dipertanyakan karena peningkatan teknologi yang berkelanjutan meningkatkan kemungkinan hasil yang diteorikan. Lebih lanjut, garis antara kolusi algoritmik yang dibantu manusia dan kolusi algoritmik otonom mungkin tidak selalu mudah ditarik, sehingga menyisakan lebih banyak area abu-abu untuk penanganan kolusi algoritmik dalam hukum antimonopoli saat ini daripada yang diyakini para kritikus.

Literatur dalam ekonomi mikro menggaungkan perpecahan ini. Bagi para ekonom, kolusi didefinisikan sebagai situasi ketika perusahaan menggunakan strategi yang menerapkan "skema penghargaan-hukuman yang dirancang untuk memberikan insentif bagi perusahaan agar secara konsisten menetapkan harga di atas tingkat persaingan".

Pertanyaan tentang apakah algoritma dapat belajar untuk terlibat dalam kolusi dan untuk "menghukum" penyimpangan dari perilaku kolusi telah dibahas baik secara teoritis, empiris, maupun dalam simulasi. Agar kolusi terjadi di pasar tanpa algoritma, kondisi struktural tertentu dapat memfasilitasi kolusi, seperti ketika ada keuntungan yang bisa diperoleh dari kolusi, terdapat sedikit pesaing, pemain di pasar simetris, hambatan masuk tinggi, terdapat interaksi berulang antar pelaku pasar, terdapat homogenitas sisi penjual, dan terdapat tingkat transparansi yang tinggi yang memungkinkan para pesaing untuk mengamati perilaku satu sama lain.

Kondisi struktural yang sama ini juga mendukung koordinasi non-kolusif, yang terkadang disebut kolusi diam-diam, koordinasi oligopolistik, atau perilaku saling bergantung, yang belum dianggap melanggar hukum AS (Bagian 1 Undang-Undang Sherman) maupun hukum Uni Eropa (Pasal 101 TFEU).

Analisis awal mengenai masalah ini diajukan oleh Donald Turner, yang menyimpulkan bahwa koordinasi diam-diam tidak boleh dianggap sebagai suatu kesepakatan karena hal ini semata-mata mencerminkan perilaku oligopoli rasional, yang bertindak demi kepentingan terbaiknya sendiri secara independen. Secara khusus, tidak ada perilaku yang dapat dikutuk dengan cara memperbaiki hasilnya, kecuali perintah yang beroperasi sebagai pengendalian harga. Yang penting, Turner mengidentifikasi sebagai tidak terpengaruh oleh kesimpulannya perilaku "yang dirancang untuk mengubah pola penetapan harga oligopoli yang tidak sempurna menjadi pola yang sempurna dengan menghilangkan ketidakpastian".

Richard Posner kemudian menganjurkan pendekatan alternatif untuk masalah di mana oligopolis "mendasarkan keputusan penetapan harga mereka sebagian pada reaksi yang diantisipasi terhadapnya" dengan "kecenderungan untuk menghindari persaingan harga yang

kuat". Posner mempermasalahkan saran Turner bahwa seorang oligopolis mungkin hanya menemukan dirinya di pasar di mana aktivitas memaksimalkan keuntungan adalah untuk mempertimbangkan respons pesaingnya, yang mengarah pada hasil yang terkoordinasi. Sebaliknya, Posner mengidentifikasi sejumlah fitur yang dibutuhkan untuk koordinasi oligopoli yang sukses, termasuk kesamaan substansial produk dan pengetahuan yang dapat diandalkan tentang harga perusahaan lain; karena fitur-fitur tersebut dan fitur lainnya muncul melalui tindakan sukarela, maka adalah mungkin untuk memperbaiki harga oligopoli dengan melarang tindakan-tindakan tersebut.

Penerapan algoritma penetapan harga dianggap mampu menghasilkan atau memperbesar beberapa kondisi yang mengarah pada hasil yang terkoordinasi, termasuk: konsentrasi pasar, kecepatan, efisiensi, transparansi, dan stabilitas. Gal, misalnya, berpendapat bahwa kecepatan algoritma memungkinkan algoritma ini untuk mendeteksi dan menghukum penyimpangan dari penetapan harga yang sangat kompetitif lebih cepat daripada manusia, sehingga membuat kolusi lebih stabil.

Algoritma penetapan harga juga memproses data dalam jumlah yang jauh lebih tinggi daripada manusia, sehingga membuatnya lebih efisien dalam mengejar strategi kolusi. Selain itu, argumen telah dibuat bahwa karena strategi algoritma "dikodekan", akan lebih mudah bagi algoritma untuk "membaca" strategi algoritma lain sehingga meningkatkan transparansi pasar.

Selain itu, kapasitas prediksi algoritma berdasarkan data pasar masa lalu dapat meningkatkan transparansi dalam hal memprediksi permintaan masa depan dengan lebih akurat. Terakhir, jika perusahaan perlu mengandalkan algoritma penetapan harga agar dapat memperoleh keunggulan kompetitif di pasar, hal ini dapat meningkatkan konsentrasi pasar, karena pengembangan atau pembelian algoritma penetapan harga yang canggih membutuhkan biaya mahal dan hanya pemain lama dan pemain besar yang memiliki sumber daya untuk bersaing.

Meskipun terdapat banyak bukti mengenai perusahaan yang menggunakan algoritma penetapan harga, hanya terdapat sedikit bukti yang menunjukkan hubungan kausal antara algoritma penetapan harga dan hasil kolusi. Namun, sebuah studi terbaru dari pasar bensin Jerman menunjukkan bahwa setelah SPBU mengadopsi teknologi penetapan harga algoritmik, kenaikan harga secara bertahap menyebabkan kenaikan harga permanen yang mengindikasikan penurunan persaingan. Dalam studi ini, penulis menunjukkan bahwa setelah perangkat lunak penetapan harga algoritmik tersedia dan diadopsi secara luas oleh SPBU Jerman, margin keuntungan di pasar non-monopoli dan di pasar duopoli di mana kedua SPBU mengadopsi penetapan harga algoritmik meningkat secara signifikan.

Karena sulitnya menemukan bukti nyata kolusi algoritmik diam-diam, dengan studi pasar bensin Jerman yang merupakan studi pertama dan satu-satunya hingga saat ini, beberapa studi telah menguji kapasitas algoritma pembelajaran untuk berkolusi dalam simulasi permainan berulang. Calvano dkk. menyediakan studi eksperimental pertama yang menunjukkan bahwa dua algoritma Q-learning dapat belajar berkolusi secara otonom dengan menerapkan skema imbalan-hukuman dalam kerangka pengaturan oligopoli Bertrand berulang. Hasilnya menunjukkan bahwa, setelah beberapa iterasi, strategi ekuilibrium

algoritma dalam permainan berulang menghasilkan tingkat harga supra-kompetitif yang berkelanjutan. Klein memberikan bukti lebih lanjut tentang kapasitas algoritma Q-learning untuk mempelajari cara berkolusi dalam situasi penetapan harga berurutan. Relevansi studi-studi ini untuk kebijakan, bagaimanapun, telah dipertanyakan karena proses pembelajaran algoritma Q-learning yang sangat lambat, yang membuat penggunaannya oleh bisnis tidak realistis.

Studi-studi ini juga telah dikritik karena sulit diperluas ke pengaturan lebih dari dua agen karena masalah dalam skalabilitas dan pembelajaran dalam lingkungan yang tidak stasioner di mana beberapa pesaing menyesuaikan perilaku mereka secara konstan. Terakhir, ketidakterapan studi-studi ini untuk skenario dengan barang-barang yang tidak homogen telah dikritik. Dengan kata lain, simulasi yang menunjukkan kemungkinan algoritma Q-learning belajar untuk berkolusi telah dikritik karena terlalu jauh dari kondisi di mana pasar nyata bekerja, sehingga mempertanyakan implikasi kebijakan dari temuan mereka.

Meskipun demikian, studi-studi tersebut secara teoritis menunjukkan bahwa dua algoritma yang diterapkan secara unilateral dan tidak dirancang untuk berkomunikasi pada akhirnya dapat belajar untuk berkolusi secara diam-diam dan menghasilkan hasil yang terkoordinasi. Selain itu, Otoritas Persaingan Swedia baru-baru ini menerbitkan laporan yang menguatkan bukti dari eksperimen sebelumnya dan menunjukkan kemungkinan hasil kolusi juga ketika berbagai algoritma pembelajaran digabungkan (Q-learning, SARSA, dan PG), dan ketika agen Q-learning diprogram secara asimetris.

Belum jelas sejauh mana temuan dari simulasi dengan algoritma Q-learning dapat diterapkan pada teknik pembelajaran mesin lainnya. Hettich memberikan bukti awal bahwa algoritma Q-learning mendalam (DQN) mempelajari cara berkolusi secara signifikan lebih cepat daripada algoritma Q-learning sederhana. Menurut studinya, kolusi menghilang dalam simulasi dengan lebih dari 7 perusahaan. Lebih lanjut, kolusi tidak terhambat ketika algoritma lain berupa algoritma sederhana berbasis aturan, tetapi terhambat jika terdapat berbagai algoritma pembelajaran dengan karakteristik yang berbeda (misalnya, perbedaan kecepatan pembelajaran). Penelitian lebih lanjut diperlukan untuk memahami dan mengukur risiko kolusi yang berasal dari DQN yang digunakan dalam penetapan harga.

Dalam ilmu komputer, ada literatur penting tentang kapasitas algoritma pembelajaran penguatan untuk bekerja sama dalam dilema tahanan berulang, yang sebanding dengan pengaturan dua perusahaan dalam oligopoli Bertrand. Literatur telah mengeksplorasi fokus pada hubungan antara pembelajaran, komunikasi, dan kerja sama. Dalam dua studi terbaru yang memodelkan situasi dilema tahanan, dua faktor penting untuk mempelajari kerja sama dengan algoritma pembelajaran mandiri diidentifikasi.

Pertama, kerja sama lebih mudah dicapai ketika tingkat koordinasi yang lebih rendah diperlukan untuk mencapai hadiah tertinggi. Kedua, kerja sama itu jauh lebih mudah dicapai jika algoritma dapat saling mengirim sinyal melalui protokol komunikasi. Menerapkan wawasan ini pada algoritma penetapan harga, literatur ini akan menunjukkan bahwa kolusi lebih mungkin terjadi semakin sedikit kompleksitas algoritma dan lingkungan tempat mereka

beroperasi. Selain itu, seperti halnya agen manusia, komunikasi antar algoritma akan meningkatkan kemungkinan kolusi.

Secara keseluruhan, ketidakpastian mengenai apakah algoritma penetapan harga dapat berkolusi dan tingkat kerugian yang dapat ditimbulkannya ketika menyelaraskan harga di pasar masih ada. Meskipun demikian, ketidakpastian ini kemungkinan besar tidak akan hilang sebelum otoritas persaingan atau pengadilan harus menangani kasus yang menimbulkan pertanyaan seputar legalitas algoritma penetapan harga yang bukan merupakan alat sederhana untuk menerapkan kartel seperti dalam kasus Amazon Poster yang dibahas di atas.

Lebih lanjut, literatur telah menunjukkan fitur-fitur dalam desain algoritma penetapan harga yang, setidaknya secara teori, seharusnya membuat kolusi algoritmik lebih atau kurang mungkin terjadi. Di bagian selanjutnya, kami membahas berbagai kemungkinan teknologi penetapan harga berbasis AI dan karakteristiknya. Selanjutnya, kami memberikan garis besar tentang cara menangani berbagai karakteristik algoritma penetapan harga berdasarkan aturan antimonopoli.

20.3 RAGAM TEKNOLOGI PENETAPAN HARGA BERBASIS AI

Penetapan harga dinamis dapat diimplementasikan dengan bantuan berbagai algoritma. Terdapat algoritma jika-maka yang sangat sederhana, atau agen berbasis aturan, tanpa kapasitas pembelajaran apa pun. Algoritma semacam itu, misalnya, dapat mengimplementasikan aturan bisnis sederhana dengan cara yang tidak sementara, yaitu tidak akan mengubah perilakunya dengan belajar dari interaksi sebelumnya.

Terdapat pengembang pihak ketiga yang menawarkan solusi penetapan harga sederhana semacam ini di pasaran, dan tampaknya merupakan bentuk yang paling umum digunakan saat ini. Algoritma yang lebih canggih yang mampu belajar dapat dikategorikan sebagai bentuk AI. Kecerdasan Buatan (AI) dapat memiliki banyak bentuk. Undang-Undang AI Uni Eropa yang baru-baru ini diusulkan mendefinisikan sistem kecerdasan buatan sebagai "perangkat lunak yang dikembangkan dengan satu atau lebih teknik dan pendekatan yang tercantum dalam Lampiran I dan dapat, untuk serangkaian tujuan yang ditentukan manusia, menghasilkan keluaran seperti konten, prediksi, rekomendasi, atau keputusan yang memengaruhi lingkungan tempat mereka berinteraksi". Lampiran I berisi tiga contoh sistem AI, yang terpenting bagi kami adalah pendekatan pembelajaran mesin.

Pembelajaran mesin mencakup berbagai program komputer yang belajar dari pengalaman. Bidang utama dalam pembelajaran mesin adalah pembelajaran terawasi, pembelajaran tanpa pengawasan, dan pembelajaran penguatan. Pembelajaran terawasi melibatkan pengajaran fungsi pemetaan pada algoritma dengan pasangan masukan-keluaran contoh yang memungkinkan algoritma tersebut pada akhirnya memprediksi hasil untuk data masukan baru secara otonom.

Pembelajaran terawasi dan semi-terawasi diterapkan di banyak bidang di mana pelatihan dilakukan dengan data berlabel, misalnya dalam kasus moderasi konten algoritmik di media sosial. Pembelajaran terawasi juga digunakan untuk regresi, dan, ketika diterapkan

dalam penetapan harga, dengan demikian dapat memungkinkan algoritma penetapan harga untuk memprediksi tren masa depan.

Di sisi lain, pembelajaran tanpa pengawasan mengharuskan algoritma hanya menerima data masukan tanpa variabel keluaran. Tugas pembelajaran algoritma kemudian adalah menemukan pola dalam data masukan. Aplikasi pembelajaran tanpa pengawasan mencakup pengenalan pola dalam data besar dan sistem rekomendasi. Pembelajaran tanpa pengawasan dapat membantu dalam pengambilan keputusan penetapan harga untuk menemukan korelasi baru atau untuk mendapatkan wawasan di seluruh pasar produk atau geografis. Pembelajaran tanpa pengawasan, misalnya, memungkinkan penjual kosmetik untuk mendapatkan wawasan baru tentang bagaimana peristiwa yang tampaknya tidak terkait, misalnya cuaca, memengaruhi permintaan kosmetik.

Terakhir, pembelajaran penguatan adalah teknik pembelajaran mesin yang telah diterapkan dalam simulasi permainan berulang yang dibahas di bagian sebelumnya. Pembelajaran penguatan melatih algoritma untuk membuat serangkaian keputusan dalam lingkungan yang kompleks untuk mencapai tujuan tertentu melalui uji coba. Pembelajaran Q adalah salah satu bentuk pembelajaran penguatan. Pembelajaran ini memungkinkan agen untuk menerapkan strategi memaksimalkan imbalan dalam lingkungan yang tidak diketahui dari waktu ke waktu. Agen belajar dengan memilih dan melakukan tindakan dalam keadaan tertentu, memperkirakan imbalan atas tindakan tersebut, dan memperbarui fungsinya berdasarkan imbalan positif atau negatif yang diterima. Proses ini kemudian diulang terus menerus hingga proses pembelajaran dihentikan. Dalam simulasi oleh Calvano dkk. dan Klein, proses ini memungkinkan algoritma untuk belajar berkolusi seiring waktu.

Bentuk pembelajaran struktural yang berbeda bergantung pada apa yang disebut jaringan saraf dalam (DNN). DNN terinspirasi oleh struktur saraf otak manusia, dan terdiri dari beberapa lapisan neuron buatan yang saling terhubung. Dalam DNN, pembelajaran dapat diawasi, tidak diawasi, atau diperkuat. DNN dapat belajar secara berbeda dan lebih cepat, dengan memodifikasi koneksi antar neuron buatan dan karena setiap lapisan dapat menjalankan tugas yang berbeda. DNN adalah teknologi AI yang mendasari pengenalan suara otomatis, pengenalan gambar, dan pemrosesan bahasa alami, dan juga dapat digunakan untuk memprediksi dan mengoptimalkan harga. Terkadang disebut sebagai "kotak hitam" karena proses yang menghasilkan keluaran sulit atau mustahil dipahami.

Alternatif untuk DNN adalah Hutan Acak yang menggabungkan prediksi dari banyak pohon keputusan. Dibandingkan dengan DNN, hutan acak membutuhkan lebih sedikit data dan secara komputasi lebih efisien. Hutan acak juga cenderung lebih transparan dibandingkan dengan DNN karena memungkinkan koneksi antara variabel dan model prediksi yang dihasilkannya.

Literatur telah menunjukkan bahwa berbagai karakteristik algoritma pembelajaran mesin dapat memengaruhi kapasitas untuk bekerja sama/berkolusi. Ini termasuk memori interaksi sebelumnya, kecepatan pembelajaran, kompleksitas algoritma (semakin kompleks, semakin kecil kemungkinannya untuk berkolusi karena perilaku sulit "dibaca" oleh algoritma lain), dan kapasitasnya untuk berkomunikasi.

20.4 KOLUSI ALGORITMIK DALAM HUKUM ANTIMONOPOLI AS

Keharusan "Kesepakatan"

Persyaratan inti untuk pelanggaran Pasal 1 Undang-Undang Sherman adalah adanya kesepakatan antara dua atau lebih pelaku ekonomi independen. Persyaratan kesepakatan juga terdapat dalam beberapa kasus berdasarkan Pasal 2, yang melarang "menggabungkan atau bersekongkol untuk memonopoli." (Karena tidak ada perbedaan yang signifikan antara konspirasi Pasal 2 dan konspirasi Pasal 1, analisis di sini tidak membahas keduanya secara terpisah.) Pertanyaan tentang kesepakatan bersifat biner:

Perilaku yang terjadi berdasarkan kesepakatan harus dianalisis legalitasnya berdasarkan kaidah akal sehat atau pendekatan analitis per se, tetapi perilaku yang tidak terjadi berdasarkan kesepakatan kebal terhadap gugatan berdasarkan Pasal 1, terlepas dari kemungkinan kerugian konsumen.

Koordinasi Non-Kesepakatan

Yang paling penting untuk analisis ini adalah kategori luas koordinasi non-kesepakatan yang termasuk dalam label "kolusi diam-diam", "koordinasi diam-diam", "perilaku saling bergantung", "paralelisme sadar", dan "koordinasi oligopolistik". Analisis ini menggunakan label "koordinasi non-kesepakatan", yang paling baik dipahami sebagai hasil yang terjadi ketika semua pesaing menyadari bahwa kepentingan terbaik independen mereka dicapai bukan melalui persaingan yang sengit, melainkan melalui koordinasi dengan pesaing. Sebuah contoh, yang sering dikutip, adalah pemilik SPBU yang bersaing yang menetapkan harga secara paralel meskipun tidak ada kesepakatan implisit maupun eksplisit di antara mereka.

Mahkamah Agung AS telah berulang kali mengakui kenyataan ini, dalam kasus-kasus termasuk *Bell Atlantic Corp v. Twombly*, dan menyatakan bahwa koordinasi antar pelaku ekonomi yang disebabkan oleh upaya mereka untuk mengejar kepentingan terbaik independen bukanlah suatu kesepakatan. Dua alasan utama mendukung pendekatan yang memperlakukan koordinasi non-kesepakatan sebagai sesuatu yang berada di luar larangan Pasal 1. Pertama, adalah masalah interpretasi undang-undang yang sederhana dan formalistik: Pasal 1 tidak melarang perilaku perusahaan yang mengejar kepentingan terbaik mereka sendiri.

Kedua, masalah penyelesaian koordinasi non-kesepakatan. Untuk melarang perilaku oligopoli, pengadilan atau regulator diwajibkan untuk memerintahkan pelaku ekonomi agar tidak mengejar kepentingan terbaik mereka sendiri, dengan misalnya tidak memperhitungkan dampak terhadap pesaing (dan kemungkinan respons mereka) terhadap keputusan penetapan harga. Aturan yang melarang penetapan harga oligopoli secara efektif akan menjadi persyaratan perilaku penetapan harga yang naif atau irasional.

Hukum AS telah berkembang dalam serangkaian kasus untuk sepenuhnya mengecualikan koordinasi non-kesepakatan dari jangkauan Undang-Undang Sherman. Hylton berpendapat bahwa pendapat Hakim Posner tahun 2015 dalam *In re Text Messaging Antitrust Litigation* merupakan pernyataan paling definitif dari prinsip ini sebagai salah satu contoh luar biasa dari pendapat tersebut: “The Sherman Act tidak membebankan kewajiban kepada

perusahaan untuk bersaing secara agresif, atau dalam hal ini sama sekali, dalam hal harga." Khususnya, *Text Messaging* menunjukkan perubahan haluan yang sepenuhnya dari pandangan Posner bahwa hukum mungkin mengambil langkah-langkah untuk menantang kolusi diam-diam.

Pengadilan telah mengembangkan sistem yang mewajibkan bukti tambahan, di luar perilaku paralel yang diamati, untuk membedakan koordinasi persetujuan dari non-persetujuan. Kerangka kerja "faktor plus" ini pada dasarnya merupakan beban pembuktian atau pembelaan, yang dimaksudkan untuk menunjukkan perilaku yang tunduk pada pertanggungjawaban berbeda dari perilaku yang tidak bersalah, alih-alih aturan hukum substantif. Sebagaimana diuraikan oleh Kovacic dkk. Faktor plus adalah fakta pembuktian yang mengurangi kemungkinan penjelasan ketidaksepakatan untuk perilaku terkoordinasi.

Jika terdapat faktor plus, bukti mungkin cukup untuk pembuktian tidak langsung atas kesepakatan, yang dapat diatasi dengan larangan dan sanksi. Namun, jika tidak ada faktor plus, tidak ada dasar untuk mengasumsikan adanya kesepakatan yang sebenarnya dari pengamatan koordinasi antar pesaing, sehingga koordinasi ketidaksepakatan yang tidak dapat diatasi menjadi kesimpulan yang diperlukan secara hukum.

Hub-and-Spoke

Kesepakatan juga dapat dicapai melalui kesepakatan paralel dengan pihak ketiga, yang bertindak sebagai "hub" dalam struktur perusahaan hub-and-spoke. Analisis perusahaan hub-and-spoke menjadi menarik jika memungkinkan untuk membuktikan adanya kesepakatan di sekitar "rim", yang merupakan garis horizontal di antara berbagai spoke. Kasus-kasus terbaru di pengadilan AS, termasuk putusan Pengadilan Banding Sirkuit Ketujuh dalam kasus Toys 'R' Us dan putusan Pengadilan Banding Sirkuit Kedua dalam kasus Apple e-Books, menemukan adanya kesepakatan horizontal di antara para pesaing ketika kesepakatan vertikal individual dengan hub dicapai dengan pemahaman, dan ketergantungan pada, kesepakatan yang sebanding dari para pesaing.

Perusahaan hub-and-spoke memiliki karakteristik yang sama dengan koordinasi non-kesepakatan, karena perusahaan-perusahaan ini tidak bergantung pada komunikasi aktual antar spoke. Namun, tidak seperti koordinasi non-kesepakatan, hub-and-spoke dapat diperbaiki melalui pelarangan berbagai perjanjian vertikal.

Memperbaiki Perilaku Terkoordinasi

Salah satu pendekatan untuk memperbaiki koordinasi non-kesepakatan adalah dengan menyerangnya secara tidak langsung, melalui penentangan terhadap perjanjian atau perilaku lain yang meningkatkan kemungkinan terjadinya koordinasi. Yang kedua adalah menyerang koordinasi secara langsung berdasarkan Pasal 5 Undang-Undang Komisi Perdagangan Federal AS, dengan larangannya terhadap "metode persaingan yang tidak adil".

Sehubungan dengan larangan perilaku yang meningkatkan kemungkinan koordinasi, pengadilan dan lembaga penegak hukum yang menafsirkan Pasal 1 melarang perjanjian antarperusahaan untuk berbagi informasi jika perjanjian tersebut meningkatkan kemungkinan koordinasi oligopoli. Clark mengategorikan praktik-praktik tersebut sebagai (1) pengumuman publik tentang rencana bisnis, (2) penyebaran data tentang harga dan hal-hal non-harga, (3)

praktik-praktik yang menghalangi diskon termasuk klausul MFN, dan (4) standarisasi syarat perdagangan. Page merangkum perlakuan hukum atas berbagi informasi antarperusahaan bergantung pada audiens dan sifat komunikasi di masa lalu, masa kini, atau masa depan.

Publikasi yang luas kepada pasar, baik konsumen maupun pesaing, memiliki fitur-fitur yang jelas dapat meningkatkan persaingan misalnya, mengomunikasikan harga bensin kepada konsumen, memungkinkan perbandingan harga saat berkendara. Komunikasi kepada pesaing, tetapi tidak kepada pasar, memfasilitasi kolusi diam-diam dan memiliki manfaat prokompetitif yang terbatas, dengan contoh-contoh tipikal termasuk komunikasi dalam asosiasi perdagangan industri. Terkait dengan karakter temporal komunikasi, informasi masa kini atau masa lalu cenderung tidak berbahaya dibandingkan informasi masa depan.

Struktur pasar juga terkait dengan kekhawatiran akan koordinasi yang tidak sesuai kesepakatan, dan dalam pengadilan dan lembaga peninjauan merger secara luas mempertimbangkan efek terkoordinasi sebagai alasan untuk menantang atau melarang merger. Pedoman Merger Horizontal 2010 mengidentifikasi sebagai perilaku terkoordinasi (1) kesepakatan eksplisit, (2) pemahaman umum yang tidak disetujui secara eksplisit tetapi tetap ditegakkan, dan (3) perilaku akomodatif paralel.

Pedoman mengakui bahwa perilaku terkoordinasi mencakup aktivitas yang bukan merupakan pelanggaran antimonopoli. Merger akan ditantang jika dengan meningkatkan konsentrasi di pasar yang rentan terhadap koordinasi, kejadian dan efek koordinasi cenderung diperburuk. Kerentanan terhadap koordinasi mencakup, antara lain, transparansi harga dan homogenitas.

Tinjauan merger merupakan perlindungan yang signifikan terhadap kemungkinan koordinasi non-kesepakatan yang biasanya tidak dapat digugat. Khususnya, Pedoman Merger Horizontal AS, yang terakhir diperbarui pada tahun 2010, pada saat publikasi ini diterbitkan, sudah siap untuk direvisi. Mengingat penekanan FTC pada pasar teknologi, dan kepemimpinan Uni Eropa dalam Draf Pedoman Horizontal (infra), kami mengantisipasi revisi akan memasukkan kekhawatiran kolusi algoritmik sebagai faktor yang perlu dipertimbangkan dalam menganalisis dampak terkoordinasi.

Memperbaiki koordinasi non-kesepakatan secara langsung berdasarkan Undang-Undang FTC secara teori dimungkinkan. Larangan luas dalam Undang-Undang FTC menargetkan "metode persaingan tidak sehat", alih-alih "kontrak, kombinasi..., atau konspirasi" yang lebih spesifik berdasarkan Undang-Undang Sherman. FTC juga tidak memiliki wewenang untuk menjatuhkan hukuman yang sangat berat, sehingga membatasi kekhawatiran bahwa hukuman pidana atau kewajiban ganti rugi tiga kali lipat akan mengurangi insentif bagi perilaku persaingan yang agresif. Namun, perhatian utama tentang bagaimana menangani perilaku tersebut secara langsung apakah dengan memerintahkan persaingan, atau memerintahkan penetapan harga dengan biaya marjinal, atau hasil lainnya masih tetap ada. Komisi Perdagangan Federal (FTC) tidak secara langsung menentang koordinasi non-perjanjian, melainkan menargetkan penuntutannya untuk memfasilitasi praktik-praktik yang cenderung mengarah pada koordinasi.

20.5 KOLUSI ALGORITMIK DALAM HUKUM ANTIMONOPOLI UNI EROPA

Koordinasi melalui Perjanjian

Pasal 101 (1) TFEU melarang "semua perjanjian antarperusahaan, keputusan asosiasi perusahaan, dan praktik bersama yang dapat memengaruhi perdagangan antar Negara Anggota dan yang bertujuan atau berdampak pada pencegahan, pembatasan, atau distorsi persaingan di pasar internal". Hal ini khususnya mencakup perjanjian penetapan harga. Bukti inti untuk membuktikan pelanggaran Pasal 101 (1) TFEU adalah adanya komunikasi antara para pihak dalam perjanjian atau praktik bersama yang mengarah pada "pertemuan pikiran" untuk menetapkan harga. Serupa dengan konteks AS, perilaku paralel semata tidak dapat dilarang berdasarkan Pasal 101 (1) TFEU.

Koordinasi Non-Perjanjian

Komisi Uni Eropa telah mampu membuktikan adanya perjanjian anti-persaingan tanpa memiliki bukti komunikasi eksplisit antara para pihak sebelumnya. Pengadilan Kehakiman Uni Eropa memutuskan dalam kasus Imperial Chemical Industries bahwa "perilaku paralel tidak dapat dengan sendirinya diidentifikasikan dengan praktik bersama, namun dapat menjadi bukti kuat praktik tersebut jika mengarah pada kondisi persaingan yang tidak sesuai dengan kondisi normal pasar, dengan mempertimbangkan sifat produk, ukuran dan jumlah perusahaan, serta volume pasar tersebut." Pengadilan melanjutkan dengan mengatakan bahwa hal ini khususnya berlaku "jika perilaku paralel tersebut sedemikian rupa sehingga memungkinkan pihak-pihak terkait untuk mencoba menstabilkan harga pada tingkat yang berbeda dari yang seharusnya dicapai oleh persaingan".

Dengan demikian, Pengadilan membuka pintu dalam putusan ini untuk mempertimbangkan bukti di luar komunikasi antara para pihak terhadap elemen kontekstual pasar yang dimaksud. Meskipun demikian, sebagaimana ditegaskan oleh putusan Pengadilan Uni Eropa dalam kasus Woodpulp, jika terdapat penjelasan yang masuk akal untuk tindakan tersebut selain kolusi, maka tidak terdapat pelanggaran Pasal 101 (1) TFEU. Lebih lanjut, Komisi Uni Eropa menyatakan dalam Pedoman Kerja Sama Horizontal bahwa mereka akan mewaspadai upaya perusahaan untuk membuat pasar lebih transparan dengan mengumumkan harga secara publik. Jika beberapa perusahaan membuat pengumuman publik secara berurutan, misalnya, hal ini dapat menjadi bukti adanya konsertasi.

Dalam Draf Pedoman Horizontal yang baru, Komisi mengakui bahwa penggunaan algoritma juga dapat meningkatkan transparansi pasar, dan dengan itu risiko terjadinya kolusi di pasar. Namun, Komisi mencatat bahwa agar algoritma dapat menghasilkan hasil tersebut, diperlukan kondisi struktural lebih lanjut, seperti "frekuensi interaksi yang tinggi, daya beli yang terbatas, dan keberadaan produk/jasa yang homogen".

Selain itu, setidaknya diperlukan pertukaran informasi tidak langsung yang dapat dibuktikan melalui perangkat algoritmik antar pesaing agar dapat dianggap sebagai bentuk koordinasi yang dapat dikenai Pasal 101 (1) TFEU. Hal ini dapat terjadi, misalnya, melalui algoritma optimasi bersama yang dipasarkan oleh satu perusahaan TI dan yang memanfaatkan data sensitif dari berbagai pesaing.

Koordinasi Hub-and-Spoke

Kemungkinan ketiga untuk membuktikan adanya koordinasi anti-persaingan yang bertentangan dengan Pasal 101 (1) TFEU adalah ketika para pihak yang bermaksud menetapkan harga tidak berkomunikasi satu sama lain, melainkan melalui pihak ketiga. Dalam hal ini, pihak ketiga, yang bertindak sebagai perantara atau hub, pada akhirnya akan menciptakan "pertemuan pikiran" di antara para pihak yang berkompeten. Hal ini terjadi, misalnya dalam kasus AC Treuhand, di mana sebuah konsultan memfasilitasi pertukaran informasi bisnis sensitif antar kliennya yang mengarah pada penyelarasan harga di pasar.

Sebagaimana disebutkan di atas, terutama jika beberapa pelaku di pasar yang sama membeli algoritma penetapan harga dari pengembang pihak ketiga yang sama, skenario hub-and-spoke dapat terjadi. Begitu ada komunikasi dari pengembang pihak ketiga kepada pembeli bahwa algoritma tersebut membantu menyelaraskan harga, hal ini mungkin cukup untuk membuktikan pelanggaran berdasarkan Pasal 101 (1) TFEU. Pada saat yang sama, fakta bahwa algoritma penetapan harga pihak ketiga ini akan memiliki elemen stokastik, dan dapat digunakan untuk menerapkan strategi penetapan harga yang berbeda yang ditentukan oleh pengguna akhir mungkin sebenarnya tidak mengarah pada penyelarasan harga apa pun.

Dominasi Kolektif

Jika terdapat peningkatan bukti hasil kolusi di pasar-pasar di mana penetapan harga algoritmik telah meluas, Pasal 102 TFEU juga berpotensi menyediakan jalan untuk menjatuhkan sanksi hukum antimonopoli. Pasal 102 TFEU melarang “segala bentuk penyalahgunaan oleh satu atau lebih perusahaan atas posisi dominan di pasar internal atau di sebagian besar pasar tersebut sejauh hal itu dapat memengaruhi perdagangan antar Negara Anggota”. Meskipun ketentuan ini paling sering diterapkan pada satu perusahaan dominan, ketentuan ini juga melarang penyalahgunaan dominasi oleh beberapa perusahaan yang secara kolektif memegang posisi dominan.

Standar pembuktian adanya dominasi kolektif telah sedikit menyatu dengan persyaratan untuk menunjukkan dampak terkoordinasi atau, dengan kata lain, kemungkinan kolusi diam-diam, dari merger. Untuk menunjukkan bahwa kondisi dominasi kolektif (atau efek terkoordinasi dari merger) dapat diwujudkan dan dipertahankan, Pengadilan Umum Uni Eropa menetapkan tiga syarat penting dalam kasus Airtours.

Pertama, tingkat transparansi di pasar tertentu harus tinggi agar memungkinkan pelaku pasar untuk saling memantau perilaku. Kedua, perlu ada mekanisme pembalasan yang mampu mencegah atau menghukum perilaku menyimpang, misalnya, penurunan harga. Ketiga, baik pelanggan maupun calon pesaing tidak dapat melawan strategi kolusi. Keberadaan kekuatan pembeli, misalnya, dapat menjadi faktor yang melemahkan kolusi diam-diam.

Selain itu, perlu ada penyalahgunaan untuk menetapkan pelanggaran Pasal 102 TFEU. Pasal 102 (a) TFEU memberikan contoh penyalahgunaan "secara langsung atau tidak langsung memaksakan harga beli atau jual yang tidak adil atau kondisi perdagangan tidak adil lainnya". Jika kondisi dominasi kolektif dapat diwujudkan melalui algoritma penetapan harga, penerapan harga yang melampaui persaingan dapat dianggap sebagai penyalahgunaan karena merupakan harga yang tidak adil. Namun, sejauh ini, hanya ada sedikit kasus hukum tentang

harga tidak adil di UE, dan bahkan lebih sedikit lagi kasus hukum tentang harga tidak adil sebagai penyalahgunaan dominasi kolektif.

20.6 ANALISIS HUKUM KOMPARATIF ATAS SITUASI PENETAPAN HARGA ALGORITMIK

Dalam hal penilaian penetapan harga algoritmik, penerapan hukum antimonopoli baik di AS maupun Uni Eropa bergantung pada pemahaman tentang apakah terdapat koordinasi melalui kesepakatan atau koordinasi non-kesepakatan. Kasus-kasus mudah muncul ketika algoritma diterapkan melalui kesepakatan tegas antara para pesaing, dengan tujuan untuk mengoordinasikan harga atau ketentuan lain yang sensitif terhadap persaingan. Hal ini dalam semua kasus seharusnya ilegal per se (Hukum AS) atau pembatasan persaingan berdasarkan objek (Hukum Uni Eropa) dan, seperti dalam kasus poster Amazon, dapat dikenakan tuntutan pidana dan penegakan hukum privat yang sesuai.

Dalam kasus koordinasi melalui algoritma, seperti halnya koordinasi non-algoritmik, muncul masalah pembuktian. Koordinasi yang diamati dapat berupa kesepakatan, yang menimbulkan perlakuan per se (di atas), atau melalui non-kesepakatan, yang memunculkan pertanyaan (1) apakah koordinasi non-kesepakatan yang dibantu oleh algoritma berbeda, dan (2) apakah koordinasi melalui algoritma dapat mencapai tingkat kesepakatan, meskipun awalnya tidak didorong oleh kesepakatan. Terkait pertanyaan kedua ini, berdasarkan kondisi hukum antimonopoli AS saat ini, penerapan algoritma yang, melalui pembelajaran mesin, berinteraksi dengan algoritma lain melalui serangkaian pertukaran dan komitmen yang dapat disamakan dengan "kesepakatan", kecil kemungkinannya menimbulkan tanggung jawab hukum. Situasi serupa juga terjadi di Uni Eropa, meskipun pemberian sinyal harga oleh algoritma penetapan harga dalam keadaan tertentu dapat diartikan sebagai bentuk tindakan bersama.

Di sinilah pengalaman dan studi empiris dapat menyoroti fitur-fitur algoritma yang tidak konsisten dengan hasil persaingan, sehingga memberikan dasar penegakan hukum yang saat ini belum ada. Bukti tambahan yang berkembang dari eksperimen yang dijalankan, misalnya, oleh Calvano dkk. dan Otoritas Persaingan Swedia, dan lainnya, yang mencoba memperkirakan, langkah demi langkah, kondisi pasar nyata tempat alat penetapan harga AI diterapkan, merupakan komponen penting. Untuk saat ini, sebagaimana diprediksi oleh teori oligopoli, simetri agen merupakan faktor yang mendorong kolusi, tetapi sebagaimana ditunjukkan oleh Otoritas Persaingan Swedia, hal tersebut bukanlah syarat mutlak bagi munculnya kolusi diam-diam algoritmik.

Selain itu, faktor-faktor tradisional dalam analisis persaingan, misalnya hambatan masuk, tetap berlaku dengan algoritma: ketika algoritma dapat memperhitungkan ancaman masuk, harga akan lebih rendah. Semua bukti awal tampaknya menunjukkan bahwa kasus-kasus kolusi algoritmik harus ditangani berdasarkan kasus per kasus. Evaluasinya akan sangat bergantung pada strategi atau kebijakan yang diprogramkan atau dikembangkan oleh algoritma.

Terkait pertanyaan pertama di atas, kekhawatiran akan hasil yang terkoordinasi memerlukan analisis tentang cara menganalisis penetapan harga algoritmik. Masalah

perbaikan yang memicu perdebatan historis mengenai tantangan koordinasi non-kesepakatan sama relevannya dalam konteks koordinasi non-kesepakatan melalui algoritma. Upaya hukum yang lebih keras yang tersedia berdasarkan hukum AS, termasuk hukuman pidana dan ganti rugi tiga kali lipat, tidak akan tepat jika diterapkan pada keputusan independen untuk menerapkan algoritma penetapan harga.

Demikian pula, analisis berdasarkan kaidah akal sehat akan sesuai, dengan menyerahkan beban pembuktian kerugian terhadap persaingan dan jika sesuai kemungkinan alternatif yang lebih longgar kepada badan penegak hukum atau penegak hukum swasta. Sanksi dalam penegakan hukum oleh pemerintah dalam kasus ini harus dibatasi pada putusan pengadilan yang berwawasan ke depan.

Dalam hukum Uni Eropa, dengan fokus utamanya pada penegakan hukum persaingan publik, sanksi yang kurang radikal dapat dijatuhkan. Kasus-kasus yang melibatkan penetapan harga algoritmik dapat diselesaikan pada awalnya melalui komitmen, di mana Komisi Uni Eropa dapat menjatuhkan upaya hukum perilaku atau meminta perubahan pada desain algoritmik. Dengan cara ini, pendekatan penegakan hukum publik Uni Eropa, yang menerapkan lebih banyak lensa regulasi daripada lensa adversarial, lebih cocok untuk kondisi ketidakpastian saat ini terkait teknologi baru yang dipermasalahkan. Namun, bahkan dengan pendekatan penegakan hukum yang hati-hati, putusan pengadilan berwawasan ke depan seperti apa yang mungkin tepat masih belum jelas. Sebagaimana Turner sampaikan, larangan perilaku memaksimalkan keuntungan secara independen merupakan hal yang bermasalah. Intervensi yang paling menjanjikan adalah menantang fitur-fitur algoritma penetapan harga yang melibatkan praktik-praktik fasilitasi yang paling mengkhawatirkan, baik itu berbagi informasi atau lainnya, yang diidentifikasi oleh Clark, Dibadj, dan Page.

Pengungkapan informasi yang sensitif terhadap persaingan, baik oleh manusia maupun algoritma, berkorelasi dengan hasil yang terkoordinasi dan harus dipertanyakan, dengan pertanyaan tentang temporalitas (masa depan, masa kini, atau masa lalu) dan keluasan (publik atau hanya diungkapkan kepada pesaing) yang menjiwai analisis. Apabila pengungkapan tersebut terjadi berdasarkan kesepakatan, pengungkapan tersebut biasanya dapat digugat. Kesepakatan lain yang meningkatkan homogenitas atau transparansi, yang memfasilitasi koordinasi non-kesepakatan melalui algoritma, juga mengkhawatirkan. Kerja sama antar pesaing dalam memilih dan memprogram algoritma harus dianggap sebagai praktik yang memfasilitasi dan tunduk pada pengawasan ketat, yang kemungkinan besar akan menimbulkan gugatan meskipun, seperti halnya diskusi praktik bisnis lainnya, mungkin ada justifikasi efisiensi yang perlu dipertimbangkan.

Seiring dengan tersedianya hasil empiris yang mendukung hal ini, pengadilan dan penegak hukum harus terbuka untuk mengajukan praduga, mengalihkan beban pembenaran penggunaan tertentu kepada perusahaan yang menggunakan algoritma. Pertimbangan efisiensi juga akan berperan dalam hukum persaingan Uni Eropa, baik untuk menentukan apakah penerapan penetapan harga algoritmik harus dikategorikan sebagai pembatasan persaingan berdasarkan efek berdasarkan Pasal 101 (1) TFEU (yang mengharuskan pembuktian

efek anti-persaingan di pasar dari tindakan tersebut) maupun apakah penerapan tersebut dapat memperoleh manfaat dari pengecualian efisiensi berdasarkan Pasal 101 (3) TFEU.

Tinjauan merger memberikan kesempatan untuk memeriksa perubahan struktur pasar atau dampak lain yang dapat memfasilitasi koordinasi melalui algoritma, yang tidak dapat digugat berdasarkan Undang-Undang Sherman. Di AS, tinjauan tersebut harus mengikuti analisis yang ada yang diuraikan dalam Pedoman Merger Horizontal Bagian 7. Secara khusus, terdapat praduga struktural sebagai pemeriksaan tingkat konsentrasi. Seiring berkembangnya bukti empiris, praduga tersebut mungkin perlu diperkuat jika argumen Gal bahwa penetapan harga algoritmik lebih mudah mengarah pada kolusi daripada penetapan harga non-algoritmik terbukti akurat.

Merger di pasar dengan riwayat kolusi atau koordinasi non-kesepakatan sudah menjadi subjek pengawasan berdasarkan riwayat tersebut, dan dalam kasus pasar yang dicirikan oleh algoritma penetapan harga, penegak harus menyelidiki apakah algoritma diadopsi sebagai sarana untuk memfasilitasi koordinasi dan apakah praktik penetapan harga algoritmik di industri tersebut sebelumnya menyebabkan hasil yang mengkhawatirkan.

Merger yang cenderung mengeliminasi "maverick", mungkin perusahaan yang terus tidak menggunakan perangkat lunak berstandar industri dalam keputusan penetapan harganya atau mungkin perusahaan yang memprogram perangkat lunaknya untuk bersaing lebih agresif daripada norma industri, juga harus tunduk pada pengawasan yang lebih ketat atas dasar tersebut. Mencerminkan konvergensi substansial dari kebijakan penegakan merger, alasan yang sama akan berlaku berdasarkan Peraturan Merger Uni Eropa.

Berdasarkan hukum Uni Eropa, permasalahan yang ditangani berdasarkan hukum merger dapat ditangani berdasarkan penyalahgunaan dominasi kolektif berdasarkan Pasal 102 TFEU. Hal ini akan memungkinkan otoritas persaingan untuk campur tangan bahkan jika merger tidak terjadi, tetapi akan mengharuskan penegak persaingan untuk menunjukkan bahwa pasar yang dimaksud memiliki kondisi struktural yang dianggap kondusif untuk kolusi berdasarkan kriteria Airtours yang dibahas di atas. Jika, dalam kasus pasar terkonsentrasi, algoritma penetapan harga memang berkontribusi pada transparansi yang lebih besar, dan tidak ada kekuatan pembeli yang signifikan atau faktor penyeimbang lainnya, pertanyaan apakah algoritma dapat melakukan pembalasan dalam kasus pemotongan harga akan menentukan apakah dominasi kolektif dapat dibangun. Agar hal ini terjadi, algoritma kemungkinan harus relatif sederhana dan membutuhkan kapasitas pembelajaran yang simetris untuk berpotensi dapat membuktikan posisi dominasi kolektif. Pertanyaannya kemudian akan tetap apakah ada penyalahgunaan kolektif dan pada tingkat mana harga supra-kompetitif akan menjadi harga yang tidak adil berdasarkan Pasal 102(a) TFEU.

Di AS, dominasi kolektif, atau "monopoli bersama," sebagai teori kerugian belum dipertimbangkan secara serius dalam lebih dari 40 tahun; satu pertanyaan adalah apakah efisiensi keputusan algoritmik dapat memberikan dasar untuk memeriksa kembali teori semacam itu di tahun-tahun mendatang. Mengingat batasan yang lebih besar dari Sherman Act relatif terhadap Pasal 101 dan 102 TFEU, analog hukum AS adalah apa yang disebut tindakan "mandiri" berdasarkan Undang-Undang FTC bagian 5 yang menantang perilaku

sebagai "metode persaingan yang tidak adil." Menggunakan Undang-Undang FTC adalah solusi umum yang dianjurkan oleh para akademisi yang mengidentifikasi masalah yang tidak dapat ditantang berdasarkan hukum yang ada di Bagian 1.

Pendekatan semacam itu paling sering diidentifikasi dengan Penuntutan *Ethyl Corp*. FTC. *Ethyl Corp*. dibatalkan pada banding ke pengadilan banding federal, yang menyatakan bahwa perilaku unilateral independen oleh firma tidak melanggar Undang-Undang FTC, sebagian berdasarkan bukti yang tidak cukup dari contoh perilaku yang menyebabkan hasil yang diamati. Kemungkinan mengidentifikasi adopsi istilah algoritmik yang secara empiris terbukti sangat merusak sebagai cara untuk memenuhi persyaratan pengadilan banding tetap ada.

20.7 SUARA PEMBUAT KEBIJAKAN DAN PROSPEK MASA DEPAN

Isu kolusi algoritmik tidak luput dari perhatian para penegak hukum dan pembuat kebijakan. Secara umum, respons kebijakan, dalam laporan publik atau pengajuan kepada komite multi-yurisdiksi, mencerminkan (1) keyakinan terhadap aturan antimonopoli yang ada untuk menyerang praktik dan dampak anti-persaingan yang disebabkan oleh penggunaan algoritma, dan (2) kehati-hatian terhadap teori-teori penegakan hukum yang baru tanpa adanya bukti nyata mengenai kerugian yang disebabkan oleh adopsi algoritma.

Berdasarkan kekuatan kekhawatiran teoretis, tetapi lemahnya bukti dampak nyata, pendekatan yang tepat akan tetap berupa kehati-hatian yang penuh dari pihak penegak hukum dan pembuat kebijakan, yang harus berupaya mengembangkan bukti dampak di tingkat pasar untuk mengembangkan perangkat regulasi mereka. Selain itu, kondisi paritas analisis yang substansial di seluruh yurisdiksi saat ini menghadirkan peluang unik untuk kerja sama antar-yurisdiksi.

Komunikasi yang ada dapat digambarkan lebih, cukup, dan kurang agresif dalam analisis dan perlakuan yang diantisipasi terhadap penetapan harga algoritmik. Laporan Autoridade da Concorrência tentang Ekosistem Digital, Big Data, dan Algoritma mungkin merupakan laporan yang paling menunjukkan perlunya pengawasan dan intervensi. Mengenai kekhawatiran yang paling dramatis, yang dicontohkan oleh pengalaman laboratorium algoritma Q-learning yang "berkolusi" tanpa intervensi manusia aktif, Digital Ecosystems dengan tepat mencatat bahwa hal ini belum terbukti terjadi dalam konteks dunia nyata, tetapi perhatian yang cermat diperlukan. Secara khusus, Digital Ecosystems mencatat tanggung jawab yang tegas atas dampak yang disebabkan oleh algoritma yang mereka terapkan dan kemungkinan adanya persyaratan untuk menguji dan memverifikasi guna memastikan kepatuhan.

Otoritas Persaingan dan Pasar Inggris (CMA) menerbitkan Laporan pada tahun 2021 tentang Algoritma. Mengenai kemungkinan risiko bagi pasar dari kolusi diam-diam algoritmik, disimpulkan bahwa bukti empiris sangat langka sehingga tidak ada kesimpulan yang jelas yang dapat ditarik. Laporan oleh Autorite de la Concurrence dan Bundeskartellamt Prancis tentang Algoritma dan Persaingan memprediksi keberhasilan aturan dan pendekatan penegakan hukum yang ada dalam sebagian besar kasus, dan mengidentifikasi kurangnya bukti untuk

masalah koordinasi otonom, mengakui juga bahwa kolusi diam-diam tidak dapat digugat berdasarkan hukum Uni Eropa.

Dalam ringkasannya, Algoritma dan Persaingan menyatakan bahwa rezim hukum persaingan yang ada sudah memadai untuk tugas tersebut, kasus-kasus sebelumnya yang melibatkan algoritma belum membebani keahlian otoritas, dan perubahan hukum harus menunggu gambaran yang lebih jelas tentang jenis kasus yang mungkin dihadapi di masa mendatang. Biro Persaingan Kanada (CCB) merupakan pelopor dalam makalah diskusinya, Big Data and Innovation, yang menyoroti kekhawatiran terkait koordinasi pasca-merger, mungkin akibat akuisisi perusahaan yang tidak konvensional, dan peningkatan transparansi melalui praktik-praktik fasilitasi. Namun, CCB tidak melihat adanya intervensi selain pengendalian harga yang tepat dan justru mencatat pengalaman penegakan hukum di mana kesepakatan kartel tradisional dicapai menggunakan perangkat digital, termasuk algoritma yang diprogram berdasarkan kesepakatan antara para kartelis.

Reaksi terbatas dari AS sejauh ini merupakan indikasi paling kecil perlunya intervensi. Pada tahun 2018, badan-badan antimonopoli AS bekerja sama dalam sebuah laporan kepada Komite Persaingan OECD di mana mereka menyoroti pandangan mereka tentang kolusi algoritmik. Badan-badan AS tersebut mencatat minat akademis terhadap hasil kolusi yang dicapai oleh algoritma, dan mencatat salah satu fitur khusus kecepatan respons dan penyesuaian yang sering disorot dalam analisis akademis tersebut. Pergeseran substansial dalam prioritas penegakan hukum setelah perubahan politik dapat diperkirakan akan memberikan perspektif yang berbeda terhadap masalah ini, tetapi kita masih kekurangan pernyataan definitif yang menunjukkan tujuan yang direvisi.

Komite multi-yurisdiksi memiliki banyak potensi dalam memfasilitasi pengembangan keahlian dan aturan secara kooperatif. Baik OECD maupun ICN telah terlibat dalam masalah koordinasi melalui algoritma. Misalnya, Laporan Algoritma dan Kolusi OECD mengkaji literatur dan pengalaman penegakan hukum terkait hasil kolusi yang dicapai oleh algoritma dan mengamati, "algoritma dapat memungkinkan perusahaan untuk mengganti kolusi eksplisit dengan koordinasi diam-diam."

Menanggapi kekhawatiran ini, Laporan OECD menyoroti kemungkinan "meninjau kembali" "gagasan kesepakatan", mungkin mempertimbangkan penyesuaian harga yang cepat yang menyebabkan hasil monopoli diperlakukan sebagai "kesepakatan", sebagaimana penetapan harga paralel oleh algoritma yang bertindak sebagai alat pemberi sinyal. Dalam makalah peninjauan ruang lingkup, "*Big Data and Cartels*", ICN mengidentifikasi kekhawatiran yang sama terkait koordinasi non-kesepakatan, yang didasarkan pada transaksi berulang yang cepat, serta kemungkinan algoritma pembelajaran mendalam yang belum terbukti mencapai kesepakatan terkomputerisasi, dan juga mengajukan pertanyaan apakah definisi kesepakatan perlu direvisi ketika algoritma beroperasi secara otonom.

Langkah awal untuk menyaring status pembelajaran menjadi amandemen pedoman menjanjikan manfaat. Hal ini akan memungkinkan penasihat hukum untuk memberikan saran yang lebih baik kepada perusahaan tentang penerapan perangkat lunak yang semakin canggih. Sebagai salah satu contoh, Pedoman Kolaborasi Pesaing yang baru direvisi oleh Biro Persaingan

Kanada mengakui "perjanjian antara pesaing untuk menggunakan algoritma penetapan harga bersama" sebagai perjanjian kartel, dengan menambahkan peringatan bahwa "perilaku yang mengarah pada paralelisme yang disengaja tidaklah cukup.

Penelitian empiris tentang dampak algoritma penetapan harga, dengan mempertimbangkan manfaat efisiensi yang substansial dan kekhawatiran yang telah diteorikan dengan baik, harus menjadi investasi prioritas pertama bagi lembaga atau kelompok multi-yurisdiksi, mungkin melalui hibah kepada tim peneliti. Banyak keahlian lembaga dikembangkan melalui investigasi kasus dan penuntutan, dan mempelajari dampak penetapan harga algoritmik sebagai bagian dari investigasi merger dan non-merger harus menjadi prioritas.

Lembaga-lembaga yang memiliki wewenang untuk melakukan penyelidikan sektoral, seperti Komisi Uni Eropa dan Otoritas Persaingan Nasional di Negara-negara Anggota Uni Eropa, serta otoritas Bagian 6 FTC AS yang analog (Komisi Perdagangan Federal AS 1981), dapat menggunakan ini kewenangan untuk memperoleh lebih banyak bukti tentang perkembangan terkini algoritma penetapan harga dan dampaknya terhadap pasar.

Upaya penegakan hukum yang mengangkat isu-isu ini harus melibatkan para ahli teknologi untuk melengkapi ketergantungan yang ada pada keahlian ekonomi dan manajemen. Setelah pengalaman dan pemahaman yang lebih mendalam dikembangkan, pembuatan peraturan atau legislasi yang menargetkan penggunaan teknologi yang paling merugikan, kemungkinan besar yang menghasilkan transparansi di antara para pesaing tetapi tidak di pasar secara umum, harus dipertimbangkan.

20.8 TANTANGAN PENEGAKAN HUKUM PADA PRICING ALGORITMIK

Perhatian akademis yang substansial dan, baru-baru ini, upaya para penegak hukum dan pembuat kebijakan dalam memahami dampak penerapan algoritma secara luas untuk menetapkan harga belum menghasilkan resolusi definitif mengenai tingkat kekhawatiran atau sifat intervensi yang tepat. Kedua sisi pengembangan teknis perangkat lunak penetapan harga, dan pembentukan perangkat untuk investigasi dan penegakan hukum, terus berkembang. Dalam bab ini, kami mengemukakan beberapa bidang minat utama bagi para akademisi, advokat, dan penegak hukum.

Pertanyaan tentang dampak penetapan harga algoritmik terhadap hasil pasar masih belum terselesaikan. Teori-teori yang diajukan oleh para pelopor di antara para akademisi, termasuk Mehra, Ezrachi, Stucke, dan Gal, telah diakui secara luas sebagai hal yang patut mendapat perhatian serius oleh otoritas antimonopoli dan pembuat kebijakan. Pertanyaan yang lebih menantang terkait penemuan dan respons otonom, yang dalam eksperimen laboratorium mengarah pada kolusi algoritmik, saat ini belum cukup terbukti secara empiris dalam pengaturan dunia nyata untuk menimbulkan lebih dari sekadar kekhawatiran yang hati-hati.

Temuan empiris juga kurang terkait dampak bersih pengambilan keputusan algoritmik yang meningkatkan efisiensi terhadap persaingan dan konsumen, yang juga mengancam hasil yang terkoordinasi. Bukti empiris tambahan dari eksperimen laboratorium dan dunia nyata,

dengan mempertimbangkan keragaman desain algoritmik dan keragaman sumber data yang dapat digunakan dalam pengembangan strategi penetapan harga algoritmik, diperlukan. Khususnya, pemahaman yang lebih baik tentang insentif dan strategi pengembang dan pengadopsi algoritma penetapan harga akan meningkatkan respons dalam kebijakan dan hukum persaingan.

Pengamatan penting dari studi perbandingan kami adalah paritas substansial dalam perlakuan di seluruh sistem antimonopoli terkemuka. Yurisdiksi di kedua sisi Atlantik umumnya tidak menyukai hasil yang terkoordinasi, tetapi menyadari tantangan atau ketidakmungkinan untuk mencegahnya, jika tidak ada kesepakatan nyata antara para pihak yang dapat dibuktikan berdasarkan larangan kartel (Bagian 1 dan Pasal 101 TFEU).

Ada jalur yang lebih jarang dilalui yang dapat ditempuh oleh penegak hukum antimonopoli di AS dan Uni Eropa, di satu sisi berdasarkan Bagian 5 Undang-Undang FTC atau melalui Pasal 102 TFEU dan dominasi kolektif. Meskipun demikian, tanpa memiliki basis bukti yang memadai yang akan menentukan apakah, bagaimana, dan kapan algoritma penetapan harga menyebabkan kerugian persaingan, tampaknya terlalu dini untuk mempertimbangkan potensi jalur penegakan hukum antimonopoli ini.

Terakhir, paritas yang diamati secara keseluruhan memberikan peluang substansial bagi kerja sama lintas yurisdiksi dalam menemukan strategi yang berhasil untuk memerangi kerugian yang belum ditangani oleh hukum secara historis. Memprioritaskan masalah-masalah ini, dan menerapkan kombinasi studi, investigasi, penuntutan, dan reduksi menjadi pedoman dan hukum, merupakan pendekatan yang tepat untuk memastikan inovasi teknologi dan penegakan hukum melayani kepentingan masyarakat.

BAB 21
PROPOSAL AI ACT E-JUSTICE EROPA: USER-FOCUSED & EFEKTIF

Abstrak e-Justice Eropa bertujuan untuk mengembangkan perangkat elektronik yang memungkinkan yurisdiksi nasional dan ECJ untuk terhubung melalui saluran digital yang andal dan aman. Strategi e-Justice 2019–2023 menggarisbawahi beberapa prinsip umum Uni Eropa baru yang dikembangkan langsung di bawah paradigma e-Justice, yang patut mendapat perhatian khusus terkait dimensi yang berfokus pada pengguna dan ramah pengguna. Mengingat tahun 2021 merupakan tahun di mana digitalisasi peradilan akan dibahas, perlu dipahami bagaimana AI akan berdampak pada bidang peradilan, tidak hanya dalam sistem peradilan MS (yurisdiksi fungsional Uni Eropa, ketika menerapkan hukum Uni Eropa), tetapi juga di ECJ, karena teknologi disruptif ini sedang dibahas.

Proposal Undang-Undang AI menekankan bahwa sistem AI yang ditujukan untuk administrasi peradilan harus diklasifikasikan sebagai berisiko tinggi, mengingat potensi dampaknya yang signifikan terhadap domain perlindungan peradilan yang efektif. Oleh karena itu, makalah ini bertujuan untuk memahami kebutuhan untuk sepenuhnya menekankan pendekatan AI yang berpusat pada manusia di bidang peradilan, sehingga perlindungan peradilan yang efektif dapat diperdalam melalui prinsip-prinsip yang berfokus pada pengguna dan mudah digunakan; dan untuk meneliti, dari sudut pandang e-Justice, bagaimana Proposal Undang-Undang AI harus lebih jauh membahas penggunaan instrumental peradilan atas sistem AI, sehingga independensi peradilan, hak-hak prosedural, dan akses ke keadilan diperhatikan dalam tatanan yurisdiksi Uni Eropa.

21.1 E-JUSTICE DAN DIGITALISASI PERADILAN: PRO-KONTRA AI

Semua lembaga Eropa terlibat dalam pengembangan e-Justice Eropa lebih lanjut. Dewan telah menanganinya sejak tahun 2007 secara terkoordinasi, pertama dengan membentuk Kelompok Kerja untuk tema tersebut dan, secara berurutan, dengan menyajikan tiga Rencana Aksi e-Justice Eropa: dari tahun 2009 hingga 2013; dari tahun 2014 hingga 2018; dan yang terakhir, yang masih dalam pengembangan, dari tahun 2019 hingga 2023.

Dalam hal ini, e-Justice Eropa “bertujuan untuk meningkatkan akses terhadap keadilan dalam konteks pan-Eropa dan mengembangkan serta mengintegrasikan teknologi informasi dan komunikasi ke dalam akses terhadap informasi hukum dan cara kerja sistem peradilan” dengan mengingat bahwa “fungsi peradilan yang efisien di Negara-negara Anggota” juga dapat berkaitan dengan penerapan komponen digital dan pendekatan teknologi baru. Khususnya dalam upaya menetapkan pedoman masa depan (yang harus dipenuhi hingga tahun 2023), Dewan, di bawah domain “Evolutivitas”, memahami Rencana Aksi “harus fleksibel berkenaan dengan perkembangan masa depan, baik yang legal maupun teknis”, menekankan bahwa “ranah teknologi hukum seperti Kecerdasan Buatan (AI) [...], misalnya, harus dipantau secara ketat, guna mengidentifikasi dan memanfaatkan peluang yang berpotensi berdampak positif pada e-Justice”.

Dalam konteks ini, lembaga Eropa yang memajukan Kecerdasan Buatan ini “dapat memberikan dampak positif pada e-Keadilan, misalnya dengan meningkatkan efisiensi dan kepercayaan” meskipun mengakui bahwa “segala pengembangan dan penerapan teknologi semacam itu di masa mendatang harus mempertimbangkan risiko dan tantangan, khususnya [...] terhadap perlindungan data dan etika”.

Dalam pengertian yang sama, Komisi Eropa baru-baru ini mencurahkan perhatiannya pada masalah ini, memahami bahwa “[k]es terhadap keadilan perlu dipertahankan dan mengimbangi perubahan, termasuk transformasi digital yang memengaruhi semua aspek kehidupan kita”. Berfokus pada Kecerdasan Buatan salah satu alat yang disamakan dalam “kotak peralatan” kelembagaan ini, Komisi Eropa memahami bahwa “munculnya teknologi baru membawa serta kebutuhan untuk penilaian berkelanjutan atas dampaknya, khususnya yang berkaitan dengan hak-hak dasar dan perlindungan data”.

Sejauh ini, meskipun memahami potensi Kecerdasan Buatan pada domain keadilan, “seperti memanfaatkan informasi dengan cara-cara baru dan sangat efisien, dan [...] mengurangi durasi proses peradilan”, lembaga Eropa itu juga menekankan bahwa “potensi ketidakjelasan atau bias juga dapat menyebabkan risiko dan tantangan bagi penghormatan dan penegakan hak-hak fundamental yang efektif, termasuk [...] hak atas pemulihan yang efektif dan pengadilan yang adil”.

Faktanya, alat dan mekanisme Kecerdasan Buatan menimbulkan beberapa pertanyaan, menuntut transparansi lebih lanjut, “pengawasan manusia, akurasi, dan ketahanan” sehingga perlindungan hak-hak fundamental, supremasi hukum dan, khususnya, tuntutan perlindungan peradilan yang efektif dipromosikan. Dalam hal ini, e-Justice Eropa, sejauh ini, telah mampu memainkan "peran ganda yang bersifat simbolis":

- Di satu sisi, "sebagai sarana untuk menerapkan sistem komunikasi yang interoperabel, sehingga artikulasi yurisdiksi dapat difasilitasi, di tingkat kelembagaan": di bawah paradigma e-Justice Eropa, Proposal Regulasi mengenai sistem e-CODEX telah diajukan sebagai sarana untuk lebih mengembangkan saluran interoperabel antar otoritas peradilan nasional selama hal tersebut juga mempertimbangkan promosi interaksi digital antara pengadilan nasional dan Mahkamah Keadilan Uni Eropa (ECJ), melalui aplikasi eCuria. Tren-tren yang konvergen ini menghasilkan pemahaman yang lebih luas tentang Prosedur Uni Eropa (berfokus pada peran ECJ dan pengadilan nasional dalam tatanan hukum dan yurisdiksi Eropa) dan suasana integrasi peradilan yang efektif, bahkan mendorong kedekatan yang lebih erat antara kerja sama peradilan dalam perkara perdata dan pidana;
- Di sisi lain, e-Justice telah bertindak “sebagai rujukan kepada perlindungan peradilan yang efektif atas hak-hak yang diberikan kepada individu oleh tatanan hukum Eropa, melalui pendekatan dan pembaruan baru”: di bawah hal terpenting tersebut, penelitian harus dilakukan agar potensi dan bahaya Kecerdasan Buatan dapat dipahami sepenuhnya, terutama dengan melakukan pembedaan yang tegas antara (a) Sistem Instrumental Kecerdasan Buatan dan (b) Sistem Keputusan Kecerdasan Buatan dalam ranah peradilan. Berdasarkan pembedaan ini, klasifikasi risiko sistem Kecerdasan

Buatan dapat dilakukan secara konkret, karena harus ada pendekatan yang jelas tentang sistem mana yang berlaku untuk bidang peradilan yang harus tunduk pada kewajiban yang lebih ketat sebagaimana dianggap berisiko tinggi. Lebih jauh lagi, sistem ini juga harus dipahami berdasarkan pendekatan Kecerdasan Buatan yang berpusat pada manusia, sehingga perlindungan hukum yang efektif harus dicapai, mempertanyakan apakah prinsip-prinsip yang berfokus pada pengguna dan mudah digunakan dapat memperdalam ketaatannya.

21.2 SISTEM KECERDASAN BUATAN YANG DITUJUKAN UNTUK ADMINISTRASI PERADILAN

Penggunaan mekanisme Kecerdasan Buatan di bidang peradilan dapat menimbulkan tantangan yang menguntungkan sekaligus merugikan bagi perlindungan peradilan yang efektif. Namun, langkah-langkah tersebut harus dirancang dan diterapkan di tingkat Eropa karena “dapat menghindari duplikasi upaya nasional dan menciptakan sinergi yang signifikan” dan “ini juga dapat memastikan interoperabilitas dan pada akhirnya mengubah proyek percontohan yang baik menjadi solusi di seluruh Uni Eropa”.

Lebih lanjut, karena sistem Kecerdasan Buatan memerlukan data yang relevan dan pelatihannya, tingkat Uni Eropa merupakan tempat yang tepat untuk mencapai penggunaannya “dengan kepatuhan penuh terhadap aturan perlindungan data pribadi”.

Bahasa Indonesia: Berangkat dari sensitivitas Komisi Eropa, meskipun keuntungan dari memperkenalkan aplikasi berbasis Kecerdasan Buatan dalam sistem peradilan sangat jelas, "ada juga risiko yang cukup besar terkait dengan penggunaannya untuk pengambilan keputusan otomatis dan 'kepolisian prediktif' / 'keadilan prediktif'".

Telah ditetapkan bahwa bagian dari "pekerjaan pengadilan dan hakim adalah untuk memproses informasi; para pihak membawa informasi ke pengadilan, transformasi terjadi selama prosedur, dan hasilnya juga merupakan informasi"; yaitu, "tidak semua pemrosesan informasi ini merupakan kustomisasi yang kompleks" sebelum proses pengadilan dan ada beberapa skenario di mana sistem Kecerdasan Buatan sedang dikembangkan.

Dalam hal ini, penggunaan umumnya dapat ditetapkan pada tiga topik utama, karena dimaksudkan untuk hanya berfokus pada sistem yang efektif dan / atau potensial yang tersedia di hadapan pengadilan dan untuk digunakan dalam proses prosedural yang berjalan di hadapan mereka: Sistem Kecerdasan Buatan pada (1) pengorganisasian informasi; (2) memobilisasi yurisprudensi yang telah ada sebelumnya; dan (3) meramalkan tren keputusan.

Sistem Kecerdasan Buatan dalam Pengorganisasian Informasi

Sistem Kecerdasan Buatan dapat digunakan untuk "mengenali pola dalam dokumen dan berkas teks". Penggunaan serupa (meskipun diekstrapolasi untuk operasi ekstrayudisial dan yudisial lebih lanjut) sudah dapat ditemukan di Amerika Serikat (AS): eDiscovery adalah "investigasi otomatis informasi elektronik untuk penemuan, sebelum dimulainya prosedur pengadilan", menggunakan pembelajaran mesin "yang [...] mampu mengekstrak bagian-bagian yang relevan dari sejumlah besar informasi" tetapi menuntut para pihak untuk menyepakati "istilah pencarian dan pengkodean yang mereka gunakan" dan bergantung pada penilaian hakim dan konfirmasi persetujuan.

Kementerian Kehakiman di Portugal, yang juga berfokus pada pengorganisasian informasi prosedural, sedang mengembangkan proyek "Magistratos", khususnya yang berfokus pada pengadilan administrasi, pajak, dan yudisial. Sistem ini bertujuan untuk menyediakan "antarmuka unik bagi para hakim (termasuk jaksa penuntut umum), yang memungkinkan pengindeksan dokumen dan informasi yang merupakan bagian dari perkara yudisial", sehingga memungkinkan "pencarian dokumen dan isinya dengan cepat". "Magistratos" merupakan proyek percontohan yang dapat menyediakan beberapa perangkat tentang cara mengorganisasikan informasi prosedural.

Lebih lanjut, dalam topik ini dan dengan mempertimbangkan kedua sistem ini dapat juga disamakan dengan pengembangan sistem yang mampu mendeteksi dugaan fakta-fakta yang bertepatan (dalam pembelaan tertulis) antara berbagai pihak dalam litigasi.

Hal ini akan membantu hakim untuk menyelesaikan fakta-fakta yang disepakati antara para pihak, sehingga tidak dapat menjadi masalah pembuktian dalam sidang putusan, misalnya. Alat ini akan menghemat waktu prosedural dan akan memungkinkan kontrol hakim mempertahankan dan memperdalam independensi dan imparsialitasnya karena ia dapat menganalisis semua pembelaan para pihak untuk memeriksa dan melengkapi pilihan algoritma dan kemungkinan para pihak untuk mempertanyakan hasilnya, dalam proses hukum yang semestinya, meminta koreksi atau pelengkap dari pilihan sistem.

Sistem Kecerdasan Buatan dalam Memobilisasi Yurisprudensi yang Sudah Ada

Kecerdasan Buatan akan berguna dalam menentukan apakah sudah ada putusan sebelumnya dari pengadilan yang berbeda atau sama yang saat ini sedang menghadapi litigasi dalam kasus dengan kerangka hukum yang serupa. Hal ini akan membantu hakim untuk lebih cepat menemukan putusan yang sudah ada dan memahami apakah landasan hukum dapat dimobilisasi dalam kasus ini.

Dalam hal ini, sistem Kecerdasan Buatan dapat memiliki peran yang lebih menentukan daripada mesin pencari berbasis teknik Boolean yang umum saat ini, yang terkadang dianggap tidak akurat untuk menghasilkan yurisprudensi yang relevan (berdasarkan istilah yang digunakan dalam penelitian hasil). Lebih lanjut, sistem ini juga akan memungkinkan hakim untuk lebih efisien mengakses yurisprudensi sebelumnya yang sebelumnya tidak akan pernah mereka ketahui.

Sistem ini akan menjadi alat instrumental bagi pekerjaan peradilan tanpa menentukan dampak independensi peradilan atau dimensi perlindungan peradilan efektif lainnya karena alat ini harus digunakan berdasarkan prinsip kendali pengguna dan akan berfokus pada pengguna serta ramah pengguna, sebagai perwujudan pendekatan yang berpusat pada manusia pada sistem Kecerdasan Buatan.

Sistem Kecerdasan Buatan dalam Memprediksi Tren Keputusan (Keadilan Prediktif)

Ini mungkin fitur yang paling terlihat dari penggunaan Kecerdasan Buatan dalam ranah peradilan karena teknologi ini "mengklaim mampu memprediksi keputusan pengadilan dan menarik banyak perhatian". Namun, ungkapan "keadilan prediktif" tidak luput dari diskusi karena "hasil dari algoritma prediksi bukanlah keadilan maupun prediktif", yang menentukan bahwa perdebatan terbaru mengarah pada pilihan istilah "prakiraan" karena "hasilnya lebih

mirip ramalan cuaca daripada fakta yang telah ditetapkan" terutama karena "proses pengadilan berisiko memiliki hasil yang tidak dapat diprediksi".

Oleh karena itu, terdapat beberapa aplikasi untuk meramalkan keputusan pengadilan, tetapi sebagian besar dikembangkan dari sudut pandang pengacara. Berfokus pada sudut pandang prosedural, beberapa pengalaman di Amerika Serikat yang diterapkan di pengadilan mungkin menunjukkan bahwa "teknologi dapat memiliki keterbatasan ketika diterapkan di lembaga peradilan".

Faktanya, salah satu isu yang paling banyak dibahas terkait Kecerdasan Buatan adalah fenomena bias dan/atau opasitas algoritma: "algoritma menyajikan hasil yang bias secara berulang" yang utamanya merupakan hasil dari kumpulan data yang digunakan dalam proses pembelajaran algoritma pembelajaran mesin. Faktanya, yang mendefinisikan pembelajaran mesin adalah kemampuan mesin untuk belajar dari dirinya sendiri melalui data yang diberikan kepadanya atau yang dapat dikumpulkannya dari interaksi sehari-hari.

Tidak sulit membayangkan apakah tren algoritma yang bias memengaruhi perangkat Kecerdasan Buatan yang digunakan di bidang peradilan dan, khususnya, hal tersebut dapat berdampak langsung atau tidak langsung pada keputusan pengadilan. Ambil contoh *Correctional Offender Management Profiling for Alternative Sanctions* (dengan akronim COMPAS): banyak studi akademis, di beberapa bidang penelitian, telah merefleksikan bagaimana alat ini dapat menyebabkan beberapa diskriminasi aktual yang memengaruhi kebebasan individu. COMPAS digunakan, dalam praktiknya, di Amerika Serikat, oleh beberapa hakim pidana untuk menilai "risiko residivisme terdakwa atau terpidana, dalam keputusan tentang penahanan pra-persidangan, hukuman atau pembebasan dini" dan menggunakan "data dari catatan kriminal dan dari kuesioner dengan 137 pertanyaan". Namun, ProPublica "sebuah ruang berita nirlaba independen" menerbitkan sebuah artikel pada 23 Mei 2016 yang mengungkap apa yang dianggapnya sebagai "bias mesin" karena COMPAS tampaknya menyebabkan tingkat residivisme yang lebih tinggi yang dikaitkan dengan orang Afrika-Amerika dibandingkan dengan orang Kaukasia. Ada beberapa publikasi yang tidak sependapat dengan kesimpulan ini, seperti NorthPointe, Inc. pengembang COMPAS yang memimpin.

Poin yang ingin disampaikan adalah bahwa bias dan/atau ketidakjelasan algoritma dapat bersifat "transversal" dan, dalam bidang peradilan, "dampak dari keputusan [bias/ketidakjelasan] ini dapat memengaruhi ranah hak-hak warga negara yang paling krusial" karena terdapat "penurunan perlindungan peradilan akibat penggunaan algoritma yang tidak memadai". Lebih jauh lagi, dengan mempertimbangkan kekhawatiran Komisi berdasarkan partisipasi luas para pemangku kepentingan, “aplikasi Kecerdasan Buatan di bidang keadilan sebagai kemungkinan kasus penggunaan berisiko tinggi” akan menjadi aplikasi yang dapat menjadi “bagian dari proses pengambilan keputusan” karena aplikasi tersebut dapat menimbulkan “dampak signifikan pada hak-hak asasi manusia”, dengan mengingat, khususnya, bahwa “ketika pembelajaran mesin digunakan, risiko hasil yang bias dan potensi diskriminasi [...] tinggi”, dengan memberikan “perhatian khusus [...] pada kualitas data pelatihan yang digunakan, termasuk representasi dan relevansinya dalam kaitannya dengan tujuan dan konteks aplikasi yang dimaksud dan bagaimana sistem ini dirancang dan dikembangkan untuk

memastikan bahwa sistem tersebut dapat digunakan dengan sepenuhnya mematuhi hak-hak fundamental".

Dengan memperhatikan ketiga topik utama ini, kita dapat menyimpulkan bahwa sistem Kecerdasan Buatan yang digunakan (1) untuk mengorganisasikan informasi dan (2) untuk memobilisasi yurisprudensi yang telah ada dapat dikategorikan sebagai Sistem Instrumental Kecerdasan Buatan di bidang peradilan karena berperan dalam mempersiapkan prosedur, informasinya, dan, secara lebih luas, rancangan segmen potensial yang akan digunakan dalam landasan hukum (berdasarkan yurisprudensi sebelumnya). Namun, sistem Kecerdasan Buatan (3) untuk meramalkan tren keputusan dapat dipahami sebagai Sistem Keputusan Kecerdasan Buatan karena dapat berdampak langsung pada putusan pengadilan dan, lebih lanjut, pada pembentukan persepsi pribadi hakim terhadap kasus yang diajukan kepadanya. Namun, dalam kedua konsep tersebut, risiko dapat ditimbulkan oleh penggunaan Kecerdasan Buatan, sehingga perlu dipahami apakah sistem tersebut dapat dikategorikan, berdasarkan Proposal Undang-Undang Kecerdasan Buatan, sebagai berisiko tinggi, yang memenuhi kewajiban yang timbul dari klasifikasi tersebut.

21.3 AI RISIKO TINGGI: PENDEKATAN MANUSIA DI PERLINDUNGAN PERADILAN

Kecerdasan Buatan secara hukum dibahas dari pendekatan etika karena berangkat dari gagasan bahwa hak-hak fundamental beroperasi sebagai hak moral dan yuridis yang kepatuhannya memberikan dasar yang menjanjikan untuk mengidentifikasi prinsip-prinsip abstrak dan nilai-nilai etika yang akan dioperasionalkan dalam konteksnya.

Kekhawatiran utama tentang evolusi dan dampak Kecerdasan Buatan pada bidang hukum adalah potensinya untuk memengaruhi "penghormatan dan penegakan hak-hak fundamental yang efektif" yang menuntut pemahaman dampaknya terhadap perlindungan peradilan yang efektif, "termasuk [...] hak atas pemulihan yang efektif dan pengadilan yang adil".

Dalam mendukung Proposal Regulasi untuk Undang-Undang Kecerdasan Buatan, Komisi menekankan bahwa "menjadi kepentingan Uni untuk mempertahankan kepemimpinan teknologi Uni Eropa dan memastikan bahwa warga Eropa dapat memperoleh manfaat dari teknologi-teknologi baru yang dikembangkan dan berfungsi sesuai dengan nilai-nilai, hak-hak dasar, dan prinsip-prinsip Uni Eropa" karena ini adalah proposal "yang didasarkan pada nilai-nilai dan hak-hak dasar Uni Eropa dan bertujuan untuk memberikan kepercayaan kepada masyarakat dan pengguna lain untuk merangkul solusi berbasis AI, sekaligus mendorong bisnis untuk mengembangkannya".

Kekhawatiran akan jaminan penuh atas perlindungan dan proklamasi hak-hak dasar sebelum teknologi-teknologi baru inilah yang membentuk pendekatan yang berpusat pada manusia terhadap Kecerdasan Buatan: "peraturan untuk AI yang tersedia di pasar Uni Eropa atau yang memengaruhi masyarakat di Uni Eropa harus berpusat pada manusia, sehingga masyarakat dapat memercayai bahwa teknologi tersebut digunakan dengan cara yang aman dan sesuai dengan hukum, termasuk penghormatan terhadap hak-hak dasar".

Proposal ini memahami bahwa "penggunaan AI dengan karakteristik spesifiknya (misalnya, ketidakjelasan, kompleksitas, ketergantungan pada data, perilaku otonom) dapat berdampak buruk terhadap sejumlah hak asasi yang tercantum dalam Piagam Hak Asasi Uni Eropa" (CFREU) dan "berusaha untuk memastikan tingkat perlindungan yang tinggi bagi hak-hak asasi tersebut dan bertujuan untuk mengatasi berbagai sumber risiko melalui pendekatan berbasis risiko yang didefinisikan secara jelas".

Dalam hal ini, Proposal untuk Undang-Undang Kecerdasan Buatan tampaknya konsisten dengan rezim hak asasi yang berasal dari CFREU dan "peraturan perundang-undangan sekunder Uni Eropa yang ada". Proposal ini membahas dimensi perlindungan hukum yang efektif secara langsung maupun tidak langsung. Faktanya, terdapat beberapa spesifikasi tentang perlindungan data pribadi yang akan memainkan peran fundamental dalam memungkinkan pengadilan untuk sepenuhnya mengakui hak akses terhadap keadilan, hak pembelaan, hak atas pengadilan yang adil dalam waktu yang wajar dan di hadapan hakim yang tidak memihak dan independen, di antara dimensi-dimensi lain dari perlindungan peradilan yang efektif. Sejauh ini, mengenai perlindungan data pribadi, proposal ini mencoba untuk melengkapi rezim Peraturan Perlindungan Data Umum (GDPR) tetapi juga beroperasi berdampingan dengan, misalnya, Direktif 2016/680 tentang pemrosesan data pribadi untuk keperluan prosedur pidana, karena bertujuan untuk memastikan "kualitas kumpulan data yang digunakan untuk pengembangan sistem AI yang dilengkapi dengan kewajiban untuk pengujian, manajemen risiko, dokumentasi, dan pengawasan manusia di seluruh sistem AI", yang berisi "aturan khusus tertentu tentang perlindungan individu sehubungan dengan pemrosesan data pribadi, terutama memberlakukan pembatasan penggunaan sistem AI untuk identifikasi biometrik jarak jauh 'waktu nyata' di ruang yang dapat diakses publik untuk tujuan penegakan hukum".

Berdasarkan tuntutan perlindungan data, Judul III dari proposal tersebut, mengenai sistem AI berisiko tinggi, "berisi aturan khusus untuk sistem AI yang menimbulkan risiko tinggi terhadap [...] hak-hak fundamental orang perseorangan"; sistem ini akan dapat masuk dan beredar di pasar Uni tetapi akan tunduk pada "kepatuhan terhadap persyaratan wajib tertentu dan penilaian kesesuaian ex-ante".

Di sisi lain, terdapat pula kekhawatiran tentang promosi perlindungan hukum yang efektif secara langsung dan dalam semua dimensinya, sebagaimana yang tertuang dalam pasal 47 CFREU, yang berfokus khususnya pada hak atas penyelesaian yang efektif dan pengadilan yang adil: karena proposal tersebut menciptakan "kewajiban untuk pengujian ex ante, manajemen risiko, dan pengawasan manusia", hal ini akan memfasilitasi promosi perlindungan hak-hak fundamental "dengan meminimalkan risiko keputusan yang dibantu AI yang keliru atau bias di bidang-bidang penting seperti [...] penegakan hukum dan peradilan".

Lebih lanjut, untuk memungkinkan "individu menjalankan hak mereka atas penyelesaian yang efektif dan transparansi yang diperlukan terhadap otoritas pengawasan dan penegakan hukum", beberapa kewajiban transparansi diprediksi tanpa memengaruhi hak kekayaan intelektual karena "akan dibatasi hanya pada informasi minimum yang diperlukan". Pertimbangan 40 dari Proposal menetapkan bahwa "sistem AI tertentu yang ditujukan untuk

administrasi peradilan [...] harus diklasifikasikan sebagai berisiko tinggi, dengan mempertimbangkan potensi dampak signifikannya terhadap demokrasi, supremasi hukum, kebebasan individu, serta hak atas penyelesaian yang efektif dan atas pengadilan yang adil". Lebih lanjut, "untuk mengatasi risiko potensi bias, kesalahan, dan ketidakjelasan, adalah tepat untuk mengkualifikasikan sebagai sistem AI berisiko tinggi yang dimaksudkan untuk membantu otoritas peradilan dalam meneliti dan menafsirkan fakta dan hukum serta dalam menerapkan hukum pada serangkaian fakta yang konkret".

Namun, klasifikasi berisiko tinggi ini dipikirkan sedemikian rupa sehingga tidak akan berlaku untuk semua sistem Kecerdasan Buatan dalam bidang administrasi peradilan: "kualifikasi semacam itu tidak boleh meluas [...] ke sistem AI yang ditujukan untuk kegiatan administratif tambahan murni yang tidak memengaruhi administrasi peradilan yang sebenarnya dalam kasus-kasus individual, seperti anonimisasi atau pseudonimisasi keputusan peradilan, dokumen atau data, komunikasi antara personel, tugas administratif atau alokasi sumber daya". Bahasa Indonesia: Karena pertimbangan digunakan, dalam hukum Uni Eropa, sebagai rujukan interpretasi untuk pasal-pasal yang diabadikan dalam tindakan legislatif yang mereka perkenalkan, ada kebutuhan, dalam pengertian ini, untuk memahami operasi mana yang dapat berada di luar cakupan klasifikasi berisiko tinggi ini dalam domain administrasi peradilan, dengan melakukan analisis konkret dari rezim hukum yang mengalir dari proposal ini. Menurut Pasal 6 (2) Proposal dan Poin 8 (a) dari Lampiran III-nya, sistem Kecerdasan Buatan pada administrasi peradilan dicirikan sebagai berisiko tinggi.

Dalam pengertian ini, sistem berisiko tinggi harus mematuhi sistem manajemen risiko: berdasarkan Pasal 9 (2) itu terdiri dari "proses berulang yang berkelanjutan yang dijalankan di seluruh siklus hidup sistem AI berisiko tinggi, yang memerlukan pembaruan sistematis secara berkala" dan, sebelum sistem ditempatkan di pasar, pemenuhan "kewajiban ketat" karena ada persyaratan untuk penilaian kesesuaian ex ante (sebagai salah satu fitur inovatifnya yang hebat): (a) "sistem penilaian dan mitigasi risiko yang memadai"; (b) "kualitas tinggi dari kumpulan data yang menjadi sumber daya sistem untuk meminimalkan risiko dan hasil yang diskriminatif"; (c) "pencatatan aktivitas untuk memastikan ketertelusuran hasil"; (d) "dokumentasi terperinci yang menyediakan semua informasi yang diperlukan tentang sistem dan tujuannya bagi otoritas untuk menilai kepatuhannya"; (e) "kormasi yang jelas dan memadai bagi pengguna"; (f) "penggunaan langkah-langkah pengawasan manusia yang tepat untuk meminimalkan risiko"; dan (g) "kotensi, keamanan, dan akurasi tingkat tinggi". Ketika sistem ini mengandalkan teknik yang melibatkan pelatihan model dengan data, terdapat persyaratan yang sangat rumit untuk pelatihan, validasi, dan pengujian kumpulan data (Pasal 10).

Dokumentasi teknis mengenai sistem harus disusun sebelum sistem dipasarkan (Pasal 11) dan catatan harus disimpan selama sistem beroperasi (Pasal 12), dengan tetap memenuhi tuntutan "transparansi [...]; pengawasan manusia; serta ketahanan, akurasi, dan keamanan". Lebih lanjut, "sebuah inovasi penting lainnya adalah mandat untuk sistem pemantauan pasca-pemasaran guna mendeteksi masalah yang sedang digunakan dan memitigasinya".

Klasifikasi risiko yang diadopsi dalam proposal ini dianggap mencapai tujuan: "menghindari beban berlebihan bagi sistem AI dan pelaku pasar terkait yang tidak mewakili risiko relevan pada tingkat hak asasi manusia". Dalam hal ini, dengan mempertimbangkan sistem Kecerdasan Buatan dalam pengorganisasian informasi, dan mempertimbangkan pengembangan proyek "Magistratos" oleh Pemerintah Portugal, perlu dipahami apakah sistem secara keseluruhan akan diklasifikasikan sebagai berisiko tinggi. Bahkan, sesuai dengan deskripsi kelembagaannya, proyek ini dianggap sebagai antarmuka yang tersedia bagi hakim dan jaksa untuk mengindeks dokumen dan informasi.

Dengan mempertimbangkan bagian sistem ini, sistem ini dapat dipahami sebagai sekadar pelepas tugas administratif dan, oleh karena itu, tidak memerlukan klasifikasi berisiko tinggi. Namun, "Magistratos" juga bertujuan untuk memungkinkan pencarian dokumen dan konten secara cepat. Oleh karena itu, jika kedua operasi utama ini dapat memengaruhi penerapan keadilan pada suatu kasus individual, sistem tersebut akan dipahami sebagai sistem berisiko tinggi dan harus memenuhi semua persyaratan yang tersirat dalam proposal tersebut.

Dalam kategori yang sama mengenai sistem pengorganisasian informasi, sebuah proposal diajukan: sebuah sistem yang dapat menganalisis dan mendeteksi dugaan fakta-fakta yang bertepatan dalam pembelaan tertulis para pihak, yang mendukung pilihan hakim tentang fakta-fakta mana yang telah disepakati antara para pihak prosedural dan tidak akan bergantung pada bukti-bukti lebih lanjut. Sebaliknya, sistem ini seharusnya diklasifikasikan sebagai berisiko tinggi sejak awal karena, jika dianggap mandiri oleh hakim, sistem ini akan membahayakan independensi hakim dan dapat memengaruhi hak atas peradilan yang adil. Dalam konteks ini, selain analisis kesesuaian ex ante dan pemantauan ex post, klasifikasi berisiko tinggi juga akan berperan penting untuk menetapkan aturan yang lebih ketat terkait transparansi, pengawasan dan akurasi manusia, serta ketahanan dan keamanan sistem.

Fitur-fitur ini yang harus dipenuhi oleh sistem Kecerdasan Buatan adalah fitur yang memfokuskan kembali bagaimana dasar pengaturan hukum untuk Kecerdasan Buatan adalah pendekatannya yang berpusat pada manusia, fitur yang memperoleh signifikansi yang lebih luas ketika ranah peradilan sedang dibahas. Memungkinkan (dan bahkan memaksakan) keterlibatan dan kendali manusia, pentingnya hakim yang independen dan tidak memihak dalam membela hak atas pengadilan yang adil dan pelaksanaan semua hak prosedural yang diberikan kepada para pihak yang berperkara menjadi sepenuhnya terpenuhi. Dalam menghadapi sistem Kecerdasan Buatan seperti itu, hal ini tidak dapat berbeda.

Sejauh ini, mengenai sistem Kecerdasan Buatan dalam mengorganisasikan informasi, sebagai Sistem Instrumental, mereka dapat tampak tidak menimbulkan masalah langsung terhadap perlindungan peradilan yang efektif, terutama karena mereka dapat dikembangkan terutama untuk meringankan beban administratif dan untuk menyajikan solusi yang sebagian besar berdampak pada fungsi administratif yang dikembangkan dalam peradilan. Namun, mereka juga dapat fokus untuk menghindari pekerjaan berulang dan / atau mekanis yang dilakukan oleh hakim. Ketika hal ini terjadi, sistem tersebut (atau bagian darinya) dapat diklasifikasikan sebagai berisiko tinggi; Karena diklasifikasikan demikian, tonik sistem ini harus

difokuskan kembali pada pendekatan antropologis Kecerdasan Buatan di mana model yang berpusat pada manusia harus diutamakan:

- Di satu sisi, memungkinkan hakim untuk sepenuhnya menjalankan tugas mereka secara independen dan tidak memihak. Hal ini hanya dapat dicapai jika mereka siap untuk penggunaan sistem Kecerdasan Buatan, melalui literasi digital aktif karena sistem ini harus "dirancang dengan kemudahan penggunaan" dan memberdayakan pengguna, yang hanya dapat dicapai jika tuntutan transparansi dipenuhi dan hakim menyadari operasi teknis yang dilakukan;
- Di sisi lain, klasifikasi berisiko tinggi juga akan memungkinkan pengacara untuk memahami intervensi sistem Kecerdasan Buatan sehingga, jika diperlukan, mereka dapat sepenuhnya menjalankan hak untuk menuntut dan hak pembelaan para pihak yang mereka wakili. Dalam hal ini, literasi digital aktif dan pasif diperlukan (i) literasi aktif bagi para pengacara karena mereka memainkan peran vital sebagai operator peradilan, yang menuntut mereka untuk juga memahami cara kerja di balik sistem Kecerdasan Buatan dan, sejauh itu, mampu mempertanyakan "penalaran"-nya dalam pembelaan dan tuduhan mereka; dan (ii) literasi pasif bagi masyarakat umum agar orang-orang yang harus pergi ke pengadilan juga menyadari penggunaan sistem Kecerdasan Buatan dan bagaimana sistem tersebut dapat memengaruhi hak-hak mereka, yang berarti mereka harus bereaksi tepat waktu. Lebih lanjut, klasifikasi risiko tinggi seharusnya meningkatkan rasa percaya terhadap sistem peradilan.

Sistem Kecerdasan Buatan dalam memobilisasi yurisprudensi kasus yang sudah ada sebelumnya dianggap sebagai sistem yang dapat mendeteksi yurisprudensi sebelumnya dan, jika memungkinkan, memajukan segmen landasan hukum yang dapat berguna untuk kasus yang tertunda di pengadilan. Dalam skenario ini, sistem Kecerdasan Buatan akan diklasifikasikan sebagai berisiko tinggi karena dapat berdampak langsung pada perilaku hakim di seluruh prosedur dan hanya akan cukup andal jika menjalani penilaian sebelum ditempatkan di pasar dan tunduk pada tuntutan transparansi kepada pengguna (berdasarkan Pasal 13 Proposal), pengawasan manusia (berdasarkan Pasal 14 Proposal) dan persyaratan akurasi, ketahanan dan keamanan siber (berdasarkan Pasal 15 Proposal).

Faktanya, perlindungan peradilan yang efektif terutama dimensi mengenai putusan yang dikeluarkan tepat waktu dapat ditingkatkan oleh sistem seperti ini; Namun, sistem ini hanya dapat berfungsi selama hakim dan pengacara para pihak mampu memahami cara kerja sistem, mengapa hasil tersebut (dan bukan yang lain) yang diangkat, dan bahwa sistem tersebut semata-mata bertindak sebagai sarana untuk mengorganisasikan cuplikan putusan sebelumnya tanpa merusak persepsi hakim secara keseluruhan terhadap litigasi dan evaluasi bebas atas bukti yang diajukan.

Lebih lanjut, cara pengorganisasian cuplikan tersebut jika sistem juga menyajikannya dengan cara ini kepada hakim harus dapat dipahami secara intuitif oleh pengguna (dengan pendekatan yang berfokus pada pengguna dan ramah pengguna) dan di bawah kendali mereka. Dalam konteks ini, hasil yang terorganisir dapat ditampilkan dalam format yang dapat

diedit, memungkinkan penolakan sebagian dan/atau keseluruhan, serta dapat disalin dan ditempel di berkas lain, tempat putusan hakim dibangun dari awal.

Terakhir, dengan berfokus pada sistem Kecerdasan Buatan dalam meramalkan tren keputusan, kekhawatiran Komisi Eropa perlu diingat: Komisi Eropa menyadari risiko yang ditimbulkan oleh pengambilan keputusan otomatis (keadilan prediktif) karena memiliki potensi paling besar dalam memengaruhi perlindungan hukum yang efektif. Sebagai sebuah Sistem Keputusan, klasifikasi risiko tinggi tidak diragukan lagi merupakan satu-satunya cara untuk mengesampingkan risiko tersebut karena sistem ini mampu "memengaruhi administrasi peradilan yang sebenarnya dalam kasus-kasus individual" (Pertimbangan 40 Proposal).

Di sini, pendekatan Kecerdasan Buatan yang berpusat pada manusia harus diingat kembali sebagai landasan etika umum bagi rezim hukum Kecerdasan Buatan di Uni Eropa. Sejauh ini, sistem keputusan tidak dapat beroperasi secara definitif atau tertanam dalam kebenaran yang mutlak dan tak terbantahkan: sistem ini memang mempermudah proses pengambilan keputusan, tetapi tidak bertindak sebagai pengganti intervensi manusia, pengawasan manusia, dan pemahaman umum tentang fungsi sistem. Tindakan-tindakan ini vital agar operator peradilan pada umumnya dapat beroperasi dan menegakkan semua dimensi perlindungan peradilan yang efektif:

- Hakim harus mengendalikan sistem, yang menuntut keterampilan digital untuk dikembangkan dan diterapkan; jika tidak, independensi mereka yang relevan dengan pemeliharaan tatanan hukum berdasarkan aturan hukum dapat terabaikan. Menurut Papan Skor Keadilan Uni Eropa 2021, meskipun beberapa Negara Anggota masih menerapkan sistem Kecerdasan Buatan dalam administrasi peradilan, “tinjauan umum ini menegaskan pengamatan bahwa reformasi peradilan membutuhkan waktu terkadang hingga beberapa tahun sejak diumumkan hingga penerapan langkah-langkah legislatif dan peraturan serta implementasi aktualnya di lapangan”. Di sisi lain, cara kerja sistem harus disajikan dalam istilah yang mudah dipahami oleh mereka, sehingga dimensi yang berfokus pada pengguna dan ramah pengguna terpenuhi. Jika tidak, operasi peradilan yang dimaksudkan bagi hakim seperti menafsirkan hukum, menerapkannya pada kasus-kasus individual, dan memberikan keadilan materiil dapat tertukar dengan operasi teknologi, yang melemahkan fungsi peradilan sebagai salah satu fungsi dasar yang mengendalikan bagaimana kekuasaan publik dijalankan dan dalam memulihkan hak-hak yang telah dilanggar di ranah hukum pihak yang berperkara;
- Para pengacara perlu memahami sistem ini, melalui kebijakan transparansi yang tersedia dan dapat dipahami oleh mereka, sehingga mereka dapat, jika demikian, menggunakan hak pembelaan dan banding. Hak-hak ini, sebelum penerapan sistem Kecerdasan Buatan dalam peradilan, perlu dilaksanakan secara lebih luas agar reaksi para pihak terkait tantangan digital yang berasal dari penggunaan teknologi disruptif dalam peradilan juga dapat didiskusikan. Hal ini menuntut interpretasi yang realistis atas landasan hukum pelaksanaan hak-hak tersebut agar argumen terkait tantangan digital juga dapat dibahas. Oleh karena itu, prinsip-prinsip yang berfokus pada

pengguna dan ramah pengguna dapat memainkan peran penengah: jika sistem Kecerdasan Buatan mudah dipahami dan dianggap memberdayakan penggunanya (hakim), sistem tersebut juga akan lebih mudah dipahami oleh pengacara, meskipun bukan pengguna langsungnya.

21.4 ETIKA AI BERPUSAT MANUSIA DALAM E-JUSTICE EROPA

Seperti yang dapat kita lihat, menganggap sistem Kecerdasan Buatan sebagai sistem yang dapat dipercaya secara etis memungkinkan pendekatan hukum yang berpusat pada manusia yang menetapkan bahwa sistem tersebut harus "memberikan ruang bagi pilihan manusia, yang hanya dapat diperoleh melalui jaminan bahwa manusia akan mampu memahami, mengawasi, dan mengendalikan proses hukum yang melekat pada fungsi Kecerdasan Buatan". Bahkan, Badan Hak Asasi Fundamental Uni Eropa (FRA) memahami perlindungan hukum yang efektif (yang tercantum dalam Pasal 47 CFREU) sebagai "salah satu hak Piagam yang paling sering digunakan dalam proses hukum", yang juga mencakup "keputusan yang diambil dengan dukungan teknologi AI".

Dalam hal ini, dan setelah memperjelas bagaimana “penggunaan AI dapat menantang hak atas penyelesaian yang efektif dengan berbagai cara”, khususnya ketika pihak-pihak yang terlibat tidak memiliki informasi dan/atau pengetahuan untuk memahami fungsi sistem: “tidak memiliki akses ke informasi ini, individu mungkin tidak dapat membela diri, menetapkan tanggung jawab atas keputusan yang memengaruhi mereka, mengajukan banding atas keputusan yang berdampak negatif terhadap mereka, atau mendapatkan pengadilan yang adil, yang mencakup prinsip kesetaraan senjata dan proses yang bersifat adversarial”.

Sebagaimana ditekankan sebelumnya, masalah-masalah ini hanya dapat diatasi jika literasi digital dipromosikan, juga terkait penggunaan sistem Kecerdasan Buatan dalam peradilan. Bahkan, sebagaimana disebutkan FRA, “menantang keputusan berdasarkan AI sangat penting untuk menyediakan akses terhadap keadilan”, yang hanya dapat terjadi jika “masyarakat menyadari bahwa AI digunakan”, “masyarakat menyadari bagaimana dan di mana harus mengajukan keluhan”; dan “sistem AI dan keputusan yang didasarkan pada AI dapat dijelaskan”.

Lebih lanjut, Kecerdasan Buatan yang berpusat pada manusia adalah pendekatan dogmatis yang “memperoleh bentuk dan relevansi baru tertentu dalam ranah hukum dan, khususnya, dalam bidang peradilan di mana hak-hak fundamental harus dipenuhi dan ditegakkan, terlepas dari tatanan hukum yang menyertainya”: oleh karena itu, selama sistem Kecerdasan Buatan “dipikirkan dan diimplementasikan, bukan untuk menggantikan operator peradilan, melainkan untuk memfasilitasi tugas-tugas mereka, terutama yang bersifat repetitif atau membandingkan yurisprudensi yang telah mapan”, hal ini mungkin tidak tampak sebagai hambatan bagi perlindungan peradilan yang efektif. Kesulitan akan muncul, tetapi harus ditangani sedemikian rupa sehingga cara kerja sistem diarahkan secara eksklusif untuk bertindak sebagai instrumen fungsi peradilan, bukan mengambil peran utama dalam pencapaiannya yang harus terus dilakukan semata-mata oleh operator peradilan, yang didukung sepenuhnya oleh informasi yang transparan dan secara digital menyadari

kekurangan dan potensi sistem. Dalam hal ini, Piagam Etika Eropa tentang penggunaan Kecerdasan Buatan dalam sistem peradilan dan lingkungannya, yang disusun oleh Komisi Eropa untuk Efisiensi Keadilan (CEPEJ) Dewan Eropa, merupakan alat yang berguna bagi Uni Eropa: faktanya, Kelompok Kerja e-Law (e-Justice) Dewan Eropa telah beberapa kali mengadakan reuni di Brussels, yang membahas topik "Penggunaan teknologi yang inovatif kecerdasan buatan", dan pada tahun 2019, Piagam Etika tersebut dipresentasikan.

Dalam hal ini, berdasarkan Pasal 6 (3) Perjanjian Uni Eropa (TEU), perlindungan hak-hak fundamental di Uni Eropa harus mengikuti standar minimum yang dikembangkan berdasarkan Konvensi Hak Asasi Manusia Eropa (ECHR) Dewan Eropa, sebagai standar interpretasi. Dalam hal ini, Piagam Etika ini telah mampu memengaruhi cara Uni Eropa memandang sistem Kecerdasan Buatan yang digunakan dalam bidang peradilan.

Sejauh ini, Piagam Etika ini diadopsi untuk menyadarkan “para pemangku kepentingan publik dan swasta yang bertanggung jawab atas desain dan penyebaran alat dan layanan kecerdasan buatan yang melibatkan pemrosesan keputusan dan data peradilan”, juga melibatkan “para pengambil keputusan publik yang bertanggung jawab atas kerangka legislatif atau peraturan, pengembangan, audit atau penggunaan alat dan layanan tersebut”.

Sejalan dengan apa yang telah kami uji sebelumnya, Piagam ini mengakui bahwa “ketika perangkat kecerdasan buatan digunakan untuk menyelesaikan sengketa atau sebagai perangkat untuk membantu pengambilan keputusan peradilan atau untuk memberikan panduan kepada publik, penting untuk memastikan bahwa perangkat tersebut tidak merusak jaminan hak untuk mengakses hakim dan hak atas pengadilan yang adil”.

Lebih lanjut, mengenai non-diskriminasi, tujuannya adalah untuk menghindari bias dalam proses peradilan yang dapat menyebabkan keputusan diskriminatif: karena metode Kecerdasan Buatan masih berjuang dengan masalah mereproduksi diskriminasi yang ada “melalui pengelompokan atau klasifikasi data yang berkaitan dengan individu dari kelompok individu”, ada kebutuhan untuk “menegakkan” langkah-langkah korektif untuk membatasi atau, jika memungkinkan, menetralkan risiko ini dan juga untuk meningkatkan kesadaran di antara para pemangku kepentingan”.

Mengenai kualitas dan keamanan, CEPEJ menggarisbawahi bahwa “data berdasarkan keputusan peradilan [...] harus berasal dari sumber yang tersertifikasi dan tidak boleh dimodifikasi sampai benar-benar digunakan oleh mekanisme pembelajaran”. Dalam hal ini, kekhawatiran ini termasuk dalam klasifikasi risiko tinggi, ketika Pasal 15 Proposal menetapkan rezim mengenai akurasi, ketahanan, dan Tuntutan keamanan siber. Hal ini khususnya berkaitan dengan bidang peradilan karena data yang andal akan mendorong hakim dan putusan yang independen dan imparsial, yang mengarah pada pemenuhan hak atas peradilan yang adil.

CEPEJ juga menggambarkan transparansi, imparsialitas, dan keadilan sebagai prinsip umum yang harus dipenuhi ketika menganalisis sistem Kecerdasan Buatan di bidang peradilan: dalam hal ini, meskipun perlu menyeimbangkan tuntutan kekayaan intelektual, tuntutan tersebut dapat dipenuhi melalui rezim rahasia dagang yang sudah berlaku di pengadilan; lebih lanjut, "praktik terbaiknya adalah mempromosikan literasi digital pengguna peradilan dan meningkatkan kesadaran akan dampak kecerdasan buatan pada bidang peradilan" karena

"sistem ini juga dapat dijelaskan dalam bahasa yang jelas dan familiar dengan mengomunikasikan [...] sifat layanan yang ditawarkan, perangkat yang telah dikembangkan, kinerja, dan risiko kesalahan".

Pendekatan serupa telah dipenuhi berdasarkan Pasal 13 Proposal, ketika menetapkan langkah-langkah apa Sistem Kecerdasan Buatan berisiko tinggi harus dipenuhi, terutama dalam membuat sistem lebih jelas bagi pengguna (lebih transparan) dan menyediakan lebih banyak informasi kepada pengguna. Terakhir, CEPEJ juga menekankan prinsip di bawah kendali pengguna: prinsip ini bertujuan menjadikan operator peradilan sebagai "aktor yang terinformasi dan mengendalikan pilihan mereka"; untuk mencapainya, "pengguna harus diinformasikan dalam bahasa yang jelas dan mudah dipahami, apakah solusi yang ditawarkan oleh perangkat kecerdasan buatan tersebut tepat atau tidak, dari berbagai pilihan yang tersedia", yang menjadi sangat penting bagi operator peradilan dan, terutama, bagi hakim dan independensi serta imparsialitas mereka; tetapi juga bagi para pihak, karena mereka harus memahami "ia berhak atas nasihat hukum dan hak untuk mengakses pengadilan".

Konstruksi ini juga menjelaskan mengapa Komisi Eropa, dalam Proposal Undang-Undang Kecerdasan Buatan, memperkenalkan Pasal 13 tentang Pengawasan Manusia dalam rezim klasifikasi risiko tinggi. Dalam hal ini, semua perkembangan ini membantu Uni Eropa untuk menetapkan klasifikasi risiko tinggi sebagai tren umum dalam ranah administrasi peradilan. Faktanya, meskipun semua perlindungan hukum yang efektif dapat diamati lebih lanjut dengan menggunakan sistem Kecerdasan Buatan, terdapat masalah-masalah yang diketahui dan tersembunyi yang dapat melemahkannya:

- Akses terhadap keadilan dan hak pembelaan dapat memperoleh manfaat dari Kecerdasan Buatan selama “literasi digital tersedia dengan baik dan metode transparansi diadopsi” karena sistem yang disruptif ini memungkinkan penggugat dan tergugat untuk mengumpulkan “sejumlah informasi kuantitatif (misalnya, jumlah keputusan yang diproses untuk mendapatkan skala) dan informasi kualitatif (asal keputusan, representasi sampel yang dipilih, distribusi keputusan di antara berbagai kriteria seperti konteks ekonomi dan sosial) yang dapat diakses oleh warga negara dan, yang terpenting, oleh para pihak dalam persidangan untuk memahami bagaimana skala telah dibangun, untuk mengukur batas-batas yang mungkin, dan untuk dapat memperdebatkannya di hadapan hakim”.
- Mengenai independensi dan imparsialitas peradilan, sistem Kecerdasan Buatan tidak boleh bergantung pada perangkat keras atau perangkat lunak yang menempatkan “hakim di bawah tekanan eksternal apa pun”. Dalam hal ini, pengawasan manusia yang berasal dari Pasal 14 Proposal, dipadukan dengan tuntutan transparansi, merupakan kunci terbaik untuk lebih mengembangkan independensi peradilan dan imparsialitasnya, terutama yang fundamental bagi pemeliharaan Uni hukum.
- Terakhir, hak atas peradilan yang adil akan terpenuhi jika para pihak mampu memahami bahwa sistem Kecerdasan Buatan hanya bertindak sebagai instrumen fungsi peradilan, tetapi tidak pernah, dalam waktu singkat, beroperasi sebagai pengganti pengadilan. Dalam hal ini, kendali manusia atas sistem yang dipadukan dengan fitur-fitur sistem

yang berfokus pada pengguna dan mudah digunakan akan memungkinkan terciptanya sensitivitas ini. Digitalisasi adalah tren baru bagi sistem peradilan e-Justice Eropa merupakan hal penting yang tidak dapat diabaikan karena dampak digital juga harus menjangkau administrasi peradilan agar dapat memanfaatkan keunggulannya. Dalam hal ini, Papan Skor Keadilan Eropa 2021 menggarisbawahi bahwa "mengenai penggunaan teknologi digital oleh pengadilan [...], mayoritas negara anggota telah memiliki berbagai perangkat digital yang [mereka] miliki", meskipun memahami "masih diperlukan kemajuan lebih lanjut mengingat [...] kecerdasan buatan dan perangkat berbasis blockchain semakin tersedia secara luas".

Dalam hal ini, komisi juga dapat memahami bahwa "akses daring ke putusan pengadilan [...] belum mengalami kemajuan, khususnya untuk publikasi putusan di instansi tertinggi" yang juga dapat memengaruhi sistem Kecerdasan Buatan yang beroperasi berdasarkan yurisprudensi sebelumnya, meskipun Papan Skor ini dapat merancang dan menganalisis beberapa "pengaturan yang berlaku di Negara Anggota yang dapat membantu menghasilkan keputusan pengadilan yang dapat dibaca mesin" sehingga "sistem peradilan yang ramah algoritma" dapat diaktifkan: dibandingkan dengan tahun 2019, "sebagian besar Negara Anggota melaporkan peningkatan pada tahun 2020", yang memungkinkan kesimpulan yang ragu-ragu bahwa "sistem peradilan yang telah menerapkan pengaturan untuk pemodelan putusan menurut standar yang memungkinkan keterbacaan mesinnya tampaknya memiliki potensi untuk mencapai hasil yang lebih baik di masa mendatang".

Waktu akan menentukan arah fundamental bagi dampak sistem Kecerdasan Buatan terhadap sistem peradilan; namun, beberapa wawasan telah dibahas, dengan mengingat bahwa pendekatan Kecerdasan Buatan yang berpusat pada manusia tidak dapat dilupakan karena kendali manusia, transparansi, akurasi, dan keamanan merupakan fitur utama yang memungkinkan perlindungan peradilan yang efektif untuk terus berfungsi, dalam sistem peradilan, sebagai dasar dan penjelasan teleologisnya. Untuk mencapainya, sistem Kecerdasan Buatan harus berfokus pada pengguna dan ramah pengguna, jika tidak, akan mempersulit literasi digital bagi operator peradilan dan masyarakat luas.

BAB 22
PENDEKATAN UNI EROPA TERHADAP AI DAN RISIKO SISTEMIK FINANSIAL

Abstrak Tulisan ini mengkaji 'Proposal Uni Eropa untuk Regulasi Parlemen Eropa dan Dewan yang Menetapkan Aturan yang Diharmonisasikan tentang Kecerdasan Buatan' ('Undang-Undang AI') dengan tujuan untuk menentukan sejauh mana undang-undang tersebut menangani risiko sistemik yang diciptakan oleh AI FinTech. Pada akhirnya, dikemukakan bahwa gagasan 'risiko tinggi' yang menjadi inti Undang-Undang AI mengabaikan risiko sistemik keuangan. Pengecualian ini tidak dapat dibenarkan dengan alasan netralitas teknologi, maupun dengan alasan proporsionalitas: risiko sistemik keuangan yang digerakkan oleh AI juga belum tercakup dalam kerangka kerja dan perangkat makroprudensial yang ada (atau yang diusulkan), dan penghilangannya dari Undang-Undang AI juga tidak dapat dibenarkan dengan memprioritaskan jenis risiko lain. Ke depannya, disarankan bahwa Undang-Undang AI Uni Eropa akan diuntungkan dengan definisi 'risiko tinggi' yang lebih luas. Diharapkan juga bahwa para pembuat kebijakan Uni Eropa akan segera mulai memperkuat perangkat makroprudensial yang ada untuk mengatasi risiko sistemik keuangan yang diciptakan oleh AI.

22.1 RISIKO SISTEMIK AI FINTECH DAN KELEMAHAN AI ACT UE

Teknologi dan keuangan telah menjadi hubungan yang tak terpisahkan. Bank-bank lama, perusahaan asuransi, dan lembaga keuangan tradisional lainnya semakin bergantung pada teknologi, banyak perusahaan baru kini berspesialisasi dalam menawarkan aplikasi dan platform keuangan berbasis teknologi, dan bahkan perusahaan teknologi informasi terbesar di dunia (yang disebut 'BigTech') telah mulai memasuki industri jasa keuangan. 'FinTech' istilah yang sering digunakan untuk menggambarkan penggunaan inovatif teknologi modern dalam penyediaan jasa keuangan tampaknya ada di mana-mana.

Dengan kemampuannya yang menjanjikan untuk meningkatkan pencarian dan pemrosesan informasi secara radikal, Kecerdasan Buatan ('AI') berpotensi merevolusi FinTech. Namun, meskipun AI membawa harapan yang signifikan bagi industri jasa keuangan, ia juga menghadirkan risiko yang penting. Yang paling jelas, FinTech berbasis AI seperti kebanyakan jenis FinTech lainnya menimbulkan risiko operasional dan siber. Lebih penting lagi, penggunaan AI dalam layanan keuangan memunculkan tantangan baru yang secara khusus melekat dalam paradigma teknologi AI saat ini seperti representasi pengetahuan, pemrosesan bahasa alami, dan pembelajaran mesin.

Yang terpenting, semakin jelas bahwa FinTech AI secara khusus menimbulkan ancaman tunggal terhadap stabilitas pasar dalam bentuk risiko sistemik keuangan. Bahasa Indonesia: Tidak mengherankan jika AI telah menjadi titik fokus yang menarik bagi para pembuat kebijakan di seluruh dunia, setelah menarik lebih dari 700 inisiatif kebijakan di lebih dari 60

yurisdiksi yang berbeda. Di Uni Eropa ('UE'), inisiatif kebijakan ini telah mencakup, khususnya, Peraturan Perlindungan Data Umum, Paket Undang-Undang Layanan Digital yang terdiri dari proposal untuk Undang-Undang Layanan Digital (Komisi Eropa 2020a) dan proposal untuk Undang-Undang Pasar Digital (Komisi Eropa 2020b) dan, baru-baru ini, Paket Legislatif AI yang mencakup Proposal ambisius untuk Peraturan tentang pendekatan Eropa untuk AI ('Undang-Undang AI') (Komisi Eropa 2021g).

Undang-undang ini secara efektif merupakan seperangkat aturan horizontal yang akan mengatur tidak hanya penggunaan AI yang semakin meningkat dalam industri jasa keuangan, tetapi semua penggunaan AI secara lebih luas. Namun, potensinya untuk mengatasi risiko yang secara khusus diciptakan oleh AI FinTech telah dicatat oleh Uni Eropa.

Tulisan ini mengkaji Undang-Undang AI yang diusulkan Uni Eropa dengan tujuan untuk menentukan sejauh mana undang-undang tersebut mengatasi risiko sistemik yang diciptakan oleh AI FinTech. Pada akhirnya, dikemukakan bahwa gagasan 'risiko tinggi' yang menjadi inti Undang-Undang AI mengabaikan risiko sistemik keuangan. Pengecualian ini tidak dapat dibenarkan dengan alasan netralitas teknologi, maupun dengan alasan proporsionalitas: risiko sistemik keuangan yang digerakkan oleh AI belum tercakup dalam kerangka kerja dan perangkat makroprudensial yang ada (atau yang diusulkan), dan penghilangannya dari Undang-Undang AI juga tidak dapat dibenarkan dengan memprioritaskan jenis risiko lain. Ke depannya, disarankan bahwa Undang-Undang AI Uni Eropa akan lebih bermanfaat jika definisi 'risiko tinggi' yang lebih luas diterapkan.

Diharapkan juga bahwa para pembuat kebijakan Uni Eropa akan segera mulai memperkuat perangkat makroprudensial yang ada untuk mengatasi risiko sistemik keuangan yang diciptakan oleh AI. Pekerjaan kami disusun sebagai berikut: Bagian 2 menentukan sejauh mana penggunaan teknologi AI di sektor keuangan dapat meningkatkan risiko sistemik; Bagian 3 menguraikan fitur-fitur dasar Undang-Undang AI Uni Eropa dan mengevaluasi kemampuannya untuk secara spesifik menangkap risiko sistemik yang diciptakan oleh AI FinTech; Bagian 4 diakhiri dengan memberikan rekomendasi untuk pengembangan regulasi dan pengawasan lebih lanjut di bidang ini.

22.2 PENGGUNAAN AI DALAM KEUANGAN DAN RISIKO SISTEMIK

Peluang dan Risiko FinTech AI

Beberapa tahun terakhir telah menyaksikan peningkatan adopsi teknologi canggih oleh sektor keuangan, dengan AI yang kokoh menjadi inti dari revolusi FinTech. Secara umum, AI digunakan baik dalam operasi front-office (meliputi prosedur yang berhubungan dengan konsumen, nasabah, dan entitas pengawas) maupun operasi back-office (melibatkan prosedur dalam kerangka organisasi perusahaan atau lembaga). Contoh FinTech bertenaga AI meliputi chatbot untuk menjawab pertanyaan klien, platform perdagangan yang menghosting atau menggunakan mekanisme perdagangan algoritmik canggih, penyediaan layanan simpanan dan pinjaman yang didukung oleh kontrak pintar yang berjalan pada protokol blockchain, dan penyampaian laporan regulasi secara otomatis oleh entitas yang diawasi. Secara khusus, AI memungkinkan pelaku keuangan untuk mengumpulkan dan mengurai data dalam jumlah

besar yang kemudian dapat digunakan untuk berbagai keperluan, mulai dari menghitung peringkat kredit dan investasi, hingga mendeteksi praktik penipuan dan ilegal. AI juga dapat digunakan untuk meningkatkan konektivitas antar agen dalam sistem keuangan dan RegTech dan SupTech yang didukung AI dapat mengubah prosedur kepatuhan, regulasi, dan pengawasan secara radikal.

Penggunaan AI dalam keuangan ini menciptakan peluang signifikan untuk meningkatkan efisiensi, keadilan, dan inklusivitas di seluruh sistem keuangan, tetapi juga menghadirkan tantangan penting. Memang, telah lama menjadi jelas bahwa banyak aplikasi teknologi modern rentan terhadap risiko siber dan operasional, menimbulkan masalah privasi data, atau dapat menjadi saluran bias algoritmik dan fitur-fitur khusus AI dapat memperburuk risiko ini. Baru-baru ini, muncul kekhawatiran khusus mengenai dampak AI FinTech terhadap risiko sistemik keuangan.

Dampak AI terhadap Dimensi Lintas Seksi dan Waktu Risiko Sistemik

Klasifikasi tradisional risiko sistemik keuangan mengacu pada dua dimensi. Pertama, 'dimensi lintas seksi (atau struktural)' yang berkaitan dengan bagaimana risiko didistribusikan dalam sistem keuangan pada suatu titik waktu tertentu. Untuk memantau dimensi ini, otoritas makroprudensial harus memperhatikan keterkaitan dan eksposur umum di pasar keuangan. Kedua, 'dimensi waktu' yang berkaitan dengan prosiklikalitas sistem keuangan dan berkaitan dengan bagaimana risiko dan kerentanan agregat terakumulasi seiring waktu dan diperkuat oleh interaksi dalam sistem keuangan serta umpan balik antara sistem keuangan dan ekonomi riil.

Baik perusahaan keuangan maupun individu cenderung menanggung risiko berlebihan saat fase kenaikan (fase boom) dan menjadi penghindar risiko saat fase penurunan (fase bust). Risiko tersembunyi dan kurang dihargai ini biasanya berkembang secara dramatis, yang berpotensi mengarah pada terwujudnya risiko sistemik. Seiring dengan semakin meluasnya penggunaan AI dalam keuangan, pertanyaan kuncinya adalah apakah teknologi berbasis AI dapat memperkuat kedua dimensi risiko sistemik ini.

Mengenai dimensi lintas sektor, interkoneksi merupakan konsep yang cukup intuitif untuk dipahami dan diterapkan dalam konteks ini. Sistem keuangan adalah jaringan lembaga keuangan dan pasar yang saling terhubung. Dalam kondisi normal, interkoneksi ini memfasilitasi pembagian risiko antar lembaga keuangan. Namun, selama periode krisis, interkoneksi yang sama dapat dengan mudah memfasilitasi penyebaran guncangan dan menghasilkan 'efek domino'. Guncangan yang melanda satu pasar, atau satu lembaga, dapat dengan cepat menyebar ke pasar dan lembaga lain yang terhubung dengannya dan berdampak pada sebagian besar sistem keuangan, atau bahkan sistem secara keseluruhan. Demikian pula, ketergantungan yang lebih besar pada teknologi di berbagai platform, perusahaan, dan mitra pihak ketiga yang saling terhubung meningkatkan interkoneksi dan konsentrasi.

Lembaga keuangan sebagian besar mengalihdayakan penggunaan teknologi AI ke sejumlah kecil penyedia teknologi dan layanan pihak ketiga. Memang 'lima yang terkenal' Amazon, Google, Microsoft, Facebook, dan Apple dan rekan-rekan mereka di Tiongkok yang

mendominasi pasar AI sebagian dengan menerapkan strategi akuisisi yang kuat dan dominasi komplementer dalam menyediakan layanan komputasi awan.

Seperti disebutkan sebelumnya, layanan AI dapat meningkatkan efisiensi pasar keuangan. Meskipun demikian, mirip dengan penyedia komputasi awan, penyedia AI, seperti BigTech yang dominan, dapat menjadi penting secara sistemik mengingat interkoneksi mereka dengan lembaga keuangan dan kurangnya pengganti yang tersedia untuk layanan yang mereka sediakan. Pada prinsipnya, penyedia layanan AI, seperti halnya infrastruktur pasar keuangan, dapat dikatakan bertindak sebagai 'pipa sistem keuangan' mengingat penyediaan infrastruktur dan layanan AI platform mereka ke pasar keuangan.

Sayangnya, saat ini, gagasan 'lembaga keuangan yang penting secara sistemik' hampir secara eksklusif diterapkan, dalam praktiknya, pada lembaga keuangan tradisional seperti bank dan perusahaan asuransi. Demikian pula, gagasan 'utilitas pasar keuangan yang penting secara sistemik' diterapkan pada infrastruktur pasar. Bahkan ketika otoritas makroprudensial domestik memiliki kewenangan penunjukan yang dapat diterapkan pada badan hukum tertentu, seperti penyedia AI, dalam grup BigTech, hal tersebut seringkali menghadapi hambatan dan batasan praktis dan hukum. Tantangan ini diperparah oleh kekhawatiran yang ada tentang dominasi pasar dan jejak sistemik BigTech mengingat pengumpulan data pengguna dan kemampuan mereka untuk mengeksploitasi efek jaringan alami.

Dalam lingkungan yang terkonsentrasi dan tanpa pengawasan regulasi langsung, ketergantungan lembaga keuangan pada penyedia AI pihak ketiga untuk layanan inti mereka dapat memperkuat risiko idiosinkratik, yang berpotensi menyebabkan gangguan di seluruh sistem. Misalnya, jika penyedia AI utama terpapar gangguan operasional yang parah, seperti ancaman siber, kegagalan teknologi informasi, proses internal, atau kegagalan kontrol, hal ini dapat menyebabkan gangguan operasional di seluruh sistem secara bersamaan. Dalam kasus ekstrem, konsentrasi pasar penyedia AI juga dapat mengakibatkan 'lock-in', di mana lembaga keuangan terlalu bergantung pada penyedia AI tertentu dan tidak dapat dengan mudah mengganti layanannya karena kurangnya penyedia alternatif yang layak dan/atau kurangnya interoperabilitas layanan.

Penyedia AI sendiri mungkin bergantung pada penggunaan dan layanan penyedia komputasi awan, sektor disruptif lain yang dapat menimbulkan risiko bagi stabilitas sistem keuangan. Penyedia layanan komputasi awan global, seperti Amazon dan Microsoft, sering kali juga menyediakan produk dan layanan AI yang dikenal sebagai Kecerdasan Buatan sebagai Layanan (‘AIaaS’). Pertumbuhan AIaaS bersifat eksponensial dan diperkirakan akan meningkat sebesar 41% selama tahun 2021–2025. Kombinasi AI dengan layanan komputasi awan dan konsentrasi penyedianya mengintensifkan risiko yang melekat pada kedua teknologi disruptif tersebut terutama karena penyedia AI ini sering kali berada di luar jangkauan regulasi dan, oleh karena itu, tidak tunduk pada regulasi mikroprudensial yang menjamin keamanan dan keandalannya. Lebih lanjut, data yang tersedia bagi otoritas makroprudensial mengenai eksposur lembaga keuangan kepada penyedia pihak ketiga tidak lengkap. Batasan regulasi mungkin belum memungkinkan otoritas makroprudensial untuk mengumpulkan data yang tepat waktu, komprehensif, dan sebanding yang dapat diagregasi untuk analisis

makroprudensial. Ketidakjelasan ini mendukung penerapan kerangka kerja regulasi yang komprehensif dan lintas batas, sebagaimana dicatat dalam Bagian 4 di bawah ini.

Dalam dimensi risiko sistemik lintas sektor, guncangan juga dapat menyebar melalui eksposur umum, yaitu eksposur lembaga keuangan terhadap sumber risiko yang sama. Misalnya, lembaga keuangan atau perusahaan lain yang menyediakan jasa keuangan dapat terpapar faktor risiko dan praktik atau model manajemen risiko yang serupa. Eksposur umum ini muncul karena kesamaan dan homogenitas di seluruh lembaga keuangan yang menciptakan kemungkinan kegagalan simultan bersama. Dengan demikian, ketergantungan pada model atau algoritma AI standar, yang dilatih pada aliran data serupa, dapat menghasilkan herding dan keseragaman prediksi dan perilaku di pasar keuangan.

Hal ini sangat mengkhawatirkan karena lembaga keuangan sudah menggunakan sistem AI untuk penetapan harga aset, pemodelan risiko kredit, dan pemantauan risiko dan akan semakin bergantung pada sistem ini, bukan sebagai sistem pelengkap tetapi sebagai pengganti sistem yang sudah ada dan dipantau manusia. Selain itu, mengadopsi sistem regulasi dan pengawasan yang dirancang khusus untuk AI secara tidak sengaja dapat menyebabkan karakteristik umum pada sistem AI, homogenitas, dan keseragaman model.

Hal ini bukan sekadar kekhawatiran teoretis karena paparan umum terhadap praktik dan model manajemen risiko serupa telah terbukti membawa bencana menjelang krisis keuangan 2007–2009, ketika standar modal minimum Komite Basel untuk bank-bank internasional sangat bergantung pada penilaian risiko lembaga pemeringkat kredit. Melihat ke belakang, menjadi jelas bahwa lembaga-lembaga ini memiliki insentif untuk meningkatkan penilaian kredit dan model mereka gagal menilai risiko secara akurat.

Penggunaan sistem AI juga dapat menyediakan lahan subur bagi pembentukan ketidakseimbangan endogen dalam sistem keuangan yang disebarkan dan diperkuat melalui siklus umpan balik antara data dan algoritma. Dalam skenario ini, sistem AI akan menghasilkan keputusan dan data algoritma yang kemudian akan digunakan untuk memperbarui model. Model-model ini akan menghasilkan lebih banyak data dan akan mengadaptasi keputusan serta prediksi mereka berdasarkan data tersebut dalam siklus umpan balik yang dinamis dan otonom. Siklus yang tidak dapat diprediksi ini, pada dasarnya, dapat berbahaya dan memperkuat risiko yang sudah ada di pasar keuangan.

Kekhawatiran tentang siklus umpan balik antara data dan algoritma sangat akut karena tiga alasan utama. Pertama, penelitian tentang efek siklus umpan balik masih dalam tahap awal, sehingga sulit untuk dipantau, apalagi dikendalikan. Hal ini sejalan dengan kekhawatiran yang lebih luas tentang kelangkaan tenaga ahli dan tantangan lembaga keuangan serta otoritas regulasi dan pengawasan untuk merekrut dan mempertahankan personel yang berkeahlian tinggi. Kedua, penyedia layanan AI menggunakan 'data alternatif' seperti data tidak terstruktur, data sintetis, dan data agregat.

Memproses 'data alternatif' dan menggunakannya untuk menginformasikan keputusan kebijakan, dalam praktiknya, jauh dari mudah. Data tidak terstruktur harus melalui proses pembersihan untuk menghilangkan kesalahan dan inkonsistensi serta memastikan penggunaannya yang efektif; data sintetis harus menangkap dan secara akurat

merepresentasikan data asli yang sebenarnya, sementara data agregat terkadang harus divalidasi tanpa mengetahui struktur granularnya. Tidak adanya standar kualitas data yang disesuaikan untuk AI semakin memperburuk ketidakpastian siklus umpan balik algoritma data. Ketiga, sistem AI dilatih berdasarkan peristiwa masa lalu.

Keluaran awal, pola, dan indikator yang ditetapkan oleh regulator dibentuk oleh manusia yang mungkin secara alami memiliki pandangan sempit dan berpandangan ke belakang tentang risiko sistemik. Oleh karena itu, terdapat risiko nyata bahwa 'pelatihan berlebihan' berdasarkan peristiwa masa lalu akan mengakibatkan jenis risiko baru terabaikan. Kekhawatiran ini mendorong para ekonom untuk memperingatkan bahaya bahwa sistem AI '... akan fokus pada jenis risiko yang paling tidak penting, yang mudah diukur sementara kehilangan risiko endogen yang lebih berbahaya. Akibatnya, hal itu akan mengotomatiskan dan memperkuat adopsi asumsi yang salah yang sudah menjadi bagian utama dari krisis saat ini.

Dengan demikian, hal itu akan membuat kepuasan diri yang dihasilkan semakin mungkin menumpuk seiring waktu.' Meskipun risiko ini mungkin tidak unik untuk AI, sebagian besar sistem AI yang diterapkan dalam layanan keuangan belum diuji untuk guncangan mendadak pada kondisi pasar, krisis keuangan, dan skenario stres lainnya. Ketika parameter data input tidak sesuai dengan kondisi ini, model mungkin perlu pelatihan ulang. Namun, pelatihan ulang adalah proses yang mahal dan, oleh karena itu, lembaga keuangan cenderung menderita bias tidak bertindak dan memilih untuk menundanya, dengan mengorbankan metode yang salah. Sebagaimana akan kita lihat di bagian selanjutnya, kebutuhan akan regulasi makroprudensial (dan kerangka hukum pendukungnya) yang dapat ‘memaksa’ lembaga keuangan serta penyedia AI untuk menginternalisasi eksternalitas negatif ini belum ditangani oleh Undang-Undang AI Uni Eropa dan saat ini belum ditangani oleh undang-undang sektoral mana pun dan tetap vital dan mendesak.

Tantangan utama lain dari sistem AI yang lazim dalam diskusi akademis adalah masalah kotak hitam. Dalam sistem algoritmik berbasis AI, data masukan dan keluaran (masuk dan keluar) dapat diamati, tetapi operasi internalnya tidak selalu dipahami dengan baik. Sebagai ilustrasi, model AI begitu kompleks sehingga bahkan pembuatnya pun seringkali tidak mampu memahami bagaimana keputusan dirumuskan atau menafsirkan penalaran yang mendukung keluaran tertentu. Oleh karena itu, pengambilan keputusan otomatis menimbulkan kekhawatiran tentang kemampuan menjelaskan atau, dengan kata lain, kekhawatiran bahwa perilaku internal model tidak dapat 'dipahami secara langsung oleh manusia (interpretabilitas)' dan penjelasannya (pembenaran) tidak dapat 'diberikan untuk faktor-faktor utama yang menyebabkan keluarannya.'

Sistem AI juga menimbulkan kekhawatiran tentang auditabilitas karena tidak selalu memungkinkan untuk melakukan evaluasi analitis dan empiris terhadap algoritma. Fitur-fitur ini dapat berdampak negatif pada kapasitas perusahaan keuangan untuk memantau kinerja algoritmik dan memastikan kepatuhan berkelanjutan terhadap persyaratan peraturan. Hal ini, pada gilirannya, dapat mengakibatkan keputusan kredit yang tidak akurat berdasarkan penilaian kelayakan kredit yang salah dan manajemen risiko kredit dan likuiditas yang tidak memuaskan. Kurangnya penjelasan juga memengaruhi kemampuan lembaga keuangan untuk

menyesuaikan strategi mereka di saat stres atau kinerja buruk, yang berpotensi menyebabkan volatilitas pasar, kekurangan likuiditas, dan bahkan kebuntuan selama gejolak keuangan.

Tanpa mengurangi pentingnya masalah kotak hitam, bahaya lain dari AI yang agak terabaikan adalah disparitas dan ketidaksesuaian antara ekspektasi dan target, di satu sisi, dan tujuan serta maksud, di sisi lain. Potensi disparitas antara tujuan regulasi dan pengoperasian sistem AI yang diprogram untuk mengoptimalkan proses harus diakui dan dipantau. Namun, kesulitan untuk meramalkan disparitas ini diilustrasikan dalam buku Yuval Harari, Homo Deus:

> Bahkan memprogram sistem dengan penjara yang tampaknya tidak berbahaya pun dapat menjadi bumerang yang mengerikan. Salah satu skenario populer membayangkan sebuah perusahaan merancang kecerdasan super buatan pertama dan memberinya tes yang tidak berbahaya, seperti menghitung pi. Sebelum siapa pun menyadari apa yang terjadi, AI mengambil alih planet ini, melenyapkan umat manusia, meluncurkan kampanye penaklukan hingga ke ujung galaksi, dan mengubah seluruh alam semesta yang dikenal menjadi komputer super yang selama miliaran tahun menghitung pi dengan semakin akurat. Lagipula. Inilah misi ilahi yang diberikan Sang Pencipta.

Meskipun skenario ini mungkin lebih tampak seperti fiksi ilmiah daripada kenyataan, poin yang perlu digarisbawahi di sini adalah bahwa tujuan regulasi, termasuk stabilitas sistem keuangan dan keamanan serta kesehatan lembaga keuangan, mungkin tidak mudah diselaraskan dan dikendalikan dalam sistem AI.

Jenis disparitas lain dapat muncul antara target optimasi AI yang lugas dan target penggunanya. Hal ini khususnya terjadi pada layanan AI siap pakai yang menawarkan model AI yang sudah dilatih oleh penyedia atau pihak lain kepada pengguna dan menghilangkan kebutuhan untuk menyiapkan, melatih, dan mengelola produk secara aktif. Layanan ini menawarkan 'abstraksi kompleksitas' kepada pengguna dan hemat biaya, tetapi juga menyerahkan kendali dan tanggung jawab layanan kepada penyedia AIaaS. Dengan demikian, penyedia tidak akan tahu banyak tentang model bisnis, praktik, dan target pengguna, dan pengguna, pada gilirannya, tidak akan tahu banyak (atau tidak tahu sama sekali) tentang pengaturan dan konfigurasi sistem AI. 'Selubung' antara penyedia AI dan pengguna meningkatkan risiko bahwa optimalisasi target tidak akan memenuhi harapan dan menghambat kemampuan untuk menyesuaikan layanan dengan kebutuhan spesifik perusahaan.

Terakhir, penggunaan AI juga dapat memengaruhi regulator dan kepatuhan perusahaan keuangan terhadap regulasi. Sebagaimana telah disebutkan sebelumnya, sistem AI semakin banyak digunakan dalam pembuatan kebijakan dan meskipun teknologi ini memiliki potensi besar, terutama dalam meningkatkan pengawasan risiko sistemik dengan mengotomatiskan analisis makroprudensial dan jaminan kualitas data, penggunaan AI untuk meningkatkan analisis makroprudensial memiliki konsekuensi.

Kurangnya kemampuan menjelaskan dan mengaudit berarti bahwa otoritas makroprudensial mungkin tidak dapat memahami bagaimana model AI sampai pada keputusan atau prediksinya, bagaimana peristiwa yang tidak diinginkan terjadi, dan bagaimana

merespons serta memitigasi risiko yang telah terwujud atau mencegah risiko muncul di masa mendatang.

Dengan demikian, otoritas makroprudensial mungkin tidak dapat mengomunikasikan alasan yang mendukung keputusan kebijakan mereka kepada publik atau parlemen, sehingga transparansi dan akuntabilitas mereka dapat terdilusi. Selain itu, meskipun AI memang dapat digunakan oleh bank untuk memaksimalkan kepatuhan regulasi mereka, misalnya, untuk optimalisasi modal dan meningkatkan profil risiko serta keamanan dan kesehatan mereka, membantu bank untuk 'memanipulasi' sistem secara lebih efisien pada akhirnya dapat meredam dampak standar regulasi kehati-hatian.

Ketika terakumulasi, perilaku strategis lembaga keuangan dapat mengakibatkan eksternalitas negatif dan mengganggu stabilitas sistem keuangan. Yang terpenting, teknik 'permainan' ini dapat meredakan tekanan bagi bank dalam jangka pendek, tetapi belum tentu dirancang dengan tujuan untuk perubahan jangka panjang yang dapat menghasilkan hasil yang lebih berkelanjutan. Ini hanyalah puncak gunung es. Pada kenyataannya, dampak risiko yang seharusnya menjadi perhatian regulator jauh melampaui sistem keuangan. Hingga saat ini, fokus regulasi pada hubungan sistem keuangan-ekonomi riil terbatas pada memastikan keberlanjutan alokasi sumber daya sistem keuangan yang efisien dalam menghadapi guncangan dan mencegah potensi dampak negatif pada ekonomi riil.

Seiring sistem AI semakin memperkuat tidak hanya sistem keuangan tetapi juga energi, militer, dan transportasi, kerusakan pada sistem tersebut akan benar-benar sistemik dan berpotensi merusak. Risiko yang melekat pada sistem AI akan berasal dari ekonomi riil dan ekosistem, yang meluas ke segmen lain dalam ekonomi dan sistem keuangan. Bahaya yang relatif jauh ini belum luput dari perhatian para pengawas. Dewan Risiko Sistemik Eropa, misalnya, mengamati bahwa 'AI dapat digunakan untuk menyerang, memanipulasi, atau merugikan perekonomian dan mengancam keamanan nasional melalui sistem keuangannya secara langsung dan/atau dampaknya terhadap perekonomian yang lebih luas.' Misalnya, algoritma dapat dimanipulasi dalam upaya mentransfer kekayaan ke kekuatan asing, melemahkan pertumbuhan ekonomi dalam upaya menciptakan keresahan, atau mengirimkan sinyal yang salah kepada unit perdagangan untuk memicu krisis sistemik’ sebuah skenario yang benar-benar merupakan kasus ‘unknown unknowns’.

Sebenarnya, masih banyak yang belum diketahui tentang sejauh mana dampak FinTech yang digerakkan oleh AI dan meskipun para pembuat kebijakan, pelaku industri, dan pakar tampak waspada terhadap banyak risiko yang diciptakan oleh AI, potensi teknologi yang digerakkan oleh AI untuk memperbesar risiko sistemik masih relatif kurang mendapat perhatian. Dengan Uni Eropa yang memimpin dalam regulasi AI, inilah saatnya untuk menilai apakah Undang-Undang AI-nya telah menangkap risiko sistemik keuangan yang diperbesar oleh FinTech AI.

22.3 PENDEKATAN UNI EROPA TERHADAP AI DAN TANTANGAN RISIKO SISTEMIK

Satu Pendekatan, Dua Pilar

Risiko yang terkait dengan pertumbuhan teknologi AI dan pengembangan beragam aplikasi AI di bidang keuangan36 dan di bidang lainnya tidak luput dari perhatian Uni Eropa. Justru sebaliknya: seiring Uni Eropa memasuki apa yang disebutnya sebagai Dekade Digital Eropa, keinginan untuk memastikan bahwa AI 'mengutamakan manusia' berada di garis depan agenda Uni Eropa.

Memang, pendekatan Uni Eropa yang baru-baru ini dipublikasikan terhadap AI mengungkapkan kekhawatiran yang jelas atas perkembangan aplikasi AI yang tak terkendali dan risikonya, dan telah memilih 'AI yang tepercaya' sebagai salah satu pilar utama kebijakan AI-nya. Pada saat yang sama, Uni Eropa tidak pernah gagal mengenali janji-janji AI, dan keinginannya untuk menciptakan lingkungan yang aman bagi pengguna, pengembang, dan pengguna AI telah diimbangi oleh rasa urgensi dalam memperkuat kemampuan Eropa untuk bersaing dalam lanskap AI global. Selain 'AI yang tepercaya', kebijakan Uni Eropa akan didukung oleh pilar kedua, yaitu 'keunggulan dalam AI'.

Gambaran yang digunakan Uni Eropa menggugah: pilar biasanya menawarkan dukungan tegak untuk superstruktur dan beberapa pilar secara intuitif menawarkan lebih banyak dukungan daripada satu. Dalam hal ini, tujuan Uni Eropa adalah untuk 'membangun Eropa yang tangguh untuk Dekade Digital' di mana, pada saat yang sama, 'masyarakat dapat menikmati manfaat AI': dengan kata lain, Uni Eropa ingin menjadi pusat kekuatan AI yang ditopang oleh inovasi dan keamanan. Namun, tidak seperti kebanyakan pilar, inovasi dan keamanan tidak selalu saling melengkapi: keduanya seringkali saling bertentangan, dan pilihan yang membuat AI lebih tepercaya dapat mengorbankan keunggulan AI (begitu pula sebaliknya).

Bisa dibilang, komplementaritas dan ketegangan di titik temu pendekatan Uni Eropa terhadap AI menjadi penjelasan yang kuat bagi banyak pilihan regulasi yang membentuk pendekatan tersebut dan khususnya, bagi opsi kunci untuk menangani berbagai jenis risiko yang diciptakan oleh AI secara berbeda, dan beberapa di antaranya bahkan tidak sama sekali. Dalam kerangka kerja ini, mengidentifikasi dan mengenali keberadaan risiko tertentu seperti risiko sistemik yang didorong oleh AI yang dibahas di bagian sebelumnya hanyalah langkah pertama dalam pembuatan kebijakan, dan langkah yang belum tentu diikuti oleh tindakan regulasi untuk memitigasi risiko tersebut. Dengan mempertimbangkan hal tersebut, artikel ini melanjutkan dengan memperkenalkan pendekatan Uni Eropa terhadap AI, membahas sejauh mana pendekatan tersebut menangkap risiko sistemik keuangan yang diperbesar oleh AI.

Pendekatan Uni Eropa terhadap AI

Telah dicatat bahwa pendekatan Uni Eropa terhadap AI bertumpu pada dua gagasan, yaitu keunggulan dan kepercayaan. Gagasan keunggulan dalam AI telah diterjemahkan menjadi kekhawatiran atas pengembangan dan adopsi AI di Eropa, pengembangan lingkungan di mana AI mampu berkembang pesat 'dari laboratorium hingga pasar', dorongan AI sebagai kekuatan untuk kebaikan dalam masyarakat, dan pembangunan kepemimpinan strategis di sektor-sektor utama. Di sisi lain, gagasan AI yang tepercaya mencerminkan kekhawatiran atas

risiko keselamatan yang spesifik untuk teknologi AI, masalah pertanggungjawaban yang berkaitan dengan AI, dan pentingnya memperbarui undang-undang keselamatan sektoral.

Dalam praktiknya, kekhawatiran ini telah mendorong Komisi Eropa untuk menerbitkan paket AI pada bulan April 2021 yang mencakup Komunikasi tentang Pembinaan Pendekatan Eropa terhadap Kecerdasan Buatan (Komunikasi) (Komisi Eropa 2021b), Rencana Terkoordinasi (yang diperbarui) dengan Negara-negara Anggota (Rencana Terkoordinasi) (Komisi Eropa 2021c), dan Undang-Undang AI yang telah dibahas sebelumnya versi kompriminya baru-baru ini telah disetujui oleh Komisi Eropa. Komunikasi tersebut menetapkan fondasi bagi pendekatan Uni Eropa terhadap AI memperluas gagasan bahwa AI membawa peluang dan risiko serta mencatat bahwa 'karakteristik tertentu dari AI ... menimbulkan risiko spesifik dan berpotensi tinggi terhadap keselamatan dan hak-hak fundamental yang tidak dapat ditangani oleh undang-undang yang ada' tetapi pada akhirnya menawarkan sangat sedikit detail tentang bagaimana Uni Eropa berencana untuk menangani apa yang disebut 'dua sisi' AI.

Sebaliknya, baik Rencana Terkoordinasi maupun Undang-Undang AI memberikan wawasan penting tentang apa yang telah disiapkan Uni Eropa untuk AI. Secara umum, Rencana Terkoordinasi (yang diperbarui) merangkum komitmen untuk mendorong kemampuan Eropa dalam bersaing di lanskap AI global, memberikan gambaran umum tentang apa yang telah dilakukan dan mengusulkan rencana untuk tindakan di masa mendatang. Yang terpenting, rencana ini mencatat pentingnya mengembangkan kerangka kebijakan untuk memastikan kepercayaan pada sistem AI dan menyoroti penerbitan Buku Putih (Komisi Eropa 2020c) yang mengusulkan Kerangka Regulasi UE tentang AI ('Buku Putih AI'). Kerangka regulasi ini akan mencakup sejumlah langkah yang mengadaptasi kerangka tanggung jawab Eropa terhadap tantangan teknologi baru (termasuk AI), beberapa revisi terhadap undang-undang keselamatan sektoral yang ada dan, khususnya, Undang-Undang AI yang disebutkan sebelumnya.

Undang-Undang AI inilah yang menawarkan gambaran paling jelas tentang bagaimana AI akan diatur di UE. Dengan panjang lebih dari 80 pasal, undang-undang ini dimaksudkan untuk mengamankan Eropa 'peran utama dalam menetapkan standar emas global' untuk AI dan menetapkan dirinya untuk mengatasi risiko yang dihasilkan oleh penggunaan teknologi AI tertentu. Secara garis besar, hal ini akan dicapai melalui kerangka regulasi horizontal yang menetapkan aturan yang harmonis untuk memperkenalkan dan menggunakan sistem AI di Uni Eropa di berbagai industri dan sektor.

Aturan horizontal lintas sektoral pada dasarnya ambisius, tetapi ambisi di balik proposal Uni Eropa ini telah diimbangi dengan kekhawatiran penting akan keseimbangan. Kekhawatiran tersebut tidaklah salah; melainkan, kekhawatiran tersebut menggambarkan ketegangan inheren antara dua pilar dalam pendekatan Uni Eropa terhadap AI keunggulan dan kepercayaan dan, secara lebih luas, ketegangan inheren antara inovasi dan keamanan yang seringkali mendasari kebijakan regulasi. Namun, penting untuk menentukan apakah Undang-Undang AI Uni Eropa menyelesaikan ketegangan ini secara memuaskan, dengan cara yang memungkinkannya mengatasi dampak AI pada sistem keuangan, yaitu dengan mencegah atau memitigasi risiko sistemik keuangan yang terbukti diperbesar oleh AI.

Kehilangan Kesempatan untuk Mengatur Risiko Sistemik yang Diperburuk oleh AI

Kerangka Regulasi untuk AI yang diusulkan Uni Eropa saat ini sangat sedikit membahas secara spesifik dampak AI terhadap sistem keuangan. Saat ini, satu proposal yang diajukan Uni Eropa dalam kerangka ini Undang-Undang AI hanya menargetkan satu aspek dari dampak tersebut: risiko diskriminasi yang ditimbulkan oleh sistem AI yang mengevaluasi kelayakan kredit perorangan ('penilaian kredit algoritmik').

Hal ini tidak serta merta merupakan kelalaian. Dampak teknologi AI terhadap sistem keuangan telah diakui secara tegas oleh Uni Eropa lebih dari sekali dan tidak serta merta berarti Uni Eropa harus menyetujui aturan baru untuk mengatasi dampak tersebut. Semuanya kembali pada ketegangan antara keunggulan dan kepercayaan inovasi dan keamanan yang diselesaikan oleh Undang-Undang AI melalui dua prinsip: netralitas teknologi dan proporsionalitas. Gagasan netralitas teknologi hadir dengan tegas dalam Komunikasi April 2021 yang meletakkan dasar bagi pendekatan Uni Eropa terhadap AI, dan mencerminkan gagasan bahwa regulasi tidak boleh memaksakan atau mendiskriminasi penggunaan teknologi tertentu.

Setiap aturan harus berfokus pada pengaturan risiko yang diciptakan oleh teknologi tertentu, alih-alih pada teknologi itu sendiri. Ini berarti dua hal bagi pendekatan Uni Eropa terhadap AI: pertama, bahwa setiap pendekatan regulasi yang disetujui oleh Uni Eropa untuk memitigasi risiko yang diciptakan oleh teknologi AI harus berbasis risiko (bukan berbasis teknologi); kedua, bahwa aturan baru hanya diperlukan jika dan sejauh mana teknologi AI menciptakan risiko yang belum ditangani secara memadai di Uni Eropa. Dengan demikian, Undang-Undang AI mendukung pendekatan regulasi berbasis risiko terhadap AI yang membatasi intervensi regulasi 'pada persyaratan minimum yang diperlukan untuk mengatasi risiko dan masalah yang terkait dengan AI' dan menyesuaikannya 'dengan situasi konkret di mana terdapat alasan yang dapat dibenarkan untuk khawatir, atau di mana kekhawatiran tersebut dapat diantisipasi secara wajar dalam waktu dekat.'

Selain itu, dicatat baik dalam Buku Putih AI Uni Eropa maupun dalam Undang-Undang AI-nya bahwa saat ini terdapat 'sejumlah besar undang-undang Uni Eropa yang ada, termasuk aturan khusus sektor, yang selanjutnya dilengkapi dengan undang-undang nasional' yang 'relevan dan berpotensi berlaku untuk sejumlah aplikasi AI yang sedang berkembang.' Aturan tersebut sepenuhnya berlaku di sektor-sektor ini, terlepas dari apakah teknologi AI terlibat, dan hanya akan memerlukan penyesuaian jika dan hanya jika aturan tersebut tidak dapat 'ditegakkan secara memadai untuk mengatasi risiko yang diciptakan oleh sistem AI.'

Demikian pula, gagasan proporsionalitas tertanam di berbagai dokumen yang menyusun pendekatan Uni Eropa terhadap AI yang berpuncak pada Pertimbangan (14) AI Undang-Undang, yang menggarisbawahi perlunya memperkenalkan ‘serangkaian aturan mengikat yang proporsional dan efektif untuk sistem AI.’ Secara lebih luas, UE menolak solusi yang ‘terlalu preskriptif,’ atau mengenakan ‘beban yang tidak proporsional’ demi solusi yang ‘memfasilitasi ... inovasi dan dengan demikian meningkatkan daya saing Eropa,’ yaitu dengan menghindari ‘pembatasan perdagangan yang tidak perlu.’ Secara khusus, proporsionalitas ini harus dicapai dengan membedakan antara berbagai tingkat risiko, dengan mengatur berbagai

aplikasi AI secara berbeda dan sesuai dengan tingkat risiko yang dirasakan dan, yang terpenting, dengan membiarkan aplikasi AI yang dianggap kurang berisiko pada dasarnya tanpa pengawasan.

Pada akhirnya, gagasan netralitas teknologi dan proporsionalitas berpadu dalam Undang-Undang AI Uni Eropa untuk memunculkan 'pendekatan berbasis risiko' terhadap AI yang secara khusus berfokus pada aplikasi 'berisiko tinggi'. Dengan kata lain, meskipun Undang-Undang AI mendukung pendekatan regulasi horizontal lintas sektoral yang ambisius, pendekatan tersebut dibatasi oleh gagasan bahwa hanya risiko AI yang tidak diatur yang perlu ditangani dan, bahkan dalam hal ini, hanya risiko AI 'tinggi'. Hasilnya adalah rezim yang hanya melarang sejumlah kecil praktik AI, dan yang hanya memberlakukan persyaratan dan kewajiban tambahan pada sistem AI yang dianggap oleh Uni Eropa sebagai 'berisiko tinggi' (dan pada peserta dalam rantai produksi dan distribusi sistem tersebut hingga pengguna akhir).

Apakah 'pendekatan risiko tinggi' terhadap AI ini secara memadai mengatasi perubahan yang dibawa oleh AI ke sistem keuangan? Sebagaimana telah disebutkan sebelumnya, hanya satu aspek dampak AI terhadap sistem keuangan yang saat ini dicakup oleh Undang-Undang AI Uni Eropa risiko diskriminasi yang ditimbulkan oleh sistem penilaian kredit algoritmik namun, pengecualian umum terhadap risiko yang ditimbulkan oleh sistem AI lain terhadap sistem keuangan dan para pelakunya hanya dapat dianggap sebagai kelalaian jika tidak dapat dibenarkan dengan tepat berdasarkan prinsip netralitas dan proporsionalitas teknologi yang mendasari pendekatan berisiko tinggi Uni Eropa terhadap regulasi AI.

Di satu sisi, tidak diragukan lagi bahwa Uni Eropa ingin menciptakan lingkungan regulasi yang mendorong 'AI yang tepercaya;' di sisi lain, tidak diragukan lagi pula bahwa undang-undang dan aturan yang terlalu luas dapat mengorbankan kemampuan Uni Eropa untuk bersaing dengan negara-negara seperti Amerika Serikat dan Tiongkok untuk mendapatkan tempat di garda terdepan pasar global teknologi AI.

Dan dapat dikatakan bahwa sebagian besar risiko yang ditimbulkan oleh AI terhadap sistem keuangan dan para pelakunya sudah tercakup secara luas oleh kerangka regulasi keuangan yang ada terutama setelah upaya terbaru untuk memperluas cakupan kerangka tersebut. Misalnya, versi terbaru dari Arahan Pasar Instrumen Keuangan Uni Eropa (‘MiFID II’) memuat serangkaian persyaratan yang secara khusus berlaku bagi perusahaan yang terlibat dalam, memfasilitasi, atau menyelenggarakan perdagangan algoritmik, dan Peraturan Penyalahgunaan Pasar Uni Eropa tahun 2014 memuat referensi eksplisit tentang manipulasi pasar yang digerakkan oleh algoritmik menunjukkan kesediaan untuk mengatasi perubahan yang digerakkan oleh algoritmik (jika tidak secara eksplisit digerakkan oleh AI).

Pada akhirnya, diskusi lengkap tentang apakah pendekatan berisiko tinggi Uni Eropa terhadap AI secara memadai mengatasi dampak AI pada sistem keuangan jauh melampaui cakupan tulisan ini. Cakupan penyelidikan kami jauh lebih sempit: apakah pendekatan berisiko tinggi terhadap AI yang didukung oleh Uni Eropa mengatasi berbagai cara di mana teknologi berbasis AI telah memperkuat risiko sistemik, dan jika tidak, adakah batasan dalam cakupan yang dibenarkan oleh netralitas teknologi atau kekhawatiran proporsionalitas?

Menentukan sejauh mana pendekatan Uni Eropa terhadap AI dan, khususnya, Undang-Undang AI-nya dapat mengatasi sumber-sumber baru risiko sistemik yang diciptakan oleh AI memerlukan analisis terhadap gagasan 'risiko tinggi' yang menjadi inti Undang-Undang AI. Menurut pasal 6 dan 7 Undang-Undang AI, klasifikasi jenis sistem AI tertentu sebagai 'risiko tinggi' pada dasarnya bergantung pada tujuan penggunaannya (lihat pasal 6(1)(a) dan 7(1)(a)) dan pada potensinya untuk menimbulkan 'risiko bahaya bagi kesehatan dan keselamatan, atau risiko dampak buruk terhadap hak-hak dasar' individu (lihat pasal 7(1)(b)) tanpa mempertimbangkan potensinya untuk menciptakan kerugian dan saluran penularan yang dapat menjangkau sistem yang lebih luas yang dihuni oleh individu-individu tersebut. Memang, fokus pada 'perlindungan ... individu' ini diperjelas dalam Pertimbangan (10) UU AI dan meresapi sebagian besar ketentuannya.

Mungkin ada sedikit penghiburan dalam kenyataan bahwa pasal 7 (2)(d) UU AI menyarankan bahwa 'penilaian risiko' yang kondusif untuk memperbarui daftar sistem 'berisiko tinggi' yang telah diidentifikasi oleh Komisi Eropa harus mempertimbangkan 'tingkat potensi [kerugian atau dampak buruk bagi individu] dalam hal intensitasnya dan kemampuannya untuk memengaruhi banyak orang,' tetapi bahkan interpretasi yang luas dari rumus ini gagal untuk menangkap sifat sebenarnya dari risiko sistemik yang, seperti yang dicatat di bagian sebelumnya, tidak hanya mencakup risiko bahwa kerugian individu dapat memengaruhi sejumlah besar agen pada saat yang sama, tetapi, yang terpenting, risiko bahwa kerugian individu dapat menyebar dari hanya satu agen individu dan menyebar ke seluruh sistem yang semakin saling terhubung atau bahwa penggunaan model yang serupa dapat mengakibatkan paparan umum yang dapat memfasilitasi penyebaran guncangan.

Oleh karena itu, tampaknya Undang-Undang AI kesulitan untuk menangkap dimensi lintas sektoral 'risiko sistemik' dalam definisinya tentang 'risiko tinggi' dan, dengan demikian, mengatasinya. Mengingat kecenderungan signifikan AI untuk memperkuat risiko sistemik yang dibahas di bagian sebelumnya, hal ini dapat dianggap sebagai kelalaian yang signifikan, tetapi dapatkah pengecualian ini dibenarkan berdasarkan prinsip netralitas teknologi atau proporsionalitas?

Netralitas teknologi mengharuskan regulator untuk mengadopsi pendekatan berbasis risiko bukan berbasis teknologi terhadap regulasi dan melakukan intervensi hanya ketika mereka mengidentifikasi risiko yang perlu dimitigasi. Kini, bagian sebelumnya telah membahas banyak sekali cara teknologi berbasis AI memperkuat risiko sistemik keuangan, dan meskipun beberapa cara ini telah ditangani secara khusus oleh regulasi terkini yaitu, MiFID II terkait perdagangan algoritmik cara lainnya belum mendapatkan perhatian regulasi yang sama. Tentu saja, dapat dikatakan bahwa risiko sistemik berbasis AI sudah tercakup dalam kerangka kerja makroprudensial yang ada, tetapi Dewan Risiko Sistemik Eropa baru-baru ini mengakui bahwa risiko sistemik keuangan berbasis teknologi (atau 'risiko siber sistemik') menciptakan ancaman yang 'memerlukan kerja lebih lanjut oleh otoritas makroprudensial.'

Dan teknologi AI seiring perkembangannya dapat menjadi sesuatu yang benar-benar baru. Alternatifnya, dapat dikatakan bahwa pengecualian pertimbangan risiko sistemik dari definisi 'risiko tinggi' dibenarkan oleh kekhawatiran proporsionalitas dan relatif kurang

pentingnya jenis risiko ini, tetapi hal itu akan bertentangan dengan konsensus yang berkembang seputar signifikansi dan, memang, prioritas yang diinginkan stabilitas keuangan sebagai tujuan regulasi.

Dari perspektif tersebut, sulit untuk memahami mengapa Uni Eropa menggunakan Undang-Undang AI-nya untuk mengatasi risiko diskriminasi yang didorong oleh AI yang diciptakan oleh fenomena seperti penilaian kredit algoritmik, sementara implikasi risiko sistemik yang lebih besar sama sekali tidak ditangani (termasuk yang berkaitan dengan penilaian kredit algoritmik). Dengan demikian, tampaknya pendekatan Uni Eropa terhadap AI, secara umum dan Undang-Undang AI yang baru, khususnya tidak hanya gagal menangkap risiko sistemik yang baru diciptakan oleh AI, tetapi juga kegagalan tersebut sulit dibenarkan berdasarkan prinsip-prinsip netralitas teknologi dan proporsionalitas yang tampaknya memandu pendekatan tersebut.

22.4 KRITIK AI ACT UE: MENGABAIKAN RISIKO SISTEMIK KEUANGAN

Seiring dunia keuangan memasuki era baru kemajuan teknologi, para regulator dan pengawas di seluruh dunia dihadapkan pada sebuah persimpangan: bagaimana sistem keuangan dapat memanfaatkan manfaat AI sekaligus mewaspadai risikonya? Jawabannya jarang jelas terkait teknologi algoritmik, tetapi hal itu tidak menghentikan Uni Eropa untuk memimpin dalam membentuk agenda regulasi global terkait AI.

Aspirasi regulasi Uni Eropa jauh melampaui sekadar mengatasi dampak AI terhadap sistem keuangan dan para pelakunya: sebuah proposal baru yang berani untuk Undang-Undang AI horizontal menargetkan aplikasi AI berisiko tinggi di berbagai industri dan sektor. Di saat yang sama, ambisi Uni Eropa telah diredam oleh kekhawatiran atas kemampuannya untuk bersaing di arena AI global: kepercayaan dan keamanan AI memang penting, tetapi keunggulan dan inovasi AI juga penting.

Pada akhirnya, pendekatan regulasi Uni Eropa terhadap AI dengan mudah mengakui dampak AI terhadap sistem keuangan dan, khususnya, risiko yang diciptakan oleh sistem penilaian kredit algoritmik tetapi membiarkan risiko sistemik keuangan yang diciptakan oleh AI seolah-olah tidak ditangani. Ini mungkin sebuah kelalaian yang signifikan: artikel ini telah menunjukkan bahwa AI telah mentransformasi sistem keuangan maupun cara sistem tersebut diatur dan diawasi, menciptakan sumber-sumber risiko sistemik baru yang sebagian besar masih kurang dipelajari.

Meskipun keputusan Uni Eropa untuk mengecualikan risiko sistemik dari definisi 'risiko tinggi' yang mendasari Undang-Undang AI yang baru dapat dibenarkan dengan alasan netralitas teknologi atau proporsionalitas, kedua alasan tersebut tidak valid. Sumber-sumber risiko sistemik baru ini juga belum ditangani oleh pendekatan regulasi dan pengawasan makroprudensial yang ada sebagaimana baru-baru ini diakui oleh Dewan Risiko Sistemik Eropa Uni Eropa dan jenis risiko ini juga tidak kalah signifikan dibandingkan risiko-risiko yang telah diidentifikasi dan dicakup oleh Undang-Undang AI.

Dapat dikatakan dengan lebih meyakinkan bahwa mungkin Undang-Undang AI Uni Eropa bukanlah jenis instrumen yang tepat untuk mengatasi jenis risiko ini. Namun, tetap saja

mengkhawatirkan bahwa aturan horizontal yang berambisi menetapkan aturan AI yang terharmonisasi di seluruh sektor dan industri dan yang bahkan mengatasi risiko yang melekat pada aktivitas dan layanan keuangan tertentu, seperti penilaian kredit algoritmik sepenuhnya mengabaikan jenis risiko spesifik yang paling jelas menunjukkan potensi ancaman terhadap stabilitas sistem tersebut. Kecenderungan untuk berfokus pada mikro alih-alih makro bukanlah hal baru, tetapi episode seperti krisis keuangan 2007–2009 telah mengajarkan kita pentingnya memprioritaskan pertimbangan gambaran besar.

Selain itu, bisa jadi Undang-Undang AI Uni Eropa justru merugikan tujuan mitigasi risiko sistemik yang diciptakan oleh AI. Yang paling jelas, Undang-Undang AI dapat berkontribusi pada anggapan keliru bahwa risiko paling signifikan yang ditimbulkan oleh AI telah ditangani baik dalam regulasi sektoral, dalam kasus perdagangan algoritmik, maupun dalam Undang-Undang AI itu sendiri, dalam kasus penilaian kredit algoritmik pada saat yang sama risiko sistemik yang didorong oleh AI justru luput dari perhatian regulator. Kedua, fakta bahwa Uni Eropa telah memilih untuk mengatasi risiko yang ditimbulkan oleh AI dengan mengusulkan kerangka kerja horizontal yang dikodifikasikan dalam sebuah Peraturan (alih-alih sebuah Direktif) memastikan bahwa persyaratan regulasi serupa akan berlaku untuk sistem AI di seluruh sektor dan di seluruh Negara Anggota, dengan ruang yang sangat kecil untuk variasi.

Sejauh persyaratan tersebut dapat mendorong pengembangan produk serupa yang tunduk pada mekanisme kontrol dan keamanan yang serupa, Undang-Undang AI dapat menciptakan tingkat keseragaman dan homogenitas yang justru merupakan sumber baru risiko sistemik. Pada akhirnya, pendekatan berbasis risiko yang menjadi inti pendekatan Uni Eropa terhadap AI merupakan upaya yang dapat dipahami untuk mengatasi kekhawatiran netralitas dan proporsionalitas teknologi yang mencerminkan ketegangan yang ada antara tujuan 'keunggulan dalam AI' dan 'AI yang tepercaya' inovasi dan keamanan. Namun, kompromi dan trade-off regulasi membutuhkan pemahaman yang jelas tentang peluang dan risiko yang muncul dari objek regulasi. Meremehkan atau mengabaikan potensi AI dalam memperkuat risiko sistemik tentu membatasi kemampuan Uni Eropa untuk mencapai keseimbangan yang tepat dalam meregulasi AI.

Ke depannya, jelas diperlukan lebih banyak penelitian tentang risiko sistemik yang ditimbulkan oleh AI. Juga jelas bahwa pendekatan berisiko tinggi yang diadopsi Uni Eropa dalam Undang-Undang AI-nya dapat memperoleh manfaat dari definisi 'risiko tinggi' yang lebih luas: definisi yang tidak hanya berfokus pada kerugian bagi individu (atau bahkan banyak individu) tetapi juga mempertimbangkan dampak struktural dan sistemik AI yang lebih luas. Selain itu, diharapkan regulator dan pengawas akan segera mulai memperkuat perangkat makroprudensial yang ada untuk memastikan mereka dapat menangani risiko sistemik baru yang ditimbulkan oleh AI. Dalam hal ini, beberapa inspirasi dapat diambil dari rezim perdagangan algoritmik Uni Eropa dan persyaratannya untuk uji ketahanan dan pemutus arus (circuit breaker).

Akhirnya, perlu digarisbawahi bahwa risiko sistemik dapat dengan mudah melintasi batas negara, dan pendekatan regulasi dan pengawasan baru yang berupaya mengatasi dampak AI terhadap sistem keuangan harus mengakui dimensi internasional dari risiko

sistemik ini. Oleh karena itu, diharapkan upaya Uni Eropa yang berjasa dalam membangun strategi AI yang inovatif dan aman pada akhirnya dapat memimpin diskusi seputar pengembangan pendekatan lintas batas yang lebih terintegrasi terhadap AI dan pendekatan yang lebih siap mengakui implikasi penting AI terhadap risiko sistemik keuangan.

BAB 23
REGULASI AI: TANTANGAN DAN REGULATORY SANDBOX

Abstrak Industri keuangan adalah yang pertama kali diajukan di mana menjadi jelas bahwa kita membutuhkan jenis regulasi baru, sebuah pendekatan evolusioner dan antisipatif yang setidaknya memiliki peluang untuk memitigasi risiko baru yang ditimbulkan oleh teknologi disruptif seperti kecerdasan buatan (AI). Pendekatan ini mengambil bentuk berbagai perangkat, namun tidak ada yang lebih menonjol daripada regulatory sandbox. Pendekatan regulasi yang relatif baru ini tersebar di berbagai sektor dan yurisdiksi, mulai dari FinTech hingga privasi dan layanan kesehatan.

Komisi Eropa mengakui potensi regulatory sandbox sebagai mekanisme peningkatan kepatuhan sekaligus sebagai cara untuk memfasilitasi inovasi dan oleh karena itu memasukkannya sebagai bagian dari rancangan peraturan tentang kecerdasan buatan (Undang-Undang AI). Dalam artikel ini, kami menganalisis potensi regulatory sandbox untuk meregulasi AI dalam format yang dibayangkan dalam Pasal 53 dan 54 dari rancangan Undang-Undang AI dan tantangan yang mungkin dihadapi pendekatan ini berdasarkan pengalaman dari regulatory sandbox sebelumnya yang melibatkan produk atau layanan AI. Kami juga bertujuan untuk menyarankan beberapa solusi yang dirancang khusus untuk memitigasi potensi kerugian dari kotak pasir regulasi untuk AI, termasuk bagaimana menyeimbangkan 'Prinsip Inovasi' yang sedang berkembang dan perlindungan hak asasi manusia.

23.1 KOTAK PASIR REGULASI AI DALAM UNDANG-UNDANG AI UE

Salah satu tokoh kunci dalam rekayasa modern, Dean Kamen, percaya bahwa "sesekali, teknologi baru, masalah lama, dan ide besar berubah menjadi inovasi". Saat ini, orang-orang terus menguji batas teknologi dan kreativitas, berjuang untuk menciptakan hal besar berikutnya dan mengubah kehidupan manusia. Persaingan yang sangat ketat ini tentu saja menarik, tetapi perubahan kehidupan sering kali menimbulkan konsekuensi yang tidak terduga dan dapat menciptakan risiko yang tidak terduga. Salah satu peran utama regulasi adalah mitigasi risiko. Menurut Prof. Karen Yeung, regulasi adalah "upaya terorganisir untuk mengelola risiko atau perilaku guna mengatasi masalah atau kekhawatiran kolektif". Namun, masalah dalam mengatur teknologi disruptif seperti AI bermula dari kombinasi sifat teknologi yang sebagian besar tidak dapat diprediksi dan dinamis dengan pendekatan legislasi tradisional yang reaksioner dan terlalu lambat untuk diadopsi dan diamandemen. Masalah besar lainnya adalah ketidakjelasan teknologi yang menyoroti perlunya melibatkan beragam orang dengan keahlian khusus dalam menyusun legislasi yang komprehensif dan melayani kebutuhan dasar setiap undang-undang, yaitu, untuk memastikan kepastian hukum.

Itulah sebabnya Buku Putih tentang AI yang telah lama dinantikan yang diadopsi oleh Komisi pada awal tahun 2020 disambut dengan kritik karena tidak mencerminkan perlunya pendekatan baru untuk mengatur teknologi baru, terutama ketika masing-masing negara anggota telah menerapkannya, terutama dalam bentuk kotak pasir regulasi. Hal ini juga

mengejutkan karena fakta bahwa kotak pasir regulasi, khususnya, telah ditunjukkan pada sejumlah kesempatan sebagai alat yang menonjol untuk memfasilitasi inovasi dan meningkatkan kepercayaan pada teknologi baru dan khususnya AI. Namun, kelalaian dalam Buku Putih tersebut diupayakan untuk diperbaiki melalui adopsi rancangan peraturan tentang penyelarasan aturan tentang kecerdasan buatan dan amandemen undang-undang persatuan tertentu (Undang-Undang AI). Judul V memberikan pandangan komprehensif pertama tentang apa yang dimaksud dengan kotak pasir regulasi untuk AI dan bagaimana penerapannya.

Bab ini bertujuan untuk menguraikan isu-isu utama yang dihadapi para pembuat kebijakan dalam upaya mereka untuk mengatur AI dan bagaimana isu-isu tersebut ditangani melalui pengenalan kotak pasir regulasi sebagai alat untuk jenis regulasi baru yang sedang berkembang. Untuk mencapai hal ini, pertama-tama kami akan menjelaskan sifat pendekatan tersebut dan mengamati apakah dan bagaimana pendekatan tersebut dapat diterapkan pada teknologi AI di berbagai sektor, mulai dari hukum keuangan hingga layanan kesehatan, dan apakah sifat multidimensinya tercermin secara memadai dalam rancangan Undang-Undang AI. Terakhir, kami akan mengidentifikasi beberapa tantangan yang dihadapi oleh alat regulasi ini dan menyimpulkan apakah alat ini memenuhi harapan sebagai terobosan dalam regulasi.

23.2 "PENJINAKAN" AI OLEH HUKUM

Sebagaimana telah disebutkan, tujuan utama pembuatan hukum adalah untuk memitigasi risiko-risiko tertentu yang timbul dari objek atau hubungan dalam masyarakat. Untuk mengilustrasikan hal ini, kita dapat melihat sebuah objek yang sangat kita kenal tetapi dulunya baru dan asing mobil. Kekhususannya dalam hal mekanika dan kendali memunculkan sejumlah kekhawatiran yang terutama berkaitan dengan kehidupan dan kesehatan manusia. Hal ini mendorong pengesahan undang-undang yang menetapkan aturan-aturan yang harus dipatuhi setiap pengemudi demi keselamatan mereka sendiri dan keselamatan pengemudi lain serta pejalan kaki. Kemudian, legislator memperoleh lebih banyak informasi yang menunjukkan perlunya aturan yang mengatur SIM dan asuransi wajib. Seiring perkembangan industri dan desain otomotif yang semakin maju, menjadi jelas bahwa produsen juga perlu diregulasi untuk memastikan bahwa mobil-mobil baru diproduksi dengan mengikuti standar keselamatan tertentu.

Dengan mobil menjadi moda transportasi yang paling umum, dampaknya terhadap lingkungan dan ruang perkotaan menjadi jelas yang memunculkan risiko-risiko tambahan, yang mengarah pada legislasi lebih lanjut dalam upaya untuk memitigasinya. Contoh sederhana ini menunjukkan hubungan antara teknologi, risiko, dan hukum. Lalu, mengapa cara legislasi tradisional tidak dapat diterapkan pada jenis teknologi lain seperti AI? Pertama, bisa dibilang teknologi AI jauh lebih rumit dan berpotensi memengaruhi masyarakat di lebih banyak domain daripada mobil. Teknologi ini sering dikategorikan sebagai teknologi disruptif4 dan karenanya memiliki risiko yang sulit diprediksi. Kedua, teknologi AI jauh lebih buram dibandingkan dengan mobil. Memang, pengguna biasa mungkin tidak tahu persis tujuan dari berbagai elemen penyusun mobil, tetapi seseorang dengan pengetahuan mekanik yang

memadai akan tahu dan karenanya memiliki prediktabilitas, dibandingkan dengan AI yang terkadang dapat bertindak dengan cara yang tidak terduga.

Perbedaan utama lainnya adalah apa yang disebut masalah kecepatan regulasi terkait AI. Masalah kecepatan ini merupakan kontras yang signifikan antara kecepatan inovasi AI dan inovasi perangkat regulasi yang digunakan untuk mengaturnya. Terakhir, agar legislasi memadai dan dapat menjalankan fungsinya dalam mitigasi risiko, objeknya perlu didefinisikan dengan jelas.

Hal ini penting karena suatu peraturan perundang-undangan yang disusun dengan baik dan bermanfaat pada akhirnya harus mencakup seluas mungkin situasi kehidupan nyata. Hal ini dipastikan melalui penggunaan terminologi hukum yang tepat dan definisi yang rinci dari setiap istilah yang digunakan dalam peraturan perundang-undangan itu sendiri, melalui rujukan ke peraturan perundang-undangan lain, penerapan aturan penafsiran hukum, atau melalui putusan pengadilan. Hal ini juga berkontribusi pada tercapainya kepastian hukum.

Kembali ke contoh kita, jika regulator ingin mengadopsi peraturan perundang-undangan yang mengatur aspek-aspek tertentu dari kendaraan bermotor, langkah pertama adalah mendefinisikan apa itu kendaraan bermotor. Misalnya, jika kita melihat Direktif 2007/46/EC, definisi tersebut langsung kita temukan di Pasal 3. Dengan membaca definisi tersebut, orang yang berakal sehat akan dengan mudah menyimpulkan bahwa kendaraannya yang beroda tiga jelas bukan kendaraan bermotor dalam lingkup Direktif tersebut dan oleh karena itu tidak tunduk pada aturannya.

Hal ini tentu saja dimungkinkan karena adanya definisi hukum yang komprehensif tentang kendaraan bermotor. Kembali ke permasalahan yang ada, maka sebelum membuat undang-undang apa pun terkait AI dan kemudian mengaturnya, harus ada definisi hukum yang dapat diterima untuk memenuhi tujuan Undang-Undang AI, tetapi juga tidak berkontribusi pada regulasi yang berlebihan terhadap subjek tersebut. Definisi ini telah menjadi topik hangat selama beberapa waktu, tidak hanya di bidang hukum, tetapi juga dalam ilmu komputer.

Isu taksonomi ini disorot dalam diskusi yang dibentuk setelah Kelompok Pakar Tingkat Tinggi untuk Kecerdasan Buatan (HLEG) mengadopsi, bersama dengan laporan tambahan tentang topik tersebut, sebuah definisi AI, yang bertujuan untuk "menghindari kesalahpahaman, mencapai pengetahuan umum tentang AI yang dapat digunakan secara bermanfaat juga oleh para ahli non-AI, dan untuk memberikan detail bermanfaat yang dapat digunakan dalam diskusi tentang pedoman etika AI dan rekomendasi kebijakan AI". Ada sejumlah isu yang diangkat mengenai definisi khusus tersebut, mulai dari beberapa pengecualian mesin yang mereplikasi diri dari cakupan definisi, melalui adopsi kriteria "diciptakan oleh manusia" hingga memiliki pendekatan "satu ukuran untuk semua" mengenai AI yang lemah dan kuat terlepas. Permasalahan yang disebutkan di atas sebagian teratasi melalui definisi yang diadopsi dalam Pasal 3(1) rancangan Undang-Undang AI yang mencakup sistem AI dan menggambarkannya sebagai "perangkat lunak yang dikembangkan dengan satu atau lebih teknik dan pendekatan yang tercantum dalam Lampiran I dan dapat, untuk serangkaian tujuan yang ditentukan manusia, menghasilkan keluaran seperti konten, prediksi, rekomendasi, atau keputusan yang memengaruhi lingkungan tempat mereka berinteraksi."

Namun, definisi baru ini juga mengungkapkan beberapa kelemahan, misalnya, tidak menunjukkan secara jelas perbedaan antara AI dan sistem AI, tidak mencerminkan upaya standardisasi yang sedang berlangsung di Uni Eropa, dan terlalu luas sehingga praktis mencakup "bahkan algoritma pencarian, pengurutan, dan perutean yang paling sederhana sekalipun" (BDVA/DAIRO 2021). Yang sangat penting bagi definisi ini adalah Lampiran I yang memuat jenis-jenis teknik dan pendekatan AI tertentu seperti pembelajaran yang diperkuat, penalaran simbolik, dll., dan yang dapat diperbarui melalui tindakan yang didelegasikan sesuai dengan Pasal 4 beserta Pasal 73 rancangan Undang-Undang AI.

Hal ini akan memungkinkan waktu reaksi yang lebih baik jika terjadi perkembangan ilmiah yang belum tercakup dalam regulasi dan dimaksudkan untuk mengatasi masalah kecepatan, meskipun mungkin masih belum cukup gesit mengingat waktu yang dibutuhkan untuk memberlakukan undang-undang yang didelegasikan. Dalam pengarahannya kepada Parlemen Eropa, Layanan Penelitian Parlemen Eropa mengakui perlunya penanganan masalah kecepatan yang lebih baik dan menyarankan "instrumen fleksibel seperti undang-undang yang didelegasikan, klausul sunset, dan legislasi eksperimental".

Perangkat regulasi baru ini bukanlah hal baru dan muncul jauh sebelum regulasi AI menjadi tugas yang harus diselesaikan. Perangkat-perangkat ini memiliki banyak nama dan sering digunakan dalam berbagai kombinasi, tetapi memiliki beberapa kesamaan: lebih dinamis dibandingkan dengan legislasi tradisional, memungkinkan partisipasi dari pemangku kepentingan yang lebih luas, dan memberikan umpan balik yang berharga kepada regulator yang memungkinkan pemahaman yang lebih baik tentang objek yang perlu diregulasi serta risiko dan manfaat yang menyertainya. Tidak diragukan lagi, salah satu perangkat yang paling banyak dibicarakan adalah kotak pasir regulasi yang akan dibahas di bagian-bagian berikut.

23.3 REGULATORY SANDBOX: INOVASI DAN REGULASI TERINTEGRASI

Istilah 'regulatory sandbox' terkadang membingungkan. Istilah ini agak mirip dengan konsep lingkungan sandbox dalam ilmu komputer. Meskipun serupa, kedua istilah tersebut tidak setara. Sandbox adalah alat pengujian, sementara regulatory sandbox adalah alat dan proses regulasi yang mengatur berbagai risiko dibandingkan dengan namanya.

Lingkup keuangan adalah area pertama di mana regulatory sandbox diuji. Krisis keuangan tahun 2008 mengakibatkan krisis regulasi global yang besar. Lingkup keuangan selalu sangat terpengaruh dan berkembang seiring dengan evolusi teknologi, yang sering disebut sebagai FinTech. Beberapa periode evolusi FinTech telah diidentifikasi, dimulai dengan penggunaan telegraf hingga mencapai apa yang saat ini dianggap sebagai FinTech 3.0.

Hal ini ditandai dengan penggunaan teknologi yang berkembang pesat, yang seringkali mengarah pada masuknya pelaku baru selain penyedia produk/jasa keuangan tradisional, atau otomatisasi proses yang mungkin menimbulkan konsekuensi yang tidak terduga dan tidak diinginkan, misalnya, bias algoritmik yang mengarah pada diskriminasi. Beragamnya cara teknologi disruptif dapat dimanfaatkan untuk tujuan FinTech menciptakan kebutuhan akan suatu bentuk regulasi.

Di sisi lain, regulasi inovasi yang berlebihan hanya untuk memastikan setiap skenario risiko yang mungkin tercakup dapat menghambat inovasi karena mengembangkan teknologi sesuai dengan persyaratan hukum yang sesuai akan memakan waktu, mahal, dan melibatkan peningkatan liabilitas. Oleh karena itu, para inovator mungkin 'berbelanja' di yurisdiksi yang kurang cepat dalam mengatur sektor keuangan.

Kekhawatiran ini menunjukkan perlunya pendekatan baru terhadap regulasi yang memposisikan regulator sebagai mitra dan pemandu, alih-alih musuh, bagi perusahaan yang ingin berinovasi. Pada tahun 2014, Otoritas Perilaku Keuangan Inggris (FCA) memulai Proyek Inovasi dan secara resmi menciptakan kotak pasir regulasi pertama. Pada bulan Oktober 2017, FCA menerbitkan "Laporan Pelajaran yang Dipetik dari Kotak Pasir Regulasi" yang menilai positif hasil penerapan kotak pasir regulasi. Hal ini mendorong pembentukan semakin banyak kotak pasir di berbagai yurisdiksi keuangan dan upaya untuk mengalihkan penggunaan alat regulasi ke sektor lain seperti perlindungan data atau penerbangan.

Potensi kotak pasir regulasi untuk meregulasi teknologi disruptif, khususnya AI, telah diakui. Sejumlah negara seperti Finlandia (Kementerian Ekonomi dan Ketenagakerjaan Finlandia dan Kelompok Pengarah Program Kecerdasan Buatan 2017) memasukkan penggunaan sandbox sebagai sarana untuk membangun kerangka hukum yang komprehensif bagi AI. Tren ini didukung oleh Uni Eropa yang memandang sandbox regulasi sebagai fasilitator inovasi dan mengakuinya sebagai alat penting dalam kegiatan regulasi AI di masa mendatang.

Tren ini semakin diperkuat dengan memasukkan sandbox regulasi ke dalam Better Regulation Toolbox Komisi Eropa dan menginstrumentalisasikannya dalam inisiatif-inisiatif lanjutan seperti sandbox blockchain pan-Eropa di masa mendatang dan rancangan Undang-Undang AI sebagai cara untuk mendorong inovasi sekaligus mendukung UKM. Pada saat yang sama, yurisdiksi lain di luar Uni Eropa telah menerapkan kotak pasir regulasi untuk menguji produk, layanan, dan model bisnis berbasis AI, baik melalui kotak pasir khusus AI maupun di bawah kerangka kerja jenis lain, misalnya, di bidang keuangan atau layanan kesehatan.

Dalam konteks ini, untuk lebih memahami sifat dan proses di balik kotak pasir regulasi, logis untuk melihat yang diterapkan di bidang FinTech karena jumlah, distribusi geografis, dan fakta bahwa di bidang tersebutlah kotak pasir regulasi pertama kali muncul. Meskipun tidak ada definisi universal untuk istilah ini, Otoritas Sekuritas dan Pasar Eropa menganggap kotak pasir regulasi sebagai "skema yang memungkinkan perusahaan menguji, sesuai dengan rencana pengujian khusus yang disepakati dan dipantau oleh fungsi khusus dari otoritas yang berwenang, produk keuangan inovatif, layanan keuangan, atau model bisnis". Definisi tersebut menyoroti beberapa poin. Kotak pasir regulasi pada dasarnya dianggap sebagai tempat pengujian untuk produk, layanan, atau model bisnis inovasi yang potensi risikonya dimitigasi, tetapi juga di mana pengawas terkait dapat memberikan kelonggaran tertentu dari aturan umum untuk tujuan pengujian.

Di sisi lain, Dewan Uni Eropa mengajukan definisi yang sedikit berbeda, dengan menyajikan kotak pasir regulasi sebagai kerangka kerja konkret yang, dengan menyediakan konteks terstruktur untuk eksperimen, memungkinkan pengujian teknologi, produk, layanan, atau pendekatan inovatif saat ini, terutama dalam konteks digitalisasi dalam jangka waktu

terbatas dan di bagian terbatas dari suatu sektor atau area di bawah pengawasan regulasi, memastikan adanya perlindungan yang memadai.

Sudah terdapat beberapa perbedaan antara dua definisi pertama. Definisi kedua lebih luas dan mencakup berbagai sektor, tetapi juga menekankan perlunya perlindungan yang memadai selama periode pengujian. Kemudian, rancangan Undang-Undang AI memberikan definisi lebih lanjut untuk kotak pasir regulasi AI tertentu dalam Pasal 53(1). Diharapkan bahwa kotak pasir regulasi AI adalah... yang ditetapkan oleh satu atau lebih otoritas kompeten Negara Anggota atau Pengawas Perlindungan Data Eropa wajib menyediakan lingkungan terkendali yang memfasilitasi pengembangan, pengujian, dan validasi sistem AI inovatif untuk jangka waktu terbatas sebelum dipasarkan atau dioperasikan sesuai rencana tertentu. Hal ini wajib dilakukan di bawah pengawasan dan arahan langsung oleh otoritas kompeten dengan tujuan memastikan kepatuhan terhadap persyaratan Peraturan ini dan, jika relevan, peraturan perundang-undangan Uni dan Negara Anggota lainnya yang diawasi dalam sandbox.

Definisi spesifik ini memberikan beberapa elemen tambahan dan baru. Pertama-tama, definisi ini secara eksplisit menekankan kemungkinan adanya kotak pasir regulasi multi-yurisdiksi. Kelayakan jenis kotak pasir ini telah dipertanyakan bahkan sebelum kita mulai membahas kotak pasir AI spesifik. Dinyatakan bahwa "fakta bahwa layanan tersebut tidak memiliki standarisasi yang terkait dengan regulasi membuat aktivitas kotak pasir tidak cocok untuk penyediaan layanan lintas batas". Belum diketahui bagaimana hambatan ini dapat diatasi.

Lebih lanjut, cakupan kotak pasir regulasi untuk AI diperluas secara signifikan, mencakup pengembangan, pengujian, dan validasi, sehingga menggabungkan fungsi tradisional kotak pasir regulasi dengan fungsi alat lain seperti pengujian dan uji coba. Penting untuk dicatat bahwa terdapat perdebatan mengenai hubungan yang tepat antara terminologi yang digunakan untuk menggambarkan 'ruang aman' yang telah didefinisikan ini untuk menguji inovasi dengan atau tanpa melibatkan otoritas tertentu. Yang disepakati adalah bahwa "terdapat hubungan inheren antara kotak pasir regulasi di satu sisi, dan pengujian serta uji coba di sisi lain" dan juga bahwa biasanya yurisdiksi "dengan pendekatan kotak pasir menempatkan aktivitas uji coba dan uji coba tertentu di dalam kotak pasir karena hal ini lebih praktis". Hal ini kemungkinan berkontribusi pada munculnya banyak istilah lain, misalnya 'laboratorium hidup', 'tempat uji coba regulasi', dll., yang digunakan sebagai sinonim dan pada akhirnya membahas "area untuk menguji coba inovasi dan regulasi". Namun demikian, definisi rancangan Undang-Undang AI tampaknya memasukkan elemen pengujian dan uji coba tertentu di samping aktivitas kotak pasir biasa, yang dapat menjadi elemen bermanfaat hanya jika benar-benar 'memfasilitasi' pengembangan inovasi dan pada akhirnya mengurangi 'waktu pemasaran' yang telah menjadi tujuan utama alat tersebut sejak awal. Namun, pertanyaan tentang cara dan tingkat elemen fasilitasi dari kotak pasir regulasi untuk AI sebagaimana dibayangkan oleh Komisi Eropa masih terbuka. Dengan kembali ke sumber aslinya dan memeriksa contoh-contoh kotak pasir untuk FinTech yang sudah ada, kita dapat menyimpulkan beberapa elemen kunci untuk keberhasilan pembuatan dan operasionalisasi kotak pasir. Pertama, kotak pasir beroperasi untuk jangka waktu terbatas dan dengan

parameter pengujian tertentu, yang memungkinkan jumlah peserta yang telah ditentukan sebelumnya.

Demi transparansi dan keadilan, persyaratan masuk kotak pasir perlu didefinisikan secara jelas dan tersedia untuk umum. Persyaratan tersebut dapat bervariasi di setiap yurisdiksi, tetapi yang paling umum adalah inovasi sejati, manfaat konsumen, dan kebutuhan akan pengujian di dalam kotak pasir. Secara umum, kotak pasir tidak terbatas hanya pada UKM, meskipun beberapa yurisdiksi memutuskan untuk menolak masuknya entitas yang diatur, hanya mendukung perusahaan yang tidak berlisensi, yang sebagian besar adalah UKM.

Di sisi lain, setelah sebuah perusahaan diterima dalam sandbox, biasanya seorang petugas kasus ditunjuk untuk menangani kasus tersebut guna memberikan keahlian regulasi dan menilai apakah perangkat sandbox untuk memfasilitasi pengujian diperlukan dalam kasus tertentu. Perangkat sandbox ini sangat banyak dan menawarkan beragam kemungkinan, mulai dari pendekatan "jangan pernah berkata tidak" yang diterapkan oleh Otoritas Moneter Singapura hingga kemudahan penegakan hukum dan surat jaminan negatif saat keluar yang ditawarkan oleh FCA. Tentu saja, ketika sandbox dibuat di domain yang sangat diatur oleh hukum Uni Eropa, masalah kelonggaran menjadi lebih rumit karena regulator nasional tidak dapat memberikan pengecualian apa pun dari aturan yang ditetapkan oleh Uni Eropa. Namun, perlu ditetapkan parameter untuk fase pengujian, misalnya, pembatasan pengungkapan, jumlah klien yang terbatas untuk menggunakan produk, layanan, atau model bisnis, dll.

Elemen penting ketiga adalah menjamin perlindungan pelanggan yang memadai selama pengujian sebagai salah satu tugas terpenting bagi regulator, dan khususnya untuk pengujian teknologi yang dapat menempatkan hak individu pada risiko yang signifikan, seperti inovasi di bidang perawatan kesehatan. Cara untuk mencapai hal ini bergantung pada kasus tertentu, tetapi mungkin yang paling umum adalah komunikasi yang jelas dan terperinci tentang sifat pengujian, yang memungkinkan konsumen untuk membuat keputusan yang tepat tentang setiap topik yang terkait dengan produk atau layanan yang diuji, dikombinasikan dengan parameter pengujian yang memitigasi risiko seperti pengujian yang dibatasi pada klien non-ritel. Perusahaan juga harus memastikan kompensasi atau tindakan ganti rugi lainnya jika terjadi kerugian yang diderita dalam konteks pengujian.

Setelah fase persiapan selesai, yang menyediakan jawaban dan solusi untuk semua pertanyaan yang dibahas di atas, fase pengujian dimulai. Fase ini mencakup komunikasi yang konstan antara perusahaan dan regulator. Dapat dikatakan bahwa fase ini paling menggambarkan sifat simbiosis dari kotak pasir regulasi. Regulator memang menjalankan perannya dalam memantau pengujian dan memastikan kompatibilitas dengan standar yang diperlukan, tetapi juga mengamati teknologi yang diuji dan memahami lebih baik cara kerjanya, potensi risikonya, dan pendekatan mana yang lebih baik untuk memitigasinya.

Terakhir, fase dan elemen terakhir adalah fase evaluasi yang mengharuskan penyerahan laporan akhir kepada otoritas, mengikuti parameter yang telah ditentukan selama fase persiapan, dan penilaian keberhasilan pengujian.

Elemen-elemen umum ini dibenarkan oleh data yang tersedia yang menunjukkan hasil kinerja tinggi dan relatif sedikit pengujian yang gagal. Namun, berbagai keunggulan model

kotak pasir regulasi inilah yang menjadikannya pilihan yang semakin populer untuk mengatur teknologi disruptif, terutama AI. Pertama, uji coba regulasi menunjukkan kesediaan regulator untuk memfasilitasi dan merangsang inovasi, yang merupakan tanda peluang bisnis yang baik di negara masing-masing. Kedua, peningkatan tingkat pengetahuan baik bagi peserta maupun regulator terlihat jelas. Hal ini memungkinkan regulator untuk menjalankan fungsi mereka dengan lebih baik dan mendapatkan wawasan penting tentang teknologi yang sedang berkembang, sehingga teknologi tersebut kurang dapat diandalkan oleh keahlian eksternal. Lebih lanjut, waktu pemasaran berkurang, dikombinasikan dengan jaminan bahwa produk/layanan baru memiliki semua standar keamanan yang memadai. Hal ini juga memungkinkan para inovator untuk mendapatkan peringatan dini tentang kemungkinan fitur bermasalah dari produk/layanan mereka, serta jaminan bahwa mereka tidak akan melanggar persyaratan peraturan yang ada selama fase pengujian.

Apakah hal itu cukup untuk menyimpulkan bahwa uji coba regulasi merupakan alat yang paling tepat dalam meregulasi AI? Jawabannya tidak sesederhana itu. Terlepas dari antusiasme yang ditunjukkan oleh negara-negara dan organisasi internasional dalam menciptakan dan menerapkan uji coba regulasi, terdapat beberapa tantangan yang perlu ditangani dan dikaji. Beberapa tantangan yang umum untuk semua jenis kotak pasir regulasi: kurangnya kerangka regulasi yang lengkap untuk produk/layanan tertentu mungkin tampak terlalu berisiko bagi konsumen untuk terlibat dalam pengujian; itu juga berarti bahwa akan ada kurangnya standardisasi. Standardisasi penting karena implikasinya terhadap implementasi lintas batas produk/layanan. Lebih lanjut, dalam beberapa kasus risiko inovasi mungkin tidak cukup signifikan untuk memerlukan regulasi apa pun. Dalam kasus seperti itu, itu hanya akan menghambat inovasi. Dalam kasus lain, inovasi mungkin tidak cukup matang untuk diuji dalam kotak pasir, sehingga pendekatan tunggu dan lihat bisa lebih cocok. Juga benar bahwa jumlah peserta yang terbatas di kotak pasir mungkin tidak memberikan sampel representatif yang diperlukan untuk sepenuhnya menentukan efek dari teknologi tertentu.

Perusahaan-perusahaan itu sendiri mungkin enggan berpartisipasi, baik karena ingin tumbuh lebih cepat, dan sandbox akan membatasi kemampuan ini maupun karena mereka tidak cukup terstimulasi oleh keleluasaan yang ditawarkan oleh regulator, terutama dalam konteks Eropa di mana perangkat sandbox jauh lebih konservatif dibandingkan dengan yurisdiksi lain.

Terakhir, perusahaan-perusahaan yang berpartisipasi mungkin enggan berpartisipasi dalam lingkungan di mana rahasia dagang berpotensi ditemukan oleh pesaing. Kategori tantangan lainnya terkait secara khusus dengan sandbox regulasi untuk AI. Rencana Terkoordinasi tentang AI menetapkan bahwa fasilitas pengujian yang direncanakan untuk AI "dapat mencakup sandbox regulasi...di area-area tertentu di mana undang-undang memberikan keleluasaan yang memadai kepada otoritas regulasi."

Hal ini agak membingungkan karena selama ini sandbox regulasi diciptakan bukan untuk menguji teknologi tertentu, melainkan inovasi di bidang tertentu, misalnya, di bidang keuangan. Pendekatan ini tidak membatasi teknologi yang digunakan, melainkan tujuan dan penerapannya. Untuk mengilustrasikan poin kami, terdapat solusi berbasis AI yang sedang

diuji dalam kotak pasir regulasi FCA sebagai bagian dari kelompok ke-5. Tujuannya adalah untuk membantu UKM yang mengajukan pinjaman dengan menggunakan AI untuk meningkatkan efektivitas penilaian kredit dan meningkatkan penilaian risiko sekaligus mengurangi biaya.

Kedua, tingkat dampak yang dapat ditimbulkan oleh teknologi AI mungkin tidak hanya berada dalam cakupan satu regulator. Misalnya, sebuah teknologi AI mungkin ditujukan untuk digunakan hanya di sektor perbankan dan dengan demikian diatur oleh otoritas keuangan, tetapi pada saat yang sama, mungkin saja teknologi tersebut memiliki implikasi yang signifikan terhadap data pribadi sehingga bantuan dari otoritas perlindungan data harus diberikan. Hal ini bermasalah dari sudut pandang organisasi dan administratif. Regulator biasanya tidak memiliki pengalaman dalam berkoordinasi satu sama lain dalam hal-hal seperti itu yang akan menyebabkan kekacauan dan inefisiensi. Perlu juga dicatat bahwa teknologi AI dirancang untuk belajar, oleh karena itu, untuk berubah. Ini berarti bahwa sebuah teknologi AI, yang keluar dari sandbox yang diberi label patuh, mungkin tidak patuh untuk waktu yang lama. Pergantian peristiwa seperti itu dapat merusak seluruh proses dan pada akhirnya kepastian hukum.

23.4 LANGKAH KE DEPAN MELALUI UNDANG-UNDANG AI

Dimasukkannya kotak pasir regulasi untuk AI dalam rancangan Undang-Undang AI menandakan pendekatan baru Uni Eropa dalam meregulasi teknologi disruptif, tetapi kita perlu mempertimbangkan semua tantangan yang diuraikan di bagian sebelumnya. Jelas bahwa meskipun banyak peluang dan keuntungan baru yang ditawarkan oleh kotak pasir regulasi dibandingkan dengan cara regulasi dan tata kelola tradisional (reaktif), kotak pasir regulasi bukanlah obat mujarab, melainkan alat dan fondasi bagi pendekatan baru dalam meregulasi masyarakat berbasis data.

Nesta telah melakukan penelitian mendetail tentang fitur-fitur yang perlu dimiliki model regulasi baru ini dan karakteristik utamanya. Berdasarkan karya Geoff Mulgan dan analisisnya tentang elemen-elemen perangkat regulasi yang sedang berkembang, Nesta menguraikan enam prinsip regulasi antisipatif baru yang harus dimilikinya, berbeda dengan pendekatan reaktif tradisional. Regulasi antisipatif harus inklusif dan kolaboratif, melibatkan publik dan berbagai pemangku kepentingan yang juga memastikan legitimasi demokratis yang lebih baik dari jenis regulasi ini. Regulasi ini juga harus berorientasi ke masa depan dan proaktif. Prinsip berikutnya bersifat iteratif, yang digambarkan sebagai "mengambil pendekatan uji-dan-kembangkan daripada selesaikan-dan-tinggalkan untuk masalah baru".

Dua prinsip terakhir berbasis hasil dan bersifat eksperimental. Keduanya menunjukkan karakter pendekatan baru terhadap legislasi yang jauh lebih pragmatis dan berorientasi pada solusi. Dengan mengikuti prinsip-prinsip ini, regulator dapat menemukan alat regulasi terbaik atau kombinasi alat untuk kebutuhan khusus mereka. Selain itu, regulator tidak boleh ragu untuk menggabungkan semua peluang yang disediakan oleh regulasi antisipatif dengan alat dan pendekatan dari mode regulasi lain seperti regulasi penasihat atau adaptif.

Penting untuk ditekankan bahwa klasifikasi ini hanyalah contoh sistem yang cocok untuk menangani teknologi yang sedang berkembang dan khususnya AI. Kita juga dapat mempertimbangkan regulasi berdasarkan apakah regulasi tersebut didasarkan pada prinsip, risiko, pasar, atau ketergantungan pada manajemen internal.

Terlepas dari klasifikasi yang dipilih, tantangan utama bagi regulator tetaplah mitigasi risiko dalam teknologi AI berisiko tinggi. Regulatory sandbox memang menawarkan kemampuan tersebut, tetapi hasilnya bergantung pada sifat risiko dan tingkatnya. Alat yang lebih konservatif adalah penerapan klausul sunset dalam regulasi, meskipun tidak akan memberikan umpan balik yang sama besarnya dengan sandbox. Alat lain yang juga dipertimbangkan adalah standardisasi. Solusi yang diusulkan adalah sistem AI yang disertifikasi aman untuk menikmati tanggung jawab hukum terbatas dibandingkan dengan yang tidak disertifikasi.

Ini hanyalah beberapa contoh dari berbagai alat yang tersedia bagi regulator saat meregulasi AI. Pilihan salah satu atau yang lain, atau bahkan kombinasi antara beberapa, harus disesuaikan semaksimal mungkin dan didukung oleh upaya berkelanjutan dalam meningkatkan kapasitas regulator dengan memberikan mereka praktik terbaik dan pengembangan keterampilan terutama mengingat sifat interdisipliner penelitian AI.

Rancangan Undang-Undang AI telah menempatkan kotak pasir regulasi sebagai pusat perhatian sebagai fasilitator inovasi utama. Pendekatan ini menimbulkan setidaknya tiga kelompok kekhawatiran yang berbeda. Pertama, isu-isu yang telah dibahas di bagian sebelumnya mengenai kotak pasir regulasi secara umum dan yang secara khusus didedikasikan untuk AI belum terselesaikan dalam versi ketentuan yang berlaku di Pasal 53.

Kedua, desain kotak pasir regulasi untuk AI, sebagaimana dijelaskan dalam teks peraturan, tampaknya tidak memberikan banyak insentif untuk bergabung dengan kotak pasir tersebut. Salah satu elemen yang biasanya menarik produk/layanan paling inovatif, yaitu perancang aturan tertentu, tidak disinggung. Sangat penting untuk "memastikan dengan jelas apakah otoritas Negara Anggota akan dapat menawarkan keringanan regulasi atau jenis pengaturan regulasi lainnya untuk eksperimen AI". Lebih lanjut, regulator nasional mungkin tidak dapat menawarkan keringanan yang signifikan karena mereka tidak memiliki kompetensi tersebut terkait ketentuan hukum Uni Eropa. Satu-satunya pengecualian aturan yang kita ketahui saat ini berasal dari teks Pasal 54 dan berkaitan dengan aturan perlindungan data pribadi. Memang, peserta dalam sandbox regulasi untuk AI akan dapat memproses data pribadi, yang "dikumpulkan secara sah untuk tujuan lain" untuk mengembangkan dan menguji "sistem AI inovatif tertentu di dalam sandbox". Namun, pengecualian terhadap aturan perlindungan data pribadi ini tunduk pada sejumlah besar persyaratan kumulatif yang bertindak sebagai jaminan terhadap hak dan kebebasan individu. Persyaratan tersebut meliputi tujuan sistem AI inovatif ("menjaga kepentingan publik yang substansial" dalam satu atau lebih bidang yang telah ditentukan sebelumnya seperti kesehatan masyarakat misalnya), data yang diperlukan untuk memenuhi persyaratan sistem AI berisiko tinggi, dan keberadaan sistem pemantauan yang efektif untuk mengidentifikasi timbulnya risiko tinggi terhadap hak-hak dasar subjek data.

Lebih lanjut, "setiap data pribadi yang akan diproses dalam konteks sandbox berada dalam lingkungan pemrosesan data yang terpisah, terisolasi, dan terlindungi secara fungsional" dan tidak boleh "ditransmisikan, dipindahkan, atau diakses dengan cara lain oleh pihak lain". Selain itu, setiap pemrosesan data pribadi tidak boleh mengarah pada tindakan atau keputusan yang memengaruhi subjek data, perlu dihapus "setelah partisipasi dalam sandbox berakhir atau data pribadi telah mencapai akhir masa penyimpanannya" dengan tunduk pada log pemrosesan yang juga memiliki batasan masa penyimpanan dan tujuan.

Ada juga persyaratan transparansi yang ketat dalam bentuk "deskripsi lengkap dan terperinci tentang proses dan alasan di balik pelatihan, pengujian, dan validasi sistem AI" dan "ringkasan singkat proyek AI yang dikembangkan dalam sandbox, tujuan dan hasil yang diharapkan dipublikasikan di situs web otoritas yang berwenang." Beban untuk memenuhi persyaratan ini tampaknya jauh lebih besar daripada keuntungan dari keringanan yang ditawarkan dalam konteks sandbox.

Kelompok kekhawatiran ketiga terkait dengan kurangnya persetujuan mengenai skalabilitas kotak pasir regulasi untuk AI serta tempatnya dalam sistem perangkat regulasi antisipatif. Dari perspektif praktis, mengadopsi kombinasi perangkat yang cerdas untuk memfasilitasi inovasi dianggap sebagai solusi terbaik, di mana kotak pasir regulasi hanyalah salah satu bagian dari keseluruhan solusi. Misalnya, pada tahun 2019, Otoritas Penerbangan Sipil Inggris (CAA) meluncurkan Pusat Inovasi dan kotak pasir regulasi yang secara khusus menargetkan inovasi AI.

Kombinasi antara pusat inovasi dan kotak pasir regulasi ini telah dipertimbangkan sebagai cara untuk memecahkan beberapa masalah skalabilitas kotak pasir. Lebih lanjut, menggabungkan berbagai perangkat regulasi antisipatif dianggap sangat bermanfaat bagi pembentukan dan pengembangan prinsip Inovasi lebih lanjut. Selain itu, Dewan Uni Eropa telah menghubungkan kotak pasir regulasi dengan klausul eksperimental, yang dipahami sebagai "ketentuan hukum yang memungkinkan otoritas yang bertugas menerapkan dan menegakkan undang-undang untuk menerapkan fleksibilitas, berdasarkan kasus per kasus, dalam hal pengujian teknologi, produk, layanan, atau pendekatan inovatif". Namun, hubungan ini saat ini tidak ada dalam rancangan undang-undang AI di mana peran kotak pasir regulasi dalam berkontribusi pada pembuatan kebijakan berbasis bukti saat ini tidak ada.

23.5 EVOLUSI KOTAK PASIR REGULASI UNTUK AI DISRUPTIF

Pada tahun 1996, Richard Susskind berpendapat bahwa "kita berada di ambang pergeseran paradigma hukum, sebuah revolusi hukum". Peristiwa seperti krisis keuangan global tahun 2008 dan pemilu AS tahun 2016 membuktikan bahwa kita membutuhkan pendekatan baru yang radikal terhadap dunia dan cara kita mengaturnya. Pendekatan ini masih dalam tahap pengembangan, dan kita masih jauh dari menyelesaikan transisi dari regulasi reaktif ke antisipatif.

Kotak pasir regulasi tentu saja merupakan langkah maju. Kotak pasir regulasi memberikan keamanan dan tingkat kendali yang relatif, membantu regulator untuk lebih memahami AI dan teknologi disruptif lainnya sebelum memutuskan apakah dan bagaimana

mereka harus diregulasi. Lagipula, "regulasi hanyalah alat. Jika bermanfaat bagi masyarakat, ia harus digunakan, jika tidak, sebaiknya dihilangkan".

Ada juga banyak pertanyaan mengenai cara mengatasi tantangan yang diuraikan dalam bab ini dan evolusi apa yang akan dilalui kotak pasir regulasi agar tetap relevan dan menjawab kebutuhan masyarakat serta sifat AI yang dinamis. Sudah ada beberapa gagasan tentang Kotak Pasir 2.0, yang menggabungkan akses ke metode penemuan inovatif dan kotak pasir terpandu, yang mencoba menyelesaikan konflik antara tingkat nasional dan supranasional/federal terkait interaksi mereka dalam kotak pasir regulasi.

Saat ini, terdapat lebih banyak pertanyaan daripada jawaban tentang cara meregulasi AI. Kekuatan global menjabarkan berbagai pendekatan dalam upaya menjadi tujuan investasi yang paling menarik, tetapi pada akhirnya pendekatan yang paling berhasil adalah pendekatan yang dapat menstabilkan 'pasir yang bergeser' dari kotak pasir regulasi dan menggunakannya sebagai landasan untuk membangun cara regulasi yang baru.

DAFTAR PUSTAKA

Ablameyko, S. V., & Ablameyko, M. S. (2021). Artificial Intelligence from an Interdisciplinary Perspective: Philosophical and Legal Aspects. ФН, 58.

Ahmed, I. (2024). Ethical Implications of Artificial Intelligence: Perspectives from Multiple Fields. Kashf Journal of Multidisciplinary Research, 1(07), 322-332.

Ahmed, S. A. A. (2022). Technology and artificial intelligence in simultaneous interpreting: A multidisciplinary approach. CDELT Occasional Papers in the Development of English Education, 78(1), 325-353.

Amann, J., Blasimme, A., Vayena, E., Frey, D., Madai, V. I., & Precise4Q Consortium. (2020). Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC medical informatics and decision making, 20(1), 310.

Arencibia-Jorge, R., Vega-Almeida, R. L., Jiménez-Andrade, J. L., & Carrillo-Calvet, H. (2022). Evolutionary stages and multidisciplinary nature of artificial intelligence research. Scientometrics, 127(9), 5139-5158.

Bernelin, M. (2016). Interdisciplinary approach: a useful but challenging tool for the regulation of science and technologies. INTERDISCIPLINARY APPROACH TO LAW IN MODERN SOCIAL CONTEXT, 62.

Bhattacharya, P. Frontiers of Artificial Intelligence, Ethics and Multidisciplinary Applications (Doctoral dissertation, Syracuse University).

Chun, K. P., Octavianti, T., Dogulu, N., Tyralis, H., Papacharalampous, G., Rowberry, R., ... & Li, C. (2025). Transforming disaster risk reduction with AI and big data: Legal and interdisciplinary perspectives. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 15(2), e70011.

Chuphal, G., Vohra, J., Sharma, S. K., Minhajuddin, M., Ahmad, W., & Baba, K. R. K. (2025). Law in the Age of Disruption: Multidisciplinary Insights on Emerging Technologies and Legal Evolution. Journal of Neonatal Surgery, 14(32s).

Covelo de Abreu, J. (2023). The “Artificial Intelligence Act” Proposal on European e-Justice Domains Through the Lens of User-Focused, User-Friendly and Effective Judicial Protection Principles. In Multidisciplinary Perspectives on Artificial Intelligence and the Law (pp. 397-414). Cham: Springer International Publishing.

De Mulder, W. (2025). Explainable artificial intelligence and the social sciences: a plea for interdisciplinary research. AI & SOCIETY, 40(4), 2951-2970.

Dhiman, M. T., Malik, M. A., Kumar, A., & Kamboj, M. G. (2025). Artificial Intelligence Synergies: A Multidisciplinary Outlook. Chyren Publication.

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., ... & Wright, R. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. International Journal of Information Management, 71, 102642.

Ejjami, R. (2024). AI-driven justice: Evaluating the impact of artificial intelligence on legal systems. Int. J. Multidiscip. Res, 6(3), 1-29.

El-Atm, S. (2023). The future is multidisciplinary: Are you ready?. LSJ: Law Society Journal, (4), 84-89.

Freitas, A. T. (2023). Data-driven approaches in healthcare: Challenges and emerging trends. Multidisciplinary perspectives on artificial intelligence and the law, 65-80.

Gravett, W. H. (2023). Judicial decision-making in the age of artificial intelligence. In Multidisciplinary Perspectives on Artificial Intelligence and the Law (pp. 281-297). Cham: Springer International Publishing.

Guerra, G. (2018). An Interdisciplinary Approach for Comparative Lawyers: Insights from the Fast-Moving Field of Law and Technology. German Law Journal, 19(3), 579-612.

Karnouskos, S. (2022). Symbiosis with artificial intelligence via the prism of law, robots, and society. Artificial Intelligence and Law, 30(1), 93-115.

Kasperiuniene, J. (2021, January). The use of artificial intelligence in social research: Multidisciplinary challenges. In World Conference on Qualitative Research (pp. 312-324). Cham: Springer International Publishing.

Keller, A., Pereira, C. M., & Pires, M. L. (2024). The European Union's approach to artificial intelligence and the challenge of financial systemic risk. Multidisciplinary Perspectives on Artificial Intelligence and the Law, 415.

Khan, A. (2024). The intersection of artificial intelligence and international trade laws: Challenges and opportunities. IIUMLJ, 32, 103.

King, T. C., Aggarwal, N., Taddeo, M., & Floridi, L. (2020). Artificial intelligence crime: An interdisciplinary analysis of foreseeable threats and solutions. Science and engineering ethics, 26(1), 89-120.

Kucuk, E. S. (2024). An Interdisciplinary Reasoning on the Legal Status of Artificial Intelligence Entities. Digital L. Rev., 6, 198.

Kyriakou, K., & Otterbacher, J. (2023). In humans, we trust: Multidisciplinary perspectives on the requirements for human oversight in algorithmic processes. Discover Artificial Intelligence, 3(1), 44.

Lareyre, F., Behrendt, C. A., Chaudhuri, A., Ayache, N., Delingette, H., & Raffort, J. (2022). Big data and artificial intelligence in vascular surgery: time for multidisciplinary cross-border collaboration. Angiology, 73(8), 697-700.

Liu, X., Yang, X., & Zhang, M. (2024). Legal Risks and Regulatory Frameworks for Artificial Intelligence. Journal of Interdisciplinary Insights, 2(4), 62-71.

Magrani, E., & da Silva, P. G. F. (2024). The ethical and legal challenges of recommender systems driven by artificial intelligence. Multidisciplinary perspectives on artificial intelligence and the law, 141.

Marwala, T., & Mpedi, L. G. (2024). Artificial intelligence and the law. In Artificial intelligence and the law (pp. 1-25). Singapore: Springer Nature Singapore.

MITAN, E., NEACȘU, D., OVREIU, S., & SAVU, D. A Multidisciplinary Approach to Trustworthy AI in the Digital Society.

Mohajer-Bastami, A., Moin, S., Ahmad, S., Ahmed, A. R., Pouwels, S., Hajibandeh, S., ... & Exadaktylos, A. K. (2025). Artificial intelligence in healthcare: applications, challenges, and future directions. A narrative review informed by international, multidisciplinary expertise. Frontiers in Digital Health, 7, 1644041.

Molè, M., Been, W., & ter Haar, B. (2024). Multidisciplinary Perspectives on Work and Digitalisation. Italian Labour Law e-Journal, 17(2), I-VI.

Mpinga, E. K., Bukonda, N. K., Qailouli, S., & Chastonay, P. (2022). Artificial Intelligence and Human Rights: Are There Signs of an Emerging Discipline? A Systematic Review. Journal of multidisciplinary healthcare, 235-246.

Muhlenbach, F., & Sayn, I. (2019, June). Artificial Intelligence and law: What do people really want?: Example of a French multidisciplinary working group. In Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law (pp. 224-228).

Nayak, P., & Çalıyurt, K. T. (Eds.). (2025). A Multidisciplinary Approach to KIIT Horizons, Volume 1: Exploring Artificial Intelligence Across Disciplines (Vol. 1). Springer Nature.

Neves, M. P., & De Almeida, A. B. (2024). Before and Beyond Artificial Intelligence: Opportunities and Challenges. Multidisciplinary Perspectives on Artificial Intelligence and the Law, 107.

Nityanath, S. A STUDY OF INTERDISCIPLINARY RESEARCH IN ARTIFICIAL INTELLIGENCE. Impact of Smart Technologies and Artificial Intelligence (AI) Paving Path Towards Interdisciplinary Research in the Fields of Engineering, Arts, Humanities, Commerce, Economics, Social Sciences, Law and Management-Challenges and Opportunities, 181.

Nweke, O. C., & Nweke, G. I. (2024). Legal and Ethical Conundrums in the AI Era: A multidisciplinary analysis. International Law Research, 13(1), 1-1.

Oliveira, A. L., & Figueiredo, M. A. (2024). Artificial Intelligence: Historical context and state of the art. Multidisciplinary Perspectives on Artificial Intelligence and the Law, 3.

Pagallo, U. (2023). Dismantling Four Myths in AI & EU Law Through Legal Information 'About'Reality. In Multidisciplinary Perspectives on Artificial Intelligence and the Law (pp. 251-261). Cham: Springer International Publishing.

Redrup Hill, E., Mitchell, C., Brigden, T., & Hall, A. (2023). Ethical and legal considerations influencing human involvement in the implementation of artificial intelligence in a clinical pathway: A multi-stakeholder perspective. Frontiers in digital health, 5, 1139210.

Régis, C., Denis, J. L., Axente, M. L., & Kishimoto, A. (2024). Human-centered AI: A multidisciplinary perspective for policy-makers, auditors, and users (p. 358). Taylor & Francis.

Sava, N. A. (2023). Artificial intelligence and public procurement–deciphering the interdisciplinary perspectives of the literature. Eur. Rev. Digit. Adm. Law, 4, 79-88.

Sousa Antunes, H., Freitas, P. M., Oliveira, A. L., Martins Pereira, C., Vaz de Sequeira, E., & Barreto Xavier, L. (2024). Multidisciplinary perspectives on artificial intelligence and the law (p. 456). Springer Nature.

Starke, C., Blanke, T., Helberger, N., Smets, S., & de Vreese, C. (2025). Interdisciplinary Perspectives on the (Un) fairness of Artificial Intelligence. Minds and Machines, 35(2), 22.

Szmodis, J. (2011). On law, from a multidisciplinary perspective. Hungarian Review, 2(04), 64-69.

Tariq, A. (2025). ETHICAL IMPLICATIONS OF ARTIFICIAL INTELLIGENCE IN DECISION-MAKING SYSTEMS: A MULTIDISCIPLINARY PERSPECTIVE. Multidisciplinary Research in Computing Information Systems, 5(4), 374-391.

Tavakoligolpaygani, S., Share, P. M. M., & Azadi, P. F. (2024). Multidisciplinary Perspectives on Artificial Intelligence and the Law.

Thanaraj, A. (2018). Studying Law in a Digital Age: Preparing Law Students for Participation in a Technologically-Advanced Multidisciplinary and Complex Professional Environment. Nottingham LJ, 27, 62.

Toto, G. A., Grilli, L., Traetta, L., Villani, R., Petito, A., & Serviddio, G. (2025). Convergence of disciplines: a systematic review of multidisciplinary development approaches in artificial intelligence. Frontiers in Digital Health, 7, 1400338.

Tran, T. L. (2025). The Use of Interdisciplinary Approach in International Law Study in the Digital Era 5.0: Highlights and Future Research Directions. PRAWO i WIĘŹ, 55(2).

Tzimas, T. (2021). Legal and ethical challenges of artificial intelligence from an international law perspective (Vol. 46). Springer Nature.

van Duuren, N., & de Pous, V. (2020). Multidisciplinary aspects of artificial intelligence. KNVI.

Vargas-Murillo, A. R., Turriate-Guzman, A. M., Delgado-Chávez, C. A., & Sanchez-Paucar, F. (2024). Transforming justice: Implications of artificial intelligence in legal systems. Academic Journal of Interdisciplinary Studies, 13(2), 433.

Wang, J., Mao, W., & Wenjie, W. (2023). The Ethics of Artificial Intelligence: Sociopolitical and Legal Dimensions. Interdisciplinary Studies in Society, Law, and Politics, 2(2), 27-32.

Wellnhofer, E. (2022). Real-world and regulatory perspectives of artificial intelligence in cardiovascular imaging. Frontiers in cardiovascular medicine, 9, 890809.

Yordanova, K., & Bertels, N. (2024). Regulating AI: Challenges and the way forward through regulatory sandboxes. Multidisciplinary perspectives on artificial intelligence and the law, 441.

MULTIDISIPLIN
ANTARA KECERDASAN BUATAN (AI) DAN ILMU HUKUM

Dr. Agus Wibowo, M.Kom, M.Si, MM.

BIO DATA PENULIS

Penulis memiliki berbagai disiplin ilmu yang diperoleh dari Universitas Diponegoro (UNDIP) Semarang. dan dari Universitas Kristen Satya Wacana (UKSW) Salatiga. Disiplin ilmu itu antara lain teknik elektro, komputer, manajemen dan ilmu sosiologi. Penulis memiliki pengalaman kerja pada industri elektronik dan sertifikasi keahlian dalam bidang Jaringan Internet, Telekomunikasi, Artificial Intelligence, Internet Of Things (IoT), Augmented Reality (AR), Technopreneurship, Internet Marketing dan bidang pengolahan dan analisa data (komputer statistik).

Penulis adalah pendiri dari Universitas Sains dan Teknologi Komputer (Universitas STEKOM) dan juga seorang dosen yang memiliki Jabatan Fungsional Akademik Lektor Kepala (Associate Professor) yang telah menghasilkan puluhan Buku Ajar ber ISBN, HAKI dari beberapa karya cipta dan Hak Paten pada produk IPTEK. Sejak tahun 2023 penulis tercatat sebagai Dosen luar biasa di Fakultas Ekonomi & Bisnis (FEB) Universitas Diponegoro Semarang. Penulis juga terlibat dalam berbagai organisasi profesi dan industri yang terkait dengan dunia usaha dan industri, khususnya dalam pengembangan sumber daya manusia yang unggul untuk memenuhi kebutuhan dunia kerja secara nyata.

YAYASAN PRIMA AGUS TEKNIK

PENERBIT :
YAYASAN PRIMA AGUS TEKNIK
Jl. Majapahit No. 605 Semarang
Telp. (024) 6723456. Fax. 024-6710144
Email : penerbit_ypat@stekom.ac.id

ISBN 978-634-7227-67-6 (PDF)
9 786347 227676