Deepfake:

Rekayasa Konten Palsu, Hasil produk AI

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

YAYASAN PRIMA AGUS TEKNIK

Rekayasa Konten Palsu, Hasil produk AI

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

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 978-634-7695-23-9 (PDF)

Deepfake : Rekayasa Konten Palsu, Hasil produk AI

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

ISBN : 978-634-7695-23-9 (PDF)

Editor :
Dr. Ir. Agus Wibowo, M.Kom, M.Si, M.M.

Penyunting :
Dr. Joseph Teguh Santoso, M.Kom.

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. (024) 6723456
Fax. 024-6710144
Email : penerbit_ypat@stekom.ac.id

Distributor Tunggal :
Universitas STEKOM
Jl. Majapahit no 605 Semarang
Telp. (024) 6723456
Fax. 024-6710144
Email : info@stekom.ac.id

KATA PENGANTAR

Puji syukur dipanjatkan ke hadirat Tuhan Yang Maha ESA atas segala Berkat, Rahmat dan Karunia, sehingga buku yang berjudul "***Deepfake : Rekayasa Konten Palsu, Hasil produk AI***" ini dapat diselesaikan dengan baik. Kehadiran buku ini dimaksudkan untuk memberikan kontribusi teoretis dan praktis dalam merespons perkembangan teknologi yang telah mendisrupsi tatanan sosial serta hukum secara fundamental.

Saat ini, masyarakat global tengah berada pada suatu era eksponensial yang belum pernah terjadi sebelumnya sebuah periodisasi di mana realitas visual dapat direkayasa secara artifisial, identitas suara seorang tokoh publik dapat dipalsukan, dan batas antara kebenaran serta disinformasi menjadi kian kabur. Teknologi *deepfake*, yang lahir dari konvergensi antara metode pembelajaran mendalam (*deep learning*) dan teknik rekayasa digital berbasis *Generative Adversarial Networks* (GANs), telah mengubah lanskap sirkulasi informasi secara paradigmatik. Fenomena manipulasi audio-visual yang dahulu hanya mampu diproduksi oleh studio sinematografi profesional dengan biaya tinggi, kini dapat diakses dan dioperasikan oleh siapa saja melalui perangkat lunak berskala personal dalam hitungan menit, bahkan tanpa memerlukan kualifikasi teknis yang mendalam.

Penulisan buku ini diinisiasi oleh sebuah kegelisahan akademik (*academic anxiety*) yang mendalam mengenai kesenjangan regulasi (*regulatoris gap*). Ketika akselerasi inovasi teknologi kecerdasan buatan bergerak secara eksponensial, instrumen hukum sering kali mengalami keterlambatan yang signifikan (*elemen lagging*), terjebak dalam rigiditas teks perundang-undangan konvensional yang dirumuskan jauh sebelum kecerdasan buatan mengintervensi ruang publik.

Di Indonesia, manifestasi empiris dari kekosongan hukum ini telah memicu eskalasi eksploitasi digital: mulai dari manipulasi rekaman pejabat publik demi motif penipuan finansial, pemanfaatan bot otomatisasi untuk memproduksi pornografi sintetis nonkonsensual (*non-consensual deepfake pornography*) yang menyasar mahasiswi selaku korban, hingga rekayasa biometrik wajah yang mengakibatkan kerugian finansial berskala masif pada sektor perbankan digital. Sementara itu, aparat penegak hukum dihadapkan pada kedaruratan metode pembuktian ilmiah (*scientific crime investigation*) karena delik-delik digital ini tidak meninggalkan impresi fisik konvensional, di sisi lain korban berada pada posisi rentan akibat ketiadaan kepastian interpretasi pasal yang spesifik.

Oleh karena itu, buku ini hadir bukan sekadar sebagai inventarisasi doktrinal atas pasal-pasal hukum positif yang berlaku, melainkan sebagai bentuk analisis kritis-analitis untuk menjawab urgensi yuridis: sejauh mana sistem hukum nasional memiliki tingkat kesiapan (*legal preparedness*) dan ketahanan (*legal resilience*) dalam menghadapi ancaman *deepfake*. Struktur pembahasan dalam buku ini dirancang secara sistematis ke dalam enam bagian yang saling berkesinambungan, meliputi: (1) konstruksi teoretis dan fondasi teknologi *deepfake* beserta implikasi sosio-psikologisnya; (2) studi kasus empiris di ranah domestik maupun global

pada periodisasi 2025–2026; (3) analisis kritis kedudukan hukum positif Indonesia beserta identifikasi *rechtsvacuum* di dalamnya; (4) kajian hukum komparatif (*comparative legal study*) terhadap regulasi global, khususnya traktat monumental *EU AI Act* dan *TAKE IT DOWN Act* Amerika Serikat; (5) anatomi pembuktian digital (*digital forensics*) serta doktrin pertanggungjawaban hukum platform penyelenggara sistem elektronik; dan (6) formulasi rekomendasi pembaruan hukum (*legal reform*) serta arah kebijakan literasi digital nasional.

Buku ini disusun menggunakan pendekatan *case-to-concept*, di mana setiap bab diorientasikan pada pembedahan skenario riil secara yuridis formal. Metodologi ini dipilih secara sengaja agar telaah teoretis hukum tidak terjebak dalam ruang abstraksi yang dogmatis, melainkan memiliki validitas dan aplikabilitas yang kuat terhadap realitas sosiologis masyarakat. Melalui pendekatan ini, para pembaca baik mahasiswa hukum, akademisi, praktisi peradilan, pembuat kebijakan (*policy makers*), maupun masyarakat umum diharapkan tidak hanya mampu mengidentifikasi hukum yang berlaku saat ini (*ius constitutum*), tetapi juga mampu memformulasikan gagasan kritis mengenai hukum yang dicita-citakan di masa depan (*ius constituendum*).

Penulis berharap semoga buku ini dapat memberikan sumbangsih yang berarti bagi akselerasi dan pengembangan ilmu hukum digital di Indonesia. Semoga segenap elemen masyarakat, akademisi, dan praktisi hukum dapat kian adaptif, responsif, dan bijaksana dalam menavigasi tantangan era kecerdasan buatan, dengan tetap teguh mengutamakan prinsip keadilan, kebenaran yuridis, serta perlindungan terhadap hak asasi manusia.

Semarang, Mei 2026
Penulis

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

DAFTAR ISI

BAB 1
DEFINISI DEEPFAKE

> *"Bayangkan Anda menerima sebuah video dari seseorang yang Anda percaya wajahnya jelas, suaranya familiar, gerak-geriknya terasa nyata. Namun kenyataannya, orang itu tidak pernah berkata atau melakukan apa pun yang tampak dalam video tersebut. Itulah deepfake: kebohongan yang dibalut wajah orang lain."*
>
> *— Gambaran metaforis tentang esensi deepfake*

Di era ketika layar menjadi jendela utama manusia melihat dunia, kepercayaan terhadap konten visual menjadi fondasi komunikasi modern. Kita cenderung mempercayai apa yang kita lihat sebuah tendensi kognitif yang telah mengakar sejak zaman prasejarah. Namun, perkembangan kecerdasan buatan telah melahirkan sebuah ancaman epistemik yang belum pernah dihadapi umat manusia sebelumnya: kemampuan untuk menciptakan kebohongan yang tampak sempurna secara visual.

Deepfake adalah salah satu manifestasi paling mengkhawatirkan dari kemajuan teknologi kecerdasan buatan. Dalam satu dekade terakhir, teknologi ini telah berkembang dari sekadar eksperimen akademis menjadi alat yang dapat diakses oleh siapa pun dengan perangkat komputer biasa. Bab ini hadir sebagai pijakan awal untuk memahami apa itu deepfake secara menyeluruh mulai dari akar kata dan konseptualnya, sejarah perkembangannya, hingga mekanisme teknologi yang bekerja di baliknya, serta potensi positif yang masih mungkin digali dari teknologi kontroversial ini. Pemahaman yang mendalam tentang deepfake bukan sekadar kepentingan teknis. Bagi para akademisi, praktisi hukum, pembuat kebijakan, dan masyarakat umum, memahami deepfake adalah prasyarat untuk meresponsnya secara tepat baik dalam dimensi regulasi, etika, maupun perlindungan hak-hak fundamental individu.

1.1 DEFINISI DEEPFAKE: DARI 'DEEP LEARNING' HINGGA 'FAKE'

Akar Kata dan Makna Terminologis

Istilah deepfake lahir dari perpaduan dua kata yang masing-masing membawa bobot makna tersendiri: deep (dalam) yang merujuk pada deep learning, dan fake (palsu) yang merujuk pada konten yang dipalsukan atau dimanipulasi. Secara etimologis, deepfake dapat didefinisikan sebagai konten audiovisual termasuk gambar, video, dan audio yang dihasilkan atau dimanipulasi menggunakan algoritma deep learning sedemikian rupa sehingga sulit dibedakan dari konten asli oleh persepsi manusia.

Secara lebih teknis, deep learning adalah sub-bidang dari machine learning (pembelajaran mesin) yang menggunakan jaringan saraf tiruan berlapis-lapis (*neural networks*) untuk memproses data dalam skala besar. Ketika teknologi ini diterapkan untuk memanipulasi identitas seseorang dalam konten visual dan auditori, hasilnya disebut sebagai deepfake. Definisi ini mencakup beberapa bentuk manipulasi, antara lain:

(1) penggantian wajah (*face swap*),
(2) animasi wajah (*face reenactment*),
(3) sintesis ucapan (*voice synthesis*), dan
(4) manipulasi ekspresi (*expression manipulation*).

Robert Chesney dan Danielle Citron, dua pakar hukum dari Universitas Texas dan Universitas Boston, adalah yang pertama kali mempopulerkan kajian akademis tentang deepfake dalam konteks hukum. Mereka mendefinisikan deepfake sebagai 'media sintetis di mana seseorang dalam gambar atau video yang ada digantikan kemiripannya dengan orang lain menggunakan teknik pembelajaran mesin dan kecerdasan buatan yang canggih.' Definisi ini menjadi salah satu rujukan paling sering dikutip dalam literatur hukum terkait deepfake.

Dimensi Konseptual Deepfake

Dalam memahami deepfake secara lebih komprehensif, penting untuk melihatnya dari beberapa dimensi:

- ☑ **Dimensi Teknis**: Deepfake adalah produk dari algoritma generatif berbasis deep learning, khususnya *Generative Adversarial Networks* (GAN) dan model difusi (diffusion models) yang akan dibahas lebih lanjut dalam bab ini.
- ☑ **Dimensi Persepsi**: Deepfake dirancang sedemikian rupa untuk menipu persepsi indrawi manusia baik penglihatan maupun pendengaran sehingga tampak autentik dan meyakinkan.
- ☑ **Dimensi Sosial-Hukum**: Deepfake mengaburkan batas antara nyata dan tidak nyata, menantang konsep-konsep fundamental dalam hukum seperti alat bukti, hak privasi, dan integritas informasi.
- ☑ **Dimensi Etis**: Deepfake menimbulkan pertanyaan mendalam tentang persetujuan (*consent*), identitas digital, dan tanggung jawab atas konten yang disebarkan.

Penting untuk dipahami bahwa deepfake tidak selalu berarti 'konten negatif.' Secara definitif, deepfake merujuk pada teknologi dan prosesnya. Namun dalam wacana publik dan hukum, istilah ini lebih sering diasosiasikan dengan penggunaan jahat penyebaran disinformasi, pelanggaran privasi, dan manipulasi politik. Ambivalensi ini menjadi salah satu tantangan dalam merumuskan regulasi yang efektif.

1.2 SEJARAH: DARI FACEAPP (2017) HINGGA GAN GENERASI TERBARU

Akar Historis: Manipulasi Gambar Sebelum AI

Manipulasi gambar bukanlah fenomena baru. Jauh sebelum era kecerdasan buatan, manusia telah memanipulasi gambar untuk berbagai tujuan dari propaganda politik hingga seni. Salah satu contoh paling awal adalah manipulasi foto Josef Stalin pada era 1930-an, di mana figur-figur yang tidak disukai rezim secara literal dihapus dari foto-foto resmi. Namun,

manipulasi semacam itu membutuhkan keahlian tinggi, waktu yang lama, dan hasilnya sering kali terdeteksi oleh mata yang terlatih.

Penemuan Photoshop oleh Adobe pada tahun 1988 menandai babak baru dalam sejarah manipulasi gambar membuatnya lebih mudah dan lebih terjangkau. Namun, teknologi ini masih membutuhkan keahlian manusia dan sebatas pada gambar statis. Manipulasi video masih merupakan domain para profesional dengan peralatan mahal.

Tabel 1.1 Tonggak Penting dalam Sejarah Deepfake

Tahun	Peristiwa Penting
2014	Ian Goodfellow dan rekan-rekannya di Universitas Montreal menerbitkan makalah seminal yang memperkenalkan konsep *Generative Adversarial Networks* (GAN). Karya ini menjadi fondasi teknologi deepfake modern.
2015	Algoritma Neural Style Transfer diperkenalkan oleh Gatys et al., memungkinkan penerapan gaya artistik gambar ke gambar lain secara otomatis menggunakan jaringan saraf konvolusi (CNN).
2017	FaceApp diluncurkan dan menjadi viral, memperkenalkan manipulasi wajah berbasis AI kepada jutaan pengguna awam. Meski bukan deepfake dalam pengertian teknis penuh, FaceApp menjadi awal kesadaran publik tentang kemampuan AI dalam manipulasi visual.
2017 (Des)	Pengguna Reddit dengan nama '*deepfakes*' memposting video manipulasi wajah selebriti. Nama pengguna tersebut kemudian menjadi nama untuk seluruh kategori konten manipulasi berbasis AI.
2018	Universitas Washington mempublikasikan *'Synthesizing Ob*ama' sebuah model yang mampu menciptakan video realistis Presiden Obama berbicara menggunakan audio yang disintesis. Ini menjadi salah satu deepfake paling awal yang mendapat perhatian akademis serius.
2018	BuzzFeed dan Jordan Peele merilis deepfake video Presiden Obama sebagai kampanye kesadaran publik tentang bahaya disinformasi berbasis deepfake.
2019	Samsung AI Research Center merilis teknologi *'Few-Shot Adversarial Learning'* yang mampu membuat deepfake dari satu foto wajah saja sebuah lompatan besar yang secara drastis menurunkan hambatan teknis pembuatan deepfake.
2019	Aplikasi DeepNude diluncurkan dan ditarik dalam waktu 24 jam setelah kontroversi besar. Aplikasi ini menghasilkan gambar telanjang sintetis dari foto perempuan berpakaian, memperlihatkan potensi penyalahgunaan deepfake yang paling berbahaya.
2020	Teknologi DALL-E dan GPT-3 dari OpenAI memperlihatkan kemampuan generatif AI yang jauh melampaui sebelumnya, membuka era baru sintesis konten multimodal.

2022	Perang Rusia-Ukraina menjadi konteks pertama di mana deepfake digunakan secara aktif dalam konflik bersenjata sebuah video deepfake Presiden Zelensky yang meminta tentara Ukraina menyerah sempat tersebar luas.
2023	Model difusi (*diffusion models*) seperti Stable Diffusion, Midjourney, dan DALL-E 3 menjadi mainstream. Kualitas gambar sintetis mencapai tingkat yang hampir tidak bisa dibedakan dari foto nyata.
2024–2025	Deepfake video real-time (*real-time deepfake*) mulai tersedia secara komersial. Teknologi ini memungkinkan penggantian wajah secara langsung dalam panggilan video, membuka ancaman baru dalam verifikasi identitas digital.

Evolusi Aksesibilitas: Dari Lab ke Smartphone

Salah satu perkembangan paling signifikan dalam sejarah deepfake adalah demokratisasi aksesnya. Jika pada tahun 2014 teknologi GAN hanya bisa dijalankan oleh peneliti dengan akses ke superkomputer, kini aplikasi deepfake tersedia secara gratis di smartphone. Tren ini mencerminkan paradoks kemajuan teknologi: semakin canggih sebuah teknologi, semakin mudah pula ia diakses oleh non-ahli.

"The democratization of deepfake technology has fundamentally altered the threat landscape. What once required state-level resources can now be accomplished by a determined individual with a consumer-grade GPU and freely available software."

— Paris & Donovan, Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence (2019)

Demokratisasi ini memiliki dua sisi yang sangat berbeda. Di satu sisi, ia memungkinkan kreativitas demokratis siapa pun dapat bereksperimen dengan teknologi canggih tanpa biaya tinggi. Di sisi lain, ia menurunkan hambatan bagi pelaku kejahatan, penyebar disinformasi, dan pelanggar privasi. Inilah mengapa regulasi deepfake harus mempertimbangkan konteks aksesibilitas ini secara serius.

1.3 CARA KERJA GENERATIVE ADVERSARIAL NETWORK (GAN) SECARA AWAM

Intuisi Dasar: Analogi Pemalsu dan Detektif

Generative Adversarial Network, atau GAN, adalah arsitektur kecerdasan buatan yang diperkenalkan oleh Ian Good fellow bersama rekan-rekannya pada tahun 2014. Untuk memahami GAN tanpa latar belakang teknis mendalam, kita dapat menggunakan sebuah analogi yang kuat: "*permainan antara pemalsu dan detektif seni*".

Bayangkan seorang pemalsu lukisan (sebut saja Pelukis Palsu) yang terus-menerus mencoba menciptakan lukisan tiruan yang terlihat seperti karya asli. Di sisi lain, ada seorang detektif seni (sebut saja Sang Ahli) yang bertugas membedakan lukisan asli dari tiruan. Kedua pihak ini berada dalam persaingan yang terus-menerus: Pelukis Palsu selalu berusaha

membuat tiruannya semakin sempurna agar berhasil menipu Sang Ahli, sementara Sang Ahli terus mengasah kemampuannya mendeteksi pemalsuan.

Dalam terminologi GAN, Pelukis Palsu disebut Generator, dan Sang Ahli disebut Discriminator. Keduanya adalah jaringan saraf tiruan yang dilatih secara bersamaan dan saling berkompetisi:

- **Generator**: Bertugas menghasilkan data sintetis (gambar, video, suara) yang tampak nyata. Generator tidak pernah melihat data asli secara langsung namun belajar dari umpan balik Discriminator.
- **Discriminator**: Bertugas membedakan data asli dari data yang dibuat oleh Generator. Ia dilatih dengan data asli dan data sintetis, kemudian memberikan penilaian: 'asli' atau 'palsu.'

Proses Pelatihan GAN Langkah demi Langkah

Proses pelatihan GAN dapat dipahami melalui tahapan berikut:

1. ***Pengumpulan Data Pelatihan***: Model GAN membutuhkan ribuan hingga jutaan contoh data asli misalnya foto wajah manusia dari berbagai sudut, ekspresi, dan kondisi pencahayaan. Semakin beragam dan banyak data pelatihan, semakin realistis hasil akhir yang dapat dihasilkan.
2. ***Generator Menghasilkan Data Awal***: Generator dimulai dengan menghasilkan gambar-gambar yang sepenuhnya acak dan tidak bermakna pada tahap ini, outputnya hanyalah derau visual (*noise*) yang tidak menyerupai wajah manusia sama sekali.
3. ***Discriminator Memberikan Penilaian***: Discriminator menerima dua jenis input gambar asli dari dataset dan gambar palsu dari Generator. Ia kemudian memberikan skor probabilitas: seberapa yakin ia bahwa gambar tersebut asli?
4. ***Umpan Balik dan Perbaikan***: Berdasarkan penilaian Discriminator, Generator memperbarui parameternya (melalui proses matematika yang disebut backpropagation untuk menghasilkan gambar yang lebih meyakinkan. Discriminator juga memperbarui parameternya agar lebih pandai mendeteksi pemalsuan.
5. ***Iterasi Berulang***: Proses ini diulang jutaan kali. Seiring waktu, Generator menjadi semakin ahli dalam menciptakan gambar realistis, sementara Discriminator menjadi semakin sensitif dalam mendeteksi anomali. Keseimbangan yang dicapai (Nash Equilibrium) menghasilkan Generator yang mampu menciptakan gambar yang hampir tidak bisa dibedakan dari aslinya.

ANALOGI SEDERHANA: GAN dalam Kehidupan Sehari-hari

Bayangkan dua orang anak yang belajar memasak:

• Anak A (Generator) terus-menerus mencoba memasak rendang semirip mungkin dengan resep asli.

• Anak B (Discriminator) terus mencicipi masakan A dan membandingkannya dengan rendang buatan chef profesional.

• Setiap kali Anak B mengatakan 'Ini palsu, kurang gurih!', Anak A memperbaiki resepnya.

• Setiap kali Anak B gagal mendeteksi perbedaan, Anak B juga belajar untuk lebih cermat.

• Setelah ribuan percobaan, Anak A menghasilkan rendang yang bahkan seorang sommelier tidak bisa bedakan dari yang asli.

Itulah esensi GAN kompetisi yang melahirkan kesempurnaan palsu.

Arsitektur Khusus untuk Deepfake Wajah

Untuk menghasilkan deepfake wajah yang spesifik, teknologi GAN dikembangkan lebih lanjut menjadi beberapa varian arsitektur yang lebih canggih:

- ☑ **Autoencoder:** Teknologi ini mengompres (*encode*) wajah seseorang menjadi representasi matematis ringkas, kemudian mendekompres (*decode*) representasi tersebut untuk merekonstruksi wajah. Untuk face swap, dua encoder dilatih pada dua wajah berbeda, tetapi hanya satu decoder yang digunakan, sehingga wajah seseorang 'ditata' mengikuti struktur wajah orang lain.
- ☑ **FaceSwap GAN (FSGAN):** Dikembangkan secara spesifik untuk penggantian wajah (face swap), mampu mempertahankan ekspresi, postur kepala, dan kondisi pencahayaan dari subjek target.
- ☑ ***First Order Motion Model* (FOMM):** Memungkinkan animasi wajah dari foto tunggal. Model ini mempelajari gerakan dari video sumber, kemudian menerapkannya pada foto diam dari orang yang berbeda.
- ☑ **Diffusion Models:** Generasi terbaru dari model generatif yang menggunakan pendekatan berbeda dari GAN. Alih-alih kompetisi Generator-Discriminator, diffusion models belajar dengan 'menghapus derau' secara bertahap dari gambar acak hingga menghasilkan gambar koheren. Model seperti Stable Diffusion dan DALL-E 3 menggunakan pendekatan ini.

Perlu dicatat bahwa perkembangan teknologi deepfake tidak berhenti pada GAN. Seiring munculnya *Large Language Models* (LLM) dan model multimodal, kemampuan untuk mensintesis konten palsu yang meyakinkan semakin melampaui kemampuan manusia dalam mendeteksinya.

1.4 PERBEDAAN DEEPFAKE, SHALLOWFAKE, DAN CHEAPFAKE

Dalam diskursus akademis dan jurnalisme, tidak semua manipulasi konten digital disebut deepfake. Penting untuk memahami tiga kategori utama manipulasi konten digital karena perbedaannya memiliki implikasi hukum, teknis, dan kebijakan yang berbeda-beda.

Tabel 1.2 Tiga Katergori Utama Manipulasi Konten

Kategori	Teknologi yang Digunakan	Karakteristik Utama

Deepfake	Deep learning, GAN, Diffusion Models	Menggunakan AI canggih; hasilnya sangat realistis; sulit dideteksi secara manual; umumnya membutuhkan komputasi signifikan (meskipun kini semakin mudah)
Shallowfake	Algoritma AI sederhana; kloning suara dasar; filter wajah	Menggunakan AI namun tidak sekompleks GAN; hasilnya kurang meyakinkan dibanding deepfake; lebih mudah dideteksi dengan pemeriksaan cermat
Cheapfake	Editing video biasa (cut, paste, speed-up, slow-down); tidak menggunakan AI	Tidak menggunakan AI sama sekali; teknik manipulasi sederhana; namun tetap mampu menyesatkan jika dikontekstualisasikan secara salah

1). Deepfake

Deepfake, sebagaimana telah dijelaskan di atas, adalah manipulasi konten yang menggunakan algoritma deep learning untuk menghasilkan atau memodifikasi konten audiovisual sehingga tampak sangat realistis. Karakteristik utamanya adalah:

- Memerlukan data pelatihan yang besar (ribuan gambar/video wajah target).
- Hasilnya dapat melewati deteksi kasual manusia.
- Dapat digunakan untuk membuat seseorang tampak mengatakan atau melakukan sesuatu yang tidak pernah dilakukannya.

Contoh: Video Presiden Zelensky yang meminta pasukan Ukraina menyerah (2022); video pornografi deepfake yang menampilkan wajah selebriti atau individu privat.

2). Shallowfake

Shallowfake adalah istilah yang kurang populer namun penting dalam literatur akademis. Shallowfake menggunakan teknologi AI yang lebih sederhana dibandingkan GAN, seperti:

- Filter wajah sederhana (seperti fitur AR di aplikasi media sosial).
- Kloning suara berbasis analisis spektral sederhana.
- Morphing gambar (pencampuran wajah) dengan algoritma tradisional.

Meskipun kurang canggih dari deepfake, shallowfake tetap dapat menyesatkan, terutama ketika dikombinasikan dengan narasi yang meyakinkan.

3). Cheapfake

Cheapfake istilah yang dipopulerkan oleh Amy Ben-David dan Hélène Murie pada 2020 merujuk pada manipulasi konten yang tidak menggunakan AI sama sekali, melainkan teknik editing video konvensional:

- Memperlambat atau mempercepat video untuk mengubah persepsi kondisi seseorang (misalnya, video politisi yang diperlambat untuk membuatnya tampak mabuk atau linglung).
- Mengambil cuplikan video di luar konteks aslinya.
- Menggabungkan dua klip video yang tidak berkaitan untuk menciptakan narasi palsu.

➲ Mengubah subtitle atau teks yang muncul bersamaan dengan video.

Dari perspektif hukum, pembedaan antara ketiga kategori ini sangat relevan. Sebagian besar diskusi legislatif berfokus pada deepfake, sementara cheapfake yang justru lebih mudah dibuat dan lebih luas persebarannya sering luput dari perhatian regulasi. Ini menciptakan celah hukum yang berbahaya.

⚠ CATATAN PENTING UNTUK AKADEMISI HUKUM

Dalam konteks pembuktian hukum, perbedaan antara deepfake, shallowfake, dan cheapfake memiliki konsekuensi yang berbeda:

• Deepfake: Memerlukan analisis forensik digital canggih untuk mendeteksi; menantang admissibility alat bukti elektronik.

• Shallowfake: Dapat dideteksi dengan tools analisis yang lebih terjangkau; namun tetap memerlukan keahlian teknis.

• Cheapfake: Dapat dideteksi dengan fact-checking kontekstual; namun sering lolos karena keterbatasan waktu dan sumber daya.

Implikasi: Regulasi hukum perlu mencakup ketiga kategori ini, tidak hanya deepfake yang canggih secara teknis.

1.5 POTENSI POSITIF: FILM, MEDIS, PENDIDIKAN, DAN AKSESIBILITAS

Teknologi ini, pada dasarnya, adalah alat dan seperti semua alat, dampaknya bergantung pada cara penggunaannya. Berbagai bidang telah menemukan aplikasi positif dari teknologi deepfake dan sintesis konten berbasis AI yang patut dipertimbangkan secara serius, khususnya dalam merumuskan regulasi yang tidak menghambat inovasi bermanfaat.

Industri Film dan Hiburan

Industri hiburan adalah salah satu sektor yang paling aktif mengadopsi teknologi deepfake secara legal dan etis. Beberapa aplikasi yang telah terbukti memberikan nilai positif antara lain:

- ☑ De-aging dan Restorasi Karakter Ikonik: Film The Irishman (2019) karya Martin Scorsese menggunakan teknologi deepfake untuk meremajakan (de-aging) penampilan Robert De Niro, Al Pacino, dan Joe Pesci di beberapa adegan. Teknologi serupa digunakan dalam Rogue One: A Star Wars Story (2016) untuk 'menghidupkan kembali' aktris Carrie Fisher yang telah meninggal dunia.
- ☑ Lokalisasi Konten Lintas Bahasa: Beberapa perusahaan media menggunakan teknologi deepfake untuk menyinkronkan gerakan bibir aktor dalam bahasa yang berbeda, menghasilkan dubbing yang jauh lebih natural dibandingkan dubbing konvensional.
- ☑ Stunt Double Digital: Teknologi deepfake memungkinkan penggantian wajah stunt double dengan wajah aktor utama secara lebih meyakinkan, mengurangi risiko fisik bagi para aktor.

- ☑ Preservasi Warisan Artistik: Museum dan lembaga budaya menggunakan deepfake untuk 'menghidupkan kembali' tokoh-tokoh sejarah dalam presentasi interaktif, memberikan pengalaman edukatif yang imersif.

Bidang Medis dan Kesehatan

Dalam bidang medis, teknologi sintesis gambar dan video berbasis AI membuka kemungkinan-kemungkinan yang sebelumnya tidak terbayangkan:

- ☑ Simulasi Prosedur Medis: Deepfake dapat digunakan untuk menciptakan simulasi visual pasien dengan berbagai kondisi medis, membantu pelatihan dokter dan tenaga medis tanpa memerlukan kasus nyata yang langka.
- ☑ Terapi Psikologis: Beberapa pendekatan terapi menggunakan avatar wajah yang disintetis untuk membantu pasien trauma menghadapi memori atau konfrontasi imajiner yang terkontrol. Teknologi deepfake memungkinkan avatar ini terlihat sangat realistis.
- ☑ Komunikasi Pasien Demensia: Penelitian di beberapa universitas menunjukkan bahwa video deepfake dari anggota keluarga yang telah meninggal atau tidak dapat hadir dapat membantu pasien Alzheimer stadium lanjut yang tidak dapat memproses informasi bahwa orang tersebut telah tiada.
- ☑ Augmentasi Data Medis: Dalam machine learning untuk diagnosis medis, deepfake digunakan untuk menghasilkan data gambar medis sintetis (CT scan, MRI) untuk melatih algoritma diagnosis AI tanpa memerlukan data pasien nyata — menjaga privasi sambil meningkatkan akurasi model.

Bidang Pendidikan

Teknologi deepfake menawarkan revolusi dalam pengalaman pendidikan:

- ☑ Pembelajaran Sejarah Imersif: Bayangkan seorang siswa dapat 'berinteraksi' dengan Soekarno, Nelson Mandela, atau Albert Einstein dalam simulasi pendidikan. Deepfake memungkinkan rekonstruksi pidato dan pernyataan tokoh sejarah berdasarkan catatan dokumentasi yang ada.
- ☑ Personalisasi Konten Pembelajaran: Video pembelajaran dapat disintetis ulang dalam berbagai bahasa atau dengan penyesuaian kecepatan bicara dan ekspresi untuk memenuhi kebutuhan individu siswa.
- ☑ Simulasi Role-Play untuk Pelatihan Profesional: Dalam pelatihan hukum, kedokteran, atau diplomasi, deepfake dapat menciptakan skenario interaktif yang sangat realistis misalnya, mahasiswa hukum dapat 'berlatih' berhadapan dengan saksi atau hakim yang disintetis secara digital.
- ☑ Aksesibilitas Materi Kelas Dunia: Kuliah dari akademisi terkemuka dapat disintetis dalam bahasa lokal, membuat konten pendidikan berkualitas tinggi lebih aksesibel bagi pelajar di seluruh dunia.

Aksesibilitas untuk Penyandang Disabilitas

Ini adalah salah satu aplikasi paling bermakna secara kemanusiaan dari teknologi deepfake:

- ☑ Komunikasi bagi Penderita ALS dan Disabilitas Bicara: Perusahaan seperti Resemble AI dan Eleven Labs telah mengembangkan teknologi kloning suara yang memungkinkan

penderita ALS (*Amyotrophic Lateral Sclerosis*) seperti yang dialami fisikawan Stephen Hawking untuk 'menyimpan' suara mereka sebelum kemampuan bicara hilang, dan menggunakannya setelah kondisi semakin parah.

- ☑ Terjemahan Bahasa Isyarat Otomatis: Deepfake memungkinkan pembuatan avatar tangan yang natural untuk menerjemahkan ucapan ke bahasa isyarat secara real-time, jauh lebih ekspresif dibanding animasi konvensional.
- ☑ Pembacaan Teks Adaptif: Konten video dapat dimodifikasi secara otomatis untuk memperlambat gerakan bibir pembicara, membantu pengguna tuli yang mengandalkan lip-reading.

PRINSIP KUNCI: NETRALITAS TEKNOLOGI

Dari kelima bidang di atas, sebuah prinsip penting muncul: teknologi deepfake, pada dasarnya, adalah netral.

Yang menentukan dampaknya adalah:

• Tujuan penggunaan (for what purpose)

• Transparansi (apakah pengguna mengetahui konten tersebut adalah sintetis?)

• Persetujuan (apakah subjek telah memberikan consent?)

• Akuntabilitas (siapa yang bertanggung jawab atas konten yang dihasilkan?)

Implikasi bagi regulasi hukum: Larangan total terhadap teknologi deepfake akan mengorbankan manfaat-manfaat di atas. Pendekatan berbasis konteks, tujuan, dan persetujuan lebih tepat secara hukum dan etis.

Rangkuman Bab

Bab ini telah mengantarkan pada pemahaman mendasar tentang deepfake dari berbagai dimensi. Beberapa poin kunci yang perlu diingat:

- ☑ Deepfake adalah perpaduan antara *deep learning* dan *fake* merujuk pada konten audiovisual yang dihasilkan atau dimanipulasi menggunakan algoritma kecerdasan buatan sehingga tampak nyata secara perseptual.
- ☑ Sejarah deepfake bermula dari penemuan GAN oleh Ian Goodfellow pada tahun 2014, berkembang pesat melalui berbagai tonggak teknologi, dan kini telah menjadi teknologi yang dapat diakses oleh publik umum melalui aplikasi smartphone.
- ☑ GAN bekerja melalui kompetisi antara Generator (pembuat konten palsu) dan Discriminator (pendeteksi kepalsuan) sebuah proses iteratif yang menghasilkan konten sintetis yang semakin sempurna.

- ☑ Deepfake berbeda dari shallowfake dan cheapfake dalam hal kompleksitas teknologi dan tingkat realisme namun ketiga kategori ini sama-sama menimbulkan tantangan bagi integritas informasi dan sistem hukum.
- ☑ Teknologi deepfake memiliki potensi positif yang signifikan di bidang hiburan, medis, pendidikan, dan aksesibilitas menunjukkan bahwa regulasi harus bersifat selektif dan kontekstual, bukan pelarangan menyeluruh.

BAB 2
TAKSONOMI DEEPFAKE: JENIS DAN MODUSNYA

Bab 1 memperkenalkan deepfake sebagai konsep dan teknologi, maka Bab 2 ini hadir sebagai peta navigasi yang lebih terperinci. Memahami berbagai jenis deepfake dan modus penyalahgunaannya bukan sekadar kepentingan akademis ini adalah kebutuhan praktis bagi siapa pun yang bergerak di bidang hukum, kebijakan publik, jurnalisme, dan penegakan hak-hak digital. Taksonomi, atau sistem klasifikasi, penting karena ancaman yang berbeda memerlukan respons yang berbeda. Sebuah undang-undang yang merespons face swap pornografi tidak otomatis relevan untuk deepfake dokumen identitas. Kebijakan platform yang menangani kloning suara tidak identik dengan yang menangani manusia sintetis. Memahami peta ini adalah langkah pertama menuju respons hukum yang presisi dan efektif.

Bab ini memetakan enam jenis utama deepfake berdasarkan cara kerjanya, disertai modus-modus penyalahgunaan nyata yang telah terdokumentasi, serta implikasi hukum awal yang akan dikembangkan lebih dalam pada bab-bab selanjutnya.

2.1 PETA TAKSONOMI DEEPFAKE: GAMBARAN UMUM

Sebelum membahas masing-masing jenis secara mendalam, penting untuk memiliki gambaran keseluruhan tentang lanskap deepfake. Taksonomi yang digunakan dalam bab ini dibangun berdasarkan tiga dimensi utama:

- Modalitas Output: Apa yang dihasilkan? (gambar wajah, video, suara, dokumen, manusia sintetis)
- Teknik Utama: Bagaimana konten tersebut dihasilkan? (face swap, lip sync, voice synthesis, full generation, text-to-video)
- Modus Penyalahgunaan: Untuk apa biasanya disalahgunakan? (penipuan identitas, disinformasi, pornografi non-konsensual, fraud finansial, dst.)

Tabel 2.1 Gambaran Umum Taksonomi Deepfake

No.	Jenis Deepfake	Output Utama	Tingkat Ancaman	Contoh Kasus Nyata
1	Face Swapping	Video wajah tertukar	🔴 Sangat Tinggi	Pornografi deepfake selebriti; fraud panggilan video CEO
2	Lip-Syncing / Puppeteering	Video ucapan dimanipulasi	🔴 Sangat Tinggi	Video Zelensky 2022; klip politisi manipulatif
3	Voice Cloning	Audio suara tiruan	🟠 Tinggi	Penipuan telepon 'suara CEO'; fraud transfer dana

4	Synthetic Humans	Manusia palsu sepenuhnya	🟠 Tinggi	Akun media sosial palsu; agen intelijen fiktif
5	Text-to-Video AI	Video dari teks (2025+)	🔴 Kritis	Sora, Kling AI; disinformasi berita tiruan
6	Deepfake Dokumen	KTP, ijazah, kontrak palsu	🟠 Tinggi	Penipuan KYC perbankan; ijazah palsu melamar kerja

Catatan penting: Tingkat ancaman dalam tabel di atas bersifat relatif dan kontekstual. Dalam konteks tertentu, deepfake dokumen bisa sama berbahayanya dengan face swap. Klasifikasi ini dimaksudkan sebagai panduan umum, bukan hierarki yang kaku.

2.2 FACE SWAPPING (PENUKARAN WAJAH)

Apa itu Face Swapping?

Face swapping adalah jenis deepfake paling awal yang dipopulerkan dan hingga kini masih menjadi yang paling dikenal secara publik. Secara teknis, face swapping adalah proses penggantian wajah satu individu dalam sebuah gambar atau video dengan wajah individu lain sedemikian rupa sehingga hasilnya tampak meyakinkan secara visual, termasuk dalam hal ekspresi, gerakan, dan pencahayaan.

Berbeda dengan penggantian wajah sederhana dalam editing foto tradisional, face swapping berbasis deep learning mampu menjaga konsistensi visual secara dinamis artinya wajah yang disisipkan dapat mengikuti gerakan kepala, perubahan ekspresi, dan variasi pencahayaan dari video asli secara real-time.

Cara kerja teknis face swapping

Proses face swapping modern umumnya melibatkan beberapa tahapan:

1. Deteksi Wajah (*Face Detection*): Algoritma secara otomatis mengidentifikasi dan mengisolasi area wajah dalam setiap frame video menggunakan model pendeteksi wajah seperti MTCNN atau RetinaFace.
2. Ekstraksi Landmark Wajah: Titik-titik kunci pada wajah (mata, hidung, mulut, kontur muka) dipetakan secara presisi biasanya 68–106 titik referensi.
3. Encoding Wajah: Wajah kedua individu (sumber dan target) dikompresi menjadi representasi matematis (*latent vector*) menggunakan encoder yang terlatih pada ribuan gambar wajah.
4. Decoding dan Transplansi: Decoder merekonstruksi wajah sumber dengan struktur geometris wajah target, menyesuaikan orientasi kepala, ekspresi, dan kondisi pencahayaan.
5. Blending dan Pasca-Produksi: Tepi wajah diblending secara halus dengan kulit sekitarnya menggunakan algoritma Poisson blending atau teknik sejenis untuk menghilangkan artefak visual yang kentara.

Tabel 2.2 Modus Penyalahgunaan Face Swapping

MODUS 1 Pornografi Non-Konsensual (NCII Non-Consensual Intimate Imagery)
Deskripsi: Wajah korban (biasanya perempuan) ditempel pada tubuh aktor porno tanpa persetujuan.
Skala: Menurut laporan Sensity AI (2023), lebih dari 95% deepfake video online adalah NCII. Platform khusus deepfake pornografi telah melayani jutaan pengguna.
Korban: Selebriti, politisi perempuan, mantan pasangan, rekan kerja, mahasiswi — siapa pun yang foto atau videonya tersedia online.
Dampak: Trauma psikologis berat, kerusakan reputasi permanen, perundungan siber, hingga pemerasan (extortion) berbasis deepfake.
Contoh Kasus: Pada 2023, deepfake NCII dari penyanyi Taylor Swift tersebar masif di platform X (Twitter), memaksa platform tersebut memblokir pencarian nama sang artis sementara.
MODUS 2 Penipuan Identitas dalam Panggilan Video (Video KYC Fraud)
Deskripsi: Pelaku menggunakan face swap real-time untuk menyamar sebagai orang lain dalam panggilan video termasuk proses verifikasi identitas (KYC) di layanan keuangan digital.
Teknik: Aplikasi seperti DeepFaceLive memungkinkan penggantian wajah secara real-time dalam panggilan Zoom, Teams, atau aplikasi banking.
Contoh Kasus: Pada Februari 2024, seorang pegawai keuangan perusahaan multinasional di Hong Kong mentransfer HK$200 juta (sekitar Rp400 miliar) setelah konferensi video dengan 'CFO' perusahaan yang ternyata adalah deepfake seluruhnya.
Implikasi Hukum: Kasus Hong Kong 2024 menjadi preseden penting menunjukkan bahwa deepfake real-time sudah cukup canggih untuk menipu dalam konteks profesional bertarget tinggi.
Relevansi Indonesia: Dengan pesatnya pertumbuhan layanan keuangan digital (neobank, P2P lending, dompet digital), ancaman video KYC fraud semakin relevan.
MODUS 3 Disinformasi dan Manipulasi Politik
Deskripsi: Face swap digunakan untuk menempatkan wajah tokoh politik di video yang menggambarkan perilaku atau pernyataan yang tidak pernah dilakukan.
Contoh: Video beredar di media sosial Indonesia menampilkan 'seorang pejabat' dalam situasi memalukan kebanyakan cheapfake, namun deepfake mulai digunakan.

Ancaman Demokrasi: Deepfake melemahkan kepercayaan publik terhadap bukti visual, bahkan ketika video tersebut asli efek 'liar's dividend' (manfaat pembohong).
Liar's Dividend: Konsep penting karena deepfake ada, politisi yang tertangkap kamera melakukan pelanggaran kini dapat mengklaim bahwa video tersebut adalah deepfake, mengaburkan akuntabilitas.

⚠ LIAR'S DIVIDEND — Ancaman Tersembunyi dari Deepfake

Konsep 'Liar's Dividend' (Chesney & Citron, 2019) merujuk pada keuntungan yang diperoleh pembohong dari keberadaan deepfake:

Ketika publik mengetahui bahwa video realistis bisa dipalsukan, mereka cenderung meragukan semua video termasuk yang asli.

Akibatnya: Seorang pejabat yang tertangkap kamera melakukan korupsi bisa mengklaim 'itu deepfake!' dan sebagian publik akan mempercayainya.

Ini berarti deepfake merusak kepercayaan publik bahkan TANPA harus digunakan secara aktif cukup dengan keberadaannya saja.

Implikasi hukum: Sistem pembuktian konvensional yang mengandalkan bukti video terancam secara fundamental.

2.3 LIP-SYNCING/PUPPETEERING (MANIPULASI UCAPAN)

Apa Itu Lip-Syncing Deepfake?

Lip-syncing deepfake disebut juga puppeteering atau talking head deepfake adalah jenis manipulasi yang lebih spesifik dari face swap. Alih-alih mengganti seluruh wajah, lip-syncing deepfake hanya memanipulasi area mulut dan gerakan bibir seseorang agar sinkron dengan audio yang berbeda dari yang asli. Hasilnya: seseorang tampak berkata sesuatu yang tidak pernah ia ucapkan, dengan ekspresi dan konteks yang tampak natural. Teknologi ini awalnya dikembangkan untuk keperluan dubbing film menyinkronkan gerakan bibir aktor dengan dialog dalam bahasa lain. Namun pemanfaatannya telah berkembang jauh melampaui tujuan awal tersebut, termasuk ke ranah manipulasi politik dan penipuan.

Teknologi di Balik Lip-Syncing

Beberapa model teknologi yang mendominasi lip-syncing deepfake:

- ***Wav2Lip***: Model open-source populer yang mampu mensinkronkan gerakan bibir pada video wajah dengan audio apa pun, dengan akurasi temporal yang tinggi. Tersedia gratis di GitHub.
- ***D-ID (Digital Humans)***: Platform komersial yang memungkinkan pengguna mengunggah foto dan teks, lalu menghasilkan video 'orang tersebut' berbicara sesuai teks yang diberikan.

- ***HeyGen***: Platform yang memungkinkan kloning avatar video pengguna merekam diri selama beberapa menit, lalu platform tersebut dapat membuat video mereka berbicara dalam bahasa dan konten apa pun yang diinginkan.
- ***Synthesizing Obama (UW, 2017)***: Makalah akademis seminal dari Universitas Washington yang mendemonstrasikan kemampuan mensintesis video Obama berbicara dari rekaman audio saja karya yang menjadi titik balik dalam kesadaran publik.

Tabel 2.3 Modus Penyalahgunaan Lip-Syncing Deepfake

MODUS UTAMA Disinformasi Politik dan Manipulasi Pemilu
Kasus Zelensky 2022: Pada Maret 2022, sebuah video deepfake beredar di media sosial menampilkan Presiden Volodymyr Zelensky tampak memerintahkan pasukan Ukraina untuk meletakkan senjata dan menyerah. Video tersebut menggunakan lip-sync deepfake pada rekaman pidato Zelensky yang asli. Meskipun terdeteksi dan dibantah dengan cepat, video ini menunjukkan potensi deepfake sebagai senjata perang informasi.
Manipulasi Debat: Klip singkat perdebatan atau wawancara politisi dapat dimanipulasi membuat mereka tampak mengakui sesuatu, mengucapkan kata kasar, atau menyatakan posisi yang bertentangan dengan pandangan aslinya.
Konteks Pemilu Indonesia: Dalam konteks demokrasi digital Indonesia dengan lebih dari 200 juta pengguna internet dan penetrasi media sosial yang tinggi, ancaman lip-sync deepfake dalam siklus pemilu sangat nyata dan perlu diantisipasi secara regulatif.
Astroturfing Digital: Tokoh-tokoh masyarakat atau pemuka agama dapat 'dibuat' tampak mendukung atau menentang kandidat atau kebijakan tertentu.
MODUS 2 Penipuan Korporat dan Business Email Compromise (BEC) Versi Video
Deskripsi: Alih-alih email palsu, penipu menggunakan video deepfake lip-sync dari eksekutif perusahaan untuk memerintahkan transfer dana atau perubahan kebijakan.
Kasus UAE 2020: Seorang manajer bank di UAE diperintahkan via telepon oleh 'direkturnya' (voice cloning + lip-sync video) untuk mentransfer $35 juta. Ini adalah salah satu kasus voice deepfake fraud terbesar yang terdokumentasi secara hukum.
Evolusi Ancaman: BEC tradisional menggunakan email kini berkembang ke audio deepfake, dan selanjutnya ke video deepfake real-time dalam panggilan konferensi.

2.4 VOICE CLONING (KLONING SUARA)

Apa itu Voice Cloning ?

Voice cloning adalah teknologi yang menggunakan AI untuk menduplikasi karakteristik unik suara seseorang termasuk intonasi, aksen, timbre, kecepatan bicara, dan pola emosional

kemudian menggunakannya untuk menghasilkan ucapan baru yang seolah-olah berasal dari orang tersebut. Berbeda dari lip-syncing yang memerlukan video sumber, voice cloning murni beroperasi pada modalitas audio.

Kemajuan teknologi voice cloning sangat dramatis dalam beberapa tahun terakhir. Jika pada 2019 kloning suara yang meyakinkan memerlukan ribuan jam rekaman audio, kini model seperti ElevenLabs dan Tortoise-TTS mampu menghasilkan kloning suara berkualitas tinggi hanya dari sampel audio tiga hingga lima menit saja.

Tabel 2.4 Teknologi Voice Cloning Terkemuka

Platform/Model	Sampel Min.	Bahasa	Akurasi	Akses
ElevenLabs	1 menit	29+ bahasa	⭐⭐⭐⭐⭐	Komersial (freemium)
Tortoise-TTS	3–5 menit	Inggris utama	⭐⭐⭐⭐	Open source
Microsoft VALL-E	3 detik	Multi	⭐⭐⭐⭐⭐	Riset (terbatas)
Resemble AI	5 menit	Multi	⭐⭐⭐⭐	Komersial
Coqui TTS	Beberapa mnt	Multi	⭐⭐⭐	Open source
OpenAI Voice Engine	15 detik	Multi	⭐⭐⭐⭐⭐	Terbatas (2024)

Yang paling mengejutkan adalah Microsoft VALL-E (2023): model ini mampu menghasilkan kloning suara yang meyakinkan hanya dari tiga detik sampel audio cukup dengan sepotong klip telepon atau video media sosial. Ini berarti suara siapa pun yang pernah berbicara di depan mikrofon secara publik dapat dikloning.

Tabel 2.5 Modus Penyalahgunaan Voice Cloning

📞 MODUS 1 **Penipuan Telepon 'Darurat Keluarga' (Grandparent Scam)**
Modus Operandi: Pelaku mengkloning suara anggota keluarga (anak, cucu, pasangan) menggunakan klip audio dari media sosial, lalu menelepon korban mengaku dalam situasi darurat dan membutuhkan uang segera.
Skala: FBI melaporkan bahwa 'grandparent scam' berbasis voice cloning telah menyebabkan kerugian lebih dari $11 juta di AS pada 2023 saja.

Kasus Nyata (2023): Seorang ibu di Arizona menerima telepon dari 'putrinya' yang menangis, mengaku diculik, dan membutuhkan tebusan $50.000. Suaranya identik ternyata voice cloning dari video media sosial sang putri.

Adaptasi Indonesia: Di Indonesia, modus ini berpotensi dikombinasikan dengan cerita 'kecelakaan', 'ditangkap polisi', atau 'rumah sakit' skenario yang sudah umum dalam penipuan telepon konvensional.

MODUS 2 Penipuan Korporat — CEO Fraud via Audio

Deskripsi: Suara eksekutif senior dikloning dan digunakan untuk menginstruksikan pegawai keuangan melakukan transfer dana darurat ke rekening tertentu.

Kasus UAE 2020: Manajer bank menerima telepon dari 'direktur' (suara kloning) yang menginstruksikan transfer $35 juta. FBI mengonfirmasi penggunaan voice cloning dalam kasus ini.

Kasus Energi 2019: CEO sebuah perusahaan energi Inggris mentransfer €220.000 setelah menerima instruksi telepon dari suara yang 'terdengar persis seperti' CEO perusahaan induk Jerman.

Mengapa Efektif: Otoritas yang melekat pada suara atasan, situasi 'darurat' yang dibuat-buat, dan instruksi untuk ti

dak memberitahu siapa pun menciptakan kondisi psikologis ideal untuk penipuan.

MODUS 3 Pelanggaran Hak Privasi dan Pencemaran Nama Baik

Podcast/Audio Palsu: Konten audio yang menampilkan 'suara' tokoh publik, ulama, pejabat, atau akademisi menyatakan pandangan yang tidak pernah mereka ungkapkan.

Pemerasan Berbasis Audio: Klip audio deepfake yang menggambarkan korban mengaku terlibat kejahatan, perselingkuhan, atau tindakan memalukan, digunakan sebagai alat pemerasan.

Manipulasi Rekaman Suara sebagai Alat Bukti: Rekaman audio yang dimanipulasi bahkan jika bukan kloning sempurna

dapat menimbulkan keraguan dalam proses hukum.

Implikasi bagi Jurnalisme: Rekaman audio wawancara dapat dipalsukan, menantang integritas jurnalisme berbasis rekaman.

2.5 SYNTHETIC HUMANS (MANUSIA SINTETIS SEPENUHNYA)

Apa itu Manusia Sintetis?

Manusia sintetis disebut juga 'full synthesis' atau 'generated personas' adalah jenis deepfake yang paling fundamental bukan manipulasi wajah atau suara orang nyata, melainkan

penciptaan manusia palsu dari nol yang tidak memiliki padanan di dunia nyata. Wajah, suara, bahkan seluruh profil digital sosialnya dapat dibuat sepenuhnya oleh AI.

Platform seperti *thispersondoesnotexist.com* (yang menggunakan StyleGAN dari NVIDIA) mampu menghasilkan foto wajah manusia yang tampak nyata secara instan dan gratis. Setiap kali halaman tersebut dimuat ulang, muncul wajah baru yang belum pernah ada di dunia lengkap dengan tekstur kulit, ekspresi, latar belakang, dan detail mata yang meyakinkan.

Penemuan dari Sophie Nightingale dan Hany Farid di atas adalah salah satu yang paling mengkhawatirkan dalam penelitian deepfake: wajah sintetis tidak hanya sulit dibedakan dari wajah asli dalam beberapa kondisi, wajah sintetis justru dinilai lebih dapat dipercaya. Ini memiliki implikasi mendalam untuk sistem kepercayaan sosial dan hukum.

Tabel 2.6 modus penyalahgunaan manusia sintetis

MODUS 1 Akun Media Sosial Palsu dan Operasi Pengaruh
Profil Lengkap: Manusia sintetis dapat dilengkapi dengan foto profil realistis (dari DALL-E atau StyleGAN), riwayat posting, jaringan pertemanan palsu, dan bahkan kepribadian yang konsisten menggunakan LLM seperti GPT-4.
Operasi Pengaruh Terkoordinasi: Pada 2019, Facebook membongkar jaringan akun palsu pro-Saudi Arabia yang menggunakan foto wajah sintetis untuk profilnya. Sebelumnya, akun menggunakan foto dicuri kini AI menghilangkan kebutuhan itu.
Skala yang Menakutkan: Dengan biaya rendah, satu operator dapat mengelola ribuan akun palsu dengan identitas unik yang masing-masing memiliki 'wajah' sendiri.
Astroturfing Kebijakan: Gerakan 'grassroots' palsu dapat diciptakan untuk mempengaruhi persepsi publik tentang suatu kebijakan atau produk.
MODUS 2 Agen Intelijen dan Spionase Digital
Kasus LinkedIn (2021): Keamanan siber menemukan profil LinkedIn atas nama 'Katie Jones', peneliti kebijakan luar negeri di Washington DC dengan jaringan koneksi ke berbagai pejabat dan think-tank. Fotonya adalah wajah sintetis. Identitasnya diduga merupakan agen intelijen asing.
Honey Trap Digital: Agen intelijen menggunakan persona sintetis yang menarik secara fisik untuk mendekati target dan membangun relasi palsu demi mendapatkan informasi sensitif.
Kesaksian Palsu: Dalam konteks hukum, identitas saksi palsu yang sepenuhnya sintetis berpotensi digunakan untuk mengelabui proses investigasi.
MODUS 3 Penipuan Romansa (Romance Scam) Berbasis AI

Modus: Pelaku membangun persona romantis lengkap wajah sintetis untuk foto profil, voice cloning untuk panggilan audio, dan teks yang dihasilkan LLM untuk percakapan untuk menjalin hubungan emosional dengan korban.
Skala Global: FTC AS melaporkan bahwa romance scam mengakibatkan kerugian $1,3 miliar pada 2022, dengan AI semakin digunakan untuk meningkatkan skala dan kualitas penipuan.
Dimensi Psikologis: Berbeda dari penipuan finansial biasa, romance scam meninggalkan luka emosional yang mendalam korban tidak hanya kehilangan uang tetapi juga mengalami pengkhianatan kepercayaan yang intens.
Implikasi Hukum: Penipuan berbasis manusia sintetis mempersulit identifikasi pelaku karena tidak ada 'wajah asli' yang bisa dilacak.

☑ POTENSI POSITIF MANUSIA SINTETIS

Di balik modus penyalahgunaannya, manusia sintetis memiliki aplikasi positif yang signifikan:

• Aktor Virtual dalam Produksi Media: Perusahaan seperti Digital Domain menciptakan aktor sintetis untuk film dan iklan, menghilangkan kebutuhan jadwal syuting yang kompleks.

• Asisten Virtual dengan Wajah: Layanan customer service dapat menggunakan avatar sintetis yang ramah dan responsif tanpa privasi manusia nyata.

• Model Pelatihan Medis: Pasien sintetis digunakan untuk melatih mahasiswa kedokteran dalam skenario yang langka atau sensitif.

• Inklusi dalam Riset: Data visual sintetis memungkinkan penelitian tentang persepsi dan bias tanpa menggunakan foto orang nyata melindungi privasi sambil memajukan ilmu pengetahuan.

2.6 TEXT-TO-VIDEO AI DEEPFAKE (GENERASI TERBARU 2025-2026)

Revolusi Text-to-Video: Dari Pikiran ke Gambar Bergerak

Jika generasi deepfake sebelumnya memerlukan video atau foto sumber sebagai bahan dasar, maka generasi terbaru teknologi AI telah melampaui batasan tersebut secara revolusioner. Text-to-video AI adalah teknologi yang dapat mengubah deskripsi teks menjadi video yang tampak nyata tanpa memerlukan sumber video apa pun. Cukup dengan mengetik: 'Presiden X sedang menerima suap dari seseorang di ruangan gelap' dan AI dapat menghasilkan video sintetis tersebut.

Teknologi ini merepresentasikan lompatan kualitatif dalam ancaman deepfake. Deepfake generasi sebelumnya masih terikat pada identitas dan rekaman yang ada seseorang harus memiliki cukup data wajah dan audio target. Text-to-video menghilangkan keterbatasan tersebut secara fundamental.

Tabel 2.7 Model Text-to-Video Terkemuka (2024-2026)

Model	Developer	Rilis	Kualitas Video	Catatan
Sora	OpenAI	2024 (terbatas)	Ultra-realistis, hingga 1 menit	Demo publik mengejutkan dunia; akses sangat terbatas; kebijakan penggunaan ketat
Kling AI	Kuaishou (China)	2024	Sangat tinggi; fisika realistis	Akses lebih terbuka; digunakan luas di Asia
Runway Gen-3	Runway ML	2024	Tinggi; kontrol presisi	Populer di kalangan kreator konten profesional
Lumiere	Google DeepMind	2024	Konsisten; natural motion	Riset; belum dirilis publik penuh
Dream Machine	Luma AI	2024	Tinggi; akses publik	Tersedia via web; populer di kalangan kreator
Wan2.1 / HunyuanVideo	Tencent/Alibaba	2025	Sangat tinggi; open source	Model open source paling canggih per 2025; dapat dijalankan lokal

Demonstrasi Kapabilitas: Mengapa Ini Berbeda?

Untuk memahami lompatan kualitatif yang diwakili text-to-video, perhatikan perbedaan berikut:

Tabel 2.8 Perbedaan Generasi Lama dengan Generasi terbaru (Text-to-Video)

Deepfake Generasi Lama vs. Text-to-Video: Perbedaan Kritis
GENERASI LAMA (Face Swap / Lip Sync):
→ Membutuhkan: Ratusan foto/video wajah target
→ Membutuhkan: Waktu pelatihan model berjam-jam
→ Output: Manipulasi pada rekaman yang sudah ada
→ Keterbatasan: Hanya dapat memanipulasi yang sudah ada

TEXT-TO-VIDEO (2024–2026):
→ Membutuhkan: Deskripsi teks saja
→ Membutuhkan: Beberapa detik hingga menit proses
→ Output: Video baru yang sepenuhnya diciptakan dari nol
→ Kemampuan: Dapat menciptakan skenario apa pun yang belum pernah terjadi
Analogi: Perbedaan antara memalsukan foto yang ada (Photoshop lama) versus menciptakan foto baru tentang peristiwa yang tidak pernah terjadi (Midjourney). Text-to-video melakukan hal yang sama tetapi untuk video bergerak.

Tabel 2.9 Modus Penyalahgunaan Text-to-Video

MODUS 1 Fabrikasi Peristiwa yang Tidak Pernah Terjadi
Deskripsi: Tidak seperti deepfake sebelumnya yang memanipulasi rekaman yang ada, text-to-video memungkinkan fabrikasi total menciptakan 'rekaman' tentang peristiwa yang tidak pernah terjadi.
Contoh Skenario: 'Polisi memukuli demonstran di depan gedung DPR' bisa dihasilkan tanpa peristiwa tersebut pernah terjadi, dengan detail lokasi, pakaian, dan konteks yang meyakinkan.
Krisis 2026 dan Seterusnya: Berbeda dari tahun sebelumnya, video sintetis 2025–2026 sudah cukup realistis untuk menipu sebagian besar penonton tanpa pemeriksaan forensik.
Implikasi Hukum: Sistem pembuktian konvensional yang mengandalkan 'rekaman CCTV' atau 'video saksi mata' menghadapi tantangan eksistensial.
MODUS 2 Disinformasi Berita dalam Skala Industri
Berita Palsu Bervisual: Text-to-video memungkinkan produksi 'liputan berita' sintetis lengkap penyiar palsu, grafis berita palsu, footage peristiwa palsu dalam hitungan menit.
Newsroom Palsu Bertenaga AI: Operasi disinformasi dapat menciptakan seluruh ekosistem 'media' palsu situs web, akun media sosial, dan konten video yang tampak seperti media berita legitim.

Amplifikasi Algoritmik: Konten video memperoleh jangkauan jauh lebih luas di media sosial dibanding teks. Video 'breaking news' palsu yang viral dapat mempengaruhi harga saham, situasi keamanan, atau hasil pemilu sebelum dapat dikonfirmasi atau dibantah.

STATUS REGULASI TEXT-TO-VIDEO DI INDONESIA (2025–2026)

Kerangka hukum Indonesia saat ini belum secara eksplisit mengatur text-to-video deepfake:

• UU ITE (UU No. 11/2008 jo. UU No. 19/2016) mencakup konten elektronik yang merugikan, namun tidak mendefinisikan konten sintetis berbasis AI.

• UU PDP (UU No. 27/2022) mengatur data pribadi, namun tidak secara spesifik mencakup sintesis wajah atau suara seseorang.

• KUHP Baru (UU No. 1/2023) mencakup pencemaran nama baik digital, namun belum spesifik pada konten generatif AI.

• Kesimpulan: Terdapat kekosongan regulasi (legal vacuum) yang signifikan yang perlu segera diisi — terutama mengingat kemudahan akses model text-to-video open source per 2025.

2.7 DEEPFAKE DOKUMEN: KTP, IJAZAH, DAN KONTRAK PALSU

Dimensi Deepfake yang Paling Diabaikan

Di antara semua jenis deepfake yang dibahas dalam buku ini, deepfake dokumen manipulasi atau fabrikasi dokumen resmi menggunakan AI adalah yang paling jarang mendapat perhatian akademis namun memiliki dampak hukum yang langsung dan konkret di Indonesia. Sementara diskusi tentang deepfake video mendominasi wacana publik, pemalsuan dokumen berbasis AI diam-diam telah menjadi instrumen penipuan yang meluas. Deepfake dokumen berbeda dari pemalsuan dokumen konvensional dalam dua hal kritis: (1) kualitas yang jauh lebih tinggi berkat AI sehingga lebih sulit terdeteksi secara visual; dan (2) skalabilitas satu operator dapat menghasilkan ratusan dokumen palsu yang masing-masing memiliki variasi unik untuk menghindari deteksi berbasis pola.

Tabel 2.10 Jenis – Jenis Deepfake Dokumen

KATEGORI 1 **Pemalsuan Dokumen Identitas — KTP, Paspor, SIM**
Teknik: Menggunakan image generation AI (Stable Diffusion, DALL-E, Midjourney) untuk mereplikasi template KTP/paspor resmi, kemudian menambahkan data dan foto orang nyata atau sintetis.

Kecanggihan Baru: Model AI dapat dilatih secara spesifik pada contoh dokumen asli untuk menghasilkan replika dengan hologram simulasi, tekstur microprinting, dan warna yang akurat.

Modus KYC Fraud: Lembaga keuangan digital yang mengandalkan verifikasi foto KTP menghadapi ancaman besar. Pelaku mengirimkan foto KTP AI-generated yang menampilkan wajah sintetis, berhasil membuka rekening palsu untuk pencucian uang atau penipuan.

Kasus Indonesia 2023: OJK melaporkan peningkatan signifikan upaya penipuan KYC pada platform pinjaman online menggunakan dokumen identitas yang dimanipulasi dengan AI.

Tantangan Verifikasi: Sistem OCR (Optical Character Recognition) konvensional yang digunakan banyak platform fintech tidak dirancang untuk mendeteksi dokumen yang dibuat AI versus dokumen asli yang difoto.

KATEGORI 2 Ijazah dan Transkrip Akademik Palsu

Teknik: Generative AI digunakan untuk menciptakan dokumen akademis yang tampak resmi lengkap dengan kop surat universitas, tanda tangan yang tampak asli, nomor seri, dan terkadang QR code palsu.

Skala: Platform gelap (dark web) kini menawarkan 'ijazah AI' dari universitas mana pun dengan harga mulai dari beberapa ratus ribu rupiah.

Dampak pada Dunia Kerja: Perekrut yang mengandalkan verifikasi visual dokumen tanpa sistem validasi resmi mudah tertipu. Ini menciptakan risiko kompetensi yang signifikan dalam berbagai profesi.

Relevansi Khusus Indonesia: Mengingat masih terbatasnya sistem verifikasi ijazah digital yang terintegrasi di Indonesia (meskipun program SICAMA Kemdikbud sedang dikembangkan), risiko ini sangat relevan.

Dimensi Hukum: Penggunaan ijazah palsu untuk melamar pekerjaan dapat dijerat dengan Pasal 263 KUHP (pemalsuan surat) dan/atau Pasal 28 ayat 1 UU ITE.

KATEGORI 3 Kontrak dan Dokumen Hukum Palsu

Teknik: AI generatif digunakan untuk menciptakan dokumen kontrak, akta notaris, surat kuasa, atau putusan pengadilan yang tampak resmi termasuk meniru tanda tangan pejabat tertentu.
Kloning Tanda Tangan: Model machine learning dapat dilatih untuk mereplikasi tanda tangan spesifik dari sampel yang tersedia online (dokumen publik, laporan tahunan, surat resmi yang dipublikasikan).
Fraud Properti: Sertifikat tanah, akta jual-beli, dan surat kuasa palsu berbasis AI digunakan dalam penipuan properti.
Penipuan Korporat: Kontrak palsu yang meyakinkan digunakan untuk mengelabui mitra bisnis agar menyerahkan uang muka atau berbagi informasi sensitif.
Dampak pada Sistem Peradilan: Ketika dokumen yang diajukan sebagai alat bukti dapat difabrikasi dengan AI, integritas seluruh sistem pembuktian dokumenter terancam.

Ekosistem *'Document-as-a-Service'* di Dark Web

Salah satu perkembangan paling mengkhawatirkan adalah munculnya ekosistem layanan pembuatan dokumen palsu berbasis AI di pasar gelap digital. Layanan-layanan ini beroperasi dengan model bisnis yang terorganisasi:

- Menawarkan 'paket' dokumen KTP, NPWP, rekening koran, slip gaji yang konsisten satu sama lain untuk membuat profil palsu yang komprehensif.
- Menyediakan antarmuka yang mudah digunakan (no-code) sehingga bahkan pelaku tanpa keahlian teknis dapat memesan dokumen palsu.
- Menerima pembayaran kripto untuk anonimitas transaksi.
- Menawarkan garansi 'dapat melewati verifikasi otomatis' platform fintech tertentu.

⚠ IMPLIKASI BAGI SISTEM HUKUM INDONESIA

Deepfake dokumen menimbulkan tantangan berlapis bagi sistem hukum Indonesia:

1. HUKUM PIDANA: Pasal 263-264 KUHP tentang pemalsuan surat perlu diinterpretasikan untuk mencakup dokumen yang difabrikasi AI bukan sekadar dokumen asli yang diubah.

2. HUKUM PERDATA: Perjanjian yang dibuat berdasarkan dokumen palsu AI — apakah dibatalkan demi hukum (null and void) atau dapat dibatalkan? Pihak mana yang menanggung kerugian?

3. HUKUM PEMBUKTIAN: Bagaimana hakim menilai keotentikan dokumen yang diajukan sebagai alat bukti ketika AI dapat menciptakan dokumen yang hampir identik dengan aslinya?

4. REGULASI FINTECH: OJK perlu memperketat standar KYC untuk mengharuskan penggunaan teknologi liveness detection yang tahan terhadap deepfake bukan sekadar unggah foto KTP.

2.8 MATRIKS ANCAMAN DEEPFAKE: SINTESIS TAKSONOMI

Setelah menelaah keenam jenis deepfake secara mendalam, tabel berikut menyajikan matriks komprehensif yang menghubungkan jenis deepfake dengan domain ancaman, teknik yang digunakan, dan kerangka hukum yang berpotensi relevan:

Tabel 2.11 Matriks Komprehensif Hubungan Antara Deepfake Dengan Domain Ancaman

Jenis Deepfake	Domain Ancaman Utama	Hak yang Dilanggar	Potensi Pasal Hukum
Face Swapping	Pornografi non-konsensual; penipuan KYC; disinformasi politik	Privasi; martabat; nama baik; harta kekayaan	Pasal. 27 ayat 1 UU ITE; Ps. 281 KUHP; Pasal. 378 KUHP
Lip-Syncing	Disinformasi politik; fraud korporat; manipulasi opini publik	Nama baik; integritas demokrasi; harta kekayaan	Pasal. 28 ayat 3 UU ITE; Pasal. 378 KUHP; Pasal. 14 UU 1/1946
Voice Cloning	Penipuan telepon; CEO fraud; pemerasan; pencemaran nama baik	Harta kekayaan; nama baik; privasi	Pasal. 378 KUHP; Pasal. 27 ayat 3 UU ITE; Pasal. 368 KUHP
Synthetic Humans	Akun palsu; spionase; romance scam; astroturfing	Integritas informasi; perlindungan data	Pasal. 35 UU ITE; UU PDP No. 27/2022
Text-to-Video	Fabrikasi peristiwa; disinformasi masif; ujaran kebencian	Integritas publik; nama baik; ketertiban	Pasal. 28 UU ITE; KUHP Baru Pasal. 263-264 (analogi)
Deepfake Dokumen	Penipuan KYC; ijazah palsu; kontrak palsu; fraud properti	Ketertiban hukum; harta kekayaan; integritas dokumen	Pasal. 263-264 KUHP; Pasal. 35 UU

			ITE; UU Dokumen Negara

Catatan: Tabel di atas bersifat panduan awal dan tidak exhaustive. Analisis pasal yang lebih mendalam termasuk yurisprudensi dan doktrin yang relevan akan dibahas dalam Bab 4 (Kerangka Hukum Pidana) dan Bab 5 (Kerangka Hukum Perdata).

Rangkuman Bab

Bab ini telah memetakan taksonomi deepfake secara komprehensif. Beberapa poin kunci untuk diingat:

- ☑ Face swapping adalah bentuk deepfake paling dikenal, dengan penyalahgunaan terbesar pada pornografi non-konsensual dan penipuan identitas berbasis video termasuk ancaman yang semakin nyata terhadap sistem KYC perbankan digital.
- ☑ Lip-syncing deepfake adalah ancaman paling langsung bagi integritas demokrasi kemampuannya membuat tokoh tampak 'berkata' apa pun yang diinginkan penyebar disinformasi menjadikannya senjata informasi yang berbahaya.
- ☑ Voice cloning kini dapat dilakukan hanya dari tiga detik sampel audio, menjadikan suara siapa pun yang pernah berbicara di depan mikrofon berpotensi dikloning dengan implikasi besar bagi penipuan telepon dan CEO fraud.
- ☑ Manusia sintetis menghilangkan kebutuhan akan identitas nyata sebagai bahan dasar manipulasi siapa pun kini dapat menciptakan persona lengkap dengan wajah, suara, dan profil digital yang sepenuhnya palsu.
- ☑ Text-to-video generasi 2024–2026 merepresentasikan lompatan revolusioner: dari memanipulasi yang ada ke menciptakan yang tidak pernah ada. Ini menantang seluruh sistem pembuktian berbasis rekaman visual.
- ☑ Deepfake dokumen adalah ancaman yang paling konkret secara hukum namun paling kurang mendapat perhatian dengan implikasi langsung bagi fintech, dunia kerja, dan sistem peradilan Indonesia.

BAB 3
PENGARUH SOSIAL DAN PSIKOLOGIS DEEPFAKE

"Deepfake bukan sekadar masalah teknologi. Ini adalah masalah manusia tentang siapa yang memiliki kekuatan untuk mendefinisikan realitas, dan siapa yang menjadi korban ketika kekuatan itu disalahgunakan."

— Kate Isaacs, pendiri Not Your Porn (kampanye anti-NCII), 2021

Ada sebuah ironi yang mencolok dalam cara dunia membicarakan deepfake. Sebagian besar perhatian akademis, legislatif, maupun jurnalistik tertuju pada aspek teknis: seberapa canggih algoritma GAN, seberapa realistis output-nya, bagaimana mendeteksinya. Sementara itu, orang-orang yang hidupnya hancur akibat deepfake harus berjuang sendiri, sering kali tanpa perlindungan hukum, tanpa dukungan sosial yang memadai, dan di bawah bayang-bayang stigma yang tidak seharusnya mereka tanggung.

Bab ini hadir untuk memperbaiki ketidakseimbangan itu. Kita akan meninggalkan domain teknis untuk sementara, dan menyelam ke dalam dimensi yang justru paling nyata dan paling mendesak: apa yang deepfake lakukan kepada manusia baik secara individual maupun kolektif. Kita akan menelaah bagaimana deepfake menghancurkan psikologi individu, bagaimana hal tersebut mengikis kepercayaan kolektif terhadap media, bagaimana deepfake menjadi alat penindasan berbasis gender dan kekuasaan, dan bagaimana deepfake mengancam fondasi demokrasi itu sendiri. Hal ini bukan pembahasan yang mudah tetapi adalah pembahasan yang tidak bisa dihindari oleh siapa pun yang serius memahami deepfake sebagai fenomena hukum dan sosial.

3.1 DAMPAK PSIKOLOGIS KORBAN: TRAUMA, STIGMA, DAN KEHILANGAN KEPERCAYAAN DIRI

Siapakah 'Korban' Deepfake?

Sebelum membahas dampaknya, penting untuk memetakan siapa yang dapat menjadi korban deepfake. Berbeda dari kejahatan konvensional yang biasanya memerlukan interaksi fisik antara pelaku dan korban, deepfake dapat menarget siapa pun yang pernah memiliki kehadiran digital foto di media sosial, video pernikahan yang diunggah, wawancara yang disiarkan di YouTube, bahkan sekadar foto di direktori perusahaan.

Tabel 3.1 Kategori Korban Kejahatan Deepfake

Kategori Korban	Kerentanan Spesifik	Jenis Deepfake yang Umum Dialami
Perempuan (umum)	Kehadiran media sosial publik; norma sosial yang menstigmatisasi	NCII deepfake pornografi; manipulasi reputasi

Tokoh publik / selebriti	Banyak rekaman wajah & suara publik tersedia	Semua jenis; terutama NCII, disinformasi, penipuan atas nama mereka
Politisi & pejabat	Kepentingan lawan politik; proses pemilu	Lip-sync disinformasi; manipulasi pernyataan
Jurnalis & aktivis	Pelaporan kritis; keterbukaan publik	Diskreditasi; konten seksual sebagai intimidasi
Perempuan muda & remaja	Aktif media sosial; minimnya kesadaran hukum	NCII; bullying berbasis deepfake di sekolah
Individu dalam konflik pribadi	Mantan pasangan; sengketa keluarga/bisnis	Deepfake sebagai alat pemerasan dan balas dendam
Profesional & akademisi	Reputasi karir sebagai target; dokumen digital	Deepfake pernyataan; pemalsuan dokumen

Trauma Psikologis: Lebih dari Sekadar Rasa Malu

Ketika para peneliti dan praktisi kesehatan mental mulai mewawancarai korban deepfake khususnya korban pornografi non-konsensual berbasis deepfake apa yang mereka temukan jauh melampaui 'rasa malu' atau 'ketidaknyamanan' yang kerap diasumsikan oleh publik umum. Dampaknya bersifat mendalam, multidimensi, dan dalam banyak kasus, berlangsung jangka panjang.

A. Gangguan Stres Pasca-Trauma (PTSD)

Penelitian yang dilakukan oleh Nicola Henry dan rekan-rekannya di Universitas RMIT Australia menemukan bahwa korban pornografi non-konsensual termasuk yang berbasis deepfake menunjukkan gejala yang konsisten dengan *Post-Traumatic Stress Disorder* (PTSD). Ini mencakup:

- Intrusi: Pikiran yang tidak terkontrol tentang konten deepfake yang tersebar; mimpi buruk; flashback ketika melihat teknologi serupa.
- Penghindaran: Meninggalkan media sosial secara total; menghindari pertemuan sosial; berhenti berpartisipasi dalam aktivitas yang sebelumnya dinikmati.
- Hipervigilans: Kecemasan berlebihan tentang potensi penyebaran lebih lanjut; memeriksa platform secara obsesif untuk menemukan salinan baru konten tersebut.
- Perubahan kognitif negatif: Perasaan malu yang mendalam; self-blame ('*mengapa saya punya media sosial?*'); perubahan fundamental dalam cara memandang diri sendiri.

B. Kehilangan Rasa Kepemilikan atas Tubuh dan Identitas Digital

Salah satu dampak psikologis paling unik dari deepfake yang membedakannya dari bentuk kejahatan seksual lain adalah pengalaman 'diambil alihnya tubuh dan identitas.' Korban deepfake NCII sering melaporkan perasaan bahwa tubuh mereka telah 'dijajah' secara digital tanpa izin; bahwa ada versi diri mereka yang hidup di ruang digital dan melakukan hal-hal yang tidak pernah mereka lakukan.

Psikolog klinis Dr. Charlotte Webb dari King's College London menyebutnya sebagai 'digital body violation' sebuah bentuk pelanggaran batas pribadi yang secara hukum belum sepenuhnya diakui, namun secara psikologis nyata dan melemahkan. Korban mengalami disonansi kognitif yang berat: mereka mengetahui konten itu tidak nyata, namun dunia (atau sebagian darinya) memperlakukannya seolah nyata.

STUDI KASUS: Sarah (nama disamarkan) Korban Deepfake NCII, Inggris 2021

Seorang perempuan berusia 28 tahun, guru sekolah dasar, menemukan bahwa mantan kekasihnya telah membuat video pornografi deepfake menggunakan fotonya dari Instagram.

Ia menggambarkan pengalamannya: 'Saya merasa seperti seluruh hidup saya hancur dalam semalam. Saya tidak bisa tidur, tidak bisa makan. Saya menghapus semua media sosial saya. Saya bahkan tidak bisa bercermin karena saya tahu ada video itu di luar sana menggunakan wajah saya.'

Setelah video tersebut ditemukan oleh orang tua murid, ia terpaksa mengundurkan diri dari pekerjaannya. Ia menjalani terapi selama 18 bulan.

Kasus ini mencerminkan pola umum: korban deepfake NCII sering menghadapi konsekuensi profesional, bukan hanya personal karena konten tersebut dapat menjangkau lingkungan kerja mereka.

C. Kehilangan Kepercayaan Diri dan Harga Diri

Penelitian dari *Cyber Civil Rights Initiative* (CCRI) Amerika Serikat organisasi yang menjadi salah satu sumber data terpenting dalam studi ini mensurvei 1.601 korban non-consensual intimate imagery. Temuan mereka, meskipun tidak semua mencakup deepfake secara spesifik, memberikan gambaran relevan ctentang dampak psikologis jangka panjang:

Tabel 3.2 Presentase Dampak Psikologis

Dampak Psikologis yang Dilaporkan	Persentase Korban
Mengalami kesulitan tidur dan kecemasan	93%
Mengalami depresi signifikan	82%
Berpikir untuk bunuh diri	51%
Kehilangan kepercayaan diri secara permanen	89%
Mengalami kesulitan dalam hubungan intim selanjutnya	73%
Berhenti dari pekerjaan atau studi	42%
Mengalami isolasi sosial	67%

D. Dampak Kumulatif dan Permanen: Konten yang Tidak Pernah Pergi

Salah satu dimensi paling menghancurkan dari deepfake dibandingkan dengan banyak bentuk kejahatan konvensional adalah sifatnya yang permanen dan berulang. Ketika sebuah deepfake disebarkan ke internet, ia dapat:

- Didownload dan disimpan oleh ribuan orang di seluruh dunia.
- Diunggah ulang berkali-kali di platform berbeda setelah dihapus dari satu platform.
- Muncul kembali bertahun-tahun kemudian dalam konteks yang sama sekali berbeda.
- Tersimpan di forum gelap dan jaringan peer-to-peer yang tidak dapat dijangkau oleh mekanisme takedown biasa.

Hal ini menciptakan apa yang para peneliti sebut sebagai secondary victimization yang berkelanjutan korban tidak hanya mengalami trauma satu kali, tetapi terus-menerus hidup dengan ancaman bahwa konten tersebut dapat muncul kembali kapan saja, dalam konteks apa pun: wawancara kerja, pertemuan pertama dengan calon pasangan, atau bahkan diperlihatkan kepada anak-anak mereka di masa depan.

Stigma Sosial: Ketika Korban Disalahkan

Dimensi psikologis deepfake tidak dapat dipisahkan dari konteks sosial yang melingkupinya khususnya budaya victim-blaming yang masih dominan. Dalam banyak kasus, ketika korban deepfake NCII mencari bantuan baik dari keluarga, teman, penegak hukum, maupun institusi respons yang mereka terima justru memperparah traumanya.

POLA VICTIM-BLAMING YANG UMUM DITEMUI KORBAN DEEPFAKE

Respons yang sering diterima korban dari lingkungan sosialnya:

• Dari keluarga: "*Makanya jangan punya foto di medsos*." atau "*Ini akibat pergaulan kamu sendiri*."

• Dari teman: "*Ya sudah, anggap saja tidak ada*." (keheningan dan penghindaran)

• Dari penegak hukum: "*Ini tidak masuk kategori kejahatan yang bisa kami proses*." / "*Apakah Anda memiliki bukti itu dilakukan secara sengaja?*"

• Dari institusi kerja: Pemecatan atau mutasi, bukan pelindungan.

• Dari platform media sosial: Proses takedown lambat; konten muncul ulang setelah dihapus.

Pola ini memiliki nama dalam literatur: *secondary victimization;* ketika respons institusional dan sosial terhadap korban justru menambah luka, bukan menyembuhkan.

Dalam konteks Indonesia, stigma berlapis ini sangat nyata. Nilai-nilai budaya yang menempatkan 'kehormatan perempuan' sebagai tanggung jawab individual bukan sebagai hak

yang dilindungi secara kolektif membuat korban deepfake NCII di Indonesia menghadapi beban yang bahkan lebih berat dibanding di negara dengan tradisi hukum perlindungan privasi yang lebih mapan.

Dampak Ekonomi dan Profesional

Di luar dampak psikologis langsung, korban deepfake menghadapi konsekuensi ekonomi yang serius dan sering kali tidak diantisipasi:

- Kehilangan pekerjaan: Konten deepfake yang menjangkau lingkungan kerja baik ditemukan oleh rekan, atasan, atau klien sering berujung pada pemecatan atau pengunduran diri paksa.
- Kerusakan reputasi profesional yang permanen: Dalam era pencarian Google, nama seseorang yang dikaitkan dengan konten deepfake dapat muncul dalam hasil pencarian selama bertahun-tahun, merusak prospek karir.
- Biaya hukum dan pemulihan: Proses hukum (jika tersedia), layanan penghapusan konten, dan terapi psikologis membutuhkan biaya yang tidak sedikit beban yang sepenuhnya ditanggung korban.
- Oportunitas yang hilang: Korban yang menarik diri dari kehidupan publik (meninggalkan media sosial, menolak promosi yang memerlukan profil publik) kehilangan peluang yang tidak mudah diukur secara finansial.

3.2 EROSI KEPERCAYAAN PUBLIK TERHADAP MEDIA DIGITAL

Kepercayaan sebagai Infrastruktur Sosial

Dalam sosiologi, kepercayaan adalah infrastruktur yang menopang hampir semua sistem sosial modern. Pasar keuangan berfungsi karena kepercayaan pada sistem verifikasi. Demokrasi berfungsi karena kepercayaan pada integritas informasi. Sistem hukum berfungsi karena kepercayaan pada kebenaran bukti. Ketika kepercayaan erosi, seluruh sistem di atasnya menjadi rentan.

Dalam konteks media digital, kepercayaan publik selama beberapa dekade terakhir telah mengalami tekanan berlapis dari berita palsu (hoaks teks), manipulasi foto, hingga kini deepfake. Setiap lapisan baru dari manipulasi ini tidak hanya menambah ketidakpercayaan secara linier; ia menghasilkan efek kumulatif yang menggerus kepercayaan terhadap seluruh kategori informasi visual.

Mekanisme Erosi Kepercayaan

Deepfake mengikis kepercayaan publik melalui beberapa mekanisme yang saling memperkuat:

A. Uncertainty Inflation; Penggelembungan Ketidakpastian

Sebelum era deepfake, ketika seseorang melihat video seseorang melakukan atau mengatakan sesuatu, default-nya adalah mempercayai bahwa hal itu benar-benar terjadi. Kini, keberadaan teknologi deepfake bahkan tanpa video palsu yang spesifik menginjeksikan ketidakpastian ke dalam setiap konsumsi konten visual. Pertanyaan 'apakah ini asli?' yang sebelumnya hanya muncul dalam situasi mencurigakan, kini menjadi respons refleksif terhadap semua konten.

Efeknya: biaya kognitif konsumsi media meningkat drastis. Publik harus berinvestasi lebih banyak waktu dan energi mental untuk memverifikasi konten sebuah beban yang secara tidak proporsional membebani individu dengan literasi media dan akses sumber daya verifikasi yang lebih rendah.

B. Asymmetric Virality; Penyebaran Asimetris

Penelitian dari MIT Media Lab yang diterbitkan dalam jurnal Science menemukan bahwa berita palsu menyebar enam kali lebih cepat di Twitter dibanding berita yang benar. Deepfake memanfaatkan fenomena ini secara lebih ekstrem:

(1) Video deepfake yang emosional dan mengejutkan memperoleh jangkauan viral jauh sebelum pemeriksaan fakta dapat dilakukan.
(2) Koreksi dan klarifikasi hampir tidak pernah mencapai audiens yang sama besar dengan konten palsu aslinya.
(3) Algoritma media sosial yang memprioritaskan keterlibatan (*engagement*) secara struktural menguntungkan konten yang membangkitkan reaksi emosional persis karakteristik deepfake manipulatif.

STUDI KASUS: Studi: Deepfake dan Penyebaran Asimetris di Indonesia

Dalam analisis yang dilakukan oleh *Centre for Innovation Policy and Governance* (CIPG) Jakarta (2022), peneliti menemukan bahwa:

• Video manipulasi (termasuk cheapfake dan deepfake) cenderung dibagikan 4–7 kali lebih banyak di WhatsApp Indonesia dibanding klarifikasi dari fact-checker.

• Faktor kunci: Pesan yang beredar melalui grup keluarga WhatsApp memperoleh tingkat kepercayaan yang sangat tinggi karena dikirim oleh orang yang dikenal dan dipercaya terlepas dari konten aslinya.

• Implikasi: Model penyebaran disinformasi di Indonesia berbeda dari negara Barat lebih banyak terjadi di platform pesan privat (WhatsApp) dibanding platform publik (Twitter/X), membuat pemantauan lebih sulit.

• Rekomendasi dari studi ini: Literasi digital berbasis komunitas dan kelompok kepercayaan lebih efektif daripada fact-checking platform-level dalam konteks Indonesia.

C. Truth Decay; Pelapukan Konsensus Faktual

Rand Corporation memperkenalkan konsep *"Truth Decay"* erosi peran fakta dan analisis empiris dalam kehidupan publik Amerika. Deepfake memperburuk Truth Decay dengan mekanisme spesifik:

➲ Partisan Content Selection: Audiens yang secara kognitif terpolarisasi cenderung menerima deepfake yang cmengonfirmasi pandangan mereka dan menolak yang menantangnya tanpa pemeriksaan faktual.

- Institutional Distrust Feedback Loop: Semakin publik tidak mempercayai institusi media arus utama, semakin mereka beralih ke sumber alternatif yang kurang memiliki standar verifikasi yang justru lebih rentan menyebarkan deepfake.
- Epistemic Cowardice: Ketidakpastian tentang apa yang nyata dan apa yang palsu dapat mendorong individu dan institusi untuk menghindari mengambil posisi yang tegas — bahkan ketika fakta jelas.

Tabel 3.3 Dampak pada Kepercayaan terhadap Institusi Spesifik

Institusi atau Domain	Mekanisme Erosi	Dampak Konkret
Media Berita	Deepfake footage yang diklaim sebagai liputan asli mengkontaminasi kepercayaan pada semua liputan video	Penurunan konsumsi berita; meningkatnya 'news avoidance'
Pengadilan dan Sistem Hukum	Alat bukti video dipertanyakan; deepfake dapat memunculkan keraguan wajar terhadap rekaman asli	Krisis admissibility alat bukti digital
Politisi dan Pemerintah	Pernyataan resmi dapat dimanipulasi atau diklaim sebagai deepfake	Menurunnya kepercayaan pada komunikasi pemerintah
Platform Media Sosial	Ketidakmampuan atau kelambatan menghapus konten deepfake	Erosi kepercayaan pada moderasi konten platform
Sistem Verifikasi Identitas	KYC digital rentan terhadap deepfake real-time	Keraguan terhadap keamanan layanan finansial digital
Sains dan Akademia	Deepfake video penelitian atau pernyataan akademisi	Komplikasi integritas komunikasi ilmiah

Paradoks Kesadaran: Mengetahui Membuat Lebih Rentan?

Ada sebuah paradoks yang menarik dalam penelitian tentang deepfake dan kepercayaan mengetahui bahwa deepfake ada tidak selalu membuat seseorang lebih baik dalam mendeteksinya dan dalam beberapa kasus, justru sebaliknya. Penelitian oleh Jevin West dan Carl Bergstrom menemukan bahwa individu dengan kesadaran tentang deepfake dapat mengalami apa yang disebut *overcorrection* mereka menjadi terlalu skeptis terhadap konten yang asli, sementara masih gagal mendeteksi deepfake yang canggih. Ini berarti kampanye kesadaran deepfake yang kurang tepat sasaran dapat secara tidak sengaja memperburuk erosi kepercayaan tanpa meningkatkan kemampuan deteksi yang sebenarnya.

3.3 DEEPFAKE DAN KRISIS "LIAR'S DIVIDEND"

Memahami Liar's Dividend Secara Mendalam

Konsep Liar's Dividend (Dividen Pembohong) diperkenalkan oleh profesor hukum Bobby Chesney (University of Texas) dan Danielle Citron (University of Virginia) dalam artikel seminar mereka di California Law Review (2019). Ini adalah salah satu kontribusi intelektual paling penting dalam studi deepfake dan hukum.

Definisi operasionalnya sederhana namun profund, karena publik kini mengetahui bahwa video realistis bisa dipalsukan, siapa pun yang tertangkap kamera melakukan sesuatu yang merugikan baik politisi korup, pejabat yang menyalahgunakan kekuasaan, atau orang biasa yang melakukan kejahatan kini memiliki senjata baru untuk mengingkari bukti: klaim 'itu adalah deepfake.'

Yang membuat konsep ini begitu mengkhawatirkan adalah bahwa ia bekerja bahkan ketika tidak ada deepfake yang digunakan. Cukup dengan keberadaan teknologi deepfake di dunia beserta liputan medianya yang masif sebuah klaim *'ini deepfake!'* telah mendapat resonansi publik yang cukup untuk menciptakan keraguan.

🎭 ANATOMI LIAR'S DIVIDEND; Bagaimana Ia Bekerja

Skenario klasik dalam 4 tahap:

TAHAP 1
BUKTI MUNCUL: Video beredar menampilkan seorang pejabat menerima suap, seorang politisi mengucapkan pernyataan rasialis, atau seorang eksekutif melakukan pelecehan.

TAHAP 2
KLAIM DEEPFAKE: Tersangka atau tim komunikasinya segera mengklaim: 'Itu adalah deepfake! Saya tidak pernah melakukan hal itu. Ini serangan politisi/pesaing/musuh saya.'

TAHAP 3
KERAGUAN PUBLIK: Sebagian publik, yang mengetahui deepfake itu nyata, menerima penjelasan tersebut dengan mudah. Media melaporkan 'kontroversi' bukan fakta. Investigasi tertunda.

TAHAP 4
IMPUNITAS DE FACTO: Meskipun video tersebut asli, tersangka lolos dari konsekuensi karena keraguan yang diciptakan cukup untuk melindungi mereka dari akuntabilitas publik maupun hukum.

Catatan kritis: Bahkan jika akhirnya terbukti video itu asli, momentum politik atau hukum yang diperlukan untuk akuntabilitas mungkin sudah hilang.

Kasus-Kasus Empiris Liar's Dividend

STUDI KASUS: Gabon, 2019. Deepfake Presiden atau Klaim Liar's Dividend?
Konteks: Presiden Ali Bongo Ondimba Gabon tidak muncul di publik selama berbulan-bulan karena stroke. Untuk meredakan spekulasi tentang kondisi kesehatannya, pemerintah merilis video Presiden Bongo yang memberikan pidato Tahun Baru.
Kontroversi: Kelompok oposisi mengklaim video tersebut adalah deepfake menuduh bahwa presiden sudah tidak mampu memimpin. Video tersebut memang memiliki sejumlah anomali visual.
Hasil: Analisis oleh beberapa pakar forensik digital tidak dapat menentukan secara pasti apakah video tersebut adalah deepfake atau bukan. Klaim tersebut cukup untuk memicu krisis politik dan upaya kudeta militer.
Pelajaran: Ini adalah kasus di mana Liar's Dividend digunakan oleh oposisi bukan oleh tersangka untuk mendelegitimasi bukti visual kehadiran pemimpin. Deepfake telah mengkontaminasi seluruh ekosistem kepercayaan terhadap bukti visual.

STUDI KASUS: Indonesia; Konteks Lokal Liar's Dividend
Meskipun belum ada kasus Liar's Dividend berprofil tinggi yang terdokumentasi secara akademis di Indonesia, pola berikut telah diamati dalam diskursus publik:
• Beberapa klip video politisi yang viral kemudian diklaim sebagai 'hasil edit' atau 'sudah diedit' oleh pendukung mereka meskipun investigasi menunjukkan klip tersebut autentik.
• Praktik ini menunjukkan bahwa budaya Liar's Dividend sudah eksis bahkan sebelum deepfake menjadi lazim deepfake hanya akan memperkuat klaimnya.
• Konteks Pemilu 2024: Dengan penetrasi media sosial yang sangat tinggi, kecanggihan teknologi deepfake yang meningkat, dan budaya politik yang polarized, Indonesia sangat rentan terhadap Liar's Dividend berbasis deepfake dalam siklus pemilu berikutnya.
• Urgensi Legislatif: Tanpa mekanisme hukum yang menetapkan otoritas teknis independen untuk memverifikasi keaslian konten yang diklaim sebagai bukti, Liar's Dividend akan menjadi semakin sulit dilawan.

Liar's Dividend dan Krisis Sistem Pembuktian

Dari perspektif hukum acara, Liar's Dividend menimbulkan tantangan yang sangat serius terhadap fondasi sistem pembuktian modern. Sistem hukum kita baik pidana maupun

perdata secara historis memberikan bobot kepercayaan yang tinggi kepada bukti audiovisual sebagai representasi realitas yang dapat diandalkan.

Tabel 3.4 Dua skenario berbahaya menjadi mungkin secara bersamaan

Skenario	Deskripsi	Implikasi Hukum
Skenario A: Bukti Asli Ditolak	Rekaman CCTV atau video saksi yang asli berhasil diragukan keasliannya oleh pihak yang mengklaim 'ini deepfake'	Beban pembuktian meningkat drastis; diperlukan expert forensik digital untuk setiap kasus
Skenario B: Bukti Palsu Diterima	Deepfake yang canggih lolos dari deteksi dan diterima sebagai alat bukti oleh pengadilan	Putusan berdasarkan bukti palsu; penghukuman yang salah (wrongful conviction)
Skenario C: Impunitas Preventif	Tersangka mengantisipasi potensi rekaman dengan menciptakan alibi bahwa 'saya sering menjadi target deepfake'	Sistem pertahanan yang mengikis akuntabilitas bahkan sebelum tindakan dilakukan

Respons terhadap Liar's Dividend: Solusi yang Mungkin

Menghadapi tantangan Liar's Dividend, berbagai solusi telah diusulkan masing-masing dengan kelebihan dan keterbatasannya:

- ☑ Deteksi Forensik Digital Terstandarisasi: Pengembangan standar internasional untuk analisis forensik konten digital termasuk standar yang diakui dalam prosedur hukum acara. Ini memerlukan investasi dalam kapasitas lembaga forensik nasional.
- ☑ Provenance Technology dan Watermarking: Teknologi seperti Content Credentials (C2PA standard) memungkinkan konten digital untuk membawa 'riwayat' pembuatannya yang terenkripsi membuktikan kapan, di mana, dan dengan perangkat apa konten tersebut dibuat.
- ☑ Lembaga Verifikasi Independen: Pembentukan lembaga teknis independen mungkin di bawah Komisi Pemilihan Umum, Komnas HAM, atau badan khusus yang memiliki kapasitas teknis dan kewenangan hukum untuk memverifikasi keaslian konten dalam konteks litigasi dan proses publik.
- ☑ Reformasi Hukum Pembuktian: Penyesuaian ketentuan hukum acara baik KUHAP maupun HIR untuk secara eksplisit mengatur prosedur autentikasi alat bukti digital, termasuk penetapan standar kualifikasi ahli forensik digital.

3.4 DAMPAK TERHADAP PEREMPUAN DAN KELOMPOK RENTAN

Deepfake sebagai Senjata Berbasis Gender

Analisis data tentang korban deepfake secara konsisten menunjukkan kesenjangan gender yang dramatis. Ini bukan sekadar statistik ini adalah cerminan dari struktur kekuasaan

yang lebih dalam tentang siapa yang dianggap berhak atas privasi, siapa yang dianggap sebagai objek seksual, dan siapa yang menanggung biaya sosial dari pelanggaran batas.

Tabel 3.5 Analisis Data Korban Deepfake

Indikator	Data / Temuan
Korban deepfake NCII yang perempuan	95–99% (Sensity AI, 2023; Deeptrace, 2019)
Pelaku deepfake NCII yang laki-laki	>90% (berdasarkan laporan platform)
Motivasi terbesar pelaku	Balas dendam (mantan pasangan), kontrol, intimidasi, eksploitasi finansial
Usia korban terbanyak	18–35 tahun (usia aktif media sosial)
Platform penyebaran terbesar	Forum khusus pornografi; Telegram; Reddit (sebelum kebijakan larangan)
Waktu rata-rata konten tetap online setelah takedown request	Beberapa jam hingga beberapa hari; konten sering muncul ulang

Angka 95–99% korban perempuan bukan kebetulan statistik. Ini mencerminkan fakta bahwa tubuh perempuan secara historis telah dikonstruksikan sebagai objek dalam budaya visual sebuah warisan patriarki yang direproduksi dan diperburuk oleh teknologi deepfake.

Mekanisme Penindasan Berbasis Deepfake

A. Revenge Porn yang Diperburuk

Sebelum deepfake, revenge porn (NCII konvensional) sudah menjadi masalah serius. Deepfake memperburuk fenomena ini dalam beberapa cara kritis:

- ✓ Tidak memerlukan konten asli: Pelaku tidak perlu memiliki foto atau video intim korban yang nyata. Cukup dengan foto wajah dari media sosial, pelaku dapat menciptakan konten intim deepfake yang tampak meyakinkan.
- ✓ Menghilangkan argumen *"kerelaan"*: Dalam NCII konvensional, pelaku kadang berargumen bahwa konten dibuat dengan 'kerelaan' (meskipun penyebarannya tidak). Deepfake mengeliminasi argumen ini tidak ada konten asli, dan tidak ada persetujuan dalam bentuk apa pun.
- ✓ Meningkatkan kualitas dan daya rusak: Deepfake berkualitas tinggi lebih sulit dikenali sebagai palsu meningkatkan potensi kerusakan reputasi.

B. Deepfake sebagai Alat Pemerasan (Sextortion)

Sextortion berbasis deepfake adalah modus yang semakin umum: pelaku mengancam akan menyebarkan deepfake intim korban kecuali korban memberikan uang, konten seksual nyata, atau kepatuhan terhadap permintaan lain. Yang membuat ini sangat berbahaya:

- ☑ Pelaku tidak perlu memiliki konten asli, ancaman saja sudah cukup jika korban percaya deepfake tersebut ada atau dapat dibuat.
- ☑ Biaya teknisnya nyaris nol, ada platform yang memungkinkan pembuatan deepfake intim dalam hitungan menit.
- ☑ Anak-anak dan remaja adalah target yang semakin umum, dengan kurangnya pemahaman tentang deepfake dan ketakutan terhadap reaksi orang tua atau sekolah.

📋 **STUDI KASUS: FBI Warning 2023 Sextortion Berbasis Deepfake Menarget Anak-Anak**
Pada Juni 2023, FBI mengeluarkan peringatan publik bahwa mereka mengamati lonjakan dramatis laporan sextortion berbasis deepfake yang menarget anak-anak dan remaja.
Modus: Pelaku mengunduh foto media sosial anak/remaja (Instagram, TikTok, Snapchat), menciptakan deepfake intim menggunakan aplikasi yang tersedia secara komersial, lalu mengancam menyebarkan konten tersebut kecuali korban mengirimkan uang atau konten seksual nyata.
Data: FBI melaporkan lebih dari 3.000 kasus dilaporkan antara Oktober 2022 dan Maret 2023 dan ini hanya yang dilaporkan, karena banyak korban (terutama anak laki-laki) tidak melapor karena malu.
Dampak: Setidaknya 20 korban anak bunuh diri terkait sextortion dalam periode yang sama sebuah tragedi yang menunjukkan batas paling ekstrem dari dampak psikologis deepfake.

Kelompok Rentan Lainnya

Jurnalis dan Aktivis Perempuan

Jurnalis dan aktivis khususnya mereka yang mengkritik kekuasaan, meliput konflik, atau mengadvokasi hak-hak minoritas menghadapi risiko deepfake sebagai alat intimidasi yang ditargetkan secara politis. Pola yang ditemukan di berbagai negara:

- ➲ Deepfake NCII digunakan untuk mendiskreditkan dan mengintimidasi jurnalis perempuan agar berhenti meliput isu-isu tertentu.
- ➲ Ancaman deepfake digunakan sebagai 'pesan' kepada komunitas wartawan untuk 'memperlunak' liputan.
- ➲ Platform tidak selalu merespons dengan cepat ketika korban adalah aktivis dari negara dengan perlindungan hak asasi yang lemah.

Kelompok LGBTQ+

Individu LGBTQ+ menghadapi risiko spesifik dari deepfake: ancaman outing (pengungkapan orientasi seksual atau identitas gender) paksa melalui konten deepfake. Dalam konteks di mana orientasi seksual masih dapat menyebabkan pengucilan keluarga, kekerasan, atau bahkan penuntutan hukum (di beberapa yurisdiksi), deepfake outing memiliki potensi merusak yang sangat besar.

Minoritas Etnis dan Agama

Deepfake dapat digunakan untuk menciptakan konten yang menampilkan anggota kelompok minoritas melakukan tindakan provokatif yang dapat memicu kekerasan komunal. Dalam konteks multietnis dan multikultural Indonesia, risiko ini sangat relevan dan harus masuk dalam pertimbangan regulasi.

Anak-Anak dan Remaja di Lingkungan Sekolah

School-based deepfake bullying penggunaan deepfake oleh sesama siswa untuk mempermalukan atau mengintimidasi korban adalah kategori yang semakin banyak dilaporkan. Karakteristiknya:

- Pelaku dan korban berada dalam komunitas sosial yang sama, memperburuk dampak psikologis dan menciptakan dinamika kekuasaan yang kompleks.
- Distribusi terjadi melalui grup WhatsApp sekolah atau akun anonim, sulit dilacak.
- Dampak psikologis pada korban yang masih dalam perkembangan identitas dan harga diri dapat sangat parah dan berdurasi panjang.

🔍 INTERSEKSIONALITAS: DEEPFAKE DAN BERLAPIS-LAPISNYA KERENTANAN

Konsep interseksionalitas (Kimberlé Crenshaw, 1989) mengajarkan bahwa berbagai bentuk penindasan tidak bekerja secara terpisah mereka saling berpotongan dan memperkuat.

Korban yang berada di persimpangan beberapa kategori kerentanan menghadapi dampak deepfake yang berlipat ganda:

• Perempuan + kelompok minoritas etnis/agama: Deepfake NCII + potensi provokasi komunal

• Remaja perempuan + lingkungan konservatif: Tekanan keluarga + stigma sosial + dampak psikologis perkembangan

• Jurnalis perempuan + liputan kritis: Intimidasi personal + tekanan pada kebebasan pers

• Aktivis LGBTQ+ + konteks hukum yang tidak protektif: Risiko outing + kekosongan perlindungan hukum

Implikasi bagi hukum: Regulasi deepfake yang efektif harus sensitif terhadap interseksi ini bukan hanya mengatur 'korban generik' tetapi mengakui bahwa dampak berbeda-beda berdasarkan posisi seseorang dalam struktur sosial.

3.5 IMPLIKASI TERHADAP DEMOKRASI DAN PEMILU

Demokrasi sebagai Sistem Berbasis Informasi

Demokrasi dalam pengertian substantifnya, bukan sekadar prosedural mensyaratkan adanya ruang publik di mana warga dapat mengakses informasi yang akurat, membentuk opini berdasarkan fakta, dan membuat pilihan politik yang terinformasi. Ini bukan sekadar ideal normatif; ini adalah prasyarat fungsional yang tanpanya proses pemilihan umum menjadi sekadar ritual tanpa substansi.

Deepfake mengancam prasyarat ini secara langsung dengan mengontaminasi ekosistem informasi yang menjadi fondasi pengambilan keputusan demokratis. Ancaman ini bukan hipotesis namun hal ini sudah termanifestasi dalam berbagai peristiwa yang terdokumentasi di seluruh dunia.

Tabel 3.6 Peta Ancaman Deepfake terhadap Proses Demokratis

Fase Pemilu / Demokrasi	Jenis Ancaman Deepfake	Contoh Skenario
Kampanye Pemilu	Lip-sync deepfake ucapan kontroversial kandidat; face swap skandal personal	Video calon presiden tampak mengaku menerima suap; deepfake skandal seksual untuk mendiskreditkan kandidat
Hari Pemungutan Suara	Deepfake deklarasi 'kemenangan' palsu; video 'kerusuhan TPS' yang difabrikasi	Video viral menampilkan kerusuhan di TPS yang tidak pernah terjadi, mendorong orang untuk tidak keluar rumah
Penghitungan Suara	Deepfake pejabat KPU tampak mengakui kecurangan; video manipulasi surat suara palsu	Dampak langsung pada kepercayaan terhadap hasil pemilu dan legitimasi pemenang
Pemerintahan Post-Pemilu	Deepfake pernyataan kebijakan kontroversial pejabat terpilih; manipulasi sidang legislatif	Kebingungan publik tentang posisi kebijakan resmi; erosi kepercayaan pada lembaga pemerintah

Proses Legislasi	Deepfake kesaksian ahli atau pernyataan anggota legislatif	Manipulasi proses pembentukan opini publik tentang RUU tertentu

Kasus-Kasus Nyata: Deepfake dalam Politik Global

📋 **STUDI KASUS: Slovakia, September 2023 Deepfake Audio Menjelang Pemilu**
Dua hari sebelum pemilihan umum Slovakia, sebuah rekaman audio beredar di media sosial yang menampilkan Michal Šimečka, pemimpin partai progresif, tampak mendiskusikan rencana untuk memanipulasi hasil pemilu.
Analisis: Pakar dari berbagai lembaga fact-checking Eropa menyimpulkan rekaman tersebut kemungkinan adalah voice cloning deepfake. Namun analisis tersebut tidak selesai sebelum hari pemungutan suara.
Hasil: Šimečka kalah tipis. Apakah deepfake mempengaruhi hasil? Tidak dapat dipastikan inilah yang membuat ancaman ini begitu berbahaya: dampaknya tidak dapat diukur secara definitif.
Pelajaran Kebijakan: Distribusi deepfake politik menjelang pemilu memerlukan mekanisme respons cepat hitungan jam, bukan hari yang saat ini belum dimiliki sebagian besar negara, termasuk Indonesia.

📋 **STUDI KASUS: Argentina, Oktober 2023 Deepfake dalam Debat Elektoral**
Menjelang pemilihan presiden Argentina, beberapa video deepfake dari kandidat Javier Milei dan Sergio Massa beredar, menampilkan keduanya membuat pernyataan kontroversial yang tidak pernah mereka ucapkan.
Konteks: Argentina memiliki ekosistem media sosial yang sangat aktif dan polarisasi politik yang tinggi kondisi yang mirip dengan Indonesia.
Respons: Argentina tidak memiliki kerangka hukum spesifik untuk deepfake pemilu pada saat itu. Respons bergantung pada platform media sosial yang tidak selalu tepat waktu.
Relevansi untuk Indonesia: Dengan intensitas Pemilu 2029 yang akan datang dan meningkatnya kemampuan teknis pelaku disinformasi, Indonesia perlu merumuskan regulasi anti-deepfake pemilu jauh sebelum siklus pemilu berikutnya.

Indonesia: Kerentanan Khusus

Indonesia memiliki kombinasi karakteristik yang membuatnya sangat rentan terhadap ancaman deepfake dalam konteks demokrasi:

Tabel 3.7 Kerentanan Deepfake di Indonesia

Faktor Kerentanan	Penjelasan	Implikasi
Penetrasi media sosial sangat tinggi	Indonesia memiliki pengguna aktif media sosial terbesar ke-4 dunia; WhatsApp adalah kanal komunikasi utama	Disinformasi berbasis deepfake dapat menyebar dengan kecepatan dan jangkauan yang luar biasa
Literasi digital yang tidak merata	Kesenjangan besar antara pengguna perkotaan dan pedesaan dalam kemampuan verifikasi informasi	Populasi yang lebih rentan menjadi target disinformasi deepfake yang lebih efektif
Polarisasi politik yang tinggi	Sejarah pemilu yang sangat polarized (2014, 2019); identitas politik yang kuat	Audiens yang terpolarisasi lebih mudah menerima deepfake yang mengonfirmasi prasangka mereka
Regulasi yang belum memadai	Tidak ada undang-undang spesifik tentang deepfake politik; UU ITE tidak dirancang untuk ancaman ini	Kekosongan hukum menciptakan impunitas bagi pelaku deepfake politik
Kapasitas forensik digital terbatas	Lembaga penegak hukum dan KPU belum memiliki kapasitas teknis memadai untuk mendeteksi deepfake secara cepat	Waktu respons yang lambat memberi deepfake waktu untuk mempengaruhi persepsi publik

Deepfake dan Krisis Legitimasi Demokrasi

Di luar dampak pada pemilu spesifik, ada ancaman jangka panjang yang lebih mendasar, deepfake berpotensi mengikis legitimasi demokrasi itu sendiri sebagai sistem pemerintahan. Legitimasi demokrasi bergantung pada kepercayaan rakyat bahwa proses pemilihan, representasi, pengambilan keputusan berjalan secara jujur dan transparan. Ketika deepfake membuat publik meragukan semua yang mereka lihat dan dengar tentang proses politik, ketika 'fakta' dapat diklaim sebagai fabrikasi dan fabrikasi diklaim sebagai fakta, landasan epistemic dari kepercayaan demokratis itu sendiri tergoyahkan.

Kekhawatiran Runciman ini bahwa ketidakpastian epistemis yang diproduksi oleh disinformasi (termasuk deepfake) dapat membuat otoritarianisme tampak lebih menarik adalah peringatan yang serius dan relevan dalam konteks konsolidasi demokrasi Indonesia yang masih berlangsung.

Kerangka Perlindungan Demokrasi dari Deepfake

Berbagai negara dan lembaga internasional telah mulai merumuskan kerangka perlindungan demokrasi dari deepfake. Elemen-elemen yang dinilai efektif meliputi:

(1) Regulasi Khusus Deepfake Pemilu: Beberapa negara bagian AS (California, Texas) dan Uni Eropa (AI Act) telah atau sedang merumuskan larangan eksplisit distribusi deepfake yang dimaksudkan untuk mempengaruhi pemilu dalam periode tertentu menjelang pemungutan suara.

(2) Kewajiban Pelabelan (Labeling Mandate): Platform media sosial diwajibkan untuk mendeteksi dan melabeli konten sintetis, khususnya yang berkaitan dengan tokoh politik dan pemilu.

(3) Pembentukan Tim Respons Cepat: Lembaga penyelenggara pemilu membentuk unit khusus yang dapat merespons dan mendiseminasikan klarifikasi terhadap deepfake dalam hitungan jam bukan hari.

(4) Kerja Sama Platform-Negara: Perjanjian antara penyelenggara pemilu dan platform media sosial untuk memprioritaskan penghapusan deepfake pemilu selama periode kritis.

(5) Pendidikan Pemilih: Program literasi digital yang secara spesifik mempersiapkan pemilih untuk mengenali dan memverifikasi konten yang meragukan dengan penekanan khusus pada deepfake.

⚠ REKOMENDASI KEBIJAKAN UNTUK INDONESIA

Berdasarkan analisis di atas, beberapa langkah mendesak yang perlu dipertimbangkan:

1. JANGKA PENDEK (sebelum siklus pemilu berikutnya):

• KPU dan Bawaslu membentuk unit pemantauan konten sintetis berbasis AI.

• Kesepakatan dengan platform media sosial (Meta, TikTok, YouTube) untuk prosedur takedown deepfake pemilu yang dipercepat.

• Literasi digital khusus deepfake masuk kurikulum pendidikan pemilih.

2. JANGKA MENENGAH:

• Revisi UU Pemilu untuk secara eksplisit mengatur dan melarang deepfake dalam kampanye.

• Pembentukan kapasitas forensik digital di lingkungan Bawaslu dan Polri.

• Kriminalisasi yang jelas terhadap distribusi deepfake dengan niat mempengaruhi pemilu.

3. JANGKA PANJANG:

• Kerangka hukum komprehensif tentang konten sintetis berbasis AI.

• Partisipasi aktif dalam pembentukan standar internasional tentang transparansi konten AI.

• Investasi berkelanjutan dalam riset dan kapasitas teknis nasional untuk deteksi deepfake.

3.6 DAMPAK TERHADAP KEPERCAYAAN INTERPERSONAL DAN HUBUNGAN SOSIAL

Di luar dampak makro terhadap demokrasi dan institusi, deepfake juga mengubah dinamika kepercayaan pada level yang paling personal dan intim hubungan antar individu.

Erosi Kepercayaan dalam Komunikasi Sehari-hari

Ketika panggilan video dengan 'anggota keluarga' bisa saja adalah deepfake real-time; ketika 'suara pasangan' di telepon bisa saja adalah kloning; ketika 'foto dari perjalanan teman' bisa saja dihasilkan AI kepercayaan yang mendasari komunikasi sehari-hari mengalami tekanan yang belum pernah ada sebelumnya, ini bukan hanya kekhawatiran teoretis. Setelah berbagai kasus penipuan voice cloning dan video fraud yang mendapat liputan luas, survei menunjukkan bahwa sebagian orang telah mulai meragukan komunikasi digital bahkan dengan orang-orang yang mereka percayai. Ini adalah pergeseran psikologis yang fundamental.

Dampak pada Profesi yang Bergantung pada Kepercayaan Digital

Beberapa profesi yang secara historis bergantung pada verifikasi visual dan audio dalam kerjanya menghadapi tantangan adaptif yang serius:

- Notaris dan Pengacara: Verifikasi identitas klien yang sebelumnya mengandalkan tatap muka atau video call kini memerlukan mekanisme tambahan.
- Jurnalis: Verifikasi keaslian rekaman sumber; konfirmasi identitas narasumber jarak jauh.
- Dokter dan Psikiater: Konsultasi telemedicine memerlukan protokol baru untuk memastikan identitas pasien.
- Penyidik dan Penegak Hukum: Verifikasi identitas saksi dan tersangka dalam konteks digital.

Rangkuman Bab

Bab ini telah menelaah dimensi sosial dan psikologis dari deepfake secara komprehensif. Gambaran yang muncul jauh lebih gelap dan lebih kompleks dari sekadar 'video palsu' yang dapat dideteksi oleh teknologi. Deepfake adalah fenomena yang:

- Menghancurkan kehidupan individu secara nyata melalui PTSD, stigma, kehilangan pekerjaan, dan dalam kasus ekstrem, mendorong korban pada keputusan bunuh diri. Dampak ini tidak proporsional menimpa perempuan dan kelompok rentan lainnya.

- Mengikis infrastruktur kepercayaan sosial melalui uncertainty inflation, penyebaran asimetris, dan melemahnya kepercayaan terhadap institusi. Ini adalah kerusakan struktural, bukan sekadar insiden individual.
- Menciptakan Liar's Dividend yang mengancam sistem pembuktian memungkinkan pelanggar hukum untuk mengklaim bahwa bukti atas pelanggaran mereka adalah deepfake, dan menciptakan keraguan terhadap rekaman yang asli.
- Menjadi senjata penindasan berbasis gender dan kekuasaan dengan 95–99% korban deepfake NCII adalah perempuan, deepfake bukan teknologi yang netral gender; ia mereproduksi dan memperburuk ketidaksetaraan struktural.
- Mengancam fondasi epistemis demokrasi dengan memungkinkan fabrikasi realitas visual, deepfake berpotensi mengikis kepercayaan publik terhadap proses demokratis hingga pada titik di mana legitimasi demokrasi itu sendiri menjadi rentan.

BAB 4
KASUS-KASUS DEEPFAKE DI INDONESIA

Tiga bab sebelumnya telah membangun fondasi konseptual yang kokoh: kita telah memahami apa itu deepfake, bagaimana deepfake bekerja, jenis-jenisnya, dan dampak sosial-psikologisnya. Kini saatnya turun ke bumi memeriksa bagaimana semua itu berwujud dalam kehidupan nyata masyarakat Indonesia. Indonesia bukan sekadar penonton dalam fenomena deepfake global. Negeri ini telah menjadi salah satu arena utama penyalahgunaan teknologi ini dengan kasus-kasus yang mencakup seluruh spektrum taksonomi yang telah kita pelajari: dari penipuan keuangan yang menyasar jutaan warga hingga kekerasan seksual berbasis digital di kampus, dari disinformasi politik hingga krisis kepercayaan terhadap sektor fintech. Yang membuat bab ini krusial adalah konteks spesifik Indonesia penetrasi media sosial yang sangat tinggi, kesenjangan literasi digital yang lebar, kerangka hukum yang masih dalam tahap penyesuaian, dan kapasitas forensik digital lembaga penegak hukum yang terus berkembang. Setiap kasus yang dianalisis di sini bukan sekadar kisah individual deepfake adalah cermin dari tantangan sistemik yang harus direspons secara hukum.

4.1 PETA KASUS DEEPFAKE INDONESIA: GAMBARAN UMUM

Sebelum menelaah masing-masing kasus secara mendalam, penting untuk memetakan lanskap keseluruhan kasus deepfake yang terdokumentasi di Indonesia. Data berikut bersumber dari laporan Bareskrim Polri, OJK/IASC, dan investigasi media terverifikasi.

Tabel 4.1 Dokumentasi Peta Kasus Deepfake di Indonesia

No.	Kasus / Modus	Periode	Kerugian / Dampak	Status Hukum
1	Deepfake Prabowo & pejabat negara penipuan bantuan pemerintah	Nov 2024–Feb 2025	11–100+ korban; Rp30–65 juta (terungkap); korban dari 20 provinsi	2 tersangka ditangkap; dijerat UU ITE & KUHP Pasal.378
2	Deepfake pornografi mahasiswi UNUD via bot Telegram	2024–April 2025	35–200+ korban dari berbagai universitas	Pelaku dipecat dari kampus; proses pidana dalam penyidikan
3	Penipuan keuangan berbasis AI/deepfake sektor fintech (IASC)	Nov 2024–Feb 2025	Rp700 miliar (3 bulan); Rp7,8 triliun kumulatif (setahun)	Ribuan rekening diblokir; investigasi lintas kasus

4	Deepfake Jokowi pidato bahasa Mandarin konteks Pemilu 2024	Oktober 2023	Disinformasi massal; dikonfirmasi Kominfo/Komdigi	Klarifikasi resmi; tidak ada penangkapan tersangka
5	Deepfake Prabowo pidato bahasa Arab kampanye pilpres	2023–2024	Disinformasi dalam konteks kampanye pilpres	Klarifikasi oleh lembaga fact-checking; tidak ada penuntutan
6	Audio deepfake Surya Paloh konteks pemilu 2024	Kampanye 2024	Disinformasi tentang hubungan Nasdem-Anies	Dibantah pihak terkait; tidak ada penangkapan

Tabel 4.1 di atas hanya mencakup kasus-kasus yang terdokumentasi secara publik. Para ahli dan lembaga fact-checking seperti Mafindo memperkirakan bahwa angka aktual kasus deepfake di Indonesia jauh lebih tinggi sebagian besar tidak dilaporkan karena korban tidak menyadari dirinya menjadi korban, malu untuk melapor, atau tidak mengetahui mekanisme pelaporan yang tersedia. Sebagaimana dicatat oleh Aribowo Sasmito, salah satu pendiri Mafindo (Masyarakat Anti Fitnah Indonesia), dalam wawancara dengan AFP (Maret 2025): timnya menemukan penipuan deepfake baru setiap minggu, seiring perangkat AI yang terus menjadi lebih mudah diakses dan terjangkau.

4.2 KASUS I: DEEPFAKE PRESIDEN PRABOWO DAN PENIPUAN BANTUAN PALSU

Kronologi Kasus

STUDI KASUS: Sindikat Penipuan Deepfake Pejabat Negara (Lampung, 2024–2025)

Sumber: Bareskrim Polri (konpers 23 Jan & 7 Feb 2025); Kompas, Tempo, Detik, CNN Indonesia, VOA Indonesia

PELAKU & MODUS:

Pelaku pertama: AMA, 29 tahun, wiraswasta, Lampung Tengah. Ditangkap 16 Januari 2025.

Pelaku kedua: JS, 25 tahun, Pringsewu, Lampung. Ditangkap 4 Februari 2025.

DPO: FA (pembuat/editor video deepfake), belum tertangkap per Februari 2025.

MODUS OPERANDI:

• Pelaku mengunduh atau memesan video deepfake yang menampilkan wajah dan suara Presiden Prabowo Subianto,

Wakil Presiden Gibran Rakabuming Raka, dan Menteri Keuangan Sri Mulyani.

• Narasi dalam video: pejabat negara seolah-olah mengumumkan program bantuan pemerintah senilai Rp50 juta dibuat, semakin sulit dideteksi, dan semakin luas dampaknya.

• JS mencari video dengan kata kunci 'Prabowo Giveaway' di Instagram, lalu mengunduh dan menyebarkan ulang.

• Korban yang menghubungi nomor tersebut diarahkan mengisi formulir pendaftaran, lalu diminta transfer dibuat, semakin sulit dideteksi, dan semakin luas dampaknya.

SKALA KASUS:

• AMA: 11 korban teridentifikasi dalam 4 bulan; kerugian ±Rp30 juta.

• JS: ±100 korban dari 20 provinsi (Desember 2024–Februari 2025); kerugian ±Rp65 juta.

• Akun @indoberbagi2025 memiliki 9.399 pengikut sebelum diblokir.

• Video pertama diunggah di akun @chandra_cchen pada 13 November 2024.

• Korban terbanyak dari Jawa Timur, Jawa Tengah, dan Papua.

PROFIL KORBAN:

Di antaranya Aryani, 56 tahun, yang menyerahkan Rp200.000 setelah melihat video deepfake.

'Saya butuh uang, tetapi sebaliknya saya diminta untuk mengirim uang. Mereka bahkan melakukan panggilan video dengan saya, seolah-olah saya berbicara langsung dengan mereka,' kata Aryani (VOA Indonesia, 2025).

Dimensi Teknis: Bagaimana Video Ini Dibuat?

Berdasarkan keterangan Dirtipidsiber Bareskrim Polri Brigjen Himawan Bayu Aji dalam konferensi pers 23 Januari 2025, video deepfake dalam kasus ini dibuat dengan memanfaatkan teknologi lip-sync AI yang memanipulasi gerakan bibir, wajah, dan suara pejabat negara. Video asli pidato atau pernyataan resmi pejabat digunakan sebagai bahan dasar, kemudian audio dan gerakan bibir diubah menggunakan aplikasi AI yang tersedia secara komersial.

Hal Yang menarik secara teknis adalah adanya pembagian kerja dalam sindikat ini: FA diduga bertindak sebagai pembuat teknis video (*deepfake creator*), sementara AMA dan JS bertindak sebagai distributor dan operator penipuan. Ini menunjukkan telah terbentuknya ekosistem *'deepfake-as-a-service'* pada skala kecil di Indonesia.

Analisis Hukum Kasus I

ANALISIS HUKUM

PASAL YANG DITERAPKAN OLEH PENYIDIK:

1. Pasal 51 ayat (1) jo. Pasal 35 UU No. 1 Tahun 2024 (Perubahan Kedua UU ITE):

Pasal 35: Setiap orang yang dengan sengaja dan tanpa hak atau melawan hukum melakukan manipulasi, dibuat, semakin sulit dideteksi, dan semakin luas dampaknya. dengan tujuan agar informasi atau dokumen tersebut seolah-olah data yang otentik. Pasal 51 ayat (1): Sanksi pidana penjara paling lama 12 tahun dan/atau denda paling banyak Rp12 miliar.

2. Pasal 378 KUHP (Penipuan):

Unsur-unsur terpenuhi:
(a)dengan maksud menguntungkan diri sendiri secara melawan hukum;
(b) menggunakan nama palsu, martabat palsu, tipu muslihat, atau rangkaian kebohongan;
(c) membujuk orang lain untuk menyerahkan sesuatu barang. Ancaman: pidana penjara paling lama 4 tahun.

ANALISIS PILIHAN DAKWAAN:

Penuntutan dengan UU ITE dan KUHP secara bersamaan mencerminkan pendekatan dakwaan berlapis yang tepat. UU ITE menjangkau aspek manipulasi konten digital; KUHP menjangkau aspek penipuan finansial.

KELEMAHAN YANG PERLU DICATAT:

• Tidak ada pasal yang secara spesifik menyebut 'deepfake' atau 'konten sintetis berbasis AI' dalam hukum positif Indonesia saat ini penyidik harus menggunakan analogi dan penafsiran ekstensif.
• FA (pembuat video) belum tertangkap, sehingga pelaku paling kritis secara teknis belum diadili.
• Tidak ada mekanisme khusus untuk pemulihan kerugian korban yang terintegrasi dalam dakwaan.

PRESEDEN HUKUM:

Kasus ini menjadi preseden penting sebagai kasus deepfake pertama yang secara eksplisit disebutkan dalam konferensi pers resmi Bareskrim Polri dan diberitakan secara luas sebagai 'kasus deepfake'.

Dimensi Sosial: Mengapa Orang Tertipu?

Pertanyaan ini penting secara akademis: mengapa ratusan orang dari 20 provinsi termasuk dari Papua dan Sulawesi dapat tertipu oleh video yang, bagi sebagian orang, tampak mencurigakan?

Jawabannya terletak pada kombinasi faktor psikologis dan kontekstual:

(1) Otoritas figur presiden: Video menampilkan wajah dan suara pemimpin tertinggi negara figur dengan otoritas tertinggi dalam hierarki kepercayaan publik Indonesia.

(2) Harapan ekonomi yang nyata: Bagi banyak warga terutama di daerah dengan tingkat kemiskinan yang lebih tinggi tawaran bantuan Rp50 juta bukan sesuatu yang secara refleks ditolak sebagai tidak mungkin, mengingat berbagai program bantuan pemerintah memang nyata ada.
(3) Kualitas video yang meyakinkan: Menurut beberapa korban, video tersebut sangat realistis dan sulit dibedakan dari video asli Presiden Prabowo.
(4) Penyebaran melalui jaringan kepercayaan: Video beredar di WhatsApp dan Instagram platform yang dipercaya dan sering disebarkan oleh orang yang dikenal.
(5) Kesenjangan literasi digital: Pemahaman tentang teknologi deepfake masih sangat rendah di segmen masyarakat yang menjadi target utama.

⚠ IMPLIKASI KEBIJAKAN: SIAPA YANG BERTANGGUNG JAWAB MELINDUNGI KORBAN?

Kasus ini memunculkan pertanyaan distribusi tanggung jawab yang penting:

• Platform media sosial (Instagram, TikTok, WhatsApp/Meta): Apakah mereka memiliki kewajiban hukum untuk mendeteksi dan menghapus konten deepfake yang jelas bertujuan menipu? Di bawah hukum Indonesia saat ini, kewajiban ini belum diatur secara spesifik.

• Pemerintah (Komdigi): Respons Komdigi dalam kasus ini lebih bersifat klarifikasi reaktif (setelah viral) daripada pencegahan proaktif. Diperlukan sistem pemantauan aktif konten deepfake yang mencatut pejabat negara.

• Bareskrim Polri: Penangkapan dilakukan setelah ada laporan artinya korban harus ada terlebih dahulu. Diperlukan mekanisme deteksi proaktif, bukan hanya responsif.

• Lembaga perlindungan konsumen: BPKN (Badan Perlindungan Konsumen Nasional) belum memiliki mandat yang jelas untuk kasus penipuan berbasis konten digital sintetis.

4.3 KASUS II: DEEPFAKE PORNOGRAFI MAHASISWI UNUD VIA BOT TELEGRAM

STUDI KASUS: Kekerasan Seksual Berbasis AI (2024–2025)

Sumber: Kompas (25 April 2025), Kumparan (27 April 2025), Suara Bali (25 & 30 April 2025), Radar Malang (29 April 2025), Bicara Network (2 Mei 2025)

PELAKU:

Sergio Lucasandro Ksatria Dwi Putra (SLKDP), mahasiswa semester 6 Program Studi Akuntansi, Fakultas Ekonomi dan Bisnis (FEB), Universitas Udayana, Bali.

MODUS OPERANDI:

• Pelaku mengambil foto perempuan secara diam-diam dari akun Instagram para korban tanpa izin.

• Foto-foto tersebut kemudian diedit menggunakan bot berbasis kecerdasan buatan (AI) di aplikasi Telegram menghasilkan gambar vulgar yang menampilkan wajah korban pada tubuh tanpa busana.

• Konten hasil editan disebarkan melalui jaringan tertutup hingga akhirnya bocor ke publik.

• Pelaku diduga mulai melakukan perbuatan ini sejak masa SMA.

TERUNGKAPNYA KASUS:

Kasus terungkap secara tidak terduga: mantan pacar pelaku mengirimkan tangkapan layar berisi foto-foto korban yang diedit kepada sejumlah korban melalui media sosial pada Kamis, 13 Maret 2025. Kasus kemudian viral setelah akun X @blankkyle506 memposting thread pada 22 April 2025.

SKALA KASUS:

• Dalam sidang etik, pelaku mengakui lebih dari 35 mahasiswi sebagai korban.

• Laporan dari berbagai sumber menyebut lebih dari 200 perempuan dari berbagai universitas

(Universitas Udayana, Universitas Tarumanagara, Universitas Bunda Mulia, dan lainnya).

• Korban berasal dari berbagai kampus di Indonesia — bukan hanya Universitas Udayana.

RESPONS INSTITUSIONAL:

• Sidang etik Fakultas Ekonomi dan Bisnis: 18 Maret 2025.

• Laporan ke Rektor: 21 Maret 2025.

• Sanksi sementara: pelaku dilarang mengakses layanan akademik.

• Pemecatan tetap (Drop Out): ditetapkan Rektor I Ketut Sudarsana pada 30 April 2025, melalui Surat Keputusan Rektor, melanggar Pasal 12 ayat 2 huruf f Permendikbudristek No.55/2024.

• Proses pidana: Satgas PPKS menyarankan korban melapor ke kepolisian.

Teknologi yang Digunakan: Bot Telegram dan Aksesibilitas Ancaman

Aspek yang paling mengkhawatirkan dari kasus UNUD adalah kesederhanaan teknologi yang digunakan. Pelaku tidak memerlukan keahlian teknis mendalam, perangkat keras canggih, atau bahkan akun berbayar di platform komersial. Deepfake cukup menggunakan bot AI di aplikasi Telegram yang dapat diakses secara gratis oleh siapa pun. Bot-bot AI undressing di Telegram (yang menghasilkan gambar vulgar dari foto orang berpakaian) telah menjadi salah

satu bentuk penyalahgunaan AI yang paling luas dan paling sedikit mendapat perhatian regulasi. Laporan Sensity AI (2019) telah mengidentifikasi ribuan pengguna bot semacam ini, dengan mayoritas korban adalah perempuan yang foto-fotonya diambil dari media sosial tanpa sepengetahuan mereka.

PROFIL ANCAMAN: BOT AI 'UNDRESSING' DI TELEGRAM

Karakteristik bot AI berbasis Telegram yang digunakan dalam kasus UNUD dan kasus serupa:

• Aksesibilitas: Dapat diakses gratis atau dengan biaya sangat rendah melalui aplikasi Telegram.

• Kemudahan penggunaan: Tidak memerlukan keahlian teknis cukup kirim foto, bot memproses dan mengembalikan gambar yang dimodifikasi dalam hitungan detik.

• Anonimitas: Telegram memungkinkan penggunaan pseudonim; pelacakan pelaku memerlukan kerja sama platform.

• Volume: Laporan Sensity AI (2019) menemukan satu bot saja telah memproses foto lebih dari 100.000 perempuan.

• Status hukum bot: Tidak ada ketentuan hukum Indonesia yang secara eksplisit melarang penyediaan atau penggunaan bot semacam ini menciptakan kekosongan hukum yang signifikan.

Catatan: Bot-bot ini menggunakan teknologi 'image inpainting' berbasis AI secara teknis berbeda dari deepfake video, namun secara hukum dan dampaknya setara karena menghasilkan konten seksual palsu dari gambar seseorang tanpa persetujuan.

Dampak Psikologis Korban: Temuan dari Kasus UNUD

Berdasarkan laporan media dan pernyataan korban yang dikutip berbagai media (dengan nama korban tidak disebutkan demi perlindungan privasi), dampak psikologis yang dialami korban mencerminkan pola yang dibahas dalam Bab 3:

- Shock dan disorientasi: *"Saya tidak menyangka foto saya dari Instagram bisa digunakan seperti itu,"* kata salah satu korban kepada media.
- Rasa malu dan penghinaan: Banyak korban menghadapi kesulitan untuk berbicara tentang pengalaman mereka, bahkan kepada keluarga terdekat.
- Kepercayaan diri yang terguncang: Beberapa korban melaporkan menjadi lebih tertutup di media sosial dan enggan berbagi foto sejak kejadian.
- Kekhawatiran tentang penyebaran lebih lanjut: Ketidakpastian tentang apakah masih ada konten yang beredar di jaringan gelap menjadi sumber kecemasan yang berkelanjutan.

Yang perlu dicatat adalah bahwa respons institusional Universitas Udayana meskipun terlambat dari sisi pencegahan tergolong responsif setelah kasus terungkap: pemecatan resmi

dalam 6 minggu setelah kasus viral adalah langkah tegas yang melampaui respons banyak institusi pendidikan dalam kasus serupa di negara lain.

Analisis Hukum Kasus II

ANALISIS HUKUM

PASAL YANG RELEVAN DAN POTENSI PENERAPANNYA:

1. UU No. 44 Tahun 2008 tentang Pornografi, Pasal 4 ayat (1):

Melarang pembuatan, penyebarluasan, dan penggunaan pornografi. Pertanyaan kritis: apakah konten deepfake NCII termasuk 'pornografi' dalam pengertian UU Pornografi? Secara harfiah, definisi pornografi dalam UU ini (Pasal 1 angka 1) mencakup 'gambar yang memuat kecabulan atau eksploitasi seksual yang melanggar norma kesusilaan dalam masyarakat' yang secara tekstual dapat mencakup deepfake NCII, meskipun kata 'sintetis' atau 'buatan AI' tidak disebutkan.

2. UU No. 12 Tahun 2022 tentang Tindak Pidana Kekerasan Seksual (TPKS), Pasal 14:

Mengkriminalisasi 'kekerasan seksual berbasis elektronik' yaitu melakukan perekaman, pengambilan, dan/atau penyebaran foto/video yang bermuatan seksual tanpa persetujuan.

ANALISIS: Ini adalah pasal yang paling tepat untuk kasus UNUD. Namun, frasa 'perekaman' dan 'pengambilan' dalam pasal ini perlu interpretasi ekstensif untuk mencakup pembuatan konten sintetis dari foto yang ada karena tidak ada 'perekaman' dalam pengertian konvensional.

3. Permendikbudristek No. 55 Tahun 2024, Pasal 12 ayat 2 huruf f:

Dasar hukum yang digunakan untuk pemecatan pelaku oleh Rektor UNUD. Ini merupakan sanksi administratif, bukan pidana.

KESENJANGAN HUKUM YANG TERIDENTIFIKASI:

• Tidak ada pasal yang secara eksplisit mengkriminalisasi PEMBUATAN (bukan sekadar penyebaran) konten seksual deepfake menggunakan AI dari foto seseorang tanpa persetujuan.

• UU TPKS Pasal 14 lebih tepat, namun masih memerlukan penafsiran untuk mencakup konten sintetis.

• Perlu dipertimbangkan: apakah mengambil foto dari media sosial publik kemudian mengeditnya menjadi konten seksual juga memenuhi unsur 'tanpa persetujuan'? Secara etis jelas; secara yuridis perlu argumentasi yang kuat.

Rekomendasi Awal:

Kasus ini menunjukkan urgensi amandemen UU TPKS atau penerbitan Peraturan Pemerintah pelaksana yang secara eksplisit mencakup kekerasan seksual berbasis konten sintetis AI.

4.4 KASUS III: PENIPUAN KEUANGAN BERBASIS AI/DEEPFAKE DI SEKTOR FINTECH

Skala Ancaman: Data OJK dan IASC

Di antara semua kasus deepfake di Indonesia, dampak ekonomi terbesar justru terjadi di sektor keuangan digital. *Indonesia Anti-Scam Centre* (IASC) lembaga yang dibentuk OJK dan mulai beroperasi pada 22 November 2024 telah mendokumentasikan kerugian dalam skala yang mengejutkan.

Tabel 4.2 Data OJK dan IASC menunjukkan total kerugian Korban Penipuan Berbasis AI

Periode	Laporan Diterima IASC	Total Kerugian Dilaporkan	Dana Berhasil Diblokir
Nov 2024–Feb 2025	42.257 laporan	Rp700,2 miliar	Rp106,8 miliar
Nov 2024–Nov 2025	343.402 laporan	Rp7,8 triliun	Rp386,5 miliar
Nov 2024–Jan 2026	432.000+ laporan	Rp9,1 triliun	Dalam proses
Rekening dilaporkan	563.558 rekening	—	106.222 rekening diblokir

Sumber data: Satgas PASTI OJK (15 November 2025) dan IASC (April 2026), sebagaimana dilaporkan oleh Selular.id dan Fintech News Indonesia.

Catatan penting: Data IASC mencakup seluruh penipuan digital, tidak hanya yang berbasis deepfake. Namun OJK secara eksplisit menyebut voice cloning dan deepfake sebagai dua modus AI-fraud yang paling marak dalam rilis resminya (November 2025). Laporan VIDA (platform identity verification) secara spesifik mencatat lonjakan 1.550% kasus penipuan berbasis AI di sektor fintech Indonesia sepanjang 2024.

Modus Spesifik Deepfake dalam Penipuan Keuangan

OJK mengidentifikasi dua modus utama AI-fraud yang menggunakan deepfake:

Modus 1: Video Call Deepfake dalam Proses KYC (Know Your Customer)

Platform keuangan digital neobank, P2P lending, dompet digital mengandalkan proses verifikasi identitas digital (e-KYC) yang biasanya melibatkan unggah foto KTP dan selfie/video singkat. Pelaku menggunakan deepfake real-time untuk:

- Menyamar sebagai orang lain dalam proses video-KYC untuk membuka rekening atas nama korban.
- Menggunakan KTP palsu berbasis AI (deepfake dokumen) yang dikombinasikan dengan foto wajah sintetis.

- Memanfaatkan rekening yang berhasil dibuka untuk pencucian uang, penipuan pinjol, atau transfer hasil kejahatan.

DATA: KENAIKAN 1.550% KASUS AI-FRAUD FINTECH INDONESIA (2024)

Laporan VIDA (2024) yang dikutip oleh Authme.com menyebutkan:
"Deepfake fraud in Indonesia saw significant growth in 2024. The VIDA report highlighted a 1,550% surge in AI-driven fraud cases related to the FinTech sector."

Konteks: Kenaikan ini terjadi bersamaan dengan pertumbuhan 35% adopsi aplikasi keuangan di kawasan Asia Pasifik (Laporan Adjust, 2025), menunjukkan bahwa ekspansi layanan keuangan digital membuka permukaan serangan (attack surface) yang lebih luas bagi penyalahgunaan deepfake.

Modus 2: Voice Cloning untuk Penipuan Transfer Dana

OJK (Satgas PASTI, November 2025) menjelaskan: pelaku kejahatan merekam dan meniru suara seseorang seperti teman, kolega, atau anggota keluarga menggunakan AI, kemudian melakukan percakapan telepon seolah-olah mereka adalah orang yang dikenal korban, untuk meminta transfer dana darurat.

STUDI KASUS: Penipuan Voice Cloning 'Suara Pejabat' Modus Korporat

Sumber: Diolah dari data Satgas PASTI OJK (2025) dan laporan media terkait penipuan berbasis AI di Indonesia

Pola penipuan yang dilaporkan kepada OJK:

• Pelaku mengidentifikasi target (biasanya pegawai keuangan atau akuntan di perusahaan).
• Mengkloning suara atasan/direktur dari rekaman publik (wawancara media, video company profile, dll.).
• Menelepon target menggunakan suara kloning, mengklaim situasi darurat dan memerintahkan transfer dana ke rekening tertentu 'untuk keperluan bisnis mendesak yang harus dilakukan sebelum akhir hari.'
• Korban, yang mempercayai suara atasannya, melakukan transfer tanpa verifikasi lebih lanjut.

Kerugian: Tidak ada kasus tunggal berkaliber tinggi yang terpublikasi di Indonesia seperti kasus UAE (USD$35 juta, 2020). Namun OJK mengindikasikan modus ini telah menyebabkan kerugian dalam miliaran rupiah secara kumulatif, tersebar dalam banyak kasus individual yang tidak selalu dilaporkan secara terpadu.

Respons Industri: Teknologi Lawan Teknologi

Menghadapi ancaman ini, beberapa lembaga keuangan digital Indonesia telah mengambil langkah proaktif:

- ☑ Allo Bank x ADVANCE.AI: Sejak 2021 menerapkan sistem verifikasi biometrik berlapis yang mencakup active liveness detection (meminta pengguna melakukan gerakan spesifik), passive liveness detection (analisis kedalaman wajah dan aliran darah), serta verifikasi dokumen ID yang dikombinasikan dengan pengenalan wajah. Sistem ini dirancang secara spesifik untuk mendeteksi deepfake dalam proses onboarding.
- ☑ Bank-bank digital lainnya: Sebagian besar neobank terkemuka Indonesia kini mengintegrasikan teknologi anti-deepfake dalam proses KYC mereka, meskipun tingkat sofistikasinya bervariasi.
- ☑ OJK: Menekankan pentingnya *'security-by-design'* membangun kemampuan deteksi penipuan ke dalam arsitektur sistem sejak awal, bukan sebagai add-on setelah fakta.

Analisis Hukum Kasus III

ANALISIS HUKUM

KERANGKA HUKUM YANG BERLAKU:

1. UU No. 4 Tahun 2023 tentang Pengembangan dan Penguatan Sektor Keuangan (P2SK):
Memberikan mandat kepada OJK untuk mengatur dan mengawasi keamanan sistem keuangan, termasuk kewajiban lembaga keuangan menerapkan sistem keamanan yang memadai.

2. POJK No. 27 Tahun 2024 tentang Aset Keuangan Digital:
Memperkuat pengawasan terhadap aset digital dan platform yang mengelolanya.

3. Pasal 378 KUHP (Penipuan) & Pasal 362 KUHP (Pencurian):
Berlaku untuk pelaku penipuan individual; namun pembuktian sangat bergantung pada kemampuan mengidentifikasi pelaku yang sering menggunakan identitas palsu.

4. UU No. 27 Tahun 2022 tentang Perlindungan Data Pribadi (UU PDP):
Penggunaan data biometrik (wajah, suara) seseorang tanpa izin untuk membuat deepfake merupakan pelanggaran UU PDP namun sanksinya bersifat administratif, bukan pidana.

KESENJANGAN KRITIS:

• Tidak ada kewajiban hukum yang eksplisit bagi platform fintech untuk menerapkan teknologi anti-deepfake dalam proses KYC ini masih bersifat rekomendasi industri.

• Standar minimum keamanan KYC yang 'deepfake-resistant' belum diformulasikan dalam regulasi OJK.

• Tanggung jawab hukum lembaga keuangan ketika nasabah menjadi korban penipuan KYC berbasis deepfake belum diatur secara jelas siapa yang menanggung kerugian?

URGENSI REGULASI:

OJK perlu menerbitkan regulasi yang secara eksplisit mewajibkan lembaga keuangan menerapkan sistem liveness detection yang tahan deepfake sebagai standar minimum KYC digital.

4.5 KASUS IV: DEEPFAKE DALAM KONTEKS PEMILU INDONESIA

Pemilu 2024 sebagai Laboratorium Disinformasi AI

Pemilihan Umum 2024 Indonesia yang mencakup pemilihan presiden, legislatif, dan kepala daerah menjadi arena pertama di mana teknologi deepfake dan manipulasi berbasis AI digunakan secara terdokumentasi dalam konteks elektoral di Indonesia. Beberapa kasus yang terdokumentasi oleh media dan lembaga fact-checking:

STUDI KASUS: Deepfake Jokowi Pidato Bahasa Mandarin: Disinformasi Elektoral

Sumber: Komdigi (klarifikasi resmi Oktober 2023); Universitas Muhammadiyah Surakarta; Antara News; CSIRT Jogjakota

KRONOLOGI:
Sebuah video beredar di media sosial menampilkan Presiden Joko Widodo (Jokowi) seolah-olah berpidato dalam bahasa Mandarin. Video tersebut sebenarnya berasal dari pidato Jokowi dalam bahasa Inggris saat menghadiri acara Gala Hosted by USINDO, US Chamber, and USABC pada 26 Oktober 2015. Video asli tersebut kemudian dimanipulasi menggunakan AI deepfake agar gerakan bibir dan audio tampak seolah Jokowi berpidato dalam bahasa Mandarin.

NARASI YANG DIBAWA:
Video deepfake ini digunakan untuk menguatkan narasi bahwa pemerintahan Jokowi dikendalikan oleh China sebuah narasi yang digunakan dalam konteks kampanye jelang Pemilu 2024.

RESPONS:
Kementerian Komunikasi dan Informatika (Kominfo, kini Komdigi) segera mengkonfirmasi video tersebut sebagai manipulasi deepfake dan mengeluarkan klarifikasi resmi.

DAMPAK:
Meskipun telah diklarifikasi, banyak pengguna media sosial yang sempat terkecoh termasuk mereka yang mengklaim cukup melek digital. Ini mengilustrasikan efek first impression dari deepfake.

STATUS HUKUM:
Tidak ada tersangka yang ditangkap dalam kasus ini.

STUDI KASUS: Audio Deepfake Surya Paloh: Narasi Perpecahan Internal Partai

Sumber: Tempo CekFakta (2024); ISEAS-Yusuf Ishak Institute; berbagai media fact-checking

KRONOLOGI: Selama masa kampanye Pemilu 2024, beredar rekaman audio yang menampilkan suara yang diklaim sebagai Ketua Umum Partai Nasional Demokrat Surya Paloh, yang seolah-olah sedang menegur Anies Baswedan.

KONTEKS:
Rekaman ini beredar dalam konteks meningkatnya spekulasi publik tentang hubungan antara Partai Nasdem dan pasangan calon Anies-Muhaimin. Konten audio tersebut berpotensi memperburuk persepsi tentang perpecahan internal koalisi.

FAKTOR KEPERCAYAAN SELEKTIF:
Studi ISEAS bersama LSI (Lembaga Survei Indonesia) yang dikutip Tempo CekFakta menemukan bahwa pemilih Indonesia menunjukkan 'keyakinan selektif' (*selective belief*) terhadap konten deepfake mereka lebih cenderung mempercayai disinformasi yang sejalan dengan kesetiaan politik mereka.

IMPLIKASI:
Temuan ini menunjukkan bahwa efektivitas deepfake politik bukan semata-mata bergantung pada kualitas teknisnya, tetapi pada polarisasi yang sudah ada dalam ekosistem informasi pemilih.

STUDI KASUS: Video Deepfake Prabowo Pidato Bahasa Arab Kampanye Pilpres

Sumber: Verihubs.com; berbagai media fact-checking; Antara News

KRONOLOGI:
Menjelang dan selama kampanye Pilpres 2024, beredar video yang menampilkan Prabowo Subianto (saat itu calon presiden) seolah-olah berpidato dalam bahasa Arab. Video ini dibuat dengan memanipulasi rekaman pidato Prabowo yang asli menggunakan teknologi deepfake lip-sync.

TUJUAN NARASI:
Konten ini berpotensi digunakan untuk berbagai narasi baik yang mendukung maupun yang menyudutkan Prabowo, tergantung konteks penyebarannya. Fenomena ini menunjukkan ambivalensi penggunaan deepfake dalam politik: konten yang sama dapat diinterpretasikan secara berbeda oleh audiens yang berbeda.

RESPONS:
Lembaga fact-checking mengkonfirmasi video tersebut sebagai manipulasi AI.

CATATAN:

Kasus ini menjadi semakin ironis ketika Prabowo kemudian memenangkan Pilpres dan menjadi Presiden lalu wajahnya digunakan kembali dalam penipuan deepfake bantuan pemerintah (Kasus I, Januari 2025).

Pola Disinformasi Deepfake dalam Pemilu Indonesia

Dari analisis kasus-kasus di atas, beberapa pola khas disinformasi deepfake dalam konteks pemilu Indonesia dapat diidentifikasi:

Tabel 4.3 Pola Disinformasi dalam Contoh Kasus

Pola	Deskripsi	Contoh Kasus
Ethnic/Religious Wedge	Deepfake digunakan untuk menguatkan narasi tentang afiliasi etnis atau agama kandidat yang kontroversial	Video Jokowi pidato Mandarin (narasi pro-China)
Internal Coalition Rupture	Audio/video yang menampilkan konflik internal koalisi yang sebenarnya tidak terjadi	Audio Surya Paloh 'menegur' Anies
Character Assassination	Konten yang menampilkan kandidat dalam situasi memalukan, melakukan kesalahan, atau mengakui sesuatu	Video berbagai kandidat dalam konteks negatif
False Policy Attribution	Video yang membuat kandidat tampak mendukung kebijakan yang sebenarnya tidak mereka dukung	Berbagai kasus yang terdokumentasi fact-checker

Analisis Hukum Kasus IV

ANALISIS HUKUM

KERANGKA HUKUM YANG RELEVAN UNTUK DEEPFAKE PEMILU:

1. UU No. 7 Tahun 2017 tentang Pemilihan Umum, Pasal 521:
Mengkriminalisasi kampanye dengan isu SARA. Deepfake yang membuat kandidat tampak berafiliasi dengan etnis atau agama tertentu secara manipulatif berpotensi dijerat pasal ini. Namun: pasal ini tidak secara eksplisit menyebut konten digital atau konten sintetis.

2. Pasal 28 ayat (2) UU ITE:
Melarang penyebaran informasi yang menimbulkan rasa kebencian berdasarkan SARA. Relevan untuk deepfake yang membawa narasi isu SARA dalam konteks pemilu.

3. PKPU No. 15 Tahun 2023 tentang Kampanye:

Mengatur larangan kampanye hitam (black campaign), namun tidak mendefinisikan deepfake atau konten AI sebagai kategori tersendiri.

KELEMAHAN FUNDAMENTAL:

• Tidak ada satu pun regulasi pemilu Indonesia yang secara eksplisit mengatur, melarang, atau memberikan sanksi terhadap penggunaan deepfake dalam kampanye.

• Bawaslu tidak memiliki mandat hukum yang jelas untuk menangani kasus deepfake sebagai pelanggaran pemilu — hanya sebagai 'pelanggaran kampanye' secara umum.

• Tidak ada prosedur yang memungkinkan Bawaslu mengajukan permintaan takedown deepfake kampanye kepada platform dalam hitungan jam.

PERBANDINGAN: California (AS) telah memiliki AB 602 dan AB 730 sejak 2019 yang secara eksplisit melarang pembuatan dan distribusi deepfake yang dimaksudkan untuk mempengaruhi pemilu. Indonesia belum memiliki regulasi setara.

4.6 MENGAPA KORBAN SULIT MELAPOR DAN APARAT SULIT MEMBUKTIKAN

Hambatan di Sisi Korban

Salah satu karakteristik paling kritis dari kasus deepfake Indonesia adalah tingginya angka under-reporting kasus yang tidak dilaporkan. Berdasarkan analisis pola dari kasus-kasus yang telah dibahas, beberapa hambatan sistemik dapat diidentifikasi:

1. Hambatan Psikologis

- ☑ Rasa malu dan stigma: Terutama dalam kasus deepfake NCII, korban sering merasa bahwa melapor akan memperluas lingkaran orang yang mengetahui konten tersebut, memperparah rasa malu daripada menguranginya.
- ☑ Self-blame: 'Mengapa saya memposting foto itu di Instagram?' pertanyaan ini menghantui korban dan menciptakan hambatan internal untuk melapor.
- ☑ Ketidakpercayaan pada respons aparat: Berdasarkan pengalaman komunitas korban kekerasan siber, banyak korban mengantisipasi respons yang tidak sensitif atau pertanyaan yang menyudutkan dari petugas yang menerima laporan.
- ☑ Trauma re-exposure: Proses pelaporan memaksa korban untuk menceritakan pengalaman traumatis berulang kali kepada penyidik, psikolog forensik, jaksa, dan mungkin pengadilan.

2. Hambatan Pengetahuan dan Akses

- ☑ Ketidaktahuan tentang mekanisme pelaporan: Banyak korban tidak mengetahui bahwa ada pasal-pasal yang relevan dalam UU ITE atau UU TPKS, atau ke mana harus melapor (Bareskrim Polri? Polda? KPAI? Komnas Perempuan?).
- ☑ Ketidakjelasan yurisdiksi: Jika korban berada di Bali dan pelaku berada di Jakarta, dan konten diunggah di server luar negeri siapa yang berwenang menangani?

- ☑ Biaya dan aksesibilitas: Proses hukum memerlukan biaya yang tidak sedikit baik langsung (pengacara, biaya forensik) maupun tidak langsung (waktu, perjalanan, cuti kerja).

3. Hambatan Institusional

- ☑ Respons tidak sensitif gender di tingkat penerimaan laporan: Data dari Komnas Perempuan menunjukkan bahwa korban kekerasan berbasis gender termasuk kasus siber sering menghadapi sikap tidak sensitif dari petugas penerima laporan.
- ☑ Minimnya unit khusus di tingkat polres/polda: Dittipidsiber Bareskrim Polri memiliki kapasitas yang lebih baik, namun akses ke unit ini terbatas bagi korban di daerah.
- ☑ Tidak ada 'one-stop' mekanisme pelaporan: Korban deepfake harus menavigasi berbagai institusi (polisi, Kominfo untuk takedown, Komnas Perempuan untuk dukungan, pengadilan untuk proses) yang tidak terintegrasi.

Hambatan di Sisi Penegakan Hukum

Di sisi aparat penegak hukum, berbagai hambatan struktural juga mempersulit proses pembuktian kasus deepfake:

Keterbatasan Kapasitas Teknis Forensik

Pembuktian kasus deepfake memerlukan keahlian forensik digital yang sangat spesifik. Analisis keaslian video deepfake membutuhkan pemahaman mendalam tentang artefak kompresi, anomali piksel, inkonsistensi pencahayaan, dan kadang-kadang akses ke model AI yang digunakan. Saat ini, kapasitas ini terkonsentrasi di Puslabfor Polri (Pusat Laboratorium Forensik) dan unit-unit tertentu di Dittipidsiber sementara kebutuhan ada di seluruh Indonesia.

Tantangan Pembuktian Dalam Hukum Acara

- ➲ Standar bukti digital: KUHAP mengatur bukti berupa 'surat dan petunjuk,' dan UU ITE mengakui dokumen elektronik sebagai alat bukti. Namun, standar autentikasi khusus untuk video yang potensially deepfake belum dirumuskan.
- ➲ Rantai bukti digital (*chain of custody*): Dalam dunia digital, video dapat disalin, diubah metadata-nya, atau dimanipulasi setelah diunduh. Menjaga integritas rantai bukti digital memerlukan prosedur khusus yang belum terstandarisasi di semua tingkat penegakan hukum.
- ➲ Identifikasi pelaku: Ketika pelaku menggunakan identitas palsu, VPN, server luar negeri, dan mata uang kripto untuk transaksi, melacak identitas aslinya memerlukan kerja sama internasional yang prosesnya panjang.

Tantangan Yurisdiksi dan Kerja Sama Internasional

- ➢ Platform media sosial: Sebagian besar platform besar (Meta/Instagram, TikTok, Telegram) beroperasi di bawah yurisdiksi hukum asing. Permintaan data pengguna harus melalui prosedur *Mutual Legal Assistance Treaty* (MLAT) yang memakan waktu berbulan-bulan.
- ➢ Server konten: Konten deepfake sering diunggah ke server di berbagai negaramembuat satu permintaan takedown tidak efektif.

- Pelaku lintas negara: Kasus penipuan deepfake sering melibatkan sindikat yang beroperasi dari berbagai negara, terutama di kawasan Asia Tenggara.

Ketiadaan Definisi Hukum yang Presisi

Ini adalah hambatan yang paling fundamental, tidak ada satu pun undang-undang Indonesia yang mendefinisikan 'deepfake' atau 'konten sintetis berbasis AI.' Akibatnya, penyidik dan penuntut harus menggunakan analogi dan penafsiran ekstensif — yang membuka ruang bagi argumen pembelaan dan ketidakpastian hukum.

TABEL PERBANDINGAN: HAMBATAN vs. SOLUSI YANG DIPERLUKAN

HAMBATAN KORBAN → SOLUSI YANG DIPERLUKAN:

- Rasa malu & stigma → Pelatihan aparat sensitif gender; jaminan anonimitas pelapor
- Tidak tahu mekanisme → Hotline terpadu deepfake nasional; edukasi publik masif
- Biaya hukum → Bantuan hukum negara untuk korban deepfake; simplifikasi prosedur
- Re-traumatisasi → Protokol trauma-informed reporting; visum digital tanpa tatap muka berulang

HAMBATAN APARAT → SOLUSI YANG DIPERLUKAN:

- Kapasitas forensik terbatas → Investasi lab forensik digital di setiap Polda; sertifikasi nasional
- Tidak ada definisi hukum → Regulasi khusus deepfake atau amandemen UU ITE yang eksplisit
- Yurisdiksi internasional → Perjanjian bilateral data-sharing dengan platform dan negara mitra
- Standar pembuktian → Revisi ketentuan alat bukti elektronik dalam KUHAP; standardisasi CoC digital
- Identifikasi pelaku → Kapasitas OSINT dan kerja sama dengan INTERPOL Cyber Crime

Peta Kesenjangan Hukum: Sintesis dari Kasus-Kasus Indonesia

Dari analisis keempat kasus di atas, sebuah peta kesenjangan hukum yang komprehensif dapat disusun. Tabel 4.4 merangkum kesenjangan yang teridentifikasi dan urgensi penanganannya:

Tabel 4.4 Kesenjangan Hukum dalam Kasus Deepfake di Indonesia

Jenis Kasus	Pasal yang Ada	Kesenjangan Kritis	Urgensi
Penipuan deepfake pejabat negara	Pasal.35+51 UU ITE; Pasal.378 KUHP	Tidak ada pasal yang menyebut deepfake secara eksplisit; pelaku teknis (creator) sulit dijerat	🔴 Sangat Tinggi

Deepfake NCII / pornografi	UU Pornografi; UU TPKS Pasal.14	Tidak mencakup konten sintetis AI secara eksplisit; bot Telegram tidak diatur	🔴 Sangat Tinggi
Deepfake fraud fintech/KYC	UU P2SK; UU PDP; KUHP	Tidak ada kewajiban minimum anti-deepfake untuk fintech; tanggung jawab lembaga keuangan tidak jelas	🟠 Tinggi
Deepfake politik/pemilu	UU Pemilu; UU ITE Pasal.28	Tidak ada regulasi deepfake pemilu; Bawaslu tidak punya mandat; tidak ada prosedur respons cepat	🔴 Sangat Tinggi
Deepfake dokumen identitas	KUHP Pasal.263-264; UU ITE Pasal.35	Tidak mencakup fabrikasi AI; KYC digital belum diwajibkan standar anti-deepfake	🟠 Tinggi

KESIMPULAN UTAMA: KEKOSONGAN REGULASI DEEPFAKE DI INDONESIA

Dari analisis seluruh kasus dalam bab ini, sebuah kesimpulan yang tidak dapat dihindari:

Indonesia saat ini tidak memiliki kerangka hukum yang dirancang secara spesifik untuk menangani deepfake. Penanganan kasus-kasus yang ada bergantung pada:

(1) interpretasi ekstensif pasal-pasal lama yang tidak dirancang untuk teknologi AI,

(2) kreativitas penyidik dan penuntut dalam menemukan pasal yang 'paling cocok',

(3) keberuntungan dalam hal apakah hakim memahami konteks teknisnya.

Ini bukan fondasi yang kokoh untuk perlindungan hukum di era di mana deepfake semakin mudah dibuat, semakin sulit dideteksi, dan semakin luas dampaknya.

Rangkuman Bab

Bab ini telah menganalisis empat kelompok kasus deepfake Indonesia yang terdokumentasi, dengan temuan-temuan kunci sebagai berikut:

- Kasus penipuan deepfake Prabowo (2025) menunjukkan bahwa Indonesia telah memiliki sindikat kejahatan terorganisir yang memanfaatkan deepfake sebagai alat penipuan massal menarget warga yang paling rentan secara ekonomi dan literasi digital.
- Kasus deepfake pornografi UNUD (2025) mengungkap ancaman yang lebih diam-diam namun sama berbahayanya: kekerasan seksual berbasis AI yang menarget perempuan menggunakan teknologi bot Telegram yang sangat mudah diakses.

- Data OJK/IASC menunjukkan kerugian akibat AI-fraud di sektor keuangan mencapai Rp9,1 triliun dalam 14 bulan angka yang menunjukkan skala ancaman deepfake terhadap sistem keuangan nasional.
- Pemilu 2024 menjadi momen pertama di mana deepfake digunakan secara terdokumentasi dalam konteks politik Indonesia menarget figur paling berpengaruh dan memanipulasi narasi dalam iklim politik yang sangat terpolarisasi.
- Hambatan sistemik dari sisi korban (stigma, ketidaktahuan, biaya) dan dari sisi aparat (kapasitas forensik, kerangka hukum, yurisdiksi) menciptakan situasi di mana sebagian besar kasus deepfake tidak pernah mencapai tahap penuntutan.
- Peta kesenjangan hukum yang teridentifikasi menunjukkan bahwa Indonesia memerlukan reformasi regulasi yang komprehensif bukan sekadar penambalan pasal demi pasal.

BAB 5
HUKUM POSITIF INDONESIA ATAS DEEPFAKE SERTA REGULASI GLOBAL

Setiap tindak pidana baru menghadapi masalah yang sama: undang-undang selalu berjalan di belakang kejahatan. Hukum dirumuskan atas dasar pengalaman masa lalu, sementara kejahatan beroperasi di masa kini dan mengeksploitasi celah masa depan. Deepfake adalah perwujudan sempurna dari dilema ini, sebuah ancaman yang lahir dari teknologi yang bergerak jauh lebih cepat dari kapasitas legislatif mana pun di dunia.

Kekosongan regulasi bukan berarti kekosongan hukum secara total. Hukum positif Indonesia, meskipun tidak pernah dirancang untuk deepfake, memiliki sejumlah ketentuan yang dapat digunakan, dengan interpretasi yang tepat, untuk menjangkau berbagai bentuk penyalahgunaan deepfake. Bab ini memetakan secara komprehensif pasal-pasal tersebut, menganalisis kekuatan dan kelemahannya, kemudian membandingkannya dengan pendekatan regulasi yang telah dilakukan oleh negara-negara lain, untuk mengidentifikasi pelajaran yang dapat dipetik bagi pembaruan hukum Indonesia. Pemahaman tentang kerangka hukum ini adalah prasyarat bagi akademisi, praktisi, maupun pembuat kebijakan yang ingin berkontribusi pada respons hukum yang efektif, berkeadilan, dan tidak menghambat inovasi yang bermanfaat.

5.1 KERANGKA ANALISIS: TIGA DIMENSI REGULASI DEEPFAKE

Sebelum menelaah pasal demi pasal, penting untuk membangun kerangka analisis yang jelas. Regulasi deepfake yang efektif harus beroperasi pada tiga dimensi secara bersamaan. Dimensi pertama adalah kriminalisasi, yang menjawab pertanyaan tentang perbuatan apa yang dilarang dan siapa yang dapat dipidana. Instrumen yang relevan mencakup hukum pidana seperti KUHP, UU ITE, UU TPKS, dan UU Pornografi.

Dimensi kedua adalah perlindungan korban, yang menyangkut hak apa yang dimiliki korban dan mekanisme apa yang tersedia untuk pemulihan. Instrumen yang relevan mencakup hukum perdata berupa gugatan ganti rugi, mekanisme takedown, serta UU PDP dan UU TPKS. Dimensi ketiga adalah tanggung jawab platform, yang mempertanyakan sejauh mana platform digital bertanggung jawab atas konten deepfake yang beredar. Instrumen yang relevan mencakup PP PTSE, hukum konsumen, dan hukum perdata.

Kerangka tiga dimensi ini akan digunakan sepanjang bab, baik dalam menganalisis hukum Indonesia maupun dalam membandingkan regulasi global. Pendekatan ini penting karena regulasi yang hanya fokus pada kriminalisasi tanpa memperhatikan perlindungan korban dan tanggung jawab platform akan tetap tidak efektif.

5.2 HUKUM POSITIF INDONESIA: ANALISIS PASAL PER PASAL

Berikut adalah analisis sistematis atas seluruh ketentuan hukum positif Indonesia yang relevan dengan penyalahgunaan deepfake, diurutkan berdasarkan hierarki dan relevansinya.

Undang-Undang Nomor 1 Tahun 2024 tentang Perubahan Kedua UU ITE

UU ITE (UU No. 11/2008 jo. UU No. 19/2016 jo. UU No. 1/2024) adalah tulang punggung utama penegakan hukum siber di Indonesia. Meski tidak menyebut deepfake secara eksplisit, beberapa pasalnya dapat digunakan secara analogi.

Pasal 27 ayat (1)

Pasal 27 ayat (1) melarang setiap orang dengan sengaja dan tanpa hak menyiarkan, mempertunjukkan, mendistribusikan, mentransmisikan, dan/atau membuat dapat diaksesnya informasi elektronik yang memiliki muatan yang melanggar kesusilaan untuk diketahui umum. Pasal ini adalah yang paling sering digunakan untuk menjerat penyebaran deepfake pornografi atau *Non-Consensual Intimate Images* (NCII). Sanksinya tercantum dalam Pasal 45 ayat (1), yakni pidana penjara paling lama 6 tahun dan/atau denda paling banyak Rp1 miliar.

Kekuatan pasal ini terletak pada ancaman pidana yang cukup berat dan kemampuannya diterapkan pada distributor serta pengunggah konten deepfake NCII. Namun, terdapat kelemahan kritis: pasal ini hanya menjangkau aspek distribusi dan transmisi, tidak menjerat pembuatan konten deepfake NCII. Selain itu, frasa "melanggar kesusilaan" bersifat multitafsir dan berpotensi digunakan untuk membatasi konten yang sesungguhnya dilindungi kebebasan berekspresi. Beban pembuktian unsur "tanpa hak" juga dapat menjadi tantangan dalam kasus deepfake yang menggunakan foto dari akun media sosial publik.

Pasal 27A (baru dalam UU No. 1/2024)

Pasal 27A melarang setiap orang dengan sengaja dan tanpa hak menyerang kehormatan atau nama baik orang lain dengan cara menuduhkan suatu hal, dengan maksud supaya hal tersebut diketahui umum dalam informasi elektronik atau dokumen elektronik. Pasal ini dapat digunakan untuk deepfake yang menciptakan konten yang merusak reputasi seseorang, misalnya deepfake yang menampilkan seseorang melakukan tindakan tercela yang tidak pernah dilakukannya.

Sanksinya adalah pidana penjara paling lama 2 tahun dan/atau denda paling banyak Rp400 juta. Kelemahan pasal ini adalah ancaman pidana yang relatif ringan dibandingkan dampak nyata deepfake. Frasa "menuduhkan suatu hal" juga memerlukan interpretasi hati-hati agar tidak mengkriminalisasi kritik atau satire politik yang sah.

Pasal 28 ayat (1)

Pasal 28 ayat (1) melarang penyebaran berita bohong dan menyesatkan yang mengakibatkan kerugian konsumen dalam transaksi elektronik. Pasal ini relevan untuk deepfake penipuan finansial, misalnya video deepfake pejabat yang menawarkan bantuan palsu seperti yang terjadi dalam kasus Prabowo 2025.

Sanksinya adalah pidana penjara paling lama 6 tahun dan/atau denda paling banyak Rp1 miliar. Unsur "mengakibatkan kerugian konsumen dalam transaksi elektronik" dapat dipenuhi dalam kasus penipuan deepfake di sektor fintech dan e-commerce, namun untuk deepfake disinformasi murni tanpa elemen penipuan finansial, pasal ini tidak langsung berlaku.

Pasal 28 ayat (2)

Pasal 28 ayat (2) melarang penyebaran informasi yang ditujukan untuk menimbulkan rasa kebencian atau permusuhan individu dan/atau kelompok masyarakat tertentu berdasarkan SARA. Pasal ini dapat digunakan untuk deepfake yang membawa narasi SARA dalam konteks politik atau konflik komunal, misalnya deepfake yang membuat tokoh publik seolah-olah menyatakan sentimen anti-kelompok tertentu.

Sanksinya adalah pidana penjara paling lama 6 tahun dan/atau denda Rp1 miliar. Perlu dicatat bahwa pasal ini telah menjadi subjek pengujian di Mahkamah Konstitusi melalui Putusan No. 105/PUU-XXII/2024 terkait potensi pembatasannya terhadap kebebasan berekspresi, sebuah dimensi yang harus dipertimbangkan dalam setiap upaya perluasan cakupannya untuk menjangkau deepfake.

Pasal 35

Di antara seluruh pasal dalam UU ITE, Pasal 35 adalah yang paling langsung relevan dengan karakteristik deepfake. Pasal ini melarang setiap orang dengan sengaja dan tanpa hak atau melawan hukum melakukan manipulasi, penciptaan, perubahan, penghilangan, atau perusakan informasi elektronik dan/atau dokumen elektronik dengan tujuan agar informasi atau dokumen elektronik tersebut dianggap seolah-olah data yang otentik.

Kata "manipulasi" dan "penciptaan" dalam pasal ini mencakup pembuatan deepfake yang tujuannya membuat orang percaya konten palsu tersebut nyata. Sanksinya adalah pidana penjara paling lama 12 tahun dan/atau denda paling banyak Rp12 miliar, yang merupakan ancaman tertinggi dalam UU ITE. Pasal ini mencakup tidak hanya distribusi tetapi juga pembuatan konten manipulatif, dan telah digunakan secara aktif oleh Bareskrim Polri dalam kasus deepfake Prabowo 2025.

Kelemahannya adalah unsur "dengan tujuan agar dianggap otentik" memerlukan pembuktian niat (mens rea) yang spesifik. Deepfake yang dibuat "untuk kesenangan" tanpa niat menipu dapat berargumen tidak memenuhi unsur ini. Selain itu, pasal ini tidak secara eksplisit menyebut teknologi AI atau deepfake sehingga interpretasi hakim menjadi penentu.

Undang-Undang Nomor 12 Tahun 2022 tentang Tindak Pidana Kekerasan Seksual (UU TPKS)

UU TPKS adalah terobosan hukum yang paling relevan untuk deepfake NCII dan konten seksual berbasis AI tanpa persetujuan korban. Pasal 14 ayat (1) memberikan perlindungan yang paling progresif untuk korban deepfake NCII. Pasal ini melarang setiap orang yang tanpa hak melakukan perekaman dan/atau pengambilan gambar atau tangkapan layar yang bermuatan seksual di luar kehendak atau tanpa persetujuan orang yang menjadi objeknya. Sanksinya adalah pidana penjara 1 hingga 4 tahun dan/atau denda Rp10 juta hingga Rp200 juta.

Kekuatan pasal ini terletak pada perlindungan eksplisit atas kehendak dan persetujuan sebagai unsur kunci. Pasal 14 ayat (2) melarang penyebaran rekaman bermuatan seksual tanpa persetujuan, sementara Pasal 14 ayat (3) melarang ancaman atau paksaan berbasis konten seksual (*sextortion*).

Namun terdapat kelemahan kritis untuk konteks deepfake. Frasa "perekaman" dan "pengambilan gambar" secara harfiah mengacu pada dokumentasi kejadian nyata, bukan

penciptaan konten sintetis AI dari foto yang ada. Dalam kasus UNUD 2025, pelaku tidak merekam korban, melainkan mengambil foto Instagram lalu membuat deepfake menggunakan bot AI. Apakah hal ini masuk dalam definisi Pasal 14 masih diperdebatkan.

Diperlukan interpretasi progresif atau amandemen eksplisit untuk secara pasti mencakup deepfake NCII. Beberapa akademisi hukum, antara lain dalam Jurnal Hukum Respublica 2024 dan UIR Law Review 2025, merekomendasikan penambahan ayat atau pasal baru yang secara eksplisit mengkriminalisasi pembuatan konten seksual sintetis berbasis AI dari gambar seseorang tanpa persetujuan.

Undang-Undang Nomor 27 Tahun 2022 tentang Perlindungan Data Pribadi (UU PDP)

UU PDP memberikan perspektif yang berbeda dalam memandang deepfake, bukan dari sudut konten yang dihasilkan, tetapi dari sudut penggunaan data pribadi seperti wajah dan suara seseorang tanpa izin. Data biometrik, termasuk citra wajah dan karakteristik suara, merupakan "data pribadi yang bersifat spesifik" berdasarkan Pasal 4 ayat (2) UU PDP. Dengan demikian, penggunaan wajah seseorang untuk membuat deepfake tanpa izin dapat dikualifikasikan sebagai pelanggaran UU PDP.

Pasal 65 melarang pengumpulan data pribadi yang bukan miliknya secara melawan hukum, sementara Pasal 66 melarang pemalsuan data pribadi yang dapat mengakibatkan kerugian bagi orang lain. Pasal 66 jo. Pasal 68 menetapkan ancaman pidana penjara paling lama 6 tahun dan/atau denda paling banyak Rp6 miliar untuk pemalsuan data pribadi termasuk citra biometrik seseorang.

Kekuatan UU PDP terletak pada kerangka perlindungan privasi yang lebih luas, melampaui sekadar konten, serta pengakuan data biometrik sebagai kategori perlindungan yang lebih ketat. Kelemahannya adalah Komisi PDP sebagai badan pengawas baru dibentuk dan kapasitasnya masih berkembang. Selain itu, terdapat pertanyaan apakah ketentuan ini menjangkau deepfake yang menggunakan data dari sumber publik, misalnya foto Instagram yang diunggah oleh pemilik sendiri.

Undang-Undang Nomor 44 Tahun 2008 tentang Pornografi

UU Pornografi mencakup "pembuatan" konten, berbeda dari UU ITE yang lebih fokus pada distribusi. Pasal 4 ayat (1) melarang setiap orang memproduksi, membuat, memperbanyak, menggandakan, menyebarluaskan, atau menyiarkan pornografi. Sanksinya adalah pidana penjara 6 bulan hingga 12 tahun dan/atau denda Rp250 juta hingga Rp6 miliar. Terdapat perdebatan akademis tentang apakah deepfake NCII dapat disebut "pornografi" dalam undang-undang ini.

Argumen yang mendukung menyatakan bahwa konten deepfake NCII memenuhi definisi "gambar yang memuat kecabulan" dan "eksploitasi seksual yang melanggar norma kesusilaan." Argumen yang menolak menyatakan bahwa definisi pornografi dalam UU ini tidak secara eksplisit mencakup konten sintetis buatan AI yang tidak merekam kejadian seksual nyata namun mensimulasikannya. Perlu juga dicatat bahwa UU Pornografi memiliki ketentuan yang berpotensi digunakan secara represif terhadap konten seni atau pendidikan.

KUHP Baru (UU Nomor 1 Tahun 2023)

KUHP Baru yang berlaku sejak 2 Januari 2026 memuat beberapa ketentuan yang relevan dengan deepfake. Pasal 263 hingga 264 tentang pemalsuan surat relevan untuk deepfake dokumen seperti KTP palsu, ijazah palsu, kontrak palsu, dan dokumen lainnya. Pasal 492 tentang penipuan digital relevan untuk penipuan deepfake finansial, dengan ancaman pidana penjara hingga 4 tahun atau denda maksimal Rp500 juta (Kategori V).

KUHP Baru tidak secara eksplisit menyebut deepfake atau konten sintetis AI. Namun sebagai kodifikasi hukum pidana umum yang berlaku penuh mulai 2026, ketentuan-ketentuan ini akan menjadi rujukan utama bagi penuntut dalam kasus yang tidak secara spesifik diatur oleh UU khusus.

Sintesis: Tiga Kekosongan Hukum Fundamental

Dari analisis di atas, tiga kekosongan hukum fundamental teridentifikasi dalam hukum positif Indonesia. Pertama, tidak ada satu pun undang-undang Indonesia yang mendefinisikan "deepfake," "konten sintetis berbasis AI," atau "manipulasi identitas digital." Ketiadaan definisi normatif ini menciptakan ketidakpastian hukum yang melemahkan seluruh upaya penuntutan.

Kedua, hukum positif Indonesia lebih kuat dalam menjerat penyebar konten deepfake daripada pembuatnya. Dalam era deepfake-as-a-service di mana pembuat dan distributor adalah entitas berbeda, ini adalah kelemahan yang sangat serius. Ketiga, tidak ada kewajiban hukum yang jelas bagi platform digital untuk mendeteksi, melabeli, atau menghapus konten deepfake secara proaktif. PP PTSE hanya mewajibkan respons terhadap laporan, bukan pencegahan aktif.

5.3 PERBANDINGAN REGULASI DEEPFAKE GLOBAL

Berbagai negara telah merespons ancaman deepfake dengan pendekatan yang berbeda-beda, mencerminkan perbedaan tradisi hukum, nilai-nilai konstitusional, dan kapasitas regulasi masing-masing. Analisis komparatif ini bertujuan mengidentifikasi elemen-elemen terbaik yang dapat dipelajari dan diadaptasi untuk konteks Indonesia.

Uni Eropa: EU Artificial Intelligence Act 2024

Uni Eropa menetapkan standar regulasi AI yang paling komprehensif melalui Regulation (EU) 2024/1689 atau EU AI Act yang mulai berlaku Agustus 2024 dengan pemberlakuan penuh pada 2 Agustus 2026. Pendekatan yang digunakan adalah berbasis risiko, mengklasifikasikan sistem AI berdasarkan tingkat risikonya dari yang tidak dapat diterima hingga minimal.

Ketentuan khusus deepfake termuat dalam Pasal 50 EU AI Act yang mewajibkan transparansi: sistem AI yang menghasilkan konten sintetis wajib menandai konten tersebut dengan cara yang dapat dideteksi secara teknis, baik melalui watermark yang dapat dibaca mesin maupun metadata. Konten deepfake untuk tujuan seni, hiburan, atau satire dikecualikan dengan syarat dilabeli secara jelas. Regulasi ini dikombinasikan dengan EU Digital Services Act 2022 yang mewajibkan platform besar melakukan penilaian risiko dan mitigasi terhadap konten berbahaya termasuk deepfake.

Kekuatan pendekatan Uni Eropa terletak pada cakupannya yang komprehensif meliputi seluruh siklus hidup AI, pendekatan berbasis hak yang mengacu pada Piagam Hak Fundamental UE, serta kewajiban transparansi aktif dari penyedia AI. Kritiknya adalah kompleksitas yang berpotensi menghambat inovasi, tantangan penegakan lintas negara anggota, dan ketidakeksplisitan dalam mengkriminalisasi deepfake NCII karena bergantung pada hukum nasional masing-masing negara.

Amerika Serikat: Regulasi Federal dan Negara Bagian

Amerika Serikat tidak memiliki undang-undang federal tunggal tentang deepfake. Regulasi berkembang di tingkat negara bagian, dengan beberapa Rancangan Undang-Undang federal yang sedang dalam proses legislatif per April 2026. Di tingkat federal, DEFIANCE Act (*Disrupt Explicit Forged Images and Non-Consensual Edits Act*) lolos Senat secara bulat pada Januari 2026 dan memberikan hak gugatan sipil bagi korban deepfake seksual non-konsensual dengan ganti rugi statutory hingga USD 150.000 atau USD 250.000 jika terkait kekerasan seksual, stalking, atau pelecehan.

Protect Elections from Deceptive AI Act yang diperkenalkan Maret 2025 melarang distribusi materi AI yang menyesatkan tentang kandidat dalam pemilu federal. Adapun NO FAKES Act yang diperkenalkan April 2025 melarang pembuatan atau distribusi replika AI atas suara atau wajah seseorang tanpa persetujuan, dengan pengecualian untuk satire, komentar, dan jurnalisme. Di tingkat negara bagian, California AB 602 dan AB 730 tahun 2019 melarang deepfake seksual dan deepfake pemilu. California AB 2655 tahun 2024 mewajibkan platform menghapus konten AI manipulatif tentang pemilu, meskipun sebagian dibatalkan secara konstitusional pada Agustus 2025. Tennessee ELVIS Act tahun 2024 memberikan perlindungan khusus terhadap penggandaan suara tanpa izin, sementara New York mewajibkan persetujuan tertulis untuk replika digital.

Kekuatan pendekatan Amerika Serikat adalah DEFIANCE Act yang menciptakan hak gugatan sipil langsung bagi korban, pendekatan berbasis hak individu, serta pengecualian untuk satire dan komentar demi menjaga kebebasan berekspresi. Kritiknya adalah fragmentasi regulasi antar negara bagian, lambatnya legislasi federal, dan ketegangan dengan Amandemen Pertama tentang kebebasan berbicara.

China: Deep Synthesis Provisions 2022 dan Labeling Rules 2025

China memiliki regulasi deepfake yang paling eksplisit secara teknis di dunia. Instrumen hukum utamanya adalah "*Provisions on the Administration of Deep Synthesis Internet Information Services*" yang berlaku Januari 2023, dikeluarkan bersama oleh Cyberspace Administration of China, Kementerian Industri, dan Kementerian Keamanan Publik. Regulasi ini dilengkapi dengan "*Measures for Labeling of AI-Generated Synthetic Content*" yang berlaku September 2025.

Ketentuan utamanya mencakup kewajiban pelabelan atas semua konten yang dihasilkan teknologi deepfake atau AI, larangan penggunaan untuk disinformasi, serta tanggung jawab penyedia layanan untuk melakukan verifikasi identitas pengguna dan menyimpan log penggunaan. Hak persetujuan diakui secara eksplisit: pembuatan deepfake yang menampilkan seseorang tanpa persetujuan dilarang.

Model China, terutama kewajiban pelabelan dan tanggung jawab platform, dinilai oleh beberapa akademisi hukum Indonesia dalam Jurnal USM Law Review 2024 sebagai referensi yang paling relevan secara teknis. Namun perlu selektif dalam mengadaptasinya: mekanisme teknisnya dapat diambil, tetapi pendekatannya terhadap kebebasan berekspresi tidak dapat diadopsi begitu saja mengingat perbedaan sistem hukum dan nilai-nilai konstitusional antara kedua negara.

Korea Selatan: Regulasi Deepfake Terkomprehensif di Asia

Korea Selatan menjadi salah satu negara yang merespons paling cepat dan komprehensif terhadap ancaman deepfake, sebagian karena gelombang kasus deepfake NCII yang menarget artis K-Pop dan perempuan biasa yang menjadi viral. Melalui amandemen Act on Special Cases concerning the Punishment of Sexual Crimes pada tahun 2020, Korea Selatan mengkriminalisasi secara eksplisit pembuatan dan distribusi deepfake seksual tanpa persetujuan dengan sanksi pidana penjara hingga 5 tahun dan/atau denda hingga 50 juta won.

Revisi Public Official Election Act pada Januari 2024 melarang deepfake yang menampilkan kandidat dalam kampanye pemilu dalam 90 hari sebelum pemungutan suara dengan sanksi penjara hingga 7 tahun dan/atau denda 50 juta won. Pembuat deepfake wajib menginformasikan bahwa konten mengandung sintetis AI bahkan untuk konten yang dibuat sebelum periode 90 hari. Model Korea Selatan tentang larangan deepfake pemilu 90 hari sebelum hari pencoblosan adalah mekanisme yang sangat relevan dan dapat diadopsi dalam amandemen UU Pemilu Indonesia.

Inggris: Online Safety Act 2023 dan Reformasi 2025

Inggris menggunakan Online Safety Act 2023 sebagai instrumen utama, yang melarang pembagian atau ancaman pembagian gambar intim deepfake tanpa persetujuan. Amandemen 2025 yang sedang dalam proses mengkriminalisasi secara langsung pembuatan gambar seksual deepfake yang eksplisit tanpa persetujuan, dengan niat menyebabkan ketakutan, distres, atau untuk kepuasan seksual, dengan ancaman hingga 2 tahun penjara.

Pendekatan Inggris yang mengkriminalisasi "niat" (intent), bukan hanya tindakan, memberikan model yang berguna untuk menjerat deepfake yang dibuat "untuk kesenangan pribadi" namun berpotensi disebarkan, bahkan sebelum penyebaran aktual terjadi.

5.4 PELAJARAN DARI PERBANDINGAN GLOBAL

Dari analisis komparatif di atas, beberapa pelajaran kunci dapat dipetik untuk menginformasikan pembaruan regulasi deepfake di Indonesia. Regulasi deepfake yang efektif harus memenuhi beberapa prinsip fundamental.

- Pertama, presisi definitif: regulasi harus mendefinisikan secara jelas apa yang dimaksud dengan "konten sintetis berbasis AI" dan "deepfake" untuk kepastian hukum. Definisi yang kabur akan selalu memberi celah bagi pelaku dan melemahkan dakwaan.
- Kedua, regulasi harus mencakup seluruh siklus pembuatan, distribusi, dan penggunaan deepfake. Memfokuskan hanya pada distribusi memberi celah bagi penyedia layanan deepfake yang menjual konten tanpa menyebarkannya.

- Ketiga, regulasi harus berbasis persetujuan (*consent-based*) di mana persetujuan dari subjek yang wajahnya atau suaranya digunakan menjadi unsur sentral, bukan sekadar "tanpa hak" yang lebih ambigu.
- Keempat, ancaman pidana harus proporsional dengan kerugian yang ditimbulkan.
- Kelima, satire, parodi, dan ekspresi artistik harus mendapat perlindungan eksplisit.
- Keenam, platform harus memiliki kewajiban aktif untuk mendeteksi dan melabeli konten deepfake, khususnya dalam konteks berita dan politik.
- Ketujuh, mekanisme pemulihan bagi korban harus mudah diakses, terjangkau, dan cepat.

Berdasarkan analisis tersebut, arsitektur regulasi yang direkomendasikan untuk Indonesia dalam jangka pendek adalah amandemen UU ITE untuk menambahkan definisi deepfake dan mengeksplisitkan kriminalisasi pembuatan konten sintetis manipulatif, amandemen UU TPKS untuk mencakup pembuatan konten seksual sintetis AI tanpa persetujuan, serta amandemen UU Pemilu untuk melarang distribusi deepfake tentang kandidat dalam jangka waktu tertentu sebelum pemilu. Dalam jangka panjang, Indonesia perlu mempersiapkan Undang-Undang Kecerdasan Buatan yang komprehensif dengan mengacu pada EU AI Act sebagai model dan menyesuaikannya dengan kapasitas institusional Indonesia.

5.5 TANTANGAN IMPLEMENTASI REGULASI DEEPFAKE DI INDONESIA

Memiliki regulasi yang baik adalah syarat perlu, namun bukan syarat cukup. Implementasi regulasi deepfake di Indonesia menghadapi beberapa tantangan struktural yang harus diantisipasi.

Tantangan Kapasitas Teknis Penegak Hukum

Mengidentifikasi apakah sebuah konten adalah deepfake memerlukan keahlian forensik digital yang sangat spesifik. Bahkan dengan regulasi yang sempurna, penegakan hukum akan tetap lemah tanpa laboratorium forensik digital di setiap Polda yang dilengkapi perangkat deteksi deepfake terkini, pelatihan berkelanjutan bagi penyidik Dittipidsiber dan jaksa penuntut tentang teknologi AI, protokol standar untuk penyitaan dan analisis bukti digital berbasis AI, serta kerjasama dengan perguruan tinggi dan lembaga riset untuk mengisi kesenjangan keahlian teknis.

Tantangan Kerja Sama Platform Internasional

Sebagian besar deepfake di Indonesia disebarkan melalui platform yang beroperasi di bawah yurisdiksi asing seperti Meta, TikTok, Telegram, dan YouTube. Penegakan hukum memerlukan mekanisme permintaan data yang dipercepat, tidak hanya melalui MLAT yang bisa memakan waktu berbulan-bulan, serta kesepakatan bilateral atau multilateral tentang kerja sama penegakan hukum siber dengan negara-negara di mana platform-platform utama berkantor pusat.

Tantangan Keseimbangan Kebebasan Berekspresi

Regulasi deepfake yang terlalu luas berpotensi menghambat ekspresi seni, satire politik, jurnalisme investigatif, dan pendidikan. Regulasi deepfake yang baru harus mendefinisikan secara eksplisit pengecualian untuk satire, parodi, edukasi, dan jurnalisme;

mensyaratkan unsur "niat merugikan" atau "niat menipu" sebagai elemen mens rea; serta memastikan mekanisme pelaporan tidak digunakan sebagai alat pelecehan terhadap konten kritik yang sah.

Tantangan Literasi Hukum dan Digital Masyarakat

Regulasi terbaik sekalipun tidak akan efektif jika masyarakat tidak tahu bagaimana menggunakannya. Diperlukan investasi paralel dalam literasi digital tentang deepfake, literasi hukum tentang hak-hak digital dan mekanisme pelaporan, serta dukungan psikologis dan hukum yang mudah diakses bagi korban deepfake.

Rangkuman Bab

Hukum positif Indonesia memiliki beberapa ketentuan yang dapat digunakan untuk menjangkau penyalahgunaan deepfake, terutama UU ITE Pasal 35 tentang manipulasi konten, UU TPKS Pasal 14 tentang kekerasan seksual berbasis elektronik, UU PDP Pasal 66 tentang pemalsuan data pribadi biometrik, dan UU Pornografi Pasal 4. Namun tiga kekosongan fundamental tetap ada: tidak ada definisi hukum untuk deepfake; kriminalisasi lebih kuat untuk distribusi daripada pembuatan; dan tanggung jawab platform belum diatur secara memadai.

Secara komparatif, Uni Eropa menetapkan standar tertinggi dengan EU AI Act yang komprehensif dan berbasis risiko. China memiliki regulasi yang paling eksplisit secara teknis. Korea Selatan paling responsif terhadap ancaman deepfake seksual dan pemilu. Amerika Serikat bergerak cepat dengan DEFIANCE Act untuk perlindungan korban.

Untuk Indonesia, pendekatan jangka pendek terbaik adalah amandemen bertahap pada UU ITE, UU TPKS, dan UU Pemilu, sambil mempersiapkan Undang-Undang AI Nasional yang komprehensif untuk jangka panjang. Regulasi saja tidak cukup: implementasi yang efektif memerlukan kapasitas forensik teknis, kerja sama platform internasional, keseimbangan kebebasan berekspresi, dan investasi literasi digital masyarakat.

BAB 6
REGULASI YANG ADA: UU ITE, KUHP, UU PDP, DAN UU TPKS

Indonesia tidak memiliki undang-undang khusus yang menyebut kata "deepfake." Tidak ada pasal yang secara tegas menyatakan bahwa membuat atau menyebarkan video manipulasi berbasis kecerdasan buatan adalah tindak pidana. Inilah kenyataan hukum yang harus dihadapi siapa pun yang ingin memahami bagaimana sistem hukum Indonesia merespons ancaman ini.

Kenyataan itu tidak berarti hukum Indonesia sepenuhnya buta terhadap deepfake. Sejumlah undang-undang yang sudah berlaku memuat ketentuan yang, dengan interpretasi yang cermat, bisa digunakan untuk menjangkau berbagai bentuk penyalahgunaan deepfake. Yang perlu dipahami adalah bagaimana masing-masing ketentuan itu bekerja, apa batas kemampuannya, dan di mana letak celah yang belum tertutupi.

Bab ini membahas enam kerangka hukum yang paling relevan, mulai dari UU ITE, KUHP baru, UU Perlindungan Data Pribadi, UU Tindak Pidana Kekerasan Seksual, UU Hak Cipta, hingga ketentuan perbuatan melawan hukum dalam KUHPerdata. Setiap kerangka hukum dibahas secara terinci: pasal yang berlaku, bagaimana penerapannya pada kasus deepfake, dan apa kelemahannya.

6.1 UU ITE: TONGGAK UTAMA HUKUM SIBER INDONESIA

Posisi UU ITE dalam Penegakan Hukum Deepfake

Undang-Undang Informasi dan Transaksi Elektronik adalah hukum siber utama Indonesia. Sejak disahkan pertama kali pada 2008, diperbaharui pada 2016, lalu diubah kembali secara substansial melalui UU No. 1 Tahun 2024, UU ITE menjadi instrumen yang paling sering digunakan oleh penyidik ketika berhadapan dengan kejahatan di ruang digital, termasuk kasus-kasus deepfake yang mulai masuk ke meja Bareskrim sejak 2024.

UU ITE tidak menyebut deepfake, kecerdasan buatan, atau konten sintetis dalam satu pasal pun. Tetapi memuat sejumlah larangan yang, secara substansi, dapat mencakup berbagai bentuk penyalahgunaan deepfake. Pemahaman tentang pasal-pasal ini penting karena dalam praktik penegakan hukum saat ini, merekalah yang menjadi andalan.

Pasal 27: Konten Asusila dan Perlindungan Kesusilaan

Pasal 27 ayat (1) melarang setiap orang dengan sengaja dan tanpa hak menyiarkan, mempertunjukkan, mendistribusikan, mentransmisikan, dan/atau membuat dapat diaksesnya informasi elektronik yang memiliki muatan yang melanggar kesusilaan untuk diketahui umum. Pasal ini adalah pasal yang paling sering digunakan untuk menjerat penyebaran deepfake pornografi. Ketika seseorang menyebarkan video deepfake yang menampilkan wajah orang lain dalam konteks seksual tanpa persetujuan, baik melalui Telegram, media sosial, maupun platform lain, tindakan itu dapat dijerat dengan pasal ini. Sanksinya adalah pidana penjara paling lama 6 tahun dan/atau denda paling banyak Rp1 miliar.

Kekuatan pasal ini terletak pada ancaman pidana 6 tahun yang termasuk berat, memberikan ruang bagi hakim untuk menjatuhkan hukuman yang sepadan dengan dampak yang dialami korban. Kelemahannya adalah pasal ini hanya menjangkau tindakan penyebaran, bukan pembuatan. Orang yang membuat deepfake pornografi namun belum menyebarkannya tidak dapat langsung dijerat pasal ini. Selain itu, pasal ini tidak membedakan antara konten asli dengan konten manipulasi AI, sehingga unsur "tanpa hak" harus dibuktikan dengan cermat oleh jaksa.

Pasal 27A yang baru dalam UU No. 1/2024 melarang setiap orang dengan sengaja dan tanpa hak menyerang kehormatan atau nama baik orang lain dengan cara menuduhkan suatu hal dengan maksud supaya hal tersebut diketahui umum dalam informasi elektronik atau dokumen elektronik. Pasal ini dapat diterapkan untuk deepfake yang menampilkan seseorang seolah-olah melakukan sesuatu yang merusak nama baiknya, misalnya video yang menampilkan tokoh publik seolah menerima suap. Sanksinya adalah pidana penjara paling lama 2 tahun dan/atau denda paling banyak Rp400 juta. Ancaman pidana Pasal 27A yang lebih ringan dibanding Pasal 27 ayat (1) menjadi pertimbangan strategis bagi jaksa dalam menentukan dakwaan utama.

Pasal 28: Penyebaran Berita Bohong dan Konten Provokatif

Pasal 28 ayat (1) melarang penyebaran berita bohong dan menyesatkan yang mengakibatkan kerugian konsumen dalam transaksi elektronik. Pasal ini relevan untuk kasus deepfake penipuan finansial, misalnya video deepfake pejabat yang mengumumkan program bantuan pemerintah palsu seperti yang terjadi dalam kasus Prabowo 2025.

Korban yang mentransfer uang karena percaya video tersebut asli mengalami kerugian dalam transaksi elektronik yang memenuhi unsur pasal ini. Sanksinya adalah pidana penjara paling lama 6 tahun dan/atau denda paling banyak Rp1 miliar. Frasa "mengakibatkan kerugian konsumen dalam transaksi elektronik" membatasi cakupan pasal ini sehingga deepfake disinformasi murni yang tidak berkaitan dengan transaksi keuangan tidak otomatis masuk kategori ini.

Pasal 28 ayat (2) melarang penyebaran informasi yang ditujukan untuk menimbulkan rasa kebencian atau permusuhan individu dan/atau kelompok masyarakat tertentu berdasarkan SARA. Deepfake yang dirancang untuk menguatkan narasi SARA, misalnya video manipulasi yang membuat seorang tokoh agama seolah mengucapkan pernyataan provokatif terhadap kelompok lain, dapat dijerat dengan pasal ini.

Sanksinya adalah pidana penjara paling lama 6 tahun dan/atau denda paling banyak Rp1 miliar. Pasal 28 ayat (2) telah menjadi pasal yang kontroversial karena potensinya untuk membatasi kebebasan berekspresi, sehingga jaksa perlu membuktikan adanya "maksud menimbulkan kebencian" dan bukan sekadar konten yang bersifat sensitif.

Pasal 35: Manipulasi Dokumen Elektronik

Di antara seluruh pasal dalam UU ITE, Pasal 35 adalah yang paling langsung relevan dengan karakteristik deepfake. Pasal ini bahkan telah digunakan secara nyata oleh Bareskrim Polri dalam menangani kasus deepfake penipuan Prabowo pada Januari dan Februari 2025.

Pasal 35 melarang setiap orang dengan sengaja dan tanpa hak atau melawan hukum melakukan manipulasi, penciptaan, perubahan, penghilangan, atau perusakan informasi elektronik dan/atau dokumen elektronik dengan tujuan agar informasi atau dokumen elektronik tersebut dianggap seolah-olah data yang otentik. Kata "manipulasi" dan "penciptaan" dalam pasal ini menjadi frasa kunci.

Deepfake pada dasarnya adalah penciptaan konten yang dirancang agar tampak autentik, persis yang dilarang pasal ini. Berbeda dari pasal-pasal lain yang fokus pada dampak seperti asusila, kebencian, dan kerugian, Pasal 35 fokus pada tindakan manipulasi itu sendiri. Sanksinya adalah pidana penjara paling lama 12 tahun dan/atau denda paling banyak Rp12 miliar, yang merupakan ancaman tertinggi dalam UU ITE.

Dalam kasus deepfake Prabowo 2025, Bareskrim menjerat tersangka dengan kombinasi Pasal 35 jo. Pasal 51 UU ITE dan Pasal 378 KUHP. Kombinasi dakwaan ini mencerminkan upaya jaksa untuk menjangkau dua dimensi sekaligus: manipulasi konten digital dan kerugian finansial. Kelemahannya adalah unsur "dengan tujuan agar dianggap seolah-olah data yang otentik" mensyaratkan pembuktian niat. Jika pembuat deepfake berdalih bahwa konten dibuat "untuk kesenangan" atau "hanya bereksperimen" tanpa niat menipu, jaksa harus mampu membantah argumen itu dengan bukti yang kuat.

6.2 KUHP BARU (UU NO. 1 TAHUN 2023, BERLAKU JANUARI 2026)

Konteks: Mengapa KUHP Baru Penting

Kitab Undang-Undang Hukum Pidana yang berlaku di Indonesia selama lebih dari satu abad akhirnya digantikan oleh KUHP Baru melalui UU No. 1 Tahun 2023, yang resmi berlaku penuh pada 2 Januari 2026. Sebagai kodifikasi hukum pidana umum yang menyeluruh, KUHP Baru memuat sejumlah ketentuan yang relevan dengan kejahatan berbasis teknologi meskipun tetap tidak menyebut deepfake secara eksplisit.

Perlu dipahami posisi KUHP Baru dalam hierarki hukum. KUHP Baru adalah hukum pidana umum (*lex generalis*), sedangkan UU ITE, UU TPKS, dan UU PDP adalah hukum pidana khusus (*lex specialis*). Dalam praktik penegakan hukum, asas lex specialis derogat legi generali berlaku, artinya UU khusus didahulukan. Namun KUHP Baru tetap relevan sebagai pelengkap ketika hukum khusus tidak menjangkau aspek tertentu dari suatu kasus.

Pasal 492: Penipuan Digital

Pasal 492 KUHP Baru mengatur penipuan dengan frasa "tipu muslihat" dan "rangkaian kebohongan" yang secara natural mencakup penggunaan video deepfake untuk meyakinkan korban bahwa mereka berinteraksi dengan orang asli. Pasal ini secara substansial sama dengan Pasal 378 KUHP lama, namun dengan penyesuaian redaksional dan sistem denda berbasis kategori. Dalam kasus deepfake penipuan finansial, pasal ini digunakan bersama UU ITE Pasal 35 untuk dakwaan berlapis. Denda Kategori V berdasarkan Pasal 79 KUHP Baru adalah Rp500 juta dengan ancaman penjara paling lama 4 tahun.

Ancaman 4 tahun penjara terasa kurang sepadan dibandingkan dengan kerugian yang bisa ditimbulkan oleh sindikat deepfake berskala besar. Inilah salah satu alasan mengapa

dakwaan berlapis dengan UU ITE Pasal 35 yang mengancam 12 tahun penjara menjadi strategi yang lebih efektif.

Pasal 263-264: Pemalsuan Dokumen

Pasal 263 hingga 264 KUHP Baru adalah dasar hukum utama untuk kasus deepfake dokumen, yaitu pembuatan KTP palsu, ijazah palsu, kontrak palsu, atau surat kuasa palsu menggunakan AI. Ketika seseorang membuat KTP palsu menggunakan AI generatif untuk keperluan penipuan KYC perbankan, Pasal 263 KUHP Baru dapat diterapkan meskipun dokumen itu tidak pernah dicetak secara fisik, karena "surat" dalam konteks digital dapat diartikan sebagai dokumen elektronik berdasarkan UU ITE.

Ancaman pidananya adalah penjara paling lama 6 tahun untuk Pasal 263 dan 8 tahun untuk Pasal 264 tentang pemalsuan surat autentik. Tantangan interpretasinya adalah KUHP Baru tidak secara eksplisit menyatakan bahwa dokumen digital termasuk dalam cakupan Pasal 263-264. Interpretasi ini bergantung pada penggabungan dengan UU ITE yang mendefinisikan dokumen elektronik sebagai alat bukti yang sah.

6.3 UU PDP NO. 27 TAHUN 2022: WAJAH DAN SUARA SEBAGAI DATA PRIBADI

Perspektif Baru: Deepfake sebagai Pelanggaran Privasi Data

Undang-Undang Perlindungan Data Pribadi hadir dengan perspektif yang berbeda dari UU ITE maupun KUHP. Jika hukum pidana melihat deepfake dari sudut tindak pidana yang merugikan orang lain, UU PDP melihatnya dari sudut hak fundamental seseorang atas datanya sendiri, termasuk data biometrik seperti wajah dan suara.

Perspektif ini penting karena membuka jalur hukum yang lebih luas: seseorang yang wajahnya digunakan untuk membuat deepfake, bahkan jika deepfake tersebut belum disebarkan, sudah mengalami pelanggaran atas haknya atas data biometriknya. Ini adalah pendekatan yang jauh lebih proaktif dibanding menunggu kerugian nyata terjadi.

Data Biometrik dalam UU PDP

UU PDP membagi data pribadi menjadi dua kategori utama. Data pribadi umum mencakup nama, alamat, usia, pekerjaan, dan status pernikahan dengan perlindungan standar. Data pribadi yang bersifat spesifik mencakup data biometrik seperti wajah, suara, dan sidik jari, serta data kesehatan, data genetika, pandangan politik, dan keyakinan agama, dengan perlindungan yang lebih ketat dan persetujuan eksplisit yang wajib dipenuhi.

Relevansi dengan deepfake sangat jelas: ketika seseorang menggunakan foto atau rekaman suara orang lain untuk membuat konten deepfake tanpa izin, mereka memproses data biometrik orang tersebut secara melawan hukum, yang merupakan pelanggaran langsung terhadap UU PDP.

Pasal-Pasal UU PDP yang Relevan

Pasal 65 melarang setiap orang memperoleh atau mengumpulkan data pribadi yang bukan miliknya secara melawan hukum. Pasal ini berlaku ketika seseorang mengumpulkan foto atau rekaman suara seseorang secara diam-diam, misalnya melalui scraping foto Instagram, untuk keperluan membuat deepfake. Pasal 66 melarang pemalsuan data pribadi dan berlaku ketika seseorang menciptakan representasi palsu dari wajah atau suara seseorang melalui AI

deepfake. Sanksinya adalah pidana penjara paling lama 5 tahun dan/atau denda paling banyak Rp5 miliar untuk Pasal 65, serta 6 tahun dan/atau denda Rp6 miliar untuk Pasal 66.

Berbeda dari UU ITE yang mensyaratkan pembuktian konten yang sudah tersebar, UU PDP dapat digunakan bahkan ketika deepfake baru dalam tahap pembuatan atau pengumpulan data. Hal ini memberi ruang untuk tindakan pencegahan yang lebih dini oleh penegak hukum. Pasal 20-21 UU PDP memberikan hak kepada subjek data pribadi untuk menarik kembali persetujuan pemrosesan data pribadinya.

Pasal ini memberikan landasan bagi korban deepfake untuk meminta penghapusan data biometriknya dari sistem yang digunakan untuk membuat konten deepfake. Namun hak ini lebih mudah diterapkan terhadap platform resmi yang beroperasi secara legal di Indonesia. Untuk pelaku yang menggunakan alat deepfake anonim atau beroperasi dari luar negeri, mekanisme ini sulit dilaksanakan.

Perlu dicatat bahwa Komisi Perlindungan Data Pribadi yang diamanatkan UU PDP masih dalam proses penguatan kapasitas institusional per tahun 2025. Penegakan UU PDP dalam kasus deepfake saat ini masih banyak bergantung pada mekanisme pengaduan ke Komdigi dan penindakan oleh Bareskrim, bukan oleh Komisi PDP yang seharusnya menjadi garda terdepan.

6.4 UU TPKS NO. 12 TAHUN 2022: PERLINDUNGAN KORBAN KEKERASAN SEKSUAL

UU TPKS sebagai Terobosan Hukum

Undang-Undang Tindak Pidana Kekerasan Seksual adalah salah satu produk legislasi paling progresif Indonesia dalam beberapa tahun terakhir. Salah satu aspek yang paling relevan dengan deepfake adalah pengakuan eksplisit UU TPKS terhadap "kekerasan seksual berbasis elektronik" sebagai bentuk tindak pidana yang berdiri sendiri.

Sebelum UU TPKS ada, korban deepfake NCII harus menggantungkan harapan pada Pasal 27 UU ITE yang fokus pada penyebaran dan tidak mengakui dimensi "kekerasan seksual" dari tindak pidana itu. UU TPKS mengubah paradigma itu dengan mengakui bahwa pembuatan dan penyebaran konten seksual tanpa persetujuan adalah bentuk kekerasan seksual, bukan sekadar pelanggaran UU siber.

Pasal 14: Kekerasan Seksual Berbasis Elektronik

Pasal 14 UU TPKS memuat tiga bentuk tindak pidana. Ayat (1) melarang perekaman dan/atau pengambilan gambar atau tangkapan layar bermuatan seksual tanpa persetujuan dengan sanksi pidana penjara 1 hingga 4 tahun dan/atau denda Rp10 juta hingga Rp200 juta. Ayat (2) melarang penyiaran, pendistribusian, atau transmisi rekaman atau gambar bermuatan seksual tanpa persetujuan dengan sanksi yang sama. Ayat (3) melarang pengancaman atau pemaksaan menggunakan rekaman atau gambar bermuatan seksual (sextortion) dengan sanksi lebih berat berupa penjara hingga 6 tahun dan/atau denda hingga Rp300 juta.

UU TPKS Pasal 14 adalah instrumen terbaik yang tersedia untuk melindungi korban deepfake NCII. Namun ada perdebatan akademis yang belum selesai: frasa "perekaman" dan "pengambilan gambar" secara harfiah mengacu pada dokumentasi kejadian nyata. Deepfake tidak merekam korban, melainkan menciptakan gambaran sintetis dari wajah korban. Untuk penyebaran (ayat 2) dan sextortion (ayat 3), Pasal 14 UU TPKS lebih kuat karena tidak

mensyaratkan bagaimana konten itu dibuat, hanya bahwa konten itu disebarkan atau digunakan untuk mengancam tanpa persetujuan.

Pasal 5: Kekerasan Seksual Nonfisik

Pasal 5 UU TPKS dapat diterapkan ketika deepfake digunakan untuk melecehkan seseorang secara seksual tanpa kontak fisik, misalnya mengirimkan deepfake pornografi kepada korban secara langsung sebagai bentuk pelecehan. Ancaman pidananya adalah 9 bulan penjara dan/atau denda paling banyak Rp10 juta. Pasal ini memperluas cakupan perlindungan ke situasi yang tidak tercakup Pasal 14, yaitu ketika deepfake dikirim langsung kepada korban (bukan disebarkan publik) sebagai alat intimidasi.

Hak Korban dalam UU TPKS

Selain aspek kriminalisasi, UU TPKS memuat ketentuan penting tentang hak-hak korban yang relevan untuk konteks deepfake. Korban berhak mendapatkan informasi tentang perkembangan kasus yang dilaporkan, layanan kesehatan dan pemulihan psikologis yang disediakan negara, pendampingan hukum secara cuma-cuma dari lembaga bantuan hukum yang ditunjuk, perlindungan identitas agar tidak dipublikasikan tanpa izin, serta ganti rugi (restitusi) dari pelaku yang dapat dimintakan dalam proses pidana. Ketentuan-ketentuan ini sangat relevan bagi korban deepfake NCII yang sering kali mengalami re-traumatisasi ketika harus melalui proses hukum.

6.5 UU HAK CIPTA NO. 28 TAHUN 2014: DEEPFAKE DAN KARYA VISUAL SESEORANG

Dimensi Hak Cipta dalam Deepfake

Deepfake menimbulkan setidaknya dua persoalan dalam hukum hak cipta: apakah seseorang memiliki hak cipta atas penampilannya sendiri sehingga penggunaan tanpa izin melanggar hak cipta, dan siapa pemilik hak cipta atas deepfake yang dihasilkan AI. Dalam bab ini, fokus diberikan pada dimensi perlindungan korban yaitu hak seseorang atas penampilan visualnya dalam konteks hukum hak cipta Indonesia.

Hak Moral atas Karya dan Penampilan

Pasal 5 UU Hak Cipta memberikan hak moral kepada pencipta untuk mempertahankan integritas karyanya dari distorsi atau modifikasi yang merugikan kehormatan dan reputasinya. Dalam konteks deepfake yang menggunakan rekaman, foto, atau penampilan seseorang sebagai bahan dasar, ada argumen bahwa pembuatan deepfake merupakan "distorsi" atau "modifikasi" dari karya visual asli yang dilindungi hak cipta. Pasal 40 mengakui bahwa karya fotografi, sinematografi, dan penampilan termasuk dalam ciptaan yang dilindungi.

Sanksi pelanggaran hak moral berdasarkan Pasal 89 adalah pidana penjara paling lama 25 tahun dan/atau denda paling banyak Rp5 miliar, yang merupakan ancaman pidana tertinggi di antara semua instrumen hukum yang dibahas dalam bab ini. Namun hingga saat ini, belum ada kasus deepfake di Indonesia yang dijerat dengan UU Hak Cipta karena pasal-pasal UU ITE dan UU TPKS masih menjadi pilihan utama penyidik.

Pertanggungjawaban Perdata dalam UU Hak Cipta

Selain jalur pidana, UU Hak Cipta juga membuka jalur gugatan perdata bagi korban deepfake. Pasal 96 UU Hak Cipta menetapkan bahwa pencipta yang dirugikan dapat

mengajukan gugatan ganti rugi kepada pengadilan niaga. Gugatan perdata berbasis hak cipta memerlukan pembuktian bahwa materi yang digunakan dalam deepfake benar-benar merupakan ciptaan yang dilindungi dan bahwa korban adalah pemilik hak cipta atas materi tersebut, sebuah syarat yang tidak selalu mudah dipenuhi.

6.6 KUHPERDATA PASAL 1365: PERBUATAN MELAWAN HUKUM SEBAGAI JALUR GANTI RUGI

PMH: Dasar Gugatan Perdata yang Fleksibel

Pasal 1365 KUHPerdata adalah ketentuan paling fundamental dalam hukum perdata Indonesia tentang pertanggungjawaban atas kerugian yang ditimbulkan oleh perbuatan melawan hukum. Pasal ini menyatakan bahwa tiap perbuatan melanggar hukum yang membawa kerugian kepada orang lain mewajibkan orang yang karena salahnya menerbitkan kerugian itu untuk mengganti kerugian tersebut.

Dalam konteks deepfake, Pasal 1365 KUHPerdata menjadi penting karena beberapa alasan. Gugatan perdata PMH tidak bergantung pada adanya ketentuan pidana yang spesifik, cukup ada perbuatan melawan hukum dan kerugian yang dibuktikan. Gugatan PMH dapat dilakukan bersamaan dengan proses pidana. Ganti rugi yang dapat diperoleh lebih fleksibel, mencakup kerugian materiil maupun immateriil.

Unsur-Unsur PMH dalam Konteks Deepfake

Untuk berhasil dalam gugatan PMH berbasis Pasal 1365, penggugat (korban deepfake) harus membuktikan empat unsur. Pertama, perbuatan melawan hukum: pembuatan dan/atau penyebaran deepfake jelas bertentangan dengan hak privasi, kehormatan, dan hak atas data biometrik korban. Kedua, kesalahan (*Schuld*): pembuatan deepfake menggunakan foto seseorang tanpa izin adalah tindakan disengaja sehingga unsur dolus terpenuhi.

Ketiga, kerugian (*Schade*): kerugian immateriil berupa trauma psikologis dan rusaknya reputasi, serta kerugian materiil berupa kehilangan pekerjaan dan biaya terapi, dapat dibuktikan. Keempat, kausalitas: harus dibuktikan bahwa deepfake yang dibuat pelaku adalah yang menyebabkan kerugian yang dialami korban.

Jenis Ganti Rugi yang Dapat Diminta

Dalam gugatan PMH terkait deepfake, korban dapat menuntut berbagai bentuk ganti rugi. Ganti rugi materiil mencakup biaya terapi psikologis, kehilangan penghasilan, biaya layanan penghapusan konten, dan biaya perkara hukum. Ganti rugi immateriil mencakup kompensasi atas penderitaan psikologis, kerusakan reputasi, dan hilangnya ketenangan hidup. Korban juga dapat meminta penghentian perbuatan (dwangsom) di mana pengadilan memerintahkan pelaku menghentikan penyebaran deepfake dan menghapus semua salinan yang ada, disertai ancaman uang paksa harian jika perintah tidak dipatuhi.

Tantangan Gugatan PMH untuk Deepfake

Meskipun Pasal 1365 KUHPerdata memberikan fleksibilitas, ada beberapa tantangan praktis yang perlu diantisipasi. Identifikasi pelaku menjadi hambatan signifikan jika pelaku beroperasi secara anonim menggunakan akun palsu atau VPN. Pengadilan Indonesia juga tidak memiliki standar yang seragam dalam menilai besaran ganti rugi immateriil, sehingga korban

deepfake perlu membuktikan penderitaan psikologis dengan bukti yang memadai seperti keterangan dokter atau psikolog. Proses gugatan perdata umumnya lebih lama dan lebih mahal dibanding pelaporan pidana, dan eksekusi putusan bisa menjadi masalah jika pelaku tidak memiliki aset yang cukup.

6.7 SINTESIS: MEMILIH INSTRUMEN HUKUM YANG TEPAT

Skema Pemilihan Instrumen Berdasarkan Jenis Kasus

Menghadapi kasus deepfake yang konkret, seorang praktisi hukum perlu memilih instrumen yang paling tepat berdasarkan karakteristik kasusnya. Untuk deepfake NCII atau pornografi, instrumen primer adalah UU TPKS Pasal 14 dikombinasikan dengan UU ITE Pasal 27 ayat (1), dengan instrumen sekunder UU Pornografi Pasal 4, serta jalur perdata melalui PMH Pasal 1365 KUHPerdata dan UU Hak Cipta. Untuk deepfake penipuan finansial, instrumen primer adalah UU ITE Pasal 35 dan 51 dikombinasikan dengan KUHP Pasal 492, dengan instrumen sekunder UU ITE Pasal 28 ayat (1). Untuk deepfake pencemaran nama baik, instrumen primer adalah UU ITE Pasal 27A, dengan instrumen sekunder KUHP Pasal 310-311.

Untuk deepfake disinformasi SARA, instrumen primer adalah UU ITE Pasal 28 ayat (2). Untuk deepfake dokumen atau identitas, instrumen primer adalah KUHP Pasal 263-264 dikombinasikan dengan UU ITE Pasal 35. Untuk pelanggaran data biometrik, instrumen primer adalah UU PDP Pasal 65-66. Untuk kasus pemerasan atau sextortion, instrumen primer adalah UU TPKS Pasal 14 ayat (3) dikombinasikan dengan KUHP Pasal 368.

Strategi Dakwaan Berlapis

Dalam praktik penegakan hukum deepfake, strategi dakwaan berlapis yang menggunakan lebih dari satu pasal dari lebih dari satu undang-undang adalah pendekatan yang paling efektif. Deepfake jarang hanya melanggar satu norma hukum. Sebuah deepfake NCII secara bersamaan melanggar UU TPKS, UU ITE, dan UU PDP. Dakwaan berlapis memastikan bahwa jika satu dakwaan gagal dibuktikan, dakwaan lain masih dapat berdiri. Selain itu, dakwaan berlapis memberi hakim ruang yang lebih luas untuk menjatuhkan hukuman yang proporsional dengan keseluruhan dampak dari tindak pidana.

Untuk kasus deepfake NCII dengan unsur penyebaran, pola dakwaan berlapis yang direkomendasikan adalah dakwaan kesatu Pasal 14 ayat (2) UU TPKS, dakwaan kedua Pasal 27 ayat (1) jo. Pasal 45 ayat (1) UU ITE, dan dakwaan ketiga Pasal 66 jo. Pasal 68 UU PDP. Untuk kasus deepfake penipuan finansial, dakwaan kesatu adalah Pasal 35 jo. Pasal 51 ayat (1) UU ITE, dakwaan kedua Pasal 492 KUHP Baru, dan dakwaan ketiga Pasal 28 ayat (1) jo. Pasal 45A ayat (1) UU ITE. Penentuan dakwaan tetap merupakan kewenangan jaksa penuntut umum berdasarkan fakta dan bukti dalam perkara konkret.

6.8 CELAH REGULASI YANG BELUM TERTUTUP

Setelah menelaah seluruh instrumen hukum di atas, ada beberapa celah yang tidak dapat disangkal keberadaannya. Celah-celah ini mencerminkan ketidaksiapan sistem hukum menghadapi teknologi yang bergerak jauh lebih cepat dari proses legislasi. Tidak ada definisi

normatif deepfake dalam satu pun undang-undang Indonesia, yang menciptakan ketidakpastian hukum yang menguntungkan pelaku dan menyulitkan penuntutan.

Sebagian besar pasal yang ada lebih kuat dalam menjerat distributor konten deepfake dibanding pembuatnya, padahal dalam era deepfake-as-a-service pembuat dan distributor bisa merupakan pihak yang berbeda. Tidak ada kewajiban aktif bagi platform digital untuk mendeteksi, melabeli, atau menghapus konten deepfake; PP PTSE hanya mewajibkan platform merespons laporan, bukan mencegah secara proaktif. Tidak ada satu pasal pun yang secara eksplisit melarang deepfake dalam konteks kampanye pemilu atau proses elektoral. Bot Telegram dan platform sejenis yang menghasilkan konten seksual deepfake dari foto siapa pun juga tidak diatur secara spesifik dalam hukum Indonesia, sebagaimana dieksploitasi secara nyata dalam kasus UNUD 2025.

Rangkuman Bab

Bab ini telah menelaah enam instrumen hukum utama yang dapat digunakan dalam konteks deepfake di Indonesia. UU ITE, khususnya Pasal 35 jo. Pasal 51, adalah instrumen paling kuat yang tersedia saat ini untuk menjerat pembuat deepfake dengan ancaman pidana tertinggi mencapai 12 tahun penjara. Pasal 27 dan 28 melengkapi dengan cakupan distribusi dan dampak sosialnya.

KUHP Baru yang berlaku Januari 2026 memperkuat instrumen penipuan digital melalui Pasal 492 dan menjangkau pemalsuan dokumen deepfake melalui Pasal 263-264. KUHP Baru berfungsi sebagai pelengkap, bukan pengganti, hukum pidana khusus. UU PDP membuka perspektif baru: wajah dan suara seseorang adalah data biometrik yang dilindungi, dan penggunaannya tanpa izin untuk membuat deepfake adalah pelanggaran hak atas data pribadi sejak tahap pengumpulan foto. UU TPKS, khususnya Pasal 14, adalah instrumen paling progresif untuk korban deepfake NCII karena mengakui dimensi "kekerasan seksual" dari tindak pidana tersebut dan menjamin hak-hak korban yang tidak diatur dalam UU ITE.

UU Hak Cipta dan KUHPerdata Pasal 1365 membuka jalur perdata yang memungkinkan korban mendapat ganti rugi finansial termasuk ganti rugi immateriil atas penderitaan psikologis. Celah regulasi yang paling mendasar adalah absennya definisi normatif deepfake dan lemahnya kriminalisasi terhadap tindakan pembuatan konten deepfake itu sendiri, bukan hanya distribusinya. Gambaran paradoksnya adalah Indonesia memiliki cukup banyak instrumen hukum yang dapat digunakan, namun tidak ada satu pun yang dirancang untuk deepfake. Reformasi legislatif bukan pilihan, melainkan keharusan.

BAB 7
KEKOSONGAN HUKUM (RECHTSVACUUM)

Hukum selalu datang terlambat. Hampir semua sistem hukum di dunia mengalami hal yang sama ketika berhadapan dengan teknologi baru. Tetapi dalam konteks deepfake, keterlambatan itu bukan sekadar soal ketinggalan waktu beberapa bulan atau setahun. Ini adalah jarak yang sudah cukup lebar untuk dimasuki oleh kejahatan yang nyata, dengan korban yang nyata, dan kerugian yang nyata, sementara hukum masih meraba-raba di pinggiran.

Bab 6 telah memetakan apa yang dimiliki Indonesia: sejumlah instrumen hukum yang bisa digunakan dengan cara analogi dan interpretasi yang kadang harus dipaksakan. Bab ini melanjutkan dengan pertanyaan yang lebih tajam: seberapa besar sesungguhnya jarak antara apa yang ada dengan apa yang dibutuhkan, di mana tepatnya letak kekosongan itu, mengapa ia terbentuk, dan apa konsekuensinya bagi korban dan bagi sistem hukum secara keseluruhan.

Rechtsvacuum, istilah hukum untuk kekosongan hukum, bukan berarti tidak ada aturan sama sekali. Istilah ini merujuk pada kondisi di mana aturan yang ada tidak dirancang untuk menjangkau situasi yang konkret ada di depan mata. Dalam konteks deepfake, Indonesia mengalami setidaknya lima rechtsvacuum yang saling berkaitan: absennya definisi teknis, masalah pembuktian forensik, pendekatan hukum yang selalu reaktif, keterbatasan kapasitas aparat, dan lemahnya perlindungan korban.

7.1 TIDAK ADA DEFINISI TEKNIS DEEPFAKE DALAM HUKUM POSITIF INDONESIA

Mengapa Definisi Itu Penting dalam Hukum

Dalam sistem hukum, definisi bukan sekadar formalitas redaksional. Definisi adalah titik awal dari seluruh proses hukum, mulai dari penyidikan hingga penuntutan dan pembuktian di pengadilan. Ketika tidak ada definisi tentang apa yang dilarang, penegak hukum terpaksa bekerja dalam ketidakpastian yang menguntungkan pelaku dan menyulitkan korban.

Bayangkan seorang penyidik yang menerima laporan kasus deepfake. Pertanyaan pertama yang harus dijawab sebelum bisa melangkah maju adalah: apakah tindakan yang dilaporkan ini merupakan tindak pidana? Tanpa definisi yang jelas tentang apa itu deepfake dalam hukum positif, jawabannya tergantung pada interpretasi, dan interpretasi yang berbeda bisa menghasilkan hasil yang berbeda tergantung hakim atau jaksa yang menanganinya. Dalam praktik penegakan hukum Indonesia saat ini, kasus deepfake diproses menggunakan analogi terhadap pasal-pasal yang tidak dirancang untuk teknologi ini. Hasilnya tidak selalu konsisten dan tidak selalu adil bagi korban.

Konsekuensi Konkret dari Absennya Definisi

Ketika tidak ada definisi normatif, setiap institusi, baik kepolisian, kejaksaan, maupun pengadilan, membangun pemahaman sendiri tentang apa yang dianggap sebagai deepfake dan tindakan apa yang dapat dipidana. Hasilnya adalah inkonsistensi yang merugikan kepastian hukum. Penyidik di Bareskrim mungkin memiliki pemahaman teknis yang lebih baik karena ada unit siber yang terlatih. Penyidik di tingkat polres daerah mungkin tidak memiliki

pemahaman yang sama, dan laporan korban bisa saja diabaikan bukan karena tidak ada dasar hukum, tetapi karena petugas tidak tahu bagaimana mengkualifikasikan perbuatan yang dilaporkan.

Absennya definisi juga menciptakan ruang yang bisa dimanfaatkan oleh kuasa hukum pelaku. Ketika jaksa menggunakan Pasal 35 UU ITE untuk menjerat pembuat deepfake, kuasa hukum pelaku bisa berargumen bahwa kliennya tidak "memanipulasi dokumen elektronik" melainkan hanya "berkreasi dengan teknologi yang tersedia." Tanpa definisi normatif yang jelas, argumen ini tidak sepenuhnya mudah dibantah. Di beberapa negara yang sudah memiliki definisi deepfake dalam hukum positifnya, argumen semacam ini jauh lebih sulit dipertahankan.

Ketiadaan definisi juga mempersulit kerjasama internasional. Permintaan bantuan hukum internasional (*Mutual Legal Assistance Treaty* atau MLAT) memerlukan definisi yang jelas tentang tindak pidana yang dimaksud. Ketika Indonesia meminta data pengguna kepada platform asing dalam rangka penyelidikan kasus deepfake, salah satu hal yang harus dijelaskan adalah tindak pidana apa yang sedang diselidiki. Tanpa definisi yang spesifik, proses ini lebih lambat dan hasilnya lebih tidak pasti.

Apa yang Harus Didefinisikan?

Sebuah definisi hukum yang memadai untuk deepfake setidaknya harus mencakup empat elemen. Pertama, substrat teknologi: definisi harus menyebut bahwa konten dibuat atau dimanipulasi menggunakan kecerdasan buatan atau algoritma pembelajaran mesin, misalnya dengan rumusan "yang dihasilkan atau dimodifikasi menggunakan kecerdasan buatan atau algoritma komputasional." Kedua, objek manipulasi: harus jelas bahwa yang dimanipulasi adalah representasi identitas seseorang, meliputi wajah, suara, atau gerak tubuh. Ketiga, efek yang dihasilkan: konten yang dihasilkan tampak autentik bagi persepsi manusia biasa tanpa alat bantu deteksi. Keempat, ketidakhadiran persetujuan: penggunaan tanpa izin dari orang yang identitasnya direpresentasikan adalah unsur yang membedakan deepfake ilegal dari yang sah.

Dengan keempat elemen ini, sebuah definisi tidak hanya menjawab pertanyaan "apa itu deepfake" tetapi juga langsung menunjukkan kapan deepfake menjadi pelanggaran hukum, yaitu ketika teknologi AI digunakan untuk merepresentasikan identitas seseorang secara palsu tanpa persetujuannya.

7.2 MASALAH PEMBUKTIAN DIGITAL FORENSIK

Pembuktian sebagai Jantung dari Penegakan Hukum

Dalam hukum pidana, keyakinan hakim dibangun dari bukti. Tanpa bukti yang kuat, proses hukum tidak bisa berjalan, atau berjalan menuju putusan yang salah. Dalam kasus deepfake, tantangan pembuktian bukan hanya tentang apakah konten yang dipermasalahkan adalah deepfake, tetapi juga tentang siapa yang membuatnya, dengan alat apa, di mana, kapan, dan dengan niat apa.

Setiap pertanyaan itu membutuhkan jawaban yang didukung oleh bukti teknis, dan jawaban itu tidak bisa diberikan hanya oleh pemeriksaan visual biasa. Dibutuhkan keahlian

forensik digital yang sangat spesifik, perangkat analisis yang tidak murah, dan prosedur yang terstandardisasi agar hasil analisis dapat diterima di pengadilan.

Tantangan Pertama: Membuktikan bahwa Konten adalah Deepfake

Langkah paling awal dalam pembuktian kasus deepfake adalah membuktikan bahwa konten yang dipermasalahkan memang merupakan hasil manipulasi AI. Deepfake yang dibuat dengan perangkat terkini sering kali meninggalkan sangat sedikit artefak visual yang bisa dideteksi dengan mata biasa. Bahkan dengan perangkat lunak analisis, hasilnya tidak selalu bersifat definitif: analisis forensik bisa mengatakan bahwa "ada kemungkinan manipulasi" atau "terdapat anomali yang konsisten dengan deepfake," tetapi jarang bisa mengatakan dengan kepastian absolut bahwa video tersebut pasti adalah deepfake.

Metode deteksi forensik yang tersedia mencakup analisis artefak kompresi untuk melihat inkonsistensi dalam pola kompresi video, deteksi anomali gerakan wajah untuk mengidentifikasi kejanggalan dalam pergerakan otot wajah atau gerakan bibir, analisis frekuensi cahaya (DCT) untuk mendeteksi inkonsistensi yang mengindikasikan penggantian area gambar, analisis metadata tentang kapan file dibuat dan apakah ada modifikasi, serta penggunaan model AI deteksi deepfake.

Setiap metode memiliki keterbatasan: deepfake berkualitas tinggi dapat meniru pola kompresi asli, model AI terbaru sudah sangat baik mensimulasikan gerakan natural, metadata bisa dimanipulasi atau dihapus, dan model deteksi sering kalah setengah langkah dari model generasi terbaru.

Tantangan Kedua: Membuktikan Identitas Pembuat

Bahkan jika berhasil dibuktikan bahwa sebuah konten adalah deepfake, pertanyaan berikutnya adalah siapa yang membuatnya. Pembuat deepfake umumnya tidak meninggalkan jejak yang jelas. Mereka menggunakan akun anonim, perangkat yang tidak terdaftar, koneksi internet melalui VPN atau jaringan proxy, dan kadang membayar layanan menggunakan kripto.

Sebagian besar platform yang digunakan untuk menyebarkan deepfake beroperasi di bawah yurisdiksi asing. Permintaan data pengguna dari Indonesia harus mengikuti prosedur MLAT yang melibatkan proses diplomatik dan bisa memakan waktu berbulan-bulan, sementara pelaku memiliki cukup waktu untuk menghapus jejaknya. Telegram secara khusus dikenal sebagai platform yang paling sulit bekerja sama dengan permintaan penegak hukum, dan banyak bot deepfake beroperasi di Telegram justru karena perlindungan yang diberikan oleh kebijakan privasi platform tersebut.

Masalah lain yang sering diremehkan adalah rantai penyimpanan bukti digital (chain of custody). Ketika seorang penyidik mengunduh konten deepfake sebagai barang bukti, harus dipastikan bahwa konten itu tersimpan dengan cara yang membuktikan keasliannya dan tidak ada perubahan sejak pertama kali diamankan.

Indonesia belum memiliki prosedur standar nasional yang mengikat tentang bagaimana bukti digital dalam kasus deepfake harus diamankan, disimpan, dan dihadirkan di pengadilan. Setiap penyidik dan jaksa menerapkan prosedurnya sendiri, yang berarti kualitas dan keandalan bukti digital bisa sangat bervariasi dari satu kasus ke kasus lain.

Tantangan Ketiga: Kualifikasi Saksi Ahli

Dalam proses persidangan, pembuktian teknis tentang keaslian konten digital biasanya dihadirkan melalui saksi ahli. Masalahnya ada di dua sisi. Pertama, Indonesia belum memiliki standar kualifikasi yang seragam dan diakui secara resmi untuk ahli forensik digital deepfake. Kedua, hakim dan jaksa umumnya bukan ahli teknologi, sehingga tidak mudah bagi pihak-pihak di persidangan untuk mengevaluasi apakah keterangan ahli akurat, dapat diandalkan, dan relevan dengan isu hukum yang sedang diputuskan.

Ketidakseimbangan sumber daya juga menciptakan masalah yang serius. Jaksa penuntut umum mengandalkan ahli dari Puslabfor Polri yang kapasitasnya terbatas dan waktu analisanya bisa sangat lama. Kuasa hukum pelaku dalam kasus yang melibatkan uang besar bisa menyewa ahli forensik digital independen dari lembaga swasta dengan perangkat lebih mutakhir. Hasilnya bisa berupa "pertempuran ahli" di mana dua ahli memberikan kesimpulan yang berbeda dari analisis yang berbeda terhadap konten yang sama. Tanpa standar metodologi yang diakui secara nasional, hakim tidak memiliki acuan yang jelas untuk menilai mana keterangan ahli yang lebih dapat dipercaya.

7.3 REGULASI YANG BERSIFAT REAKTIF, BUKAN PREVENTIF

Dua Filosofi Regulasi

Ada dua pendekatan fundamental dalam merancang regulasi untuk ancaman baru. Pendekatan reaktif membiarkan masalah terjadi terlebih dahulu, kemudian membuat aturan untuk menanganinya. Pendekatan preventif mengantisipasi masalah sebelum terjadi luas dan membuat aturan yang mencegahnya dari awal. Hukum yang ada di Indonesia, baik UU ITE, UU TPKS, maupun UU PDP, semuanya lahir dari pengalaman masalah yang sudah ada, bukan dari antisipasi masalah yang akan datang. Konsekuensinya adalah bahwa regulasi selalu bekerja dalam posisi mengejar ketertinggalan, sementara teknologi sudah melangkah jauh ke depan.

Pola Keterlambatan Regulasi Indonesia

Pola keterlambatan ini dapat dilihat secara historis. Penipuan SMS (*smishing*) mulai terjadi di awal 2000-an, tetapi UU ITE baru muncul pada 2008, dengan jarak 5 hingga 8 tahun. Pornografi daring yang muncul pertengahan 2000-an baru direspons dengan UU Pornografi 2008, dengan jarak 3 hingga 5 tahun. Revenge porn atau NCII yang marak sejak 2013-2016 baru mendapat respons legislatif melalui UU TPKS 2022, dengan jarak 6 hingga 9 tahun.

Hoaks berbasis foto manipulasi yang muncul sejak 2016-2018 hingga kini belum memiliki regulasi spesifik, dengan jarak lebih dari 7 tahun dan belum selesai. Deepfake video dan audio yang mulai signifikan di Indonesia sejak 2022-2024 juga belum memiliki regulasi spesifik, sementara deepfake text-to-video yang muncul 2024-2025 sama sekali belum tersentuh regulasi.

Pola yang terlihat konsisten: regulasi di Indonesia muncul rata-rata 5 sampai 9 tahun setelah ancaman pertama kali muncul secara signifikan. Untuk teknologi yang berkembang secepat AI generatif, jarak waktu seperti itu sudah cukup untuk menimbulkan kerusakan yang sangat luas.

Mengapa Pendekatan Reaktif Tidak Memadai untuk Deepfake

Kelemahan pendekatan reaktif menjadi lebih serius ketika berhadapan dengan deepfake karena beberapa alasan yang saling menguatkan. Dari sisi kecepatan evolusi teknologi, deepfake yang ada hari ini sudah jauh lebih canggih dari deepfake dua tahun lalu, dan regulasi yang dibuat untuk merespons masalah yang sudah terjadi sering kali sudah tertinggal pada saat disahkan karena proses legislasi sendiri membutuhkan waktu.

Dari sisi kerusakan yang sulit dipulihkan, ketika deepfake pornografi sudah tersebar di internet, tidak ada regulasi yang bisa benar-benar memulihkan kondisi sebelumnya. Konten bisa dihapus dari satu platform, tetapi sudah diunduh dan disimpan oleh ribuan orang, sudah diunggah ulang di platform lain, dan sudah masuk ke forum yang tidak terjangkau oleh mekanisme takedown biasa. Regulasi preventif yang melarang dan mencegah pembuatan konten sejak awal jauh lebih efektif daripada regulasi reaktif yang hanya bisa menindak setelah konten sudah beredar.

Dari sisi efek pencegahan yang hilang, salah satu fungsi terpenting dari hukum pidana adalah efek deterrence: ancaman sanksi yang cukup berat membuat orang berpikir dua kali sebelum melakukan pelanggaran. Efek ini hanya bekerja jika orang mengetahui bahwa perbuatan yang mereka pertimbangkan adalah ilegal. Ketika tidak ada regulasi yang jelas melarang deepfake, banyak orang, termasuk yang bukan berniat jahat, tidak menyadari bahwa yang mereka lakukan bisa bermasalah secara hukum.

Apa yang Seharusnya Dilakukan: Pendekatan Preventif

Pendekatan preventif untuk regulasi deepfake tidak berarti melarang semua teknologi AI generatif. Yang dimaksud adalah membangun kerangka regulasi yang menetapkan standar minimum yang harus dipenuhi oleh platform dan penyedia layanan AI sebelum layanan mereka boleh beroperasi, mewajibkan pelabelan konten sintetis secara proaktif agar pengguna dapat mengetahui bahwa yang mereka lihat bukan rekaman asli, mewajibkan verifikasi identitas untuk layanan yang berpotensi disalahgunakan, serta menciptakan mekanisme pengaduan yang cepat dan responsif sebelum konten berbahaya terlanjur tersebar luas.

7.4 KETERBATASAN KAPASITAS APARAT PENEGAK HUKUM MENGHADAPI AI CRIME

Kesenjangan antara Kejahatan dan Kemampuan Penindakan

Regulasi yang baik di atas kertas tidak akan menghasilkan penegakan hukum yang efektif jika aparat yang bertugas menegakkannya tidak memiliki kapasitas untuk melakukannya. Ini adalah salah satu kelemahan paling nyata dalam sistem hukum Indonesia menghadapi deepfake.

Ada kemajuan nyata yang telah dilakukan: Dittipidsiber Bareskrim Polri telah berhasil menangkap pelaku dalam beberapa kasus deepfake yang penting. Tetapi kemajuan itu masih terkonsentrasi di tingkat nasional, sementara kasus deepfake terjadi di seluruh wilayah Indonesia dengan kapasitas penegakan yang sangat tidak merata.

Peta Kapasitas: Kesenjangan Pusat dan Daerah

Di tingkat Dittipidsiber Bareskrim Polri, kapasitas relatif memadai karena ada unit siber terlatih, perangkat forensik, dan kerjasama internasional terbatas. Namun unit ini mengalami

kelebihan kapasitas karena semua kasus tingkat nasional bermuara di sini, sehingga respons bisa lambat untuk kasus yang dinilai kurang prioritas.

Subdit Siber Polda di tingkat provinsi cukup bervariasi. Polda besar di Jawa umumnya lebih siap, sedangkan Polda luar Jawa masih berkembang dengan perangkat forensik yang terbatas. Di tingkat Sat Reskrim Polres kabupaten/kota, kondisinya umumnya belum memadai karena kebanyakan tidak memiliki unit siber khusus dan mengandalkan penyidik umum yang tidak terlatih untuk kejahatan digital. Di tingkat kejaksaan dan pengadilan, pemahaman teknis tentang deepfake dan AI sangat terbatas dan hakim sangat bergantung pada keterangan ahli.

Tiga Defisit Kapasitas yang Paling Kritis

Defisit pertama adalah defisit perangkat forensik. Mendeteksi deepfake memerlukan perangkat lunak khusus yang tidak murah dan terus harus diperbarui mengikuti perkembangan teknologi. Puslabfor Polri sebagai laboratorium forensik utama kepolisian memiliki perangkat yang terus ditingkatkan, tetapi kapasitas analitisnya belum sebanding dengan volume kasus yang masuk. Penyidik di daerah yang menerima laporan kasus deepfake harus mengirim barang bukti ke Jakarta atau Surabaya untuk dianalisis, sebuah proses yang bisa berlangsung berbulan-bulan.

Defisit kedua adalah defisit sumber daya manusia terlatih. Memiliki perangkat saja tidak cukup. Dibutuhkan tenaga ahli yang memahami cara menggunakan perangkat, menginterpretasikan hasilnya, dan mempresentasikan kesimpulan analisis di pengadilan dengan cara yang dapat dipahami oleh hakim dan jaksa yang tidak berlatar belakang teknis. Pelatihan khusus tentang forensik deepfake untuk aparat penegak hukum Indonesia masih sangat terbatas.

Defisit ketiga adalah defisit prosedur standar. Bahkan dengan perangkat dan personel yang ada, efektivitasnya terbatas oleh ketiadaan prosedur standar yang mengikat. Bagaimana cara mengamankan video deepfake sebagai barang bukti? Format apa yang harus digunakan? Hash apa yang harus dibuat untuk menjamin integritas barang bukti? Pertanyaan-pertanyaan ini belum terjawab dalam Peraturan Kapolri atau prosedur baku Bareskrim. Kuasa hukum pelaku yang cerdas bisa mempersoalkan validitas barang bukti digital berdasarkan ketidaksesuaian prosedur pengamannya.

Keterbatasan dalam Merespons AI yang Terus Berkembang

Di luar masalah kapasitas yang bersifat teknis, ada tantangan yang lebih mendasar: kecerdasan buatan berkembang dengan kecepatan yang berbeda dari kecepatan pelatihan dan pengembangan kapasitas institusi pemerintah. Ketika aparat selesai dilatih untuk mendeteksi deepfake dengan model AI generasi saat ini, model generasi berikutnya sudah tersedia. Artefak yang bisa dideteksi oleh perangkat forensik saat ini mungkin sudah tidak ada dalam deepfake yang dibuat dengan teknologi enam bulan ke depan. Ini adalah perlombaan senjata yang strukturnya menguntungkan pihak yang menyerang, bukan pihak yang bertahan.

Data yang tersedia dari lembaga bantuan hukum dan organisasi masyarakat sipil menunjukkan bahwa sebagian besar kasus deepfake yang dilaporkan ke kepolisian tidak pernah mencapai tahap penuntutan di pengadilan. Alasan yang paling sering dikemukakan mencakup sulitnya mengidentifikasi pelaku yang beroperasi secara anonim, lamanya proses

permintaan data ke platform asing, terbatasnya kapasitas analisis forensik digital, dan bukti yang dikumpulkan tidak cukup kuat untuk memenuhi standar "bukti yang cukup" yang diatur dalam KUHAP. Konsekuensinya adalah korban yang melapor sering kali tidak mendapat keadilan dari jalur pidana.

7.5 PERLINDUNGAN KORBAN YANG LEMAH

Korban yang Ditinggalkan oleh Sistem

Dalam pembahasan tentang deepfake, ada kecenderungan untuk berfokus pada pelaku dan sanksi. Tetapi ada pihak yang sering terlupakan: korban yang sudah terlanjur menderita, yang membutuhkan bantuan sekarang, dan yang sering kali tidak mendapatkannya dari sistem yang ada.

Perlindungan korban dalam konteks deepfake mencakup tiga hal yang seharusnya berjalan bersamaan: penghapusan konten yang cepat untuk membatasi penyebaran lebih lanjut, dukungan psikologis untuk menghadapi trauma, serta akses ke keadilan baik pidana maupun perdata untuk memastikan pertanggungjawaban pelaku. Sistem Indonesia hari ini lemah di ketiganya.

Tidak Ada Lembaga Khusus Penanganan Korban Deepfake

Indonesia tidak memiliki lembaga yang secara spesifik dibentuk untuk menangani korban deepfake. Korban harus menavigasi berbagai institusi yang tugasnya tumpang tindih, belum terintegrasi, dan tidak satu pun yang dirancang khusus untuk kasus seperti ini. Bareskrim dan Polda fokus pada penindakan pelaku, bukan pemulihan korban. Komnas Perempuan mendokumentasikan kekerasan berbasis gender termasuk digital tetapi tidak memiliki kewenangan penindakan. KPAI terbatas pada korban yang masih anak-anak. Komdigi/BRTI menangani pemblokiran konten tetapi prosesnya memakan waktu dan tidak menjangkau platform asing yang tidak kooperatif. LBH dan LSM memiliki kapasitas yang sangat terbatas dan tidak ada di semua daerah. Komisi PDP masih dalam proses penguatan institusional.

Yang absen adalah satu lembaga yang memiliki kewenangan komprehensif: bisa menerima laporan, membantu proses takedown secara cepat, memberikan pendampingan psikologis, memfasilitasi akses ke bantuan hukum, dan memantau perkembangan kasus pidana yang dilaporkan. Lembaga seperti ini ada di beberapa negara, misalnya StopNCII.org di Inggris atau Cyber Civil Rights Initiative di Amerika Serikat, tetapi belum ada padanannya di Indonesia.

Hambatan Akses ke Keadilan

Hambatan geografis menjadi masalah utama karena unit siber yang paling mampu menangani kasus deepfake terkonsentrasi di Jakarta dan beberapa kota besar. Korban dari daerah terpencil yang ingin melaporkan kasus kepada unit yang memiliki kapasitas menanganinya harus menempuh perjalanan jauh atau mengandalkan unit polres setempat yang kapasitasnya sangat terbatas. Internet yang membawa deepfake tidak mengenal batas geografis, tetapi akses ke penegakan hukum yang berkualitas tidak tersebar merata.

Hambatan ekonomi juga signifikan karena proses hukum membutuhkan uang. Menyewa pengacara, mengurus berkas administrasi, menghadiri persidangan, mendapatkan keterangan ahli, semuanya memiliki biaya. Dalam kasus deepfake NCII di mana kerugiannya bersifat psikologis dan reputasional tanpa nilai finansial yang jelas, banyak korban yang menyimpulkan bahwa biaya dan energi yang diperlukan untuk mengejar keadilan tidak sebanding dengan kemungkinan hasilnya.

Hambatan psikologis juga harus dihadapi. Proses pelaporan dan penyidikan kasus deepfake, terutama yang bersifat seksual, bisa sangat memberatkan korban. Korban harus menceritakan pengalaman traumatisnya berulang kali kepada berbagai pihak. UU TPKS sudah mengamanatkan pendekatan yang sensitif terhadap pengalaman korban, tetapi implementasinya di lapangan masih tidak konsisten.

Mekanisme Takedown yang Terlalu Lambat

Satu kebutuhan yang paling mendesak bagi korban deepfake adalah penghapusan konten yang cepat. Setiap jam konten berbahaya dibiarkan beredar adalah kerugian tambahan bagi korban karena lebih banyak orang melihatnya, lebih banyak yang mengunduh dan menyimpannya, dan lebih sulit untuk benar-benar membersihkannya dari internet.

Mekanisme takedown yang ada di Indonesia memiliki beberapa jalur: laporan langsung ke platform, laporan ke Komdigi untuk pemblokiran URL, atau laporan ke Bareskrim. Tetapi tidak satu pun dari jalur ini dirancang untuk kecepatan. Proses verifikasi laporan, persetujuan internal platform, hingga tindakan teknis pemblokiran bisa memakan waktu dari beberapa jam hingga beberapa hari. Di Uni Eropa, Digital Services Act mewajibkan platform besar memproses laporan konten ilegal dalam waktu 24 jam untuk konten yang berpotensi membahayakan. Ketentuan serupa belum ada dalam hukum Indonesia.

Skenario yang sering dilaporkan oleh korban deepfake di Indonesia menggambarkan betapa lambatnya mekanisme yang ada. Pada hari pertama, korban menemukan konten deepfake dirinya beredar dan melapor ke platform melalui tombol "report" tanpa konfirmasi. Pada hari kedua hingga ketiga, konten masih ada dan korban melapor ke Bareskrim.

Pada hari keempat hingga ketujuh, konten di Instagram akhirnya dihapus setelah sudah diunduh oleh ratusan pengguna dan diunggah ulang, sementara konten di Telegram masih ada karena platform tidak merespons. Pada minggu kedua, Bareskrim memulai penyidikan dan mengirim surat ke Telegram yang tidak direspons, sementara Komdigi memblokir URL spesifik tetapi konten terus muncul di URL baru. Hasilnya: konten tidak pernah benar-benar hilang dari internet, pelaku tidak teridentifikasi karena anonimitas platform, dan korban hidup dengan kecemasan bahwa konten kapan saja bisa muncul lagi.

7.6 SINTESIS: PETA KEKOSONGAN HUKUM YANG KOMPREHENSIF

Dari kelima dimensi yang telah dibahas, absennya definisi, masalah pembuktian, pendekatan reaktif, keterbatasan kapasitas aparat, dan lemahnya perlindungan korban, sebuah gambaran yang komprehensif tentang rechtsvacuum deepfake Indonesia menjadi jelas. Kelima kekosongan ini bukan masalah yang berdiri sendiri-sendiri. Semua saling

berkaitan dan saling memperkuat sehingga membuat setiap satu kekosongan menjadi lebih serius karena keempat kekosongan lainnya juga ada.

Tidak adanya definisi teknis berdampak langsung pada penegakan yang tidak konsisten dan memperkuat kelemahan kapasitas aparat yang tidak punya pegangan jelas. Masalah pembuktian forensik diperburuk oleh absennya prosedur standar dan berdampak langsung pada akses korban ke keadilan. Regulasi yang reaktif memperburuk semua kekosongan lain karena sistem selalu bekerja dalam kondisi ketertinggalan.

Kapasitas aparat yang terbatas diperburuk oleh absennya definisi dan prosedur standar serta berdampak langsung pada perlindungan korban. Lemahnya perlindungan korban merupakan akumulasi dari semua kekosongan lain dan menjadi umpan balik yang mendorong under-reporting.

Kelima kekosongan menciptakan lingkaran setan yang memperpanjang dirinya sendiri: tidak ada definisi teknis membuat aparat kesulitan mengkualifikasikan perbuatan, kapasitas terbatas untuk menyidik dengan standar yang jelas menyebabkan banyak kasus tidak bisa dibuktikan di pengadilan, pelaku tidak mendapat sanksi sehingga tidak ada efek jera dan kasus terus meningkat, korban tidak terlindungi dan tidak mau melapor, data kasus yang tercatat sedikit membuat tekanan untuk membuat regulasi baru terasa kurang mendesak, sehingga tidak ada definisi teknis dalam regulasi baru dan lingkaran pun berulang. Memotong lingkaran ini memerlukan intervensi di lebih dari satu titik sekaligus.

7.7 JALAN KELUAR: ARAH YANG HARUS DITEMPUH

Intervensi Jangka Pendek (6 bulan - 1 tahun)

Dalam jangka pendek, diperlukan penerbitan Surat Edaran atau Peraturan Kapolri yang menetapkan prosedur standar penyidikan kasus deepfake, termasuk cara pengamanan barang bukti digital, prosedur permintaan data ke platform, dan kualifikasi minimum ahli forensik digital yang bisa dimintai keterangan.

Penguatan unit siber di tingkat Polda dengan perangkat forensik digital yang memadai dan personel yang terlatih secara khusus untuk kejahatan berbasis AI juga sangat mendesak. Selain itu, perlu segera dibentuk layanan aduan terpadu untuk korban deepfake yang mengintegrasikan laporan ke kepolisian, permintaan takedown ke Komdigi, dan rujukan ke layanan dukungan psikologis.

Intervensi Jangka Menengah (1 - 3 tahun)

Dalam jangka menengah, amandemen UU ITE perlu dilakukan untuk memasukkan definisi deepfake dan ketentuan yang secara eksplisit mengkriminalisasi pembuatannya, bukan hanya distribusinya. Amandemen UU TPKS diperlukan untuk secara eksplisit mencakup konten seksual sintetis berbasis AI, bukan hanya rekaman yang diambil dari kejadian nyata.

Penguatan kewenangan Komisi PDP agar dapat bergerak lebih cepat dalam kasus pelanggaran data biometrik yang berkaitan dengan pembuatan deepfake juga menjadi prioritas. Perjanjian bilateral atau multilateral dengan negara-negara di mana platform utama berkantor pusat perlu segera dijalin untuk mempercepat kerja sama dalam kasus deepfake yang melintasi batas yurisdiksi.

Intervensi Jangka Panjang (3 - 5 tahun)

Dalam jangka panjang, Indonesia memerlukan undang-undang kecerdasan buatan yang komprehensif yang mengatur pengembangan dan penggunaan AI generatif secara menyeluruh, termasuk kewajiban transparansi, standar keamanan, dan mekanisme akuntabilitas untuk penggunaan AI dalam produksi konten.

Pembentukan lembaga pengawas AI yang independen dengan mandat dan kewenangan yang jelas serta kapasitas teknis yang memadai juga harus disiapkan. Investasi jangka panjang dalam pendidikan hukum dan literasi digital untuk mempersiapkan generasi hakim, jaksa, dan penyidik yang memahami teknologi AI secara memadai menjadi fondasi jangka panjang yang tidak bisa diabaikan.

Rangkuman Bab

Bab ini telah menganalisis secara kritis lima dimensi rechtsvacuum yang dihadapi sistem hukum Indonesia dalam merespons deepfake. Absennya definisi teknis deepfake dalam hukum positif Indonesia menciptakan ketidakpastian yang menguntungkan pelaku dan menyulitkan penuntutan, dan ini adalah kekosongan yang paling mendasar yang harus diisi pertama kali.

Pembuktian dalam kasus deepfake jauh lebih kompleks dari kasus digital biasa: membuktikan bahwa konten adalah deepfake memerlukan analisis forensik khusus, membuktikan identitas pembuat sering kali harus melewati platform yang tidak kooperatif, dan chain of custody bukti digital belum memiliki prosedur standar di Indonesia.

Regulasi Indonesia secara historis bersifat reaktif, muncul setelah masalah sudah terjadi secara masif. Untuk teknologi yang berkembang secepat AI generatif, pendekatan ini tidak memadai karena kerusakan yang ditimbulkan sering kali tidak bisa dipulihkan. Kapasitas aparat penegak hukum menghadapi AI crime sangat tidak merata, terkonsentrasi di tingkat nasional dan kota besar, sementara kasus deepfake terjadi di seluruh wilayah Indonesia. Perlindungan korban deepfake sangat lemah: tidak ada lembaga khusus, mekanisme takedown yang lambat, hambatan akses keadilan yang berlapis, dan prosedur yang belum sensitif terhadap pengalaman korban.

Kelima kekosongan ini saling berkaitan dan menciptakan lingkaran setan yang memperpanjang dirinya sendiri. Memotong lingkaran ini memerlukan intervensi yang simultan di beberapa titik, bukan penambalan satu per satu. Rechtsvacuum yang dibahas dalam bab ini bukan masalah yang tidak bisa diselesaikan. Yang paling menentukan adalah kehendak politik untuk memprioritaskan masalah ini: mengakui bahwa korban deepfake adalah korban yang nyata, bahwa kerugian yang mereka alami adalah serius, dan bahwa sistem hukum memiliki kewajiban untuk merespons secara serius pula.

BAB 8
PERTANGGUNG JAWABAN PIDANA DAN PERDATA PELAKU DEEPFAKE

Siapa yang harus mempertanggungjawabkan kerugian yang ditimbulkan oleh deepfake? Pertanyaan ini terdengar sederhana, tetapi jawabannya tidak pernah sederhana dalam praktik. Dunia deepfake sering kali melibatkan lebih dari satu pihak yang berperan: ada yang membuat konten, ada yang menyebarkan, ada yang menyediakan platform, dan ada yang meminta atau membayar pembuatannya. Menentukan siapa yang bertanggung jawab dan seberapa besar memerlukan analisis yang cermat tentang peran masing-masing pihak dalam rantai kejahatan itu.

Bab ini membahas dua jalur pertanggungjawaban yang bisa ditempuh, yaitu pidana dan perdata. Keduanya bisa berjalan bersamaan, keduanya memiliki karakteristik yang berbeda, dan keduanya menawarkan bentuk pemulihan yang berbeda pula. Jalur pidana bertujuan memberikan sanksi kepada pelaku dan efek jera bagi masyarakat. Jalur perdata bertujuan memberikan ganti rugi kepada korban dan memulihkan apa yang hilang akibat perbuatan pelaku.

Pemahaman yang baik tentang dua jalur ini bukan hanya penting secara akademis. Ini adalah pengetahuan praktis yang dibutuhkan oleh siapa pun yang akan mendampingi korban deepfake, baik sebagai pengacara, konselor hukum, maupun advokat masyarakat sipil.

8.1 UNSUR-UNSUR PIDANA: NIAT, PERBUATAN, AKIBAT, DAN KAUSALITAS

Prinsip Dasar: Tidak Ada Pidana tanpa Unsur yang Terpenuhi

Dalam hukum pidana, seseorang tidak bisa begitu saja dinyatakan bersalah hanya karena ada korban atau ada kerugian. Harus ada serangkaian unsur yang terpenuhi secara kumulatif sebelum pengadilan bisa menjatuhkan pidana. Prinsip ini disebut asas legalitas: nullum crimen sine lege, nulla poena sine crimine, yang berarti tidak ada kejahatan tanpa aturan hukum yang melarangnya, dan tidak ada pidana tanpa kejahatan yang terbukti.

Dalam konteks deepfake, prinsip ini berarti jaksa harus membuktikan setiap unsur dari pasal yang didakwakan. Tidak ada unsur yang bisa diasumsikan atau dianggap terbukti sendirinya hanya karena konten deepfake itu ada dan seseorang dirugikan olehnya. Pemahaman tentang unsur-unsur ini krusial karena di situlah letak titik terlemah dan terkuat dalam setiap kasus.

Unsur Pertama: Niat (Mens Rea)

Niat atau kesengajaan adalah unsur subjektif yang paling sering diperdebatkan dalam kasus deepfake. Hukum pidana Indonesia mengenal dua bentuk kesalahan: kesengajaan (dolus) dan kelalaian (culpa). Hampir semua pasal yang relevan dengan deepfake, termasuk UU ITE Pasal 35, UU TPKS Pasal 14, dan UU PDP Pasal 66, mensyaratkan kesengajaan, bukan sekadar kelalaian.

Yang membuat pembuktian mens rea dalam kasus deepfake menantang adalah bahwa niat itu tidak bisa dilihat secara langsung. Jaksa harus merekonstruksinya dari fakta-fakta

objektif: apa yang dilakukan pelaku, bagaimana cara pelaku melakukannya, apa yang pelaku katakan kepada orang lain tentang perbuatannya, dan apa konsekuensi yang seharusnya sudah dapat pelaku perkirakan.

Kesengajaan dalam hukum pidana terbagi menjadi beberapa bentuk. Dolus directus atau niat langsung terjadi ketika pelaku memang menghendaki terjadinya akibat dari perbuatannya, misalnya seseorang yang membuat deepfake NCII dengan tujuan eksplisit menyebarkan dan mempermalukan korban. Dolus eventualis atau niat tidak langsung terjadi ketika pelaku tidak menghendaki akibatnya tetapi menyadari kemungkinan akibat itu terjadi dan tetap menerimanya, misalnya seseorang yang membuat deepfake "untuk kesenangan sendiri" namun menyimpannya di folder yang bisa diakses orang lain. Culpa atau kelalaian terjadi ketika pelaku tidak menghendaki dan tidak menyadari kemungkinan akibat buruk, padahal seharusnya menyadari. Culpa umumnya tidak cukup untuk pasal-pasal deepfake yang mensyaratkan kesengajaan.

Dalam praktik, kuasa hukum pelaku sering berargumen bahwa kliennya membuat deepfake hanya untuk "bereksperimen dengan teknologi" atau sekadar "iseng", sebagai upaya menghindari pembuktian dolus directus. Jaksa perlu mampu menunjukkan bahwa setidaknya dolus eventualis terpenuhi: pelaku mengetahui bahwa tindakannya bisa merugikan orang yang wajahnya digunakan, dan pelaku menerima kemungkinan risiko tersebut.

Unsur Kedua: Perbuatan (Actus Reus)

Unsur perbuatan adalah tindakan konkret yang dilarang oleh hukum. Dalam konteks deepfake, perbuatan ini bisa berbentuk beragam tergantung peran pelaku dalam rantai kejahatan. Bentuk pertama adalah membuat konten deepfake, yaitu menggunakan perangkat lunak AI untuk menghasilkan gambar, video, atau audio yang memanipulasi identitas seseorang.

Bentuk kedua adalah menyebarkan konten deepfake, yaitu mengunggah, membagikan, atau membuat konten tersebut dapat diakses oleh pihak lain. Bentuk ketiga adalah mengancam menggunakan konten deepfake, yaitu mengomunikasikan kepada korban bahwa konten tersebut ada dan akan disebarkan sebagai bentuk tekanan atau pemerasan. Bentuk keempat adalah memesan atau membiayai pembuatan deepfake, yaitu bertindak sebagai pihak yang meminta dan mendanai pembuatan konten oleh orang lain.

Hal penting yang perlu dipahami adalah bahwa setiap bentuk perbuatan ini bisa memenuhi unsur yang berbeda dari pasal yang berbeda. Membuat deepfake bisa dijerat dengan Pasal 35 UU ITE atau Pasal 14 ayat (1) UU TPKS. Menyebarkannya bisa dijerat dengan Pasal 27 UU ITE atau Pasal 14 ayat (2) UU TPKS. Mengancam menggunakan deepfake bisa dijerat dengan Pasal 14 ayat (3) UU TPKS tentang sextortion. Pemahaman tentang diferensiasi ini sangat penting bagi jaksa dalam menyusun dakwaan yang tepat sasaran.

Unsur Ketiga: Akibat (Gevolg)

Beberapa pasal dalam hukum pidana mensyaratkan terjadinya akibat tertentu sebagai unsur delik. Pasal 28 ayat (1) UU ITE, misalnya, mensyaratkan adanya "kerugian konsumen dalam Transaksi Elektronik" sebagai akibat dari penyebaran berita bohong. Tanpa akibat itu terbukti, pasal tersebut tidak bisa diterapkan. Jenis delik seperti ini disebut delik materiil.

Sementara itu, pasal lain bersifat formil, artinya cukup perbuatannya saja yang dilarang tanpa perlu membuktikan akibat spesifik. Pasal 35 UU ITE adalah contoh pasal formil: manipulasi dokumen elektronik sudah bisa dipidana bahkan jika belum ada yang tertipu oleh deepfake tersebut. Perbedaan antara pasal materiil dan formil ini memiliki implikasi besar dalam strategi dakwaan. Jaksa yang menghadapi kesulitan membuktikan akibat konkret dari sebuah deepfake lebih baik menggunakan pasal formil seperti Pasal 35 UU ITE sebagai dakwaan utama.

Unsur Keempat: Kausalitas

Kausalitas adalah hubungan sebab-akibat antara perbuatan pelaku dan kerugian yang dialami korban. Dalam kasus deepfake, pembuktian kausalitas bisa sangat kompleks ketika kerugian korban terjadi melalui rantai yang panjang. Misalnya, deepfake dibuat oleh pihak A, diunggah oleh pihak B ke platform C, dilihat oleh rekan kerja korban D, yang kemudian memberitahu atasan E, yang akhirnya memecat korban. Apakah pihak A bertanggung jawab atas pemecatan korban? Secara logis ya, tetapi secara hukum hal ini memerlukan argumentasi kausalitas yang solid dan tidak mudah dipatahkan.

Dalam kasus deepfake penipuan, analisis kausalitas menjadi lebih rumit lagi ketika ada beberapa pelaku dengan peran berbeda. Pembuat deepfake (A) dapat dijerat Pasal 35 UU ITE tentang manipulasi informasi elektronik ditambah KUHP Pasal 492 tentang penipuan, jika dapat dibuktikan bahwa pelaku mengetahui konten akan digunakan untuk menipu. Penyebar deepfake (B) dapat dijerat Pasal 28 ayat (1) UU ITE tentang penyebaran berita bohong yang merugikan, baik secara terpisah maupun bersama pelaku A sebagai turut serta. Platform umumnya tidak bertanggung jawab secara pidana kecuali terbukti mengetahui dan membiarkan konten berbahaya tetap ada setelah dilaporkan.

8.2 PELAKU TUNGGAL VS. PELAKU BERSAMA: TURUT SERTA DAN PEMBANTUAN

Mengapa Doktrin Penyertaan Penting dalam Kasus Deepfake

Kejahatan deepfake jarang dilakukan oleh satu orang seorang diri dari awal hingga akhir. Lebih sering ada pembagian peran yang cukup kompleks: satu orang mengumpulkan foto korban, orang lain membuat deepfake-nya, orang lain lagi yang menyebarkan, dan mungkin ada pihak lain yang membiayai atau memesan keseluruhan proses itu. Kondisi ini menciptakan pertanyaan penting: siapa yang bisa dijerat, dan dengan ancaman pidana berapa?

Hukum pidana Indonesia mengatur tanggung jawab dalam perbuatan yang dilakukan bersama-sama melalui doktrin penyertaan (*deelneming*) yang diatur dalam Pasal 55 dan 56 KUHP. Pemahaman tentang doktrin ini penting karena menentukan siapa yang bisa dijerat dan dengan ancaman pidana berapa.

Pasal 55 KUHP: Pelaku, Penyuruh, dan Turut Serta

Pasal 55 KUHP mengatur tentang siapa yang dianggap sebagai pelaku (dader) dalam suatu tindak pidana. Semua yang masuk dalam kategori Pasal 55 diancam dengan pidana yang sama seperti pelaku tunggal, tanpa pengurangan. Hal ini penting untuk dipahami karena artinya pemesan kejahatan mendapat ancaman yang sama besarnya dengan yang secara fisik melakukan kejahatan.

Pertama adalah pleger atau pelaku utama, yaitu orang yang secara langsung melakukan perbuatan pidana, misalnya orang yang secara langsung membuat atau mengunggah konten deepfake. Kedua adalah doen pleger atau penyuruh, yaitu orang yang menyuruh orang lain melakukan perbuatan pidana, misalnya orang yang memesan dan membayar orang lain untuk membuat deepfake korban.

Ketiga adalah medepleger atau turut serta, yaitu orang yang bersama-sama dengan orang lain melakukan perbuatan pidana, misalnya dua orang yang bersama-sama merencanakan dan menjalankan kampanye deepfake secara terkoordinasi. Ketiga kategori ini mendapat ancaman pidana yang sama seperti pelaku tunggal tanpa pengurangan apapun.

Pasal 56 KUHP: Pembantuan

Berbeda dari Pasal 55, Pasal 56 KUHP mengatur tentang pembantuan, yaitu pihak yang turut berperan dalam terlaksananya kejahatan namun dengan peran yang lebih bersifat pendukung dari pelaku utama. Pembantuan dibagi menjadi dua bentuk, dan penting dicatat bahwa ancaman pidananya dikurangi sepertiga dari ancaman pokok.

Pembantuan sebelum kejahatan terjadi ketika seseorang memberikan kesempatan, sarana, atau keterangan untuk melakukan kejahatan. Contohnya dalam kasus deepfake adalah orang yang memberikan foto korban kepada pembuat deepfake, atau orang yang mengajarkan cara menggunakan perangkat deepfake untuk tujuan jahat.

Pembantuan saat kejahatan berlangsung terjadi ketika seseorang memberikan bantuan teknis atau lainnya pada saat kejahatan sedang terjadi. Contohnya adalah orang yang membantu menyebarkan tautan deepfake selagi konten masih aktif, atau operator teknis yang membantu proses pembuatan deepfake secara langsung.

Dalam praktik penegakan hukum kasus deepfake, doktrin pembantuan sering kali menjadi alat yang berguna untuk menjangkau pihak-pihak yang perannya tidak langsung tetapi tetap berkontribusi pada terjadinya kejahatan. Seseorang yang "hanya membagikan" deepfake kepada rekan-rekannya bisa dijerat sebagai pembantu meskipun tidak membuat konten tersebut. Pengurangan sepertiga ancaman pidana tetap berarti hukuman yang berat jika pasal pokoknya sudah mengancam 12 tahun penjara seperti Pasal 35 UU ITE.

Ilustrasi Sindikat Deepfake: Siapa Menjawab Pasal Berapa

Untuk memahami bagaimana doktrin penyertaan bekerja dalam praktik, perhatikan skenario berikut yang berbasis pola kasus nyata Indonesia. Roni adalah mantan pacar korban yang memesan dan membayar pembuatan deepfake. Sebagai penyuruh (doen pleger), Roni dijerat dengan Pasal 55 jo. Pasal 35 dan 51 UU ITE dengan ancaman pidana yang sama seperti pembuat, ditambah Pasal 55 jo. Pasal 14 UU TPKS dan KUHP Pasal 368 tentang pemerasan jika ada ancaman kepada korban.

Budi adalah freelancer jasa deepfake yang ditemukan di media sosial dan menjadi pembuat deepfake (pleger). Budi dijerat dengan Pasal 35 jo. Pasal 51 UU ITE sebagai pelaku utama, Pasal 14 ayat (1) UU TPKS, dan Pasal 66 jo. Pasal 68 UU PDP karena memalsukan data biometrik korban. Cici adalah teman Roni yang ikut menyebarkan deepfake ke grup WhatsApp.

Cici dijerat dengan Pasal 27 ayat (1) jo. Pasal 45 ayat (1) UU ITE dan Pasal 14 ayat (2) UU TPKS. Statusnya bisa sebagai medepleger (Pasal 55) atau pembantu (Pasal 56) tergantung

sejauh mana Cici mengetahui rencana keseluruhan. Doni adalah admin grup Telegram tempat deepfake itu akhirnya disebarkan luas. Doni dapat dijerat dengan Pasal 27 ayat (1) UU ITE jika terbukti mengetahui dan membiarkan konten tetap ada, dan Pasal 56 KUHP sebagai pembantu jika terbukti aktif memfasilitasi penyebaran.

Pelajaran utama dari ilustrasi ini adalah bahwa setiap peran memiliki pasal dan ancaman yang berbeda. Strategi dakwaan harus cermat dalam mengkualifikasikan peran masing-masing terdakwa agar dakwaan tidak mudah gugur di pengadilan.

Pertanggungjawaban Korporasi

Pertanyaan yang semakin relevan seiring berkembangnya ekosistem deepfake komersial adalah apakah perusahaan atau badan hukum bisa dijerat pidana jika layanan mereka digunakan untuk membuat deepfake berbahaya. KUHP Baru (UU No. 1/2023) telah mengadopsi konsep pertanggungjawaban korporasi secara lebih komprehensif dibanding KUHP lama. Pasal 45 KUHP Baru menegaskan bahwa korporasi dapat menjadi subjek tindak pidana dan dapat dipidana, dengan sanksi berupa pidana denda yang besarannya lebih tinggi dari pidana denda untuk orang perseorangan.

Dalam konteks deepfake, hal ini membuka kemungkinan untuk menjerat perusahaan yang secara sadar menyediakan layanan deepfake yang digunakan untuk kejahatan. Misalnya, platform yang mengetahui bahwa layanannya sering digunakan untuk membuat konten NCII namun tidak mengambil tindakan pencegahan apapun berpotensi dijerat sebagai korporasi yang bertanggung jawab. Namun pembuktian "pengetahuan" dan "kebijakan korporasi" yang diperlukan untuk pertanggungjawaban korporasi masih merupakan tantangan yang tidak mudah dalam praktik peradilan Indonesia.

8.3 TANGGUNG JAWAB PLATFORM DIGITAL SEBAGAI INTERMEDIARY

Posisi Platform dalam Hukum Indonesia

Platform digital seperti Instagram, Telegram, TikTok, dan YouTube berada di posisi yang unik dalam ekosistem deepfake. Platform-platform tersebut bukan pembuat konten dan bukan penyebar aktif dalam pengertian konvensional, tetapi menyediakan infrastruktur yang memungkinkan deepfake disebarkan kepada jutaan orang. Pertanyaan tentang sejauh mana platform harus bertanggung jawab adalah salah satu isu paling kompleks dalam hukum digital saat ini.

Dalam hukum Indonesia, platform digital dikategorikan sebagai Penyelenggara Sistem Elektronik (PSE) yang tunduk pada PP No. 71 Tahun 2019 tentang Penyelenggaraan Sistem dan Transaksi Elektronik. Peraturan ini menetapkan sejumlah kewajiban bagi PSE, termasuk kewajiban merespons laporan konten ilegal dan kewajiban bekerja sama dengan pemerintah dalam penegakan hukum.

Prinsip Safe Harbor dan Batasannya

Banyak platform mengklaim perlindungan safe harbor, yaitu prinsip bahwa intermediary tidak bertanggung jawab atas konten yang dibuat oleh penggunanya selama tidak secara aktif berperan dalam pembuatan konten dan merespons laporan penyalahgunaan dengan baik. Prinsip ini dikenal dalam hukum Amerika Serikat melalui Section 230

Communications Decency Act dan dalam hukum Uni Eropa melalui e-Commerce Directive. Indonesia tidak memiliki ketentuan safe harbor yang eksplisit dan komprehensif seperti itu.

PP PTSE menetapkan kewajiban yang lebih langsung: platform harus merespons permintaan penurunan konten yang melanggar hukum Indonesia dalam waktu yang ditentukan. Kegagalan merespons dalam waktu itu bisa mengakibatkan sanksi administratif, termasuk pemblokiran akses platform di seluruh Indonesia. Tanggung jawab platform bersifat berjenjang tergantung pada kondisinya. Ketika platform tidak mengetahui adanya konten deepfake berbahaya, umumnya tidak ada pertanggungjawaban pidana karena asas mens rea tidak ada niat jika tidak ada pengetahuan.

Ketika platform menerima laporan tetapi tidak merespons dalam waktu yang ditentukan, platform bertanggung jawab secara administratif dan bisa dikenai sanksi serta pemblokiran berdasarkan PP PTSE dan ketentuan Komdigi. Ketika platform secara aktif merekomendasikan atau mendistribusikan konten deepfake berbahaya melalui algoritmanya, platform berpotensi bertanggung jawab secara pidana dan/atau perdata berdasarkan UU ITE Pasal 27 dan PMH Pasal 1365 KUHPerdata. Ketika platform menyediakan fitur yang secara spesifik memfasilitasi pembuatan deepfake berbahaya, platform berpotensi bertanggung jawab sebagai pembantu berdasarkan Pasal 56 KUHP, meskipun teori ini belum teruji di pengadilan Indonesia.

Kewajiban Platform yang Seharusnya Ada

Dalam konteks kekosongan regulasi yang ada, perlu ditegaskan apa yang seharusnya menjadi kewajiban platform digital dalam merespons deepfake. Pertama, deteksi proaktif: platform besar harus menginvestasikan teknologi untuk mendeteksi konten deepfake, terutama yang bersifat seksual, sebelum konten itu tersebar luas. Kedua, batas waktu respons yang mengikat: laporan konten deepfake berbahaya harus direspons dalam waktu yang jelas, misalnya 24 jam untuk konten seksual dan 48 jam untuk kategori lain.

Ketiga, transparansi data: platform harus menyediakan mekanisme yang memudahkan korban mengidentifikasi konten yang menampilkan dirinya tanpa izin. Keempat, kerja sama dengan penegak hukum: platform harus memiliki prosedur yang jelas dan cepat untuk merespons permintaan data dalam konteks penyidikan tanpa harus selalu menunggu proses MLAT yang panjang untuk kasus-kasus yang mendesak.

8.4 GUGATAN PERDATA PMH BERBASIS PASAL 1365 KUHPERDATA

Mengapa Jalur Perdata Diperlukan

Proses pidana dan proses perdata melayani tujuan yang berbeda. Pidana bertujuan memberikan sanksi kepada pelaku demi ketertiban umum; ganti rugi kepada korban hanyalah tujuan sekunder dan mekanismenya tidak selalu ada. Perdata bertujuan memulihkan kerugian yang diderita korban secara langsung; ini adalah jalur yang paling langsung menuju kompensasi finansial dan pemulihan.

Bagi korban deepfake yang menderita kerugian materiil maupun immateriil, jalur perdata melalui gugatan Perbuatan Melawan Hukum (PMH) berdasarkan Pasal 1365 KUHPerdata adalah salah satu mekanisme pemulihan yang paling komprehensif. Jalur ini bisa

ditempuh bersamaan dengan pelaporan pidana, atau secara terpisah jika proses pidana tidak berjalan.

Konstruksi Gugatan PMH: Unsur Pertama, Perbuatan Melawan Hukum

Mahkamah Agung Indonesia dalam berbagai putusannya telah mengadopsi definisi PMH yang luas: tidak hanya melanggar undang-undang tertulis, tetapi juga melanggar hak subjektif orang lain, kewajiban hukum pelaku, kesusilaan, dan kepatutan. Ini memberikan fleksibilitas yang sangat penting dalam kasus deepfake.

Pembuatan dan/atau penyebaran deepfake yang menampilkan seseorang tanpa persetujuannya memenuhi beberapa kategori PMH sekaligus. Perbuatan itu melanggar hak subjektif korban atas privasi dan integritas identitas digitalnya. Perbuatan itu juga melanggar kewajiban umum untuk tidak menciptakan konten yang merugikan orang lain.

Selain itu, perbuatan tersebut bertentangan dengan kesusilaan dan kepatutan yang berlaku dalam masyarakat, sekaligus melanggar ketentuan perundang-undangan, yaitu UU PDP, UU ITE, dan UU TPKS, yang semuanya melarang penggunaan data pribadi tanpa izin dan pembuatan konten yang merugikan.

Unsur Kedua: Kesalahan

Dalam gugatan PMH, kesalahan bisa berupa kesengajaan atau kelalaian. Ini berbeda dari hukum pidana yang umumnya mensyaratkan kesengajaan, sehingga jalur perdata lebih fleksibel. Bahkan jika sulit dibuktikan bahwa pelaku sengaja ingin merugikan korban, bisa diargumentasikan bahwa pelaku lalai dalam memperkirakan akibat dari perbuatannya.

Dalam kasus deepfake, kesengajaan relatif mudah diargumentasikan: seseorang yang mengambil foto orang lain dari media sosial dan menggunakannya untuk membuat konten manipulatif jelas melakukan tindakan yang disengaja. Yang mungkin diperdebatkan adalah apakah pelaku sengaja ingin merugikan korban. Namun hal ini tidak selalu diperlukan; cukup dibuktikan bahwa pelaku mengetahui atau seharusnya mengetahui bahwa tindakannya berpotensi merugikan.

Unsur Ketiga: Kerugian

Kerugian dalam gugatan PMH bisa bersifat materiil maupun immateriil. Korban deepfake hampir selalu mengalami keduanya, dan keduanya bisa dimintakan ganti rugi. Kerugian materiil langsung mencakup kehilangan pekerjaan akibat deepfake diketahui oleh atasan, biaya terapi psikologis, biaya bantuan hukum, dan biaya layanan penghapusan konten. Pembuktiannya dilakukan dengan surat PHK, kwitansi biaya pengobatan, dan faktur honorarium pengacara.

Kerugian materiil tidak langsung mencakup kehilangan peluang bisnis atau promosi, kerusakan jaringan profesional, dan penurunan penghasilan jangka panjang, yang dibuktikan melalui kontrak yang gagal, proyeksi penghasilan, dan keterangan rekan bisnis. Kerugian immateriil mencakup trauma psikologis, rasa malu, gangguan tidur, kecemasan, dan kerusakan hubungan sosial dan keluarga, yang dibuktikan melalui keterangan psikolog atau psikiater serta kesaksian orang-orang terdekat.

Unsur Keempat: Kausalitas

Penggugat harus membuktikan bahwa kerugian yang dialaminya adalah akibat langsung dari perbuatan tergugat. Dalam kasus deepfake, rantai kausalitas kadang panjang: dari pembuatan deepfake, penyebaran, kemudian dampak pada kehidupan korban. Argumentasi kausalitas harus menelusuri rantai ini secara logis dan dapat dibuktikan.

Tantangan muncul ketika ada faktor-faktor lain yang berkontribusi pada kerugian. Misalnya, jika korban dipecat bukan hanya karena deepfake tetapi juga karena alasan lain yang tidak berkaitan, pengacara tergugat akan mencoba memecah kausalitas ini untuk meminimalkan tanggung jawab kliennya. Jaksa dan pengacara korban perlu mempersiapkan argumentasi yang kuat untuk menunjukkan bahwa deepfake adalah penyebab dominan atau paling signifikan dari kerugian yang terjadi.

Besaran Ganti Rugi: Bagaimana Hakim Menentukannya

Satu pertanyaan yang sering muncul dari korban adalah berapa ganti rugi yang bisa didapatkan. Jawabannya tergantung pada berbagai faktor yang ada dalam kewenangan diskresi hakim. Untuk kerugian materiil, hakim akan menilai bukti-bukti yang diajukan dan menentukan besaran yang terbukti. Untuk kerugian immateriil, hakim memiliki diskresi yang lebih luas. Pengadilan Indonesia dalam berbagai perkara PMH telah mengabulkan tuntutan ganti rugi immateriil, meskipun nilainya sering lebih rendah dari yang dituntut dan variasinya cukup besar antar pengadilan.

Praktik di beberapa yurisdiksi lain memberikan kepastian yang lebih baik bagi korban. Amerika Serikat dengan DEFIANCE Act menetapkan ganti rugi statutory minimum bagi korban deepfake seksual non-konsensual. Indonesia belum memiliki mekanisme seperti ini, tetapi ini adalah sesuatu yang bisa dipertimbangkan dalam reformasi legislatif ke depan untuk memberikan kepastian hukum yang lebih baik bagi korban.

Dalam menyusun gugatan PMH untuk kasus deepfake, beberapa hal perlu diperhatikan oleh pengacara korban. Kerugian harus didokumentasikan sedini mungkin melalui screenshot, keterangan saksi, rekam medis psikologis, dan bukti kehilangan pekerjaan atau kontrak. Gugatan materiil dan immateriil harus dipisahkan secara jelas dalam petitum dengan uraian yang spesifik dan dapat diverifikasi untuk setiap item kerugian. Perlu dipertimbangkan untuk mengajukan permohonan sita jaminan (*conservatoir beslag*) terhadap aset tergugat agar putusan ganti rugi bisa dieksekusi. Untuk ganti rugi immateriil, keterangan ahli dari psikolog klinis yang mendampingi korban jauh lebih efektif daripada argumen abstrak tentang penderitaan.

8.5 HAK KORBAN: GANTI RUGI, PENGHAPUSAN KONTEN, DAN PEMULIHAN NAMA BAIK

Tiga Dimensi Pemulihan bagi Korban Deepfake

Pemulihan yang komprehensif bagi korban deepfake tidak cukup jika hanya berfokus pada satu dimensi. Ada tiga hal yang idealnya harus didapatkan korban secara bersamaan: kompensasi finansial atas kerugian yang sudah terjadi, penghapusan konten yang membatasi kerusakan lebih lanjut, dan pemulihan nama baik yang membangun kembali reputasi yang rusak. Ketiga hal ini saling melengkapi dan tidak bisa sepenuhnya digantikan satu sama lain.

Ganti Rugi: Dua Jalur Utama

Ganti rugi adalah bentuk pemulihan yang paling sering menjadi fokus dalam diskusi hukum, tetapi dalam praktiknya sering yang paling sulit didapatkan korban. Jalur pertama adalah restitusi dalam proses pidana. UU TPKS secara eksplisit mengatur hak korban atas restitusi yang dibebankan kepada pelaku. Restitusi bisa mencakup ganti rugi materiil dan/atau immateriil yang nilainya ditetapkan oleh hakim dalam putusannya. Mekanisme ini lebih mudah karena terintegrasi dalam proses pidana yang sudah berjalan dan korban tidak harus mengajukan gugatan terpisah.

Tantangan dari jalur restitusi adalah pengajuan restitusi harus diajukan sebelum putusan dijatuhkan, jaksa harus memfasilitasinya, dan hakim tidak wajib mengabulkan semua yang diminta. Jika pelaku tidak memiliki aset yang cukup, restitusi yang ditetapkan dalam putusan pun tidak bisa dieksekusi secara efektif. Jalur kedua adalah gugatan perdata terpisah.

Jika proses pidana tidak menghasilkan restitusi yang memadai atau jika proses pidana tidak berjalan sama sekali, korban bisa mengajukan gugatan PMH terpisah di pengadilan perdata. Keuntungannya adalah tidak bergantung pada hasil proses pidana, meskipun prosesnya lebih panjang dan biayanya lebih besar.

Penghapusan Konten: Mekanisme dan Tantangannya

Bagi banyak korban deepfake, penghapusan konten adalah prioritas yang jauh melampaui ganti rugi. Selama konten masih beredar, trauma terus berlangsung dan kerusakan terus bertambah. Ada beberapa mekanisme yang tersedia. Permintaan langsung ke platform adalah jalur tercepat karena semua platform besar memiliki mekanisme pelaporan konten yang melanggar kebijakannya, tetapi tidak selalu efektif, terutama untuk platform yang kurang responsif seperti Telegram.

Laporan ke Komdigi memberikan wewenang untuk memblokir akses ke URL yang memuat konten melanggar hukum Indonesia. Prosesnya lebih lambat dari pelaporan langsung ke platform tetapi lebih mengikat untuk platform yang beroperasi di Indonesia. Putusan pengadilan untuk penghapusan adalah mekanisme yang paling mengikat dalam gugatan PMH, korban bisa meminta hakim memerintahkan tergugat menghapus semua konten deepfake dan membayar uang paksa (dwangsom) harian jika perintah tidak dipatuhi. Ini memakan waktu tetapi memberikan perlindungan hukum paling kuat. Permintaan melalui Bareskrim dalam konteks penyidikan pidana juga efektif karena platform umumnya lebih responsif terhadap permintaan dari lembaga penegak hukum dibandingkan dari individu.

Pemulihan Nama Baik

Kerusakan reputasi yang ditimbulkan oleh deepfake sering kali lebih susah dipulihkan dibanding kerugian finansial. Uang bisa diganti, tetapi kepercayaan orang terhadap seseorang tidak bisa dibeli kembali dengan mudah. Ada beberapa mekanisme pemulihan nama baik dalam sistem hukum Indonesia.

Permohonan pernyataan penyesalan publik bisa diminta dalam gugatan perdata, di mana korban meminta hakim memerintahkan pelaku membuat pernyataan penyesalan yang dipublikasikan di media, baik cetak maupun daring. Mekanisme ini jarang diminta tetapi bisa sangat efektif untuk memulihkan persepsi publik. Hak jawab atau hak koreksi berlaku jika

konten deepfake disebarkan melalui media massa yang terdaftar, menggunakan hak jawab sebagaimana diatur dalam UU Pers.

Publikasi putusan pengadilan yang menyatakan konten deepfake sebagai pelanggaran hukum bisa digunakan sebagai cara "meluruskan catatan" bagi orang-orang yang mungkin sudah melihat deepfake tersebut. Selain jalur hukum, korban kadang perlu mengambil langkah komunikasi aktif untuk menjelaskan situasinya kepada komunitas yang terpengaruh, terutama jika deepfake sudah menyentuh lingkungan profesional atau akademis korban.

Hak Korban dalam UU TPKS: Rangkuman

UU TPKS memberikan serangkaian hak kepada korban yang berlaku juga untuk korban deepfake NCII. Dalam proses hukum, korban berhak mendapatkan informasi tentang perkembangan kasus, pendampingan hukum secara cuma-cuma, perlindungan identitas agar tidak dipublikasikan tanpa izin, sidang tertutup untuk kasus kekerasan seksual, dan hak mengajukan restitusi dalam proses pidana.

Dalam hal pemulihan, korban berhak atas layanan kesehatan fisik dan psikologis dari negara, rehabilitasi sosial, serta pemberdayaan ekonomi jika diperlukan. Dari segi keamanan, korban berhak atas perlindungan dari pelaku dan ancaman selanjutnya, serta hak untuk tidak dikonfrontasikan langsung dengan pelaku selama proses hukum berlangsung.

8.6 SINTESIS: SKEMA PERTANGGUNGJAWABAN YANG KOMPREHENSIF

Memahami siapa yang bisa dijerat, dengan dasar apa, dan melalui jalur mana adalah inti dari pengetahuan praktis dalam penanganan kasus deepfake. Skema berikut merangkum hal tersebut secara menyeluruh.

Pembuat deepfake (pleger) yang secara langsung membuat konten manipulatif tanpa izin dapat dijerat secara pidana melalui UU ITE Pasal 35 jo. Pasal 51, UU TPKS Pasal 14 ayat (1), UU PDP Pasal 66, dan UU Pornografi Pasal 4. Secara perdata, pembuat deepfake dapat digugat PMH berdasarkan Pasal 1365 KUHPerdata untuk ganti rugi materiil dan immateriil.

Penyebar deepfake yang mendistribusikan konten tanpa izin dapat dijerat secara pidana melalui UU ITE Pasal 27 ayat (1) dan Pasal 28, serta UU TPKS Pasal 14 ayat (2). Secara perdata, penyebar dapat digugat PMH dan bertanggung jawab secara tanggung renteng dengan pembuat. Pemesan atau penyuruh yang memerintahkan pembuatan deepfake dijerat berdasarkan Pasal 55 KUHP jo.

semua pasal yang berlaku untuk pembuat, dengan ancaman pidana yang sama. Secara perdata, pemesan biasanya adalah pihak dengan aset terbesar sehingga menjadi sasaran gugatan yang paling efektif. Pembantu kejahatan yang memfasilitasi pembuatan atau penyebaran dijerat berdasarkan Pasal 56 KUHP dengan pengurangan sepertiga ancaman. Secara perdata, tanggung jawab pembantu bersifat proporsional dengan perannya.

Platform digital yang gagal merespons laporan umumnya tidak dapat dipidana tetapi dapat dimintai pertanggungjawaban perdata melalui gugatan PMH sebagai turut tergugat dalam gugatan perdata korban, terutama jika terbukti tidak merespons laporan takedown dalam waktu yang wajar.

Rangkuman Bab

Bab ini telah menelaah secara komprehensif sistem pertanggungjawaban pidana dan perdata dalam kasus deepfake. Pembuktian unsur pidana dalam kasus deepfake, terutama niat (mens rea), adalah tantangan utama. Jaksa harus mampu merekonstruksi niat dari fakta-fakta objektif, karena niat tidak bisa dilihat secara langsung. Perbedaan antara dolus directus, dolus eventualis, dan culpa menentukan strategi dakwaan yang paling efektif.

Doktrin penyertaan melalui Pasal 55 dan 56 KUHP memungkinkan semua pihak dalam rantai kejahatan deepfake dijerat, dari pembuat, penyebar, pemesan, hingga pembantu. Ancaman pidananya berbeda berdasarkan perannya: pelaku dan penyuruh mendapat ancaman yang sama, sedangkan pembantu mendapat pengurangan sepertiga.

Tanggung jawab platform digital bersifat berjenjang: tidak ada pidana jika tidak ada pengetahuan, ada tanggung jawab administratif jika gagal merespons laporan, dan ada potensi tanggung jawab perdata jika terbukti memfasilitasi penyebaran secara aktif. Gugatan PMH berdasarkan Pasal 1365 KUHPerdata memberikan jalur perdata yang fleksibel untuk mendapatkan ganti rugi materiil dan immateriil, dan bisa ditempuh bersamaan dengan proses pidana atau secara terpisah.

Pemulihan korban yang komprehensif memerlukan tiga hal sekaligus: ganti rugi finansial, penghapusan konten, dan pemulihan nama baik. Sistem hukum Indonesia menyediakan mekanisme untuk ketiganya, meski dengan efektivitas yang masih perlu ditingkatkan melalui reformasi legislatif dan penguatan kapasitas institusional.

BAB 9
UNI EROPA: EU AI ACT DAN STANDAR HUKUM DEEPFAKE

Dari seluruh kerangka regulasi yang ada di dunia saat ini, EU AI Act adalah yang paling ambisius, paling komprehensif, dan paling banyak dipelajari. Regulasi ini bukan sekadar regulasi tentang deepfake, melainkan regulasi tentang seluruh ekosistem kecerdasan buatan, dengan deepfake sebagai salah satu dari sekian banyak ancaman yang hendak dikendalikan. Namun justru karena sifatnya yang menyeluruh itulah EU AI Act menjadi referensi yang tidak bisa diabaikan oleh negara mana pun yang serius ingin membangun regulasi AI yang efektif.

Bab ini menguraikan EU AI Act secara sistematis, bukan untuk mengkopikannya mentah-mentah ke konteks Indonesia, tetapi untuk memahami logika di baliknya: mengapa pendekatan berbasis risiko dipilih, bagaimana kewajiban transparansi bekerja, seberapa besar sanksi yang ditetapkan, dan mengapa pengaruhnya sudah terasa jauh sebelum regulasi ini berlaku penuh. Dari pemahaman itu, kita dapat menilai mana yang relevan dan realistis untuk diadaptasi dalam konteks hukum Indonesia.

Yang membuat bab ini penting bukan hanya kandungan EU AI Act itu sendiri, tetapi juga fenomena yang disebut Brussels Effect, yaitu kecenderungan regulasi Uni Eropa untuk menjadi standar de facto global karena ukuran pasar dan kekuatan institusional UE yang mendorong perusahaan multinasional mengadopsi standarnya secara sukarela demi efisiensi operasional.

9.1 MENGAPA EU AI ACT LAHIR DAN APA YANG INGIN DICAPAI

Perjalanan Legislatif yang Panjang

EU AI Act bukanlah produk legislasi yang lahir dalam semalam. Proses pembuatannya dimulai ketika Komisi Eropa menerbitkan White Paper on Artificial Intelligence pada Februari 2020, sebuah dokumen konsultasi publik yang mengajukan pertanyaan mendasar tentang bagaimana Uni Eropa harus menghadapi revolusi AI. Dari situ lahir proposal resmi pada April 2021, yang kemudian melalui proses negosiasi panjang antara Komisi Eropa, Parlemen Eropa, dan Dewan Eropa.

Proses negosiasi itu tidak mudah. Salah satu titik perdebatan paling sengit adalah tentang pengawasan massal menggunakan AI oleh aparat penegak hukum, sebuah isu yang jauh lebih sensitif di Eropa mengingat sejarah Perang Dunia Kedua dibanding di banyak negara lain. Setelah lebih dari tiga tahun proses legislatif, EU AI Act disetujui oleh Parlemen Eropa pada Maret 2024 dengan 523 suara mendukung dan mulai berlaku resmi pada 1 Agustus 2024.

Perjalanan implementasinya dilakukan secara bertahap. Pada 2 Februari 2025 mulai berlaku Bab I tentang definisi dan Bab II tentang praktik AI terlarang. Pada 2 Agustus 2025 mulai berlaku ketentuan tentang *General Purpose AI* (GPAI) dan model berdampak sistemik. Pada 2 Agustus 2026 seluruh regulasi berlaku penuh, termasuk ketentuan tentang AI berisiko tinggi dan transparansi konten. Jadwal implementasi bertahap ini penting dipahami karena beberapa ketentuan sudah berlaku dan mempengaruhi perusahaan teknologi global saat ini.

Tiga Tujuan Utama

EU AI Act dirancang untuk mencapai tiga tujuan yang saling berkaitan. Tujuan pertama adalah melindungi hak-hak fundamental warga, termasuk privasi, kebebasan berekspresi, dan non-diskriminasi, dari ancaman yang ditimbulkan oleh sistem AI yang tidak bertanggung jawab. Tujuan kedua adalah mendorong inovasi yang dapat dipercaya dengan menciptakan kepastian hukum bagi pengembang dan pengguna AI yang beroperasi secara bertanggung jawab.

Tujuan ketiga adalah memposisikan Uni Eropa sebagai pemimpin global dalam pengembangan AI yang etis, sebuah ambisi geopolitik yang tidak bisa diabaikan. Ketiga tujuan ini mencerminkan ketegangan yang inheren dalam regulasi AI: bagaimana melindungi masyarakat dari bahaya tanpa menghambat inovasi yang bermanfaat. Cara EU AI Act menjawab ketegangan itu adalah melalui pendekatan berbasis risiko yang menjadi inti dari seluruh arsitektur regulasinya.

9.2 STRUKTUR EU AI ACT: PENDEKATAN BERBASIS RISIKO

Logika di Balik Pendekatan Berbasis Risiko

Alih-alih melarang AI secara kategoris atau membiarkannya tanpa regulasi sama sekali, EU AI Act mengambil jalan tengah yang pragmatis: mengklasifikasikan sistem AI berdasarkan tingkat risiko yang ditimbulkannya, kemudian menetapkan kewajiban yang proporsional dengan level risiko itu. Semakin tinggi risiko sebuah sistem AI, semakin ketat regulasi yang berlaku untuknya. Pendekatan ini masuk akal karena tidak semua AI sama berbahayanya: AI yang merekomendasikan playlist musik berbeda secara fundamental dari AI yang memutuskan apakah seseorang layak mendapat pinjaman bank.

Sistem klasifikasi risiko dalam EU AI Act terdiri dari empat level. Level pertama adalah Unacceptable Risk atau risiko yang tidak dapat diterima, yang mencakup sistem AI yang mengancam nilai-nilai fundamental, hak asasi, atau keselamatan. Contohnya adalah sistem penilaian sosial oleh pemerintah dan manipulasi bawah sadar. Sistem dalam kategori ini dilarang sepenuhnya tanpa pengecualian.

Level kedua adalah High Risk atau risiko tinggi, yang mencakup sistem AI yang berdampak signifikan pada kehidupan manusia dalam sektor kritis seperti rekrutmen, pemberian kredit, pendidikan, penegakan hukum, dan infrastruktur kritis. Sistem dalam kategori ini wajib menjalani penilaian risiko, transparansi, pengawasan manusia, dan pendaftaran di database UE.

Level ketiga adalah Limited Risk atau risiko terbatas, yang mencakup sistem AI yang menimbulkan risiko khusus terkait manipulasi atau penipuan, seperti chatbot, deepfake, dan AI generatif. Sistem dalam kategori ini wajib memenuhi kewajiban transparansi. Level keempat adalah Minimal Risk atau risiko minimal, yang mencakup sistem AI dengan risiko rendah seperti filter spam dan rekomendasi konten sederhana. Sistem ini tidak memiliki kewajiban khusus dan boleh beroperasi secara bebas.

Deepfake dalam berbagai bentuknya, baik video, audio, maupun gambar sintetis, masuk dalam kategori Limited Risk. Ini berarti deepfake tidak dilarang sepenuhnya oleh EU AI

Act, tetapi harus memenuhi kewajiban transparansi yang spesifik. Pengecualian terjadi ketika deepfake digunakan dengan cara yang masuk kategori Unacceptable Risk, misalnya untuk memanipulasi perilaku orang tanpa sepengetahuan mereka secara subliminal.

General Purpose AI (GPAI): Kategori Baru yang Relevan

Salah satu penambahan paling signifikan dalam negosiasi akhir EU AI Act adalah ketentuan tentang General Purpose AI (GPAI), yaitu model AI yang bisa digunakan untuk berbagai keperluan yang berbeda, seperti GPT-4, Claude, Gemini, Llama, atau Stable Diffusion. Model-model ini sebelumnya sulit dikategorikan karena tidak dirancang untuk satu aplikasi spesifik saja.

EU AI Act menetapkan kewajiban khusus bagi penyedia GPAI, termasuk menyusun dokumentasi teknis, mematuhi aturan hak cipta, dan mempublikasikan ringkasan data pelatihan. Untuk GPAI yang memiliki dampak sistemik, yaitu model yang dilatih dengan lebih dari 10 pangkat 25 FLOP sebagai ukuran komputasi, ada kewajiban tambahan seperti penilaian risiko yang lebih ketat dan pelaporan kepada regulator.

Ketentuan ini langsung relevan dengan deepfake karena model-model GPAI seperti Stable Diffusion, Midjourney, atau Sora adalah yang paling sering digunakan untuk membuat konten sintetis. Dengan mengatur GPAI, EU AI Act secara tidak langsung juga mengatur kemampuan pembuatan deepfake di tingkat infrastruktur teknologi.

9.3 KEWAJIBAN TRANSPARANSI DAN PELABELAN DEEPFAKE

Ketentuan Inti Article 50

Article 50 EU AI Act adalah pasal yang paling langsung dan spesifik berbicara tentang deepfake dan konten sintetis. Pasal ini menetapkan kewajiban transparansi yang berlaku bagi penyedia dan pengguna sistem AI yang menghasilkan konten sintetis. Pemahaman yang baik tentang Article 50 penting karena mencerminkan pilihan desain regulasi yang sangat spesifik: alih-alih melarang deepfake, Uni Eropa memilih untuk membuatnya harus diungkapkan kepada publik.

Article 50 ayat (1) mengatur tentang chatbot: penyedia sistem AI yang berinteraksi dengan manusia wajib memastikan sistem tersebut memberi tahu pengguna bahwa mereka berinteraksi dengan AI, kecuali jika sudah jelas dari konteks. Article 50 ayat (2) mengatur tentang konten sintetis termasuk deepfake: penyedia sistem AI yang menghasilkan konten audio, gambar, video, atau teks sintetis wajib memastikan output ditandai dengan cara yang dapat dideteksi secara mesin (*machine-readable*). Ini mencakup penggunaan watermark digital, metadata, atau kode yang tertanam dalam file.

Article 50 ayat (3) mengatur kewajiban pengguna deepfake: orang yang menggunakan AI untuk menghasilkan atau memanipulasi konten, termasuk deepfake gambar, audio, atau video yang tampak nyata, wajib mengungkapkan bahwa konten tersebut dihasilkan secara artifisial. Ada dua pengecualian: pertama, untuk tujuan penegakan hukum yang sah; kedua, untuk konten seni atau satire yang tidak menimbulkan risiko penipuan material. Article 50 ayat (4) mengatur kewajiban platform: platform yang menyebarkan konten AI-generated secara

besar-besaran wajib mendeteksi dan menandai konten sintetis tersebut menggunakan teknologi yang tersedia.

Poin penting yang perlu dipahami adalah bahwa kewajiban pelabelan berlaku secara otomatis, tidak tergantung pada niat atau apakah konten itu berbahaya. Deepfake yang dibuat untuk tujuan hiburan pun harus dilabeli. Ini adalah pendekatan yang berbeda dari regulasi yang hanya menarget konten berbahaya.

Mekanisme Teknis Pelabelan

EU AI Act tidak menetapkan standar teknis pelabelan yang spesifik secara detail; itu diserahkan kepada proses standardisasi lanjutan. Namun regulasi ini menegaskan bahwa pelabelan harus dilakukan dengan cara yang dapat dideteksi secara mesin (*machine-readable*), bukan hanya secara visual yang mudah dihapus.

Ada beberapa teknologi yang sedang dikembangkan untuk memenuhi persyaratan ini. Pertama, watermark digital: tanda tersembunyi yang disisipkan dalam piksel gambar atau sinyal audio, yang tidak terlihat secara kasat mata dan sulit dihapus secara tidak sengaja, meskipun bisa hilang melalui konversi format atau kompresi ulang. Kedua, C2PA Content Credentials (*Coalition for Content Provenance and Authenticity*): metadata kriptografis yang menyertai file dan berisi riwayat pembuatan konten, yang dapat diverifikasi secara independen menggunakan standar terbuka. Ini adalah standar yang paling maju dan sudah mulai diterapkan oleh Adobe, Microsoft, dan Google.

Ketiga, model AI deteksi deepfake: algoritma yang mendeteksi pola khas AI generatif dalam konten, yang bekerja tanpa memerlukan watermark di sisi produsen tetapi akurasinya tidak 100% dan sering kalah langkah dengan model generasi terbaru. Keempat, metadata platform: label yang ditambahkan oleh platform saat konten diunggah jika terdeteksi sebagai konten AI, yang bekerja di level distribusi bukan produksi dan bergantung pada kebijakan masing-masing platform.

Pengecualian untuk Seni dan Satire

Article 50 memuat dua pengecualian penting dari kewajiban pelabelan: penegakan hukum yang sah dan konten seni atau satire. Pengecualian ini mencerminkan keseimbangan yang EU AI Act coba capai antara perlindungan publik dan kebebasan berekspresi. Pengecualian untuk seni dan satire adalah yang paling sensitif secara hukum karena menimbulkan pertanyaan tentang siapa yang memutuskan apakah sebuah konten adalah satire atau manipulasi.

EU AI Act menyerahkan penilaian ini kepada konteks. Pengecualian tidak berlaku jika konten sintetis tersebut "menimbulkan risiko penipuan material", artinya satire yang terasa meyakinkan sebagai berita nyata tidak bisa berlindung di balik pengecualian ini. Ini adalah pilihan yang memang tidak sempurna tetapi mencerminkan kesulitan fundamental dalam meregulasi ekspresi kreatif. Bagi Indonesia yang sedang merancang regulasi serupa, pertanyaan tentang bagaimana mendefinisikan batas satire yang sah adalah salah satu tantangan desain regulasi yang paling sulit.

Implikasi Article 50 untuk Indonesia

Ketentuan Article 50 EU AI Act memiliki implikasi langsung bagi Indonesia meskipun Indonesia bukan anggota UE. Ada empat mekanisme implikasi yang perlu dipahami. Pertama, perusahaan teknologi global seperti Adobe, Microsoft, Meta, Google, dan OpenAI yang melayani pengguna Indonesia sudah mulai atau akan mengimplementasikan C2PA dan ketentuan pelabelan untuk memenuhi EU AI Act. Pengguna Indonesia ikut terdampak secara langsung meskipun tidak ada regulasi domestik yang mengharuskan.

Kedua, platform yang menyebarkan konten di UE harus mendeteksi dan menandai deepfake. Karena platform yang sama juga beroperasi di Indonesia, mekanisme deteksi itu bisa diperluas ke pasar Indonesia dengan relatif mudah. Ketiga, Indonesia bisa mengadopsi kewajiban pelabelan serupa dalam regulasi domestik dengan memanfaatkan standar teknis C2PA yang sudah dikembangkan untuk EU AI Act, tanpa harus memulai dari nol. Keempat, tantangan utamanya adalah bahwa standar pelabelan yang efektif memerlukan kerja sama dari produsen perangkat AI, platform distribusi, dan browser; sebuah ekosistem yang tidak mudah dibangun tanpa kewenangan regulasi yang kuat.

9.4 SANKSI: DENDA HINGGA 35 JUTA EURO ATAU 7 PERSEN OMZET GLOBAL

Skema Sanksi Berjenjang

Salah satu aspek EU AI Act yang paling banyak mendapat perhatian dan yang paling mempengaruhi perilaku perusahaan teknologi global adalah skema sanksinya. Besaran denda yang ditetapkan dirancang cukup besar untuk terasa menyakitkan bahkan bagi perusahaan teknologi terbesar di dunia. Sanksi dalam EU AI Act bersifat berjenjang sesuai tingkat keparahan pelanggaran.

Pelanggaran paling serius, yaitu melanggar ketentuan tentang praktik AI terlarang dalam Article 5, diancam dengan denda maksimal 35 juta euro atau 7 persen dari total omzet global tahunan, mana yang lebih tinggi. Kategori pelanggaran ini mencakup sistem AI yang memanipulasi perilaku manusia, penilaian sosial, dan pengawasan massal yang tidak sah.

Pelanggaran tingkat menengah, yaitu melanggar ketentuan tentang AI berisiko tinggi, GPAI, dan transparansi termasuk kegagalan memenuhi Article 50 tentang pelabelan deepfake, diancam dengan denda maksimal 15 juta euro atau 3 persen dari total omzet global tahunan, mana yang lebih tinggi. Pelanggaran administratif, yaitu memberikan informasi yang salah kepada regulator, diancam dengan denda maksimal 7,5 juta euro atau 1,5 persen dari total omzet global tahunan, mana yang lebih tinggi. EU AI Act juga memberikan kelonggaran bagi usaha kecil dan menengah dengan batas bawah yang lebih rendah dari semua kategori.

Mengapa Angka Ini Bermakna Secara Global

Untuk memahami mengapa sanksi ini efektif secara global, perlu dilihat konteksnya dengan cermat. Denda maksimal 35 juta euro terdengar sangat besar bagi perusahaan kecil, tetapi bagi perusahaan seperti Google, Meta, atau Microsoft, angka nominal itu relatif kecil dibandingkan pendapatan mereka. Yang benar-benar menyakitkan adalah persentase omzet. Google, misalnya, memiliki pendapatan global sekitar 280 miliar dolar AS pada 2023. Tujuh persen dari angka itu adalah sekitar 19,6 miliar dolar AS, jauh lebih besar dari 35 juta euro. Ini

berarti bahwa untuk perusahaan besar, sanksi berdasarkan persentase omzetlah yang relevan, dan angka itu cukup besar untuk mendorong perubahan perilaku yang nyata.

Perbandingan dengan GDPR sangat menarik. GDPR (*General Data Protection Regulation*) yang berlaku sejak 2018 memiliki denda maksimal 4 persen dari omzet global. EU AI Act melampaui itu dengan 7 persen untuk pelanggaran terberat. Ini menunjukkan bahwa UE memandang pelanggaran regulasi AI sebagai lebih serius dari sekadar pelanggaran privasi data. Bagi perusahaan besar, kepatuhan terhadap EU AI Act jelas jauh lebih murah daripada menghadapi sanksi persentase omzet tersebut. Desain sanksi ini disengaja untuk menciptakan kalkulasi ekonomi yang mendorong kepatuhan.

Penegakan: Lembaga Pengawas Dua Tingkat

EU AI Act membentuk dua lembaga pengawas baru yang bekerja secara berjenjang. Pertama, AI Office di tingkat Uni Eropa yang bertanggung jawab mengawasi kepatuhan terhadap ketentuan tentang GPAI dan model AI berdampak sistemik secara lintas batas negara anggota. Kedua, National Competent Authorities di setiap negara anggota yang bertanggung jawab mengawasi ketentuan yang berlaku di tingkat nasional masing-masing.

Model pengawasan dua tingkat ini layak dicermati oleh Indonesia. Struktur serupa, dengan adaptasi terhadap sistem pemerintahan Indonesia, bisa menjadi template untuk lembaga pengawas AI nasional yang perlu dibentuk. Koordinasi antara lembaga nasional dan mekanisme internasional adalah elemen krusial yang perlu dirancang sejak awal agar penegakan hukum tidak terhambat oleh masalah yurisdiksi.

9.5 PRAKTIK AI TERLARANG: BERLAKU SEJAK FEBRUARI 2025

Apa yang Dilarang dan Mengapa

Dari seluruh EU AI Act, Bab II tentang Prohibited AI Practices adalah yang berlaku paling awal, yaitu sejak 2 Februari 2025. Ini mencerminkan penilaian UE bahwa praktik-praktik dalam kategori ini sudah cukup jelas berbahaya untuk dilarang tanpa menunggu periode transisi lebih lama. Berlakunya ketentuan ini jauh sebelum regulasi secara keseluruhan berlaku penuh pada Agustus 2026 adalah sinyal politik yang kuat tentang prioritas regulasi UE.

Larangan pertama menyangkut manipulasi bawah sadar. Sistem AI yang menggunakan teknik subliminal, manipulatif, atau menipu yang beroperasi di bawah batas kesadaran seseorang untuk mempengaruhi perilaku dengan cara yang bisa merugikan mereka dilarang sepenuhnya. Dalam konteks deepfake, larangan ini mencakup video deepfake yang dirancang khusus untuk memanipulasi respons emosional korban secara tidak sadar, misalnya deepfake yang menggunakan teknik persuasi subliminal tersembunyi.

Larangan kedua menyangkut eksploitasi kerentanan. Sistem AI yang mengeksploitasi kerentanan seseorang berdasarkan usia, disabilitas, atau situasi sosial ekonomi untuk mempengaruhi perilaku dengan cara yang merugikan dilarang. Dalam konteks deepfake, larangan ini mencakup deepfake yang menarget kelompok rentan seperti lansia dengan skenario penipuan yang memanfaatkan wajah figur otoritas yang dipercaya, misalnya deepfake dokter atau pejabat pemerintah. Larangan ketiga menyangkut social scoring atau penilaian sosial oleh entitas publik yang dapat merugikan individu berdasarkan perilaku mereka.

Meskipun tidak langsung terkait deepfake, larangan ini mencerminkan filosofi regulasi yang melarang AI digunakan untuk mengendalikan atau menilai perilaku sosial warga negara.

Larangan keempat menyangkut identifikasi biometrik real-time di ruang publik oleh aparat penegak hukum. Ada pengecualian sempit untuk kejahatan serius dan terorisme, tetapi penggunaan umum sistem identifikasi biometrik real-time dilarang. Larangan ini relevan karena teknologi pengenalan wajah secara real-time juga menjadi fondasi banyak sistem deepfake. Larangan kelima menyangkut penginferensian emosi, yaitu sistem AI yang menyimpulkan emosi seseorang dari ekspresi wajah dalam konteks tempat kerja dan pendidikan. Relevansinya tidak langsung tetapi berkaitan dengan teknologi analisis wajah yang juga digunakan dalam pembuatan dan deteksi deepfake.

Implikasi Larangan-Larangan Ini

Larangan-larangan dalam Article 5 mencerminkan garis merah yang UE anggap tidak bisa dikompromikan: penggunaan AI untuk memanipulasi, menipu, atau mengendalikan manusia tanpa sepengetahuan dan persetujuan mereka. Bagi deepfake, ini berarti bahwa deepfake yang dirancang secara spesifik untuk memanipulasi perilaku seseorang secara tidak sadar tidak hanya melanggar aturan transparansi Article 50, tetapi masuk dalam kategori yang dilarang sepenuhnya dengan sanksi terberat.

Secara praktis, ketentuan ini menimbulkan pertanyaan menarik: apakah deepfake disinformasi yang disebarkan menjelang pemilu masuk dalam kategori "manipulasi bawah sadar" yang dilarang oleh Article 5? Jawabannya tergantung pada interpretasi frasa "subliminal atau manipulatif", dan ini kemungkinan akan menjadi area yang diperdebatkan dalam penerapan hukum ke depan. Bagi Indonesia yang merancang regulasi serupa, bagaimana mendefinisikan batas antara "manipulasi yang dilarang" dan "persuasi yang sah" adalah pertanyaan desain hukum yang harus dijawab dengan hati-hati.

9.6 CODE OF PRACTICE ON AI-GENERATED CONTENT

Apa Itu Code of Practice dan Mengapa Diperlukan

EU AI Act adalah kerangka hukum yang mengikat, tetapi tidak bisa beroperasi sendiri. Regulasi tingkat tinggi memerlukan panduan implementasi yang lebih teknis dan spesifik, panduan yang bisa diperbarui lebih cepat dari proses legislasi formal dan yang melibatkan keahlian industri secara aktif. Di sinilah Code of Practice berperan.

Code of Practice on AI-Generated Content adalah panduan yang dikembangkan bersama antara regulator UE, perusahaan teknologi, peneliti, dan masyarakat sipil. Targetnya adalah memberikan panduan operasional tentang bagaimana kewajiban Article 50 tentang pelabelan konten AI, termasuk deepfake, diimplementasikan dalam praktik sehari-hari. Pendekatan partisipatif ini berbeda dari model regulasi top-down yang lebih umum di banyak negara dan dimaksudkan untuk menghasilkan standar yang lebih mudah diimplementasikan karena melibatkan mereka yang akan menerapkannya.

Proses Pengembangan dan Target Finalisasi

AI Office Uni Eropa memimpin proses pengembangan Code of Practice dengan melibatkan ratusan organisasi, dari perusahaan teknologi besar seperti Google, Meta, dan

OpenAI, hingga akademisi, LSM, dan perwakilan masyarakat sipil. Proses ini berlangsung melalui beberapa putaran konsultasi dan diskusi kelompok kerja yang tersruktur.

Target finalisasi Code of Practice on AI-Generated Content dijadwalkan pada Mei hingga Juni 2026. Bagi Indonesia, perkembangan ini sangat relevan karena Code of Practice yang dihasilkan akan menjadi panduan implementasi teknis yang komprehensif dan dapat dipelajari langsung tanpa harus mengembangkan panduan serupa dari nol.

Area-area yang dibahas dalam Code of Practice mencakup standar teknis watermarking dan C2PA untuk menentukan format dan metode pelabelan yang tahan terhadap manipulasi, prosedur takedown untuk konten AI berbahaya dengan batas waktu respons yang mengikat platform, kewajiban pelaporan bagi penyedia GPAI kepada AI Office, standar pengujian keamanan untuk model AI sebelum dirilis, dan hak-hak pengguna atas konten sintetis termasuk hak untuk mengetahui, menolak, dan mendapat pemulihan atas penggunaan data biometriknya.

9.7 BRUSSELS EFFECT: PENGARUH REGULASI UE TERHADAP STANDAR GLOBAL

Memahami Brussels Effect

Brussels Effect adalah fenomena yang pertama kali dianalisis secara sistematis oleh Anu Bradford dari Columbia Law School dalam bukunya *The Brussels Effect* (2020). Konsepnya sederhana tetapi implikasinya luar biasa: karena pasar UE sangat besar dan sangat terpadu, perusahaan multinasional yang ingin beroperasi di dalamnya harus memenuhi standar regulasinya. Karena jauh lebih efisien untuk menerapkan satu standar yang berlaku di semua pasar daripada standar yang berbeda-beda di setiap negara, perusahaan akhirnya menerapkan standar UE di seluruh dunia, termasuk di negara-negara yang tidak memiliki regulasi serupa.

Contoh paling terkenal adalah GDPR yang berlaku sejak 2018. Meskipun GDPR hanya berlaku di UE, regulasi tersebut secara efektif menjadi standar global privasi data karena hampir semua perusahaan teknologi besar yang beroperasi di UE juga beroperasi di seluruh dunia dan memilih untuk menerapkan standar GDPR secara universal daripada memilah-milah standar per negara.

Brussels Effect dalam Konteks EU AI Act

EU AI Act berpotensi menghasilkan Brussels Effect yang serupa atau bahkan lebih besar dari GDPR karena beberapa alasan. Pertama, pasar digital UE dengan 450 juta konsumen terlalu besar untuk diabaikan oleh perusahaan teknologi global mana pun. Kedua, biaya membangun sistem yang berbeda untuk UE dan non-UE jauh lebih tinggi dari biaya menerapkan standar UE secara universal. Ketiga, standar EU AI Act terutama tentang pelabelan dan transparansi bersifat teknis dan bisa diimplementasikan di level infrastruktur yang berlaku lintas pasar secara serentak.

Ada empat mekanisme utama bagaimana Brussels Effect dari EU AI Act berdampak bagi Indonesia. Melalui mekanisme market access requirement, perusahaan yang ingin beroperasi di UE harus memenuhi EU AI Act, sehingga platform yang melayani Indonesia dan UE secara bersamaan akan menerapkan standar EU AI Act juga di Indonesia. Melalui mekanisme single global product standard, lebih efisien membuat satu produk yang memenuhi standar tertinggi,

sehingga produk AI yang dijual ke Indonesia akan sudah memiliki fitur pelabelan deepfake yang diharuskan EU AI Act.

Melalui mekanisme regulatory arbitrage prevention, standar global mencegah perusahaan melarikan diri ke yurisdiksi yang lebih longgar, sehingga Indonesia tidak bisa menjadi surga regulasi untuk platform deepfake berbahaya jika platform itu juga beroperasi di UE. Melalui mekanisme reputational spillover, merek global tidak ingin dipersepsi memiliki standar berbeda di negara berbeda, sehingga konsumen Indonesia bisa memanfaatkan mekanisme perlindungan yang perusahaan terapkan karena tekanan EU AI Act.

Brussels Effect Sudah Berlangsung Sekarang

Brussels Effect dari EU AI Act tidak perlu menunggu berlakunya secara penuh pada Agustus 2026 karena sudah berlangsung saat ini. Beberapa contoh nyata yang dapat diamati langsung oleh pengguna Indonesia. Meta (Facebook/Instagram) sudah mengumumkan akan menerapkan pelabelan konten AI secara global, bukan hanya di UE, yang sebagian didorong oleh antisipasi terhadap EU AI Act.

Google mengintegrasikan C2PA Content Credentials dalam produk-produknya termasuk Google Images dan YouTube, yang berarti pengguna di Indonesia pun dapat melihat label konten AI pada konten yang relevan. Adobe sudah mengimplementasikan Content Authenticity Initiative berbasis C2PA di seluruh produknya termasuk Photoshop dan Firefly yang digunakan secara luas di Indonesia. OpenAI menerapkan watermarking pada gambar yang dihasilkan DALL-E sebagai antisipasi terhadap kewajiban EU AI Act.

Indonesia tidak memiliki regulasi deepfake yang komprehensif, sudah merasakan sebagian dampak dari EU AI Act melalui perubahan perilaku platform global. Ini adalah argumen yang kuat untuk mengadopsi elemen-elemen EU AI Act dalam regulasi domestik karena infrastruktur teknis untuk implementasinya sudah sebagian dibangun oleh perusahaan teknologi global tanpa perlu dimulai dari nol.

9.8 PELAJARAN DARI EU AI ACT UNTUK INDONESIA

Apa yang Bisa dan Tidak Bisa Diadopsi

Mengadopsi EU AI Act secara utuh bukan pilihan yang realistis atau bahkan yang tepat untuk Indonesia. EU AI Act lahir dari konteks institusional, kapasitas regulasi, dan nilai-nilai konstitusional UE yang berbeda dari Indonesia. Yang bisa dan sebaiknya dilakukan adalah mengidentifikasi prinsip-prinsip dan mekanisme spesifik yang relevan untuk diadaptasi secara selektif.

Pendekatan berbasis risiko adalah elemen yang sangat relevan dan realistis untuk diadopsi Indonesia karena memungkinkan regulasi proporsional yang tidak menghambat inovasi bermanfaat. Elemen ini bisa menjadi prinsip dasar regulasi AI Indonesia tanpa perlu mengkopi seluruh sistem klasifikasi EU AI Act secara verbatim.

Kewajiban pelabelan deepfake yang diatur dalam Article 50 juga sangat relevan dan teknologinya sudah tersedia melalui standar C2PA. Elemen ini bisa diadopsi dalam regulasi Komdigi atau amandemen UU ITE dengan relatif cepat karena tidak memerlukan pembangunan infrastruktur teknis baru dari pihak Indonesia; infrastrukturnya sudah dibangun

oleh platform global. Larangan manipulasi bawah sadar yang termuat dalam Article 5 relevan secara prinsip untuk Indonesia tetapi memerlukan definisi yang lebih operasional agar bisa ditegakkan secara efektif dalam konteks sistem hukum Indonesia.

Sementara itu, beberapa elemen EU AI Act belum realistis untuk diadopsi dalam jangka pendek. Denda sebesar 7 persen omzet global ideal secara konsep tetapi tidak realistis tanpa infrastruktur penegakan yang kuat terlebih dahulu. AI Office yang independen sangat dibutuhkan tetapi memerlukan investasi institusional yang besar dan bisa dimulai secara bertahap sebagai unit di Komdigi atau BRIN. Kewajiban GPAI relevan tetapi Indonesia belum memiliki perusahaan GPAI domestik besar, sehingga bisa diterapkan secara parsial pada penggunaan model GPAI asing di Indonesia.

Strategi Adopsi Bertahap

Berdasarkan analisis di atas, strategi adopsi yang realistis untuk Indonesia bisa dilakukan dalam tiga tahap yang berkesinambungan.

Tahap pertama mencakup adopsi prinsip dan mekanisme teknis dalam rentang waktu 2025 hingga 2026. Langkah konkretnya adalah mewajibkan platform digital yang beroperasi di Indonesia untuk menerapkan pelabelan konten AI menggunakan standar C2PA atau yang setara. Hal ini tidak memerlukan undang-undang baru karena bisa dilakukan melalui Peraturan Menteri Komdigi yang mewajibkan platform PSE untuk mendeteksi dan menandai konten deepfake. Karena platform global sudah menerapkan ini untuk kepatuhan EU AI Act, memperluas kewajiban yang sama ke Indonesia hanyalah masalah kebijakan, bukan masalah teknis.

Tahap kedua mencakup reformasi legislatif dalam rentang waktu 2026 hingga 2027. Langkah konkretnya adalah mengamandemen UU ITE untuk memasukkan definisi konten sintetis berbasis AI dan kewajiban transparansi yang eksplisit, mengembangkan UU Hak Cipta untuk mengakomodasi pertanyaan tentang kepemilikan dan tanggung jawab atas konten yang dihasilkan AI, serta membentuk unit pengawas AI di bawah Komdigi dengan mandat dan anggaran yang jelas sebagai cikal bakal lembaga pengawas yang lebih independen.

Tahap ketiga mencakup pembangunan kerangka AI nasional yang komprehensif dalam rentang waktu 2027 hingga 2030. Langkah konkretnya adalah mengembangkan Undang-Undang Kecerdasan Buatan Nasional yang mengadopsi pendekatan berbasis risiko EU AI Act namun disesuaikan dengan kapasitas institusional dan prioritas pembangunan Indonesia, membentuk lembaga pengawas AI yang independen dengan kapasitas teknis yang memadai, serta berpartisipasi aktif dalam forum-forum internasional tentang standar AI global agar Indonesia tidak hanya menjadi penerima pasif dari standar yang ditetapkan orang lain.

9.9 EU AI ACT SEBAGAI CERMIN, BUKAN BLUEPRINT

EU AI Act adalah pencapaian regulasi yang luar biasa sebagai hasil dari proses demokratis yang melibatkan ratusan pemangku kepentingan selama bertahun-tahun. Namun perlu diingat bahwa regulasi tersebut merupakan produk dari konteks Eropa: kapasitas institusional yang kuat, tradisi hak asasi yang mapan, dan anggaran regulasi yang memadai. Kondisi-kondisi tersebut berbeda dari kondisi Indonesia saat ini.

Indonesia tidak harus mengkopi EU AI Act. Yang perlu dilakukan adalah belajar dari logikanya: mengapa pendekatan berbasis risiko lebih efektif dari larangan kategoris, mengapa transparansi lebih berkelanjutan dari sekadar kriminalisasi, dan mengapa melibatkan industri dalam pembuatan standar menghasilkan regulasi yang lebih mudah diimplementasikan. Logika-logika inilah yang merupakan warisan terpenting dari EU AI Act untuk negara-negara yang sedang merancang regulasi AI mereka.

EU AI Act adalah cermin yang berguna untuk melihat ke mana arah regulasi AI global bergerak. Indonesia harus memutuskan sendiri seberapa cepat dan ke arah mana ingin melangkah, tetapi melangkah tanpa peta sama sekali bukan pilihan yang bisa dipertahankan di tengah ancaman deepfake yang nyata dan terus berkembang. Setiap hari tanpa regulasi yang memadai adalah hari di mana korban deepfake tidak memperoleh perlindungan yang layak mereka dapatkan.

Rangkuman Bab

Bab ini telah menguraikan EU AI Act secara sistematis dari lahirnya hingga implikasi globalnya. EU AI Act lahir dari proses legislatif panjang sejak 2020 dan berlaku penuh pada Agustus 2026. Regulasi tersebut merupakan regulasi AI paling komprehensif di dunia dan menjadi referensi yang tidak bisa diabaikan oleh siapa pun yang serius membangun regulasi AI.

Pendekatan berbasis risiko adalah inti dari EU AI Act yang mengklasifikasikan sistem AI berdasarkan tingkat bahayanya dan menetapkan kewajiban yang proporsional. Deepfake masuk dalam kategori Limited Risk yang tunduk pada kewajiban transparansi, bukan larangan total. Article 50 mewajibkan pelabelan konten sintetis yang dapat dideteksi secara mesin, dan teknologi implementasinya melalui standar C2PA sudah tersedia dan sudah mulai diterapkan oleh platform global.

Skema sanksi yang besar hingga 35 juta euro atau 7 persen omzet global sudah mendorong perubahan perilaku perusahaan teknologi global bahkan sebelum EU AI Act berlaku penuh. Prohibited AI Practices yang berlaku sejak Februari 2025 melarang penggunaan AI untuk manipulasi bawah sadar, eksploitasi kerentanan, dan penilaian sosial, mencerminkan garis merah yang UE anggap tidak bisa dikompromikan dalam regulasi kecerdasan buatan.

Brussels Effect berarti pengaruh EU AI Act sudah terasa di Indonesia meskipun bukan anggota UE karena platform global menerapkan standar EU AI Act secara universal demi efisiensi operasional. Indonesia dapat dan seharusnya memanfaatkan momentum ini untuk menetapkan kewajiban serupa dalam regulasi domestik. Strategi adopsi bertahap yang dimulai dari mekanisme teknis, dilanjutkan dengan reformasi legislatif, dan diakhiri dengan kerangka AI nasional yang komprehensif adalah pendekatan yang paling realistis dan efektif untuk konteks Indonesia saat ini.

BAB 10
AMERIKA SERIKAT, ASIA, DAN NEGARA-NEGARA PIONIR REGULASI

"Teknologi selalu berlari lebih cepat dari hukum. Tugas kita bukan mengejarnya, melainkan memastikan kita berlari ke arah yang benar."

parafrase dari diskursus kebijakan teknologi kontemporer

Regulasi deepfake tidak muncul dari kekosongan. Ia lahir dari akumulasi kegelisahan yang tumbuh perlahan, kemudian meledak dalam bentuk skandal-skandal yang cukup nyata untuk memaksa para pembuat kebijakan bergerak. Seorang politisi yang suaranya dikloning untuk menyebarkan pernyataan palsu. Seorang perempuan yang wajahnya ditempelkan pada konten pornografi lalu disebar ke rekan-rekan kerjanya. Seorang remaja yang deepfake-nya beredar di grup-grup tertutup sekolahnya. Kasus-kasus seperti ini dan ribuan yang tidak pernah dilaporkan adalah tekanan riil yang akhirnya menggerakkan roda legislasi.

Hal yang menarik dari peta regulasi global deepfake adalah ia tidak berkembang secara linier atau seragam. Tidak ada satu model yang kemudian diadopsi oleh semua negara. Sebaliknya, kita menyaksikan semacam percobaan paralel: masing-masing yurisdiksi merespons dengan instrumen, filosofi, dan prioritas yang berbeda-beda dipengaruhi oleh tradisi hukumnya, tekanan politiknya, dan konteks sosialnya masing-masing. Bab ini memetakan lima yurisdiksi yang paling signifikan sebagai pionir regulasi: Amerika Serikat, China, Korea Selatan, Inggris, dan Prancis. Keberhasilan dan keterbatasan masing-masing memberikan pelajaran yang tidak bisa diabaikan oleh siapa pun yang serius memikirkan tata kelola teknologi di masa depan.

Perlu dicatat sejak awal bahwa "regulasi deepfake" adalah istilah yang lebih luas dari yang tampaknya. Ia mencakup setidaknya tiga domain berbeda yang kerap tumpang tindih: deepfake seksual atau intim (yang menargetkan individu, sering berbasis gender), deepfake politik (yang menargetkan proses demokrasi), dan deepfake komersial atau penipuan (yang menargetkan transaksi ekonomi). Tidak semua yurisdiksi mengatur ketiganya, dan prioritas yang dipilih masing-masing negara mencerminkan kekhawatiran sosial yang berbeda-beda.

10.1 MENGAPA REGULASI BARU INI DIBUTUHKAN: KEGAGALAN KERANGKA HUKUM LAMA

Sebelum membahas undang-undang spesifik, penting untuk memahami mengapa kerangka hukum yang sudah ada dianggap tidak memadai. Ini bukan pertanyaan retoris jawabannya menentukan arah dan bentuk regulasi baru yang kemudian dibangun. Hukum pencemaran nama baik *(defamation law)* adalah kandidat pertama yang tampaknya bisa merespons masalah deepfake. Jika seseorang membuat deepfake yang menampilkan orang lain melakukan sesuatu yang tidak pernah mereka lakukan, bukankah itu fitnah? Secara intuitif, ya. Tapi hukum pencemaran nama baik di sebagian besar yurisdiksi mensyaratkan bahwa konten yang

disengketakan harus berupa pernyataan faktual bukan sekadar gambar atau video. Selain itu, dalam tradisi hukum Amerika, standar pembuktian untuk tokoh publik sangat tinggi (*actual malice*), sementara untuk individu biasa prosesnya tetap panjang dan mahal. Lebih fundamental lagi: deepfake seksual yang tidak disebarkan secara publik tapi tetap membuat korban trauma tidak selalu memenuhi syarat untuk tuntutan pencemaran nama baik.

Hukum hak cipta juga pernah dicoba sebagai instrumen, terutama ketika deepfake menggunakan gambar atau rekaman yang sudah ada. Tapi ini menghadapi masalah mendasar: yang dilanggar dalam deepfake seksual bukan hak cipta si korban atas gambar dirinya (karena gambar aslinya mungkin diambil dari media sosial atau sumber publik lainnya), melainkan hak atas privasi, otonomi tubuh, dan martabat. Hak cipta tidak dirancang untuk melindungi hal-hal itu.

Undang-undang privasi memberikan pijakan yang sedikit lebih kuat, terutama di yurisdiksi yang memiliki kerangka privasi komprehensif. Tapi bahkan di sana, deepfake menimbulkan pertanyaan yang belum dijawab: apakah wajah seseorang adalah "data pribadi" yang dilindungi? Apakah pembuatan deepfake dari foto publik seseorang merupakan pemrosesan data yang tidak sah? Jawabannya berbeda-beda tergantung yurisdiksi, dan bahkan di tempat di mana jawabannya "ya", mekanisme penegakannya sering tidak siap menghadapi skala dan kecepatan penyebaran konten deepfake.

Kekosongan inilah yang kemudian mendorong pembuat kebijakan untuk mempertimbangkan legislasi spesifik deepfake. Tidak selalu mudah untuk meyakinkan kolega-kolega di parlemen atau kongres bahwa masalah ini memerlukan undang-undang baru ada selalu argumen bahwa hukum yang ada sudah cukup, atau bahwa regulasi baru akan menimbulkan efek dingin pada kebebasan berbicara. Argumen-argumen ini sering kali disuarakan oleh koalisi yang cukup beragam: dari libertarian yang murni khawatir soal sensor, hingga pelobi industri teknologi yang tidak ingin platform mereka dibebani kewajiban baru.

Tapi tekanan dari kelompok advokasi korban, dari akademisi hukum, dan dari publik yang semakin sadar akan bahaya nyata deepfake, akhirnya lebih berat. Dan ketika satu yurisdiksi mulai bergerak, yang lain mulai merasakan tekanan untuk tidak tertinggal.

10.2 AMERIKA SERIKAT: DARI CHAOS FEDERALISME KE HUKUM FEDERAL PERTAMA

Struktur Masalah: Mengapa AS Tertatih-tatih

Amerika Serikat adalah kasus yang paling kompleks di antara lima yurisdiksi yang kita bahas bukan karena kurangnya kesadaran akan masalahnya, melainkan karena arsitektur konstitusional negara itu sendiri. Sistem federalisme Amerika berarti bahwa negara-negara bagian memiliki kewenangan legislatif yang signifikan di bidang-bidang yang tidak secara eksplisit diserahkan kepada pemerintah federal. Hukum pidana dan hukum perdata untuk sebagian besar kejahatan terhadap individu termasuk kejahatan berbasis teknologi secara tradisional merupakan domain negara bagian.

Akibatnya, ketika masalah deepfake mulai mendapat perhatian serius sekitar tahun 2018-2019, yang bergerak pertama adalah negara-negara bagian. Virginia menjadi yang pertama melarang distribusi deepfake pornografi pada 2019. California menyusul dengan aturan yang melarang deepfake dalam konteks pemilu. Texas mengkriminalisasi distribusi deepfake seksual. Dalam beberapa tahun berikutnya, gelombang legislasi negara bagian ini berkembang menjadi lebih dari 45 undang-undang sebuah angka yang, tergantung cara pandangnya, bisa dibaca sebagai bukti responsivitas sistem federal, atau sebagai bukti betapa tidak terkoordinasinya respons Amerika terhadap masalah nasional bahkan global ini.

Dari sisi praktis, fragmentasi ini menimbulkan masalah nyata. Sebuah deepfake yang dibuat di satu negara bagian dan disebarkan ke negara bagian lain melalui platform internet yang merupakan skenario paling umum tiba-tiba masuk ke wilayah abu-abu yurisdiksi. Korban di negara bagian yang belum memiliki undang-undang spesifik deepfake menghadapi jalur hukum yang jauh lebih sempit. Dan platform-platform besar, yang beroperasi secara nasional bahkan global, menghadapi tumpukan kewajiban yang berbeda-beda dari setiap negara bagian.

TAKE IT DOWN Act: Hukum Federal Pertama

TAKE IT DOWN Act yang ditandatangani Presiden pada Mei 2025 adalah momen yang signifikan, meski bukan tanpa kontroversi. Undang-undang ini menetapkan dua hal utama: pertama, ia mengkriminalisasi pembuatan dan distribusi konten intim non-konsensual yang dihasilkan atau dimodifikasi oleh AI (deepfake seksual); kedua, ia mewajibkan platform online untuk menghapus konten semacam itu dalam waktu yang ditentukan setelah menerima laporan dari korban. Ini adalah kali pertama pemerintah federal Amerika secara eksplisit mengatur deepfake seksual.

Nama undang-undang ini TAKE IT DOWN adalah akronim yang sedikit dipaksakan *(Tools to Address Known Exploitation by Immobilizing the Technological Infrastructure for Deepfake Now)*, tapi ia mencerminkan semangat legislasinya: respons cepat, orientasi pada penghapusan konten, dan fokus pada eksploitasi seksual. Penting untuk dipahami bahwa undang-undang ini tidak mengatur semua jenis deepfake dan tidak menyentuh deepfake politik, deepfake keuangan, atau deepfake satire. Cakupannya disengaja lebih sempit, tampaknya sebagai strategi untuk memaksimalkan kemungkinan lolos dari kongres yang terpecah.

Dari sisi korban, undang-undang ini memberikan dua hal yang sebelumnya tidak ada: mekanisme penghapusan yang lebih jelas dan lebih cepat, serta dasar hukum federal untuk penuntutan pidana. Tapi beberapa pengkritik mencatat bahwa ketentuan penghapusan bisa disalahgunakan jika prosesnya terlalu mudah dan tidak memerlukan verifikasi yang memadai, ia bisa menjadi alat bagi pihak yang tidak beritikad baik untuk menyensor konten yang sebenarnya sah. Ini adalah ketegangan yang belum sepenuhnya diselesaikan oleh teks undang-undang.

Pertanyaan tentang penegakan juga tetap terbuka. Hukum pidana federal memerlukan sumber daya investigatif yang memadai, dan FBI serta jaksa federal sudah sibuk dengan banyak prioritas lain. Apakah TAKE IT DOWN Act akan benar-benar menghasilkan penuntutan yang

signifikan, atau akan menjadi undang-undang simbolis yang jarang ditegakkan, adalah pertanyaan yang baru bisa dijawab dalam beberapa tahun ke depan.

DEFIANCE Act dan Jalur Perdata

DEFIANCE Act (*Disrupt Explicit Forged Images and Non-Consensual Edits Act*) mengambil pendekatan yang berbeda dari TAKE IT DOWN Act. Alih-alih mengkriminalisasi perilaku, ia membuka jalur perdata bagi korban: seseorang yang menjadi subjek deepfake seksual non-konsensual dapat menggugat pelaku untuk mendapatkan ganti rugi. Ini penting karena jalur pidana, meski lebih dramatis secara simbolis, sering kali tidak memenuhi kebutuhan konkret korban proses pidana mengutamakan kepentingan negara dalam menghukum pelaku, bukan kepentingan individu korban dalam mendapatkan pemulihan.

Jalur perdata yang dibuka DEFIANCE Act memungkinkan korban untuk menuntut kompensasi finansial, termasuk untuk kerugian emosional dan reputasional. Ini teoritis sangat berarti, tapi ada kendala praktis yang tidak kecil: banyak pelaku deepfake beroperasi secara anonim atau pseudonim, dan mengidentifikasi mereka secara hukum bisa memerlukan proses discovery yang panjang dan mahal. Untuk korban dengan sumber daya terbatas, jalur ini tetap aksesibel hanya di atas kertas.

ELVIS Act: Ketika Suara Menjadi Properti Hukum

ELVIS Act Tennessee (*Ensuring Likeness Voice and Image Security Act, 2024*) layak mendapat perhatian tersendiri karena ia mengangkat dimensi deepfake yang sering diabaikan: kloning suara. Undang-undang ini memperluas perlindungan hak publisitas Tennessee yang sebelumnya melindungi nama, potret, dan penampilan seseorang untuk secara eksplisit mencakup suara yang direplikasi oleh AI.

Tennessee bukan negara bagian acak dalam konteks ini. Nashville adalah ibu kota industri musik Amerika, dan industri itu sudah merasakan ancaman nyata dari AI yang mampu mengkloning suara artis untuk menghasilkan lagu-lagu baru tanpa izin atau kompensasi. ELVIS Act memberikan artis dan siapa pun, bukan hanya artis dasar hukum untuk menuntut penggunaan suara mereka yang tidak sah oleh AI.

Implikasinya lebih luas dari yang tampak. Suara adalah identitas dalam cara yang berbeda dari wajah. Ia bisa digunakan untuk autentikasi (banyak sistem perbankan menggunakan verifikasi suara), untuk penipuan (deepfake audio sudah digunakan dalam penipuan CEO fraud bernilai jutaan dolar), dan untuk manipulasi emosional (membayangkan menerima pesan suara dari anggota keluarga yang sudah meninggal yang ternyata dihasilkan oleh AI). ELVIS Act mungkin lahir dari kekhawatiran industri musik, tapi ia menyentuh kerentanan yang jauh lebih universal. Poin ini penting untuk diingat kembali ketika kita nanti mendiskusikan kerangka regulasi yang lebih komprehensif di Bab 12.

Lanskap 45+ Undang-Undang Negara Bagian

Di balik undang-undang federal dan ELVIS Act, ada lebih dari 45 undang-undang negara bagian yang masing-masing mengatur aspek-aspek berbeda dari deepfake. Peta ini sangat tidak

seragam: beberapa negara bagian hanya mengatur deepfake seksual, beberapa hanya deepfake pemilu, beberapa mencakup keduanya, dan beberapa yang lebih ambisius mencoba mendefinisikan "deepfake" secara luas untuk mencakup semua bentuk manipulasi media berbasis AI.

California, misalnya, punya beberapa undang-undang deepfake yang berbeda: satu untuk konten pemilu, satu untuk deepfake seksual, dan satu yang mewajibkan platform untuk mengungkapkan jika konten diproduksi oleh AI. New York menggabungkan perlindungan deepfake dengan hak publisitas yang sudah ada. Illinois memiliki *Biometric Information Privacy Act* (BIPA) yang sudah lama ada dan kini diterapkan dalam konteks deepfake yang menggunakan data biometrik wajah atau suara.

Dari perspektif akademik, lanskap ini adalah bahan kajian yang sangat kaya. Ia memungkinkan analisis komparatif tentang mana pendekatan yang lebih efektif apakah jalur pidana atau perdata, apakah definisi sempit atau luas, apakah kewajiban platform atau hanya kewajiban individu pembuat konten. Tapi dari perspektif korban dan praktisi hukum, kerumitan ini adalah beban nyata yang harus ditanggung setiap hari.

10.3 CHINA: REGULASI SISTEMATIS DALAM EKOSISTEM DIGITAL YANG TERKONTROL

Konteks: Tata Kelola Digital China

Cara memahami regulasi deepfake China, seseorang perlu terlebih dahulu memahami arsitektur tata kelola digital China secara lebih luas karena regulasi deepfake di sana bukan fenomena yang berdiri sendiri, melainkan bagian dari sistem yang lebih besar dan lebih koheren. China memiliki apa yang sering disebut sebagai "sovereignty internet" atau "Firewall Besar": sebuah ekosistem digital yang sebagian besar tertutup dari platform-platform global besar seperti Google, Facebook, atau YouTube, dan disi oleh ekuivalen domestik seperti Baidu, WeChat, dan Douyin (versi TikTok yang beroperasi di China).

Dalam ekosistem ini, platform tidak hanya tunduk pada regulasi pemerintah mereka secara aktif berkolaborasi dalam penegakan regulasi tersebut. Identifikasi pengguna dengan nama asli (*real-name registration*) sudah menjadi persyaratan standar di sebagian besar platform China sejak bertahun-tahun lalu. Ini berarti anonimitas yang membuat penegakan hukum deepfake begitu sulit di Barat adalah masalah yang jauh lebih kecil di China setidaknya secara teoritis.

Administrative Provisions on Deep Synthesis

Administrative Provisions on Deep Synthesis Technology, yang dikeluarkan oleh *Cyberspace Administration of China* (CAC) dan mulai berlaku pada Januari 2023, adalah salah satu kerangka regulasi deepfake paling komprehensif di dunia. "Deep synthesis" adalah terminologi regulasi China untuk mencakup seluruh spektrum teknologi generatif AI yang menghasilkan atau memodifikasi konten media teks, gambar, audio, video, dan kombinasinya. Regulasi ini menetapkan beberapa kewajiban inti.

- ☑ Pertama, konten yang dihasilkan atau dimodifikasi secara signifikan oleh teknologi deep synthesis harus diberi label yang jelas baik yang dapat dibaca oleh manusia maupun yang dapat dibaca oleh mesin. Ini adalah persyaratan teknis yang cukup serius: tidak cukup hanya menambahkan teks kecil "konten AI" di sudut layar, tapi harus ada tanda yang tertanam dalam metadata konten itu sendiri.
- ☑ Kedua, penyedia layanan deep synthesis harus melakukan verifikasi identitas pengguna yang menggunakan layanan mereka.
- ☑ Ketiga, ada larangan eksplisit terhadap penggunaan teknologi deep synthesis untuk menghasilkan konten yang "membahayakan keamanan nasional" atau "merusak kepentingan nasional" sebuah klausul yang luas dan berpotensi fleksibel dalam penerapannya.

Satu aspek yang sering luput dari perhatian adalah bahwa regulasi ini juga mewajibkan platform untuk membangun mekanisme pelaporan dan penghapusan konten deep synthesis yang melanggar. Ini meletakkan tanggung jawab signifikan pada platform berbeda dari pendekatan AS yang lebih berfokus pada pelaku individual.

Pelabelan Wajib: Sebuah Arsitektur Transparansi

Persyaratan pelabelan wajib dalam regulasi China layak dibahas lebih dalam karena ia mencerminkan filosofi regulasi yang berbeda dari kebanyakan pendekatan Barat. Alih-alih hanya melarang penggunaan berbahaya setelah fakta, pelabelan wajib berupaya membangun infrastruktur transparansi yang bersifat preventif: jika semua konten AI diberi tanda yang dapat diverifikasi, maka konten yang tidak berlabel secara otomatis dapat diduga sebagai upaya penipuan.

Secara teknis, implementasi pelabelan ini melibatkan standar metadata seperti C2PA (*Coalition for Content Provenance and Authenticity*) sebuah standar industri yang juga sedang dikembangkan secara paralel oleh perusahaan-perusahaan teknologi Barat. Yang menarik adalah bahwa China, melalui regulasinya, menjadikan standar semacam ini sebagai kewajiban hukum jauh sebelum Barat mengadopsinya secara serius. Ini adalah contoh di mana intervensi regulasi mendahului dan mungkin membentuk praktik industri, bukan sekadar mengikutinya.

Apakah sistem pelabelan ini efektif dalam praktiknya? Tidak selalu mudah untuk menilai dari luar. Pengawasan independen terhadap implementasi regulasi China sangat terbatas, dan laporan yang tersedia umumnya berasal dari sumber-sumber yang kepentingannya tidak selalu netral baik yang terlalu kritis (media Barat dengan bias tertentu) maupun yang terlalu positif (sumber-sumber resmi China). Yang bisa dikatakan dengan cukup yakin adalah bahwa ekosistem platform China yang lebih terkontrol secara struktural lebih mudah diberi mandat teknis dibanding ekosistem platform global yang terfragmentasi.

Ketegangan yang Tidak Terselesaikan

Regulasi deep synthesis China tidak bebas dari ketegangan internal. Klausul tentang "kepentingan nasional" dan "keamanan nasional" memberikan otoritas sangat luas kepada

pemerintah untuk menentukan konten AI mana yang boleh dan tidak boleh beredar. Dari perspektif hak kebebasan berekspresi, ini adalah kerentanan serius: regulasi yang seharusnya melindungi individu dari bahaya deepfake secara bersamaan memberikan alat kepada negara untuk mengontrol narasi politik.

Ini bukan masalah yang unik untuk China banyak regulasi konten di berbagai negara menghadapi ketegangan yang sama antara perlindungan individu dan kontrol negara. Tapi di China, ketegangan ini lebih tajam karena absennya mekanisme check and balance yang independen. Akademisi yang mengkaji regulasi ini perlu memisahkan setidaknya dua hal: apakah regulasinya secara teknis efektif dalam mengatasi bahaya deepfake (pertanyaan empiris), dan apakah ia bisa diadopsi oleh negara lain tanpa membawa serta konteks otoritariannya (pertanyaan normatif dan institusional).

10.4 KOREA SELATAN: KRIMINALISASI SEBAGAI PILIHAN UTAMA

Latar Belakang Sosial: Dari "Molka" ke Deepfake

Korea Selatan memiliki sejarah yang kelam dengan apa yang disebut "molka" rekaman tersembunyi yang ditempatkan di toilet umum, kamar pas, atau tempat-tempat pribadi lainnya untuk merekam perempuan tanpa izin mereka. Skandal molka mencapai puncak perhatian publik sekitar 2018-2019, ketika ratusan ribu perempuan turun ke jalan dalam demonstrasi yang disebut sebagai salah satu gerakan *#MeToo* terbesar di Asia. Pemerintah terpaksa merespons dengan memperkuat hukum perekaman tersembunyi dan membentuk satuan tugas khusus.

Ketika teknologi deepfake kemudian memungkinkan pembuatan konten seksual palsu tanpa memerlukan rekaman tersembunyi cukup dengan foto publik seseorang Korea Selatan sudah memiliki kesadaran kolektif yang cukup kuat untuk merespons dengan serius. Tidak perlu waktu lama bagi legislator untuk bergerak, terutama setelah laporan-laporan tentang deepfake seksual yang menargetkan perempuan biasa (bukan hanya selebritas) menjadi berita utama.

Kriminalisasi Pembuatan: Sebuah Langkah Doktrinal

Langkah paling signifikan yang diambil Korea Selatan dalam regulasi deepfake adalah mengkriminalisasi pembuatan konten deepfake seksual bukan hanya distribusinya. Ini adalah posisi yang lebih keras dibanding sebagian besar yurisdiksi lain, dan implikasi doktrinnya layak dieksplorasi.

Dalam hukum pidana konvensional, titik intervensi biasanya adalah tindakan yang menyebabkan bahaya distribusi konten, bukan pembuatannya. Logikanya: jika konten tidak disebarkan, tidak ada pihak yang dirugikan. Tapi Korea Selatan, dalam regulasi deepfake seksualnya, bergerak lebih ke hulu.

Argumen yang mendasarinya adalah bahwa pembuatan deepfake seksual seseorang sudah merupakan pelanggaran terhadap martabat dan otonomi orang tersebut terlepas dari apakah konten itu akhirnya disebarkan atau tidak. Bagi korban yang mengetahui bahwa seseorang telah membuat deepfake seksual dirinya, kerusakan psikologis sudah terjadi bahkan sebelum distribusi.

Ini adalah argumen yang kuat secara moral, tapi menimbulkan tantangan pembuktian yang serius. Bagaimana penegak hukum mengetahui bahwa seseorang telah membuat deepfake jika konten itu tidak disebarkan? Jawabannya sering kali adalah: melalui laporan dari korban yang mengetahui eksistensinya (misalnya karena pelaku mengancam akan menyebarkannya), atau melalui operasi penegakan hukum aktif. Tapi ini berarti banyak kasus yang tidak akan pernah terdeteksi. Tampaknya kriminalisasi pembuatan lebih berfungsi sebagai sinyal normativif bahwa tindakan itu sendiri dianggap salah daripada sebagai mekanisme penegakan yang komprehensif.

Penegakan dan Realitas Lapangan

Korea Selatan juga cukup agresif dalam hal penegakan, setidaknya dibandingkan banyak yurisdiksi lain. Kepolisian nasional membentuk unit khusus untuk kejahatan berbasis teknologi termasuk deepfake seksual, dan ada koordinasi dengan platform-platform yang beroperasi di Korea Selatan untuk memfasilitasi penghapusan konten dan identifikasi pelaku.

Tapi kenyataan di lapangan tetap kompleks. Banyak konten deepfake yang menargetkan perempuan Korea beredar di platform-platform asing yang tidak memiliki kehadiran hukum di Korea Selatan. Telegram, misalnya, sudah lama menjadi saluran utama penyebaran konten molka dan deepfake, dan keterbatasan yurisdiksi terhadap platform asing adalah hambatan nyata yang dirasakan oleh penyidik. Kasus-kasus yang berhasil dituntut sebagian besar adalah kasus di mana pelaku dapat diidentifikasi dan berada di wilayah yurisdiksi Korea Selatan.

Meski demikian, Korea Selatan telah membangun reputasi sebagai yurisdiksi yang tidak main-main dalam hal ini, dan tekanan hukum yang dirasakan tampaknya memiliki efek pencegahan tertentu meski efek ini sulit diukur secara empiris. Satu yang bisa diamati adalah bahwa kampanye kesadaran publik yang menyertai regulasi ini, dikombinasikan dengan ancaman pidana yang nyata, tampaknya mengubah persepsi sosial tentang deepfake seksual sebagai sesuatu yang "tidak berbahaya" menjadi sesuatu yang serius dan dapat dihukum.

10.5 INGGRIS: ONLINE SAFETY ACT DAN AMBISI YANG LUAS

Online Safety Act 2023: Regulasi Platform dalam Skala Besar

Online Safety Act 2023 adalah salah satu upaya regulasi platform digital paling ambisius yang pernah dilakukan oleh demokrasi liberal Barat. Undang-undang ini tidak semata-mata tentang deepfake, Undang-undang mencakup spektrum yang sangat luas: dari perlindungan anak terhadap konten berbahaya daring, hingga misinformasi, ujaran kebencian, dan berbagai bentuk konten ilegal lainnya. Deepfake intim adalah salah satu dari banyak isu yang dirangkumnya dalam sebuah kerangka tunggal yang kompleks.

Pendekatan Inggris ini mencerminkan strategi regulasi yang berbeda dari, katakanlah, Amerika: alih-alih membuat undang-undang sektoral yang terpisah untuk setiap masalah (deepfake, ujaran kebencian, dll.), Inggris mencoba membangun kerangka umum yang mengatur tanggung jawab platform secara holistik. Ide dasarnya adalah bahwa platform besar seperti

Facebook, YouTube, atau TikTok memiliki kewajiban sebagai pengasuh ruang publik digital bukan hanya sebagai pipa pasif yang mengalirkan konten pengguna.

Kriminalisasi Deepfake Intim: Apa yang Dicakup dan Apa yang Tidak

Dalam konteks deepfake spesifik, Online Safety Act mengkriminalisasi pembuatan konten deepfake intim tanpa persetujuan. Perhatikan batasannya: "intim" undang-undang ini tidak mengkriminalisasi semua deepfake, melainkan hanya yang bersifat intim atau seksual yang dibuat tanpa izin subjeknya. Ini adalah batasan yang disengaja dan cukup ketat.

Pertanyaan yang segera muncul adalah apa yang dimaksud dengan "intim" dalam konteks ini. Undang-undang menyediakan definisi yang cukup teknis, tapi seperti kebanyakan definisi hukum, ia meninggalkan ruang interpretasi yang akan diisi oleh yurisprudensi yang berkembang dalam beberapa tahun ke depan. Apakah deepfake yang menampilkan seseorang dalam pakaian renang termasuk "intim"? Bagaimana dengan deepfake yang memanipulasi ekspresi wajah seseorang untuk terlihat sedang mengalami kesenangan seksual tanpa menampilkan konten eksplisit? Batas-batas ini akan diuji di pengadilan.

Yang tidak dicakup oleh ketentuan deepfake intim Online Safety Act adalah sama pentingnya: deepfake politik, deepfake penipuan komersial, deepfake yang digunakan untuk pelecehan yang tidak bersifat seksual, dan deepfake satire. Ini bukan berarti tindakan-tindakan tersebut sepenuhnya tidak diatur ada ketentuan lain dalam Online Safety Act maupun hukum Inggris yang mungkin berlaku tapi tidak ada kriminalisasi spesifik deepfake untuk domain-domain itu.

Tantangan Implementasi: OFCOM dan Beban Regulasi

Office of Communications (Ofcom) ditunjuk sebagai regulator yang bertanggung jawab atas implementasi Online Safety Act. Ini berarti Ofcom tidak hanya menetapkan standar teknis tentang bagaimana platform harus menangani konten berbahaya, tapi juga memiliki kewenangan untuk mengenakan denda yang sangat besar hingga 10% dari pendapatan global kepada platform yang tidak mematuhi.

Skala tugas ini luar biasa. Online Safety Act berlaku untuk ribuan platform yang beroperasi di Inggris, dari raksasa global seperti Meta dan Google hingga platform-platform kecil dengan sedikit pengguna. Ofcom harus memprioritaskan sumber dayanya, dan platform-platform besar dengan jutaan pengguna Inggris adalah prioritas utama yang jelas. Ini berarti platform-platform yang lebih kecil yang sering kali justru menjadi tempat konten berbahaya berlindung setelah dihapus dari platform besar mendapat perhatian yang jauh lebih sedikit.

Pertanyaan tentang efektivitas jangka panjang Online Safety Act masih terbuka. Undang-undang ini mendapat kritik dari berbagai arah: dari kelompok kebebasan sipil yang khawatir tentang potensi sensor berlebihan, dari industri teknologi yang menganggap beban kepatuhannya terlalu berat, dan dari kelompok advokasi yang menganggap perlindungannya masih kurang kuat. Mungkin kritik dari semua arah ini justru merupakan tanda bahwa undang-undang telah mencapai semacam keseimbangan meski keseimbangan yang tidak memuaskan siapa pun sepenuhnya.

10.6 PRANCIS: INTEGRASI KE DALAM KODIFIKASI YANG ADA

Article 226-8-1 KUHP: Pendekatan Kodifikasi

Prancis memilih jalur yang secara teknis paling elegan tapi mungkin paling kurang dramatis: menyisipkan ketentuan baru ke dalam *Code Pénal* (KUHP) yang sudah ada, tepatnya melalui Article 226-8-1. Ketentuan ini mengkriminalisasi pembuatan dan distribusi deepfake yang dibuat tanpa persetujuan subjeknya, dengan ancaman hukuman penjara hingga dua tahun dan denda hingga 60.000 euro.

Pilihan untuk menggunakan KUHP yang ada bukan sekadar kebiasaan birokrasi. Ia mencerminkan filosofi hukum Eropa Kontinental khususnya tradisi hukum Romawi-Jermanik yang mendominasi sistem hukum Prancis yang mengutamakan kodifikasi: gagasan bahwa hukum sebaiknya tersusun dalam satu atau beberapa kitab yang komprehensif dan sistematis, bukan dalam tumpukan undang-undang sektoral yang terpisah-pisah. Dari perspektif ini, menambahkan ketentuan deepfake ke dalam KUHP adalah cara yang tepat secara doktrin untuk mengintegrasikan norma baru ke dalam sistem yang sudah ada.

Secara praktis, ini berarti jaksa Prancis yang menangani kasus deepfake bekerja dalam kerangka prosedural yang sudah familiar ketentuan KUHP yang sama, standar pembuktian yang sama, mekanisme banding yang sama seperti kejahatan-kejahatan lain. Ini bisa menjadi keuntungan (lebih sedikit ketidakpastian hukum) atau kerugian (kerangka yang dirancang untuk kejahatan konvensional mungkin tidak sepenuhnya fit untuk kejahatan digital).

Ancaman Hukuman dan Deterensinya

Ancaman dua tahun penjara dan denda 60.000 euro untuk deepfake tanpa persetujuan adalah angka yang cukup serius. Untuk konteks: ini lebih tinggi dari denda maksimum untuk beberapa pelanggaran lalu lintas berat di Prancis, dan sebanding dengan hukuman untuk beberapa bentuk pencurian. Sinyal normatifnya jelas: hukum Prancis menyamakan pelanggaran deepfake dengan pelanggaran properti dan ketertiban umum yang serius.

Tapi seperti semua ancaman hukuman, efek deterensinya bergantung pada probabilitas pendeteksian dan penuntutan, bukan hanya pada beratnya ancaman itu sendiri. Jika seseorang yakin bahwa deepfake yang mereka buat tidak akan pernah terdeteksi atau dikaitkan kepada mereka karena mereka beroperasi secara anonim atau menggunakan platform asing maka ancaman dua tahun penjara tidak akan mengubah perilaku mereka. Ini adalah masalah yang dihadapi semua sistem hukum pidana dalam domain kejahatan digital, dan Prancis tidak kebal terhadapnya.

Prancis dalam Konteks Eropa yang Lebih Luas

Regulasi Prancis tidak bisa dipahami sepenuhnya tanpa mempertimbangkan konteks Eropa yang lebih luas. *General Data Protection Regulation* (GDPR) sudah menetapkan bahwa wajah dan biometrik adalah data pribadi yang dilindungi dan ini memberikan pijakan hukum tambahan bagi individu yang wajahnya digunakan tanpa izin dalam deepfake, bahkan di luar ketentuan KUHP.

Digital Services Act (DSA) Uni Eropa juga menetapkan kewajiban bagi platform untuk mengelola konten ilegal, termasuk deepfake yang melanggar hukum nasional anggota.

Prancis, sebagai salah satu anggota terkuat Uni Eropa, juga memainkan peran dalam membentuk regulasi Eropa di level supranasional. Kepedulian Prancis terhadap "kedaulatan digital" (*souveraineté numérique*) dan perlindungan identitas individu dari eksploitasi komersial atau kriminal adalah tema yang konsisten dalam posisi negosiasi Prancis di Brussels. Dalam hal ini, regulasi deepfake Prancis bukan berdiri sendiri, namun regulasi Prancis adalah bagian dari visi yang lebih besar tentang bagaimana Eropa ingin mengatur ruang digitalnya.

Rangkuman Bab

Lima Dimensi Perbandingan

Membandingkan lima yurisdiksi ini secara bermakna memerlukan kerangka analisis yang lebih sistematis dari sekadar menyandingkan daftar undang-undang. Setidaknya ada lima dimensi yang relevan untuk perbandingan:

(1) cakupan substantif: jenis deepfake apa yang diatur;
(2) instrumen hukum: pidana, perdata, atau administratif;
(3) alokasi tanggung jawab: pelaku individual, platform, atau keduanya;
(4) mekanisme penegakan: siapa yang berwenang dan dengan sumber daya apa; dan
(5) ekosistem kelembagaan: apakah ada regulator khusus, pengadilan khusus, atau mekanisme alternatif.

Dalam dimensi cakupan substantif, China adalah yang paling luas Administrative Provisions on Deep Synthesis mencakup semua jenis konten AI generatif, bukan hanya yang bersifat seksual atau berbahaya. Amerika Serikat, sebaliknya, memiliki cakupan federal yang paling sempit (TAKE IT DOWN Act hanya menyentuh deepfake seksual), meski agregat undang-undang negara bagian lebih luas. Korea Selatan dan Inggris sama-sama fokus pada deepfake intim/seksual di tingkat kriminalisasi utamanya. Prancis dalam KUHP-nya tidak membatasi pada jenis tertentu, tapi praktiknya undang-undang ini paling sering diperdebatkan dalam konteks konten seksual.

Dalam dimensi instrumen hukum, Korea Selatan dan Prancis paling tegas menggunakan pidana sebagai instrumen utama. Amerika Serikat paling beragam—ada pidana federal (TAKE IT DOWN Act), perdata federal (DEFIANCE Act), dan campuran di level negara bagian. China paling menonjolkan instrumen administratif melalui kewajiban platform dan sistem pelabelan. Inggris mencoba memadukan semuanya dalam Online Safety Act: ada kriminalisasi, ada kewajiban platform, ada pengawasan regulator.

Masalah Platform dan Batas Yurisdiksi

Satu masalah yang dihadapi semua yurisdiksi ini tanpa pengecualian adalah batas yurisdiksi terhadap platform asing. Deepfake yang dibuat di satu negara dan disebarkan melalui platform yang berkantor di negara lain adalah skenario yang sangat umum, dan semua sistem hukum yang ada menghadapi kesulitan nyata untuk menanganinya.

Amerika Serikat, ironisnya, memiliki pengaruh tidak proporsional dalam hal ini karena sebagian besar platform media sosial global berkantor di sana terutama di California. Ini berarti perubahan kebijakan AS, atau tuntutan hukum yang berhasil terhadap platform AS, bisa memiliki efek global yang jauh melampaui batas jurisdiksi formal AS. Platform-platform seperti Meta, Google, atau X (Twitter) yang tunduk pada hukum AS secara tidak langsung membawa norma-norma hukum AS ke pengguna di seluruh dunia melalui kebijakan layanan mereka.

China, dengan ekosistem platform domestiknya yang sebagian besar terpisah dari internet global, tidak menghadapi masalah ini dalam bentuk yang sama. Tapi sebagai gantinya, pengguna China yang menggunakan layanan VPN untuk mengakses platform asing sebuah praktik yang tersebar luas meski secara teknis ilegal berada di luar jangkauan regulasi domestik.

Korea Selatan, Inggris, dan Prancis semuanya menghadapi versi masalah yang sama: konten yang tidak diinginkan disebarkan melalui platform-platform yang secara formal tidak tunduk pada yurisdiksi mereka. Telegram adalah contoh yang sering disebut platform yang secara konsisten menolak untuk bekerja sama dengan permintaan penegakan hukum dari pemerintah-pemerintah Eropa dan Asia. Situasi ini tampaknya belum memiliki solusi hukum yang memuaskan, dan ini adalah keterbatasan yang perlu diakui secara jujur oleh siapa pun yang mengevaluasi keefektifan regulasi deepfake yang ada.

Keseimbangan antara Perlindungan dan Kebebasan Berekspresi

Setiap sistem regulasi deepfake menghadapi ketegangan yang sama: bagaimana melindungi individu dari bahaya nyata tanpa menciptakan mekanisme sensor yang bisa disalahgunakan atau yang menghambat ekspresi sah. Ketegangan ini tidak memiliki solusi sempurna, hanya bisa dikelola dengan lebih atau kurang baik.

Amerika Serikat, dengan tradisi Amandemen Pertama yang sangat kuat, secara historis paling berhati-hati dalam mengatur konten berdasarkan isinya. Bahkan TAKE IT DOWN Act yang tampaknya jelas fokus pada konten berbahaya mendapat tantangan dari kelompok kebebasan sipil yang khawatir tentang potensi penyalahgunaan mekanisme penghapusannya. Prancis dan Eropa pada umumnya lebih nyaman dengan regulasi konten, berangkat dari tradisi yang menyeimbangkan kebebasan berbicara dengan perlindungan martabat dan privasi. China, dalam spektrum yang berbeda secara kualitatif, menempatkan ketertiban sosial dan keamanan nasional di atas kebebasan berekspresi individual.

Korea Selatan berada di posisi yang menarik: ia adalah demokrasi liberal yang nyata, dengan budaya kebebasan berekspresi yang aktif, tapi juga memiliki kedaruratan moral yang kuat terhadap kejahatan seksual berbasis teknologi akibat pengalaman traumatik dengan molka. Ini menghasilkan kebijakan yang cukup keras dari sisi kriminalisasi tapi berjalan dalam kerangka institusional demokratis.

Arah Peta Regulasi Deepfake Global Bergerak

Peta regulasi deepfake yang telah kita telusuri dalam bab ini memberikan gambaran yang jauh dari sederhana. Tidak ada satu model terbaik yang bisa langsung diadopsi oleh semua negara.

Setiap sistem regulasi lahir dari konteks kelembagaan, tradisi hukum, dan tekanan sosial yang spesifik dan konteks itu tidak bisa dipisahkan dari regulasinya.

Yang bisa kita identifikasi sebagai tren umum adalah: pertama, pergeseran dari pendekatan reaktif (merespons kasus individual) ke pendekatan sistemik (membangun infrastruktur hukum dan teknis yang preventif). China dengan pelabelan wajibnya adalah contoh paling jelas dari pendekatan sistemik, tapi tren ini juga terlihat dalam diskusi tentang provenance teknologi seperti C2PA yang semakin mendapat dukungan di Amerika Serikat dan Eropa.

Kedua, ada kecenderungan yang semakin kuat untuk menempatkan platform sebagai pihak yang memiliki tanggung jawab, bukan hanya individu yang membuat konten. Online Safety Act Inggris, Administrative Provisions China, dan bahkan TAKE IT DOWN Act Amerika semuanya mewajibkan platform untuk mengambil tindakan tertentu bukan hanya bergantung pada individu korban untuk mencari keadilan sendiri. Ini adalah pergeseran yang signifikan dari paradigma awal internet yang memandang platform semata-mata sebagai pipa pasif.

Ketiga, meski kelima yurisdiksi yang kita bahas sudah bergerak, mereka bergerak secara tidak terkoordinasi. Tidak ada kerangka internasional yang mengatur deepfake secara komprehensif berbeda dari, misalnya, hak kekayaan intelektual yang memiliki perjanjian internasional yang cukup matang. Upaya-upaya di level Perserikatan Bangsa-Bangsa atau forum-forum multilateral lainnya masih dalam tahap sangat awal. Selama koordinasi internasional ini absen, bahaya tetap ada bahwa regulasi nasional yang paling ketat pun akan dikelabui oleh pelaku yang beroperasi dari yurisdiksi yang lebih permisif.

Bab-bab berikutnya akan mengeksplorasi implikasi dari lanskap regulasi ini untuk Indonesia secara spesifik termasuk pertanyaan tentang regulasi mana yang paling relevan untuk diadaptasi, tantangan kelembagaan apa yang harus diantisipasi, dan bagaimana komunitas akademik dapat berkontribusi pada pembentukan kebijakan yang berbasis bukti. Sebelum ke sana, ada baiknya kita catat bahwa regulasi, seberapa pun baiknya, adalah satu layer dari respons yang diperlukan. Pendidikan, literasi media, norma sosial, dan kapasitas teknis adalah layer-layer lain yang tidak kalah pentingnya dan justru sering kali lebih tahan lama dari undang-undang yang bisa diubah setiap periode legislatif.

Satu hal yang tampak cukup jelas dari studi komparatif ini: negara-negara yang memiliki fondasi kelembagaan yang kuat kepercayaan publik terhadap institusi hukum, kapasitas teknis penegak hukum, dan ekosistem akademik kebijakan yang aktif lebih mampu mengimplementasikan regulasi deepfake yang efektif, terlepas dari model spesifik yang dipilih. Regulasi yang baik, pada akhirnya, hanya sebaik institusi yang mengimplementasikannya.

BAB 11
HAK ASASI MANUSIA DAN DEEPFAKE: PERSPEKTIF HUKUM INTERNASIONAL

"Hak asasi manusia bukan daftar ketentuan yang menunggu dilanggar. Ia adalah janji tentang jenis kehidupan yang pantas diterima setiap orang."

parafrase dari diskursus filsafat hukum HAM kontemporer

Ada godaan intelektual yang cukup kuat untuk mendekati masalah deepfake semata-mata sebagai masalah teknologi atau masalah pidana sesuatu yang diselesaikan dengan membangun detektor yang lebih canggih, atau dengan mengesahkan undang-undang yang lebih keras. Bab ini berargumen bahwa pendekatan itu, meski tidak salah, tidak cukup. Deepfake bukan hanya masalah teknologi atau hukum pidana nasional. Ia adalah masalah hak asasi manusia dan ketika kita menempatkannya dalam kerangka itu, implikasinya berubah secara signifikan.

Kerangka HAM tidak sekadar menyediakan bahasa yang berbeda untuk menyebut masalah yang sama. Ia menyediakan seperangkat kewajiban yang berbeda, dengan subjek yang berbeda. Dalam hukum pidana, yang dipersoalkan adalah apakah seseorang melanggar undang-undang dan layak dihukum. Dalam kerangka HAM, yang dipersoalkan adalah apakah negara telah memenuhi kewajibannya untuk melindungi, menghormati, dan memenuhi hak-hak warganya dan jika tidak, apa konsekuensinya. Perbedaan ini bukan terminologis semata; ia menentukan siapa yang bertanggung jawab, kepada siapa, dan melalui mekanisme apa.

Bab ini memetakan setidaknya lima lini argumen HAM yang relevan untuk masalah deepfake: hak privasi yang dilindungi Pasal 17 ICCPR, hak atas martabat dan reputasi, perlindungan khusus perempuan dan anak melalui CEDAW dan Konvensi Hak Anak, keseimbangan antara kebebasan berekspresi dan perlindungan korban, serta yang mungkin paling operasional kewajiban negara dalam kerangka due diligence, regulasi, dan remediasi. Tidak semua argumen ini sama kuatnya, dan beberapa di antaranya masih dalam proses pembentukan dalam yurisprudensi HAM internasional. Tapi bersama-sama, mereka membentuk sebuah kasus yang cukup kuat bahwa deepfake adalah isu HAM yang serius, bukan sekadar isu teknologi atau keamanan.

11.1 KERANGKA KONSEPTUAL: MENGAPA HAM RELEVAN UNTUK DEEPFAKE

Dari Bahaya Individual ke Kewajiban Struktural

Ketika seseorang menjadi korban deepfake wajahnya ditempelkan pada konten seksual, suaranya dikloning untuk menyebarkan kebohongan, atau identitasnya dimanipulasi untuk tujuan penipuan yang pertama kali terlihat adalah bahaya yang sangat personal dan individual. Orang ini

dirugikan oleh orang lain yang menggunakan teknologi. Ini adalah narasi yang mudah dipahami dan yang mendorong respons hukum pidana: temukan pelakunya, hukum, selesai.

Tapi ada lapisan lain yang tersembunyi di bawah narasi itu. Mengapa korban tidak memiliki mekanisme yang efektif untuk mendapatkan keadilan? Mengapa platform tempat deepfake disebarkan tidak diwajibkan untuk bertindak cepat? Mengapa tidak ada sistem yang memungkinkan korban membuktikan bahwa konten itu dipalsukan? Pertanyaan-pertanyaan ini menunjuk bukan pada kegagalan individual pelaku, melainkan pada kegagalan struktural sistem dan kegagalan struktural sistem adalah domain kewajiban negara dalam kerangka HAM.

Hukum HAM internasional membedakan antara tiga tingkatan kewajiban negara:

(1) kewajiban untuk menghormati (*respect*) negara tidak boleh secara aktif melanggar hak;
(2) kewajiban untuk melindungi (*protect*) negara harus mencegah pihak ketiga melanggar hak; dan
(3) kewajiban untuk memenuhi (*fulfil*) negara harus mengambil langkah-langkah positif untuk mewujudkan hak.

Deepfake terutama menguji kewajiban perlindungan, apakah negara telah cukup mengatur aktor-aktor swasta (pembuat deepfake, platform distribusi, pengembang teknologi) agar tidak melanggar hak orang lain?

Ini adalah pertanyaan yang jauh lebih sulit dari sekadar "apakah ada undang-undang yang melarang deepfake." Ia menuntut evaluasi menyeluruh terhadap seluruh ekosistem regulasi, penegakan, dan akses keadilan. Dan dalam evaluasi itu, hampir semua negara termasuk yang sudah memiliki undang-undang deepfake seperti yang dibahas di Bab 10 masih menemukan diri mereka kurang.

Sifat Universal HAM dan Tantangan Teknologi

Salah satu prinsip dasar hukum HAM internasional adalah universalitas: hak-hak itu berlaku untuk semua orang, di mana saja, tanpa memandang kewarganegaraan, status, atau konteks teknologi. Prinsip ini menghadapi uji coba yang menarik ketika berhadapan dengan teknologi baru seperti deepfake karena teknologi ini tidak pernah diantisipasi oleh para penyusun instrumen HAM internasional utama seperti UDHR (1948), ICCPR (1966), atau CEDAW (1979).

Tapi absennya antisipasi spesifik ini bukan berarti instrumen-instrumen itu tidak berlaku. Kerangka HAM dirancang setidaknya dalam teorinya untuk bersifat adaptif: prinsip-prinsip yang ia tetapkan cukup luas untuk merespons ancaman-ancaman baru terhadap martabat dan otonomi manusia, meski bentuk spesifik ancaman itu tidak pernah dibayangkan oleh para pendirinya. Tugas para akademisi dan praktisi hukum adalah melakukan interpretasi yang bertanggung jawab: tidak memaksakan teks lama ke masalah baru secara artifisial, tapi juga tidak menolak relevansi kerangka yang ada hanya karena ia tidak menyebut "deepfake" secara eksplisit.

Dalam hal ini, para Pelapor Khusus (*Special Rapporteurs*) PBB telah memainkan peran penting. Laporan-laporan dari Pelapor Khusus tentang Privasi, tentang Kekerasan terhadap Perempuan, dan tentang Kebebasan Berekspresi dalam beberapa tahun terakhir secara bertahap

mulai mengintegrasikan diskusi tentang deepfake dan teknologi AI ke dalam kerangka HAM yang ada sebuah proses interpretasi yang hidup dan terus berkembang.

11.2 HAK PRIVASI: ICCPR PASAL 17 DAN BIOMETRIK SEBAGAI DATA SENSITIF

Pasal 17 ICCPR: Teks dan Interpretasinya

Pasal 17 *International Covenant on Civil and Political Rights* (ICCPR) menyatakan bahwa tidak seorang pun boleh dikenai campur tangan sewenang-wenang atau tidak sah atas privasi, keluarga, rumah, atau korespondensinya, maupun serangan yang tidak sah atas kehormatan dan reputasinya. Ini adalah formulasi yang cukup luas *"campur tangan sewenang-wenang atau tidak sah"* dan interpretasinya telah berkembang secara signifikan sejak teks itu diadopsi pada 1966.

Komite HAM PBB (*Human Rights Committee*), yang berwenang menginterpretasi ICCPR, dalam General Comment No. 16 menegaskan bahwa hak privasi mencakup perlindungan terhadap pengumpulan dan penyimpanan informasi personal yang tidak sah. Lebih relevan lagi untuk konteks deepfake: Komite telah menegaskan bahwa negara harus memastikan bahwa informasi tentang kehidupan pribadi seseorang tidak sampai ke tangan pihak yang tidak berwenang, dan bahwa ada mekanisme efektif untuk memastikan ini.

Bagaimana ini berlaku untuk deepfake? Deepfake yang menggunakan wajah atau suara seseorang tanpa izin mereka adalah bentuk penggunaan data pribadi yang tidak sah. Wajah seseorang terutama sebagai titik referensi untuk sistem pengenalan wajah adalah informasi biometrik. Dan biometrik, dalam pandangan yang semakin diterima luas baik di akademisi maupun dalam regulasi seperti GDPR, adalah kategori data yang paling sensitif karena ia melekat permanen pada tubuh seseorang dan tidak bisa diganti seperti kata sandi atau nomor kartu kredit.

Biometrik sebagai Data Sensitif: Mengapa Ini Penting

Pergeseran konseptual yang paling signifikan dalam hukum privasi dekade terakhir adalah pengakuan bahwa tidak semua data pribadi sama sensitifnya. Data biometrik wajah, sidik jari, iris mata, voiceprint, bahkan gaya berjalan menempati kategori tersendiri karena tiga alasan yang saling terkait.

Pertama, data biometrik bersifat unik dan tidak tergantikan. Jika kata sandi bocor, Anda bisa menggantinya. Jika wajah Anda digunakan tanpa izin untuk membuat deepfake, tidak ada yang bisa Anda ganti. Wajah Anda tetap wajah Anda, dan teknologi pengenalan wajah yang semakin canggih berarti biometrik Anda terus menjadi kunci akses ke berbagai sistem perbankan, imigrasi, perangkat pribadi. Kedua, data biometrik sering dikumpulkan tanpa sadar atau tanpa persetujuan eksplisit. Foto-foto yang Anda unggah ke media sosial adalah, dalam satu pengertian, database biometrik wajah Anda yang diakses secara publik dan model-model AI dapat menggunakannya untuk melatih sistem pengenalan wajah atau untuk menghasilkan deepfake tanpa sepengetahuan Anda. Ketiga, pelanggaran biometrik memiliki konsekuensi asimetris: kejahatan berbasis biometrik baik deepfake maupun penipuan identitas biometrik lainnya jauh lebih sulit untuk dipulihkan dibanding kejahatan berbasis informasi konvensional.

Dalam kerangka Pasal 17 ICCPR, penggunaan data biometrik seseorang untuk membuat deepfake tanpa izin mereka adalah campur tangan yang tidak sah atas privasi mereka. Ini bukan interpretasi yang dipaksakan, penggunaan data biometrik mengikuti logika yang sudah diterima dalam yurisprudensi HAM tentang privasi data. Yang diperlukan adalah negara-negara yang meratifikasi ICCPR untuk mengakui secara eksplisit bahwa kewajiban mereka berdasarkan Pasal 17 mencakup perlindungan terhadap penyalahgunaan biometrik dalam konteks AI, termasuk deepfake.

Privasi sebagai Prasyarat Hak-Hak Lain

Satu argumen yang sering kurang mendapat perhatian adalah bahwa privasi bukan hanya hak yang berdiri sendiri, namun privasi adalah prasyarat bagi hak-hak lain. Ketika seseorang menjadi target deepfake seksual, pelanggaran privasi yang terjadi sering kali memicu serangkaian pelanggaran hak yang lain: ia mungkin kehilangan pekerjaan (pelanggaran hak atas kehidupan yang layak), mengalami trauma psikologis yang parah (pelanggaran hak atas kesehatan fisik dan mental), menarik diri dari ruang publik (pelanggaran hak berpartisipasi dalam kehidupan publik), atau mengubah perilaku daring secara drastis karena takut (efek dingin pada kebebasan berekspresi).

Cascade of harms ini efek berantai dari satu pelanggaran privasi adalah argumen kuat mengapa hak privasi dalam konteks deepfake layak mendapat perhatian HAM yang serius. Ia juga menjelaskan mengapa solusi yang hanya berfokus pada penghapusan konten (*remove the deepfake*) tidak cukup: bahkan setelah konten dihapus, kerusakan terhadap hak-hak lain bisa sudah terjadi dan tidak bisa dikembalikan.

11.3 HAK ATAS MARTABAT DAN REPUTASI: ICCPR PASAL 17 DAN 19

Martabat sebagai Fondasi Hukum HAM

Martabat manusia (*human dignity*) adalah konsep yang menempati posisi sangat khusus dalam hukum HAM internasional. Ia disebutkan dalam preambul UDHR, ICCPR, dan ICESCR sebagai dasar dari seluruh bangunan hak asasi manusia bukan sebagai hak yang berdiri sendiri, melainkan sebagai fondasi yang menopang semua hak yang lain. Ini berarti argumen bahwa deepfake melanggar martabat manusia bukan argumen yang bisa dengan mudah diabaikan; ia menyentuh inti dari sistem HAM internasional.

Apa yang dimaksud dengan martabat dalam konteks deepfake? Satu bisa berargumen bahwa martabat mencakup setidaknya dua dimensi yang relevan: dimensi otonom (hak untuk menentukan representasi diri sendiri di dunia) dan dimensi relasional (hak untuk diperlakukan sebagai subjek yang memiliki nilai intrinsik, bukan sebagai objek yang bisa dimanipulasi untuk kepentingan orang lain). Deepfake melanggar keduanya. Ia mengambil alih representasi seseorang wajah, suara, tubuh dan menempatkannya dalam narasi yang sama sekali tidak dipilih oleh orang itu. Dan ia memperlakukan korban sebagai bahan mentah untuk konten orang lain, bukan sebagai manusia yang memiliki otonomi atas citranya sendiri.

Argumen martabat ini paling kuat dalam kasus deepfake seksual, di mana tubuh seseorang atau representasi digitalnya digunakan tanpa izin untuk konten seksual. Tapi ia tidak terbatas pada konteks itu. Deepfake yang menampilkan seseorang mengucapkan kata-kata yang tidak pernah mereka katakan, atau melakukan tindakan yang tidak pernah mereka lakukan, dalam konteks apapun politik, profesional, sosial juga melanggar martabat dalam pengertian ini.

Reputasi sebagai Hak yang Dapat Dilanggar

Pasal 17 ICCPR secara eksplisit menyebut perlindungan terhadap *"serangan yang tidak sah atas kehormatan dan reputasi"* sebuah ketentuan yang sering diabaikan karena teks utama pasal ini lebih banyak dibahas dalam konteks privasi. Tapi ketentuan reputasi ini sangat relevan untuk deepfake.

Ketika deepfake menampilkan seorang politisi mengumumkan kebijakan yang tidak pernah ia buat, atau seorang profesional mengakui perilaku tidak etis yang tidak pernah terjadi, reputasi mereka dirugikan secara nyata bahkan jika deepfake itu kemudian terbukti palsu. Psikologi persepsi menunjukkan bahwa kesan pertama sangat persisten: studi tentang misinformasi konsisten menunjukkan bahwa bahkan setelah seseorang diberitahu bahwa sebuah informasi adalah palsu, kepercayaan residual terhadap informasi itu sering bertahan.

Deepfake mengeksploitasi kelemahan kognitif ini secara sangat efektif. Dari perspektif hukum HAM, perlindungan reputasi berdasarkan Pasal 17 ICCPR menuntut negara untuk menyediakan mekanisme remediasi yang efektif bukan hanya penghapusan konten, tapi juga koreksi yang memadai dan kompensasi atas kerugian reputasional. Ini adalah standar yang belum dipenuhi oleh sebagian besar sistem hukum yang ada.

Pasal 19 ICCPR dan Ketegangan Internal

Pasal 19 ICCPR melindungi kebebasan berekspresi, tapi dengan klausa pembatasan yang penting: pembatasan terhadap kebebasan berekspresi diperbolehkan jika (a) diatur dalam hukum, (b) diperlukan untuk tujuan tertentu termasuk perlindungan hak orang lain, dan (c) proporsional dengan tujuan yang hendak dicapai. Ini berarti ada ketegangan internal dalam ICCPR sendiri antara Pasal 17 (perlindungan martabat dan reputasi) dan Pasal 19 (kebebasan berekspresi).

Bagaimana ketegangan ini diselesaikan dalam konteks deepfake? Komite HAM PBB dalam berbagai general comment-nya telah menegaskan bahwa kebebasan berekspresi tidak mencakup hak untuk membuat pernyataan yang secara faktual keliru dan merugikan orang lain, dan bahwa perlindungan terhadap ujaran yang melukai reputasi orang lain adalah tujuan yang sah untuk membatasi kebebasan berekspresi, asalkan batasan itu proporsional. Deepfake yang bertujuan merusak reputasi atau melanggar privasi jelas masuk dalam kategori ekspresi yang dapat dibatasi berdasarkan Pasal 19 ayat (3).

Yang lebih rumit adalah deepfake satire atau deepfake artistik deepfake yang tidak bertujuan menipu tapi menggunakan wajah atau suara seseorang untuk tujuan komentar sosial atau artistik. Di sini ketegangan antara Pasal 17 dan Pasal 19 menjadi lebih nyata dan tidak memiliki jawaban yang mudah. Yurisprudensi tentang ini masih dalam tahap sangat awal, dan

kemungkinan besar akan berkembang berbeda di yurisdiksi yang berbeda. Poin ini penting untuk diingat kembali ketika kita nanti mendiskusikan soal batas-batas regulasi yang proporsional.

11.4 PERLINDUNGAN PEREMPUAN: CEDAW DAN KEKERASAN BERBASIS GENDER

Data yang Tidak Bisa Diabaikan

Sebelum masuk ke analisis hukumnya, ada satu fakta empiris yang perlu ditetapkan dengan jelas: deepfake seksual adalah masalah yang sangat berdimensi gender. Berbagai penelitian yang dilakukan antara 2019 hingga 2024 secara konsisten menunjukkan bahwa sekitar 90 hingga 96 persen konten deepfake seksual menargetkan perempuan sementara pembuatnya didominasi oleh laki-laki. Ini bukan kebetulan teknologi; ia mencerminkan pola kekerasan berbasis gender yang sudah lama ada, yang kini mendapatkan instrumen baru yang lebih mudah diakses dan lebih sulit dilacak.

Angka-angka ini penting karena mereka menentukan kerangka analisis yang tepat. Deepfake seksual yang menargetkan perempuan bukan hanya pelanggaran privasi individual ia adalah ekspresi dari sistem ketidaksetaraan gender yang lebih luas, dan oleh karena itu ia layak dianalisis sebagai bentuk kekerasan berbasis gender, bukan hanya sebagai kejahatan teknologi generik. Pembedaan ini punya konsekuensi hukum yang nyata.

CEDAW dan Kewajiban Negara terhadap Kekerasan Berbasis Gender

Convention on the Elimination of All Forms of Discrimination against Women (CEDAW) adalah instrumen HAM internasional utama yang mengatur hak-hak perempuan. Meski teks aslinya (diadopsi 1979) tidak secara eksplisit membahas teknologi digital, Komite CEDAW badan perjanjian yang mengawasi implementasi konvensi telah secara konsisten mengembangkan interpretasi yang responsif terhadap konteks kontemporer.

General Recommendation No. 35 (2017) dari Komite CEDAW adalah dokumen kunci yang memperluas pemahaman tentang kekerasan berbasis gender untuk mencakup kekerasan yang dimediasi teknologi. Rekomendasi ini menegaskan bahwa kekerasan berbasis gender yang dilakukan melalui teknologi informasi dan komunikasi termasuk yang dilakukan oleh aktor non-negara adalah tanggung jawab negara untuk dicegah, diselidiki, dihukum, dan dipulihkan. Negara tidak bisa berdalih bahwa pelakunya adalah individu swasta dan bukan agen negara untuk menghindari tanggung jawab CEDAW.

Implikasi untuk deepfake seksual cukup langsung: negara yang meratifikasi CEDAW dan hampir semua negara di dunia sudah meratifikasinya, termasuk Indonesia memiliki kewajiban spesifik untuk memastikan bahwa perempuan terlindungi dari deepfake seksual. Kewajiban ini mencakup: memiliki kerangka hukum yang mengkriminalisasi atau setidaknya memberikan remediasi yang efektif terhadap deepfake seksual; memastikan akses perempuan terhadap keadilan ketika mereka menjadi korban; dan mengambil langkah-langkah preventif termasuk pendidikan dan perubahan norma sosial.

Deepfake sebagai Kontinum Kekerasan

Konsep yang berguna untuk memahami deepfake seksual dalam kerangka CEDAW adalah kontinum kekerasan (*continuum of violence*) ide bahwa berbagai bentuk kekerasan terhadap perempuan, meski berbeda dalam manifestasinya, terhubung dalam sebuah kontinum yang mencerminkan pola kekuasaan dan kontrol yang sama. Deepfake seksual bukan fenomena yang sepenuhnya baru; ia adalah iterasi teknologi dari praktik-praktik yang sudah lama ada: distribusi foto intim tanpa persetujuan (revenge porn), ancaman publikasi konten seksual untuk memeras atau mengontrol (*sextortion*), dan penghinaan seksual sebagai alat kontrol sosial.

Menempatkan deepfake dalam kontinum ini memiliki implikasi regulasi yang penting. Ia berarti bahwa solusi terhadap deepfake seksual tidak bisa diisolasi dari respons yang lebih luas terhadap kekerasan berbasis gender termasuk reformasi budaya, pendidikan tentang persetujuan, dan penguatan mekanisme perlindungan bagi korban kekerasan seksual secara umum. Regulasi spesifik deepfake diperlukan, tapi ia tidak cukup sendiri.

Perlu dicatat juga bahwa deepfake seksual sering digunakan bukan hanya untuk menyakiti secara langsung, tapi sebagai alat kontrol dan intimidasi. Ancaman "saya akan membuat deepfake-mu" memiliki efek yang sama destruktifnya dengan deepfake itu sendiri, memaksa perempuan untuk mengubah perilaku, menarik diri dari posisi publik, atau menerima tuntutan pelaku. Ini adalah bentuk kekerasan psikologis dan kontrol koersif yang diakui dalam kerangka CEDAW, bahkan ketika deepfake-nya sendiri tidak pernah dibuat.

11.5 PERLINDUNGAN ANAK: KONVENSI HAK ANAK DAN TANTANGAN KHUSUS

Anak sebagai Kelompok yang Memerlukan Perlindungan Khusus

Convention on the Rights of the Child (CRC) mengakui bahwa anak-anak memerlukan perlindungan khusus karena kerentanan khusus mereka. Pasal 16 CRC melindungi privasi anak, Pasal 34 mewajibkan negara untuk melindungi anak dari semua bentuk eksploitasi seksual, dan Pasal 17 mewajibkan negara untuk memastikan bahwa anak memiliki akses terhadap informasi yang sesuai dan terlindungi dari materi yang merugikan.

Dalam konteks deepfake, anak-anak menghadapi setidaknya tiga ancaman yang berbeda:

- pertama, mereka bisa menjadi korban langsung deepfake seksual kasus-kasus di mana gambar-gambar anak yang diunggah oleh orang tua ke media sosial dimanipulasi menjadi konten seksual sudah terdokumentasi dan mengkhawatirkan;
- kedua, mereka bisa menjadi korban deepfake yang digunakan untuk perundungan (bullying) di lingkungan sekolah atau sosial mereka;
- ketiga, mereka bisa terpapar pada deepfake sebagai konsumen baik konten deepfake seksual yang tidak seharusnya mereka lihat, maupun informasi palsu yang dibuat dengan teknologi deepfake.

Optional Protocol dan CSAM Generatif

Optional Protocol on the Sale of Children, Child Prostitution and Child Pornography (OPSC) yang merupakan tambahan pada CRC secara eksplisit mengharuskan negara untuk mengkriminalisasi pornografi anak, termasuk representasi visual yang "memperlihatkan anak terlibat dalam perilaku seksual nyata atau tersimulasi." Pertanyaan kritis yang muncul dengan perkembangan AI generatif adalah: apakah konten yang sepenuhnya dihasilkan oleh AI yang tidak melibatkan anak nyata dalam pembuatannya termasuk dalam cakupan OPSC?

Perdebatan ini masih berlangsung di kalangan akademisi hukum dan pembuat kebijakan. Satu posisi berargumen bahwa jika tidak ada anak nyata yang dieksploitasi dalam pembuatan konten, maka fondasi utama larangan CSAM (*child sexual abuse material*) perlindungan anak dari eksploitasi langsung tidak terpenuhi.

Posisi yang berlawanan, yang tampaknya semakin dominan, berargumen bahwa CSAM generatif berbahaya karena beberapa alasan independen, menormalkan dan bahkan mempromosikan eksploitasi seksual anak, hal tersebut bisa digunakan sebagai alat grooming, dan dalam beberapa kasus, pembuatannya melibatkan penggunaan gambar anak nyata sebagai materi pelatihan. Sebagian besar yurisdiksi yang sudah mempertimbangkan masalah ini memilih untuk mengkriminalisasi CSAM generatif, terlepas dari perdebatan doktrin yang belum selesai.

Deepfake Bullying dan Kerentanan Remaja

Satu dimensi yang sering luput dari perhatian regulasi adalah penggunaan deepfake untuk perundungan di kalangan remaja. Berbeda dari deepfake seksual yang menargetkan orang dewasa, deepfake bullying tidak selalu bersifat seksual, bisa berupa manipulasi wajah seseorang menjadi gambar yang memalukan, atau penciptaan rekaman suara palsu yang menampilkan seseorang mengucapkan hal-hal yang merusak reputasinya di kalangan teman sebaya.

Teknologi yang diperlukan untuk membuat deepfake semacam ini semakin mudah diakses, dan ada laporan yang mengkhawatirkan tentang aplikasi deepfake yang dipasarkan secara implisit untuk tujuan seperti ini. Anak-anak yang menjadi korban deepfake bullying mengalami kerusakan psikologis yang serius termasuk kecemasan, depresi, dan dalam kasus ekstrem, pikiran untuk menyakiti diri sendiri.

Kewajiban negara berdasarkan CRC mencakup memastikan bahwa lingkungan sekolah dan sosial anak aman dari bentuk-bentuk kekerasan ini. Ini berarti tidak hanya regulasi teknologi, tapi juga kebijakan sekolah yang jelas, pelatihan guru, dan program pendidikan tentang penggunaan AI yang bertanggung jawab yang dimulai sejak dini. Mengintegrasikan literasi deepfake ke dalam kurikulum pendidikan adalah, satu bisa berargumen, bukan pilihan tapi kewajiban berdasarkan CRC Pasal 17.

11.6 KEBEBASAN BEREKSPRESI VS. PERLINDUNGAN KORBAN: MENCARI KESEIMBANGAN

Kebebasan Berekspresi Bukan Hak Mutlak

Setiap diskusi serius tentang regulasi deepfake dalam kerangka HAM maupun di luar itu akan bertemu dengan argumen kebebasan berekspresi. Argumennya sering disajikan sebagai dilema: jika kita melarang deepfake, kita membatasi kebebasan berekspresi; jika kita tidak melarangnya, kita membiarkan korban dirugikan. Tapi framing ini adalah oversimplifikasi yang perlu diatasi.

Kebebasan berekspresi dalam hukum HAM internasional bukan hak mutlak. Pasal 19 ayat (3) ICCPR secara eksplisit mengizinkan pembatasan yang (a) diatur dalam hukum, (b) diperlukan untuk tujuan-tujuan yang sah termasuk perlindungan hak orang lain dan (c) proporsional. Ini berarti bahwa pertanyaan yang benar bukan "haruskah kita membatasi kebebasan berekspresi demi melindungi korban deepfake" tetapi "apakah pembatasan spesifik yang diusulkan memenuhi standar legalitas, legitimasi, dan proporsionalitas yang diatur oleh hukum HAM internasional."

Uji Proporsionalitas dalam Konteks Deepfake

Uji proporsionalitas adalah alat analisis yang paling berguna untuk mengevaluasi regulasi deepfake dari perspektif HAM. Ia menuntut kita untuk bertanya: apakah larangan ini adalah cara yang paling tidak membatasi untuk mencapai tujuan perlindungan yang sah? Apakah ada cara lain yang lebih tidak membatasi kebebasan berekspresi tapi sama efektifnya dalam melindungi korban?

Untuk deepfake seksual non-konsensual, uji proporsionalitas relatif mudah dilalui: larangan pembuatan dan distribusi konten ini adalah cara yang tepat sasaran untuk melindungi individu dari bahaya nyata, dan sulit membayangkan argumen yang secara serius mempertahankan deepfake seksual non-konsensual sebagai ekspresi yang layak dilindungi. Tidak ada orang yang bisa berargumen dengan wajah lurus bahwa hak untuk membuat konten seksual palsu dari wajah orang lain tanpa izin adalah ekspresi yang dilindungi ICCPR.

Untuk deepfake satire dan deepfake artistik, analisisnya jauh lebih kompleks. Satire tentang tokoh publik termasuk yang menggunakan representasi visual atau audio yang dimanipulasi memiliki tradisi panjang dalam kebebasan berekspresi dan memainkan peran penting dalam wacana demokratis. Caricature adalah bentuk visual dari deepfake yang sangat sederhana, dan ia telah lama dianggap sebagai ekspresi yang dilindungi. Di mana batas antara satire yang dilindungi dan manipulasi yang berbahaya? Ini adalah pertanyaan yang tidak memiliki jawaban universal dan yang harus dijawab kasus per kasus dalam yurisprudensi yang berkembang.

Efek Dingin dan Siapa yang Sebenarnya Membungkam Siapa

Argumen efek dingin (*chilling effect*) bahwa regulasi deepfake akan membuat orang takut untuk berekspresi secara sah perlu diperiksa dengan lebih hati-hati dari yang biasanya dilakukan dalam perdebatan kebijakan. Argumen ini sering digunakan oleh kelompok yang khawatir tentang sensor berlebihan, dan kekhawatiran itu bukan tanpa dasar. Regulasi yang terlalu luas memang bisa menciptakan efek dingin yang nyata.

Tapi ada efek dingin yang lain yang jarang dibahas secara seimbang: efek dingin yang diciptakan oleh deepfake itu sendiri terhadap korban dan terhadap kelompok yang paling rentan. Perempuan yang mengetahui bahwa gambar mereka bisa kapan saja dimanipulasi menjadi deepfake seksual mungkin menarik diri dari platform publik, mengurangi kehadiran daring, atau menghindari posisi yang menempatkan mereka di mata publik. Ini juga adalah efek dingin yang nyata pengebirian ekspresi bukan oleh negara, tapi oleh ancaman kekerasan privat yang difasilitasi oleh teknologi.

Analisis HAM yang seimbang harus mempertimbangkan kedua efek dingin ini secara bersama-sama, bukan hanya yang pertama. Dari perspektif ICCPR, negara yang gagal melindungi perempuan dari deepfake seksual, sehingga perempuan menarik diri dari ruang publik, sebenarnya juga gagal melindungi kebebasan berekspresi perempuan itu sebuah ironi yang sering tidak terlihat dalam perdebatan yang mengframing masalah ini semata-mata sebagai regulasi vs. kebebasan.

11.7 KEWAJIBAN NEGARA: DUE DILIGENCE, REGULASI, DAN REMEDIASI

Due Diligence sebagai Standar Kewajiban

Standar due diligence adalah konsep sentral dalam hukum HAM internasional yang mengatur kewajiban negara terhadap tindakan aktor non-negara. Standar ini pertama kali diartikulasikan secara eksplisit dalam konteks kekerasan terhadap perempuan oleh Pelapor Khusus PBB Radhika Coomaraswamy pada 1990-an, dan sejak itu telah berkembang menjadi standar umum yang berlaku untuk berbagai konteks.

Dalam konteks deepfake, due diligence negara berarti setidaknya empat hal yang beroperasi secara bersamaan. Negara harus mencegah mengambil langkah-langkah untuk mencegah terjadinya pelanggaran hak melalui deepfake, termasuk melalui regulasi, pendidikan, dan pengawasan platform. Negara harus menyelidiki ketika pelanggaran terjadi, negara harus memiliki kapasitas untuk menyelidikinya secara efektif, yang berarti penyidik yang terlatih, alat forensik digital yang memadai, dan prosedur yang tidak menghambat korban untuk melapor. Negara harus menghukum jika investigasi menemukan pelaku, harus ada sanksi yang proporsional dan efektif. Dan negara harus memulihkan korban harus mendapatkan remediasi yang efektif, yang mencakup lebih dari sekadar penghapusan konten.

Kegagalan di salah satu dari empat elemen ini berarti negara tidak memenuhi standar due diligence, dan dapat diminta pertanggungjawabannya di bawah mekanisme HAM internasional yang relevan. Perlu dicatat bahwa standar ini tidak mensyaratkan negara untuk menjamin bahwa tidak akan pernah ada deepfake itu adalah standar yang tidak realistis. Yang dituntut adalah bahwa negara mengambil langkah-langkah yang wajar dan efektif; standarnya adalah reasonableness, bukan perfection.

Kewajiban Regulasi: Apa yang Harus Diatur

Kewajiban regulasi negara dalam konteks deepfake mencakup beberapa domain yang saling terkait. Yang paling jelas adalah kewajiban untuk memiliki kerangka hukum yang memadai undang-undang atau peraturan yang melarang atau memberikan remediasi terhadap deepfake berbahaya, dengan definisi yang cukup jelas untuk memberikan kepastian hukum tapi cukup luwes untuk mencakup perkembangan teknologi yang tidak terantisipasi.

Yang sering kurang mendapat perhatian adalah kewajiban untuk mengatur platform. Sebagian besar deepfake berbahaya tidak hanya dibuat oleh individu, disebarkan melalui platform yang memiliki sumber daya dan kapasitas untuk mendeteksi dan menghapus konten semacam itu, tapi sering kali tidak memiliki insentif yang cukup kuat untuk melakukan itu tanpa tekanan regulasi. Kewajiban due diligence negara mencakup memastikan bahwa platform tidak menjadi fasilitator pasif atau aktif dari pelanggaran hak berbasis deepfake.

Ada juga kewajiban yang lebih jarang dibahas: mengatur pengembang teknologi deepfake itu sendiri. Apakah perusahaan yang mengembangkan dan menjual alat deepfake memiliki kewajiban untuk membangun safeguard yang mencegah penggunaan berbahaya? Argumen dari kerangka Prinsip-Prinsip Panduan PBB tentang Bisnis dan HAM (UN Guiding Principles on Business and Human Rights, atau UNGPs) adalah ya, bisnis memiliki tanggung jawab untuk menghormati HAM, yang berarti melakukan due diligence HAM sebelum dan selama operasi mereka termasuk dalam desain dan pemasaran produk.

Akses terhadap Remediasi yang Efektif

Hak atas remediasi yang efektif adalah hak yang sering paling sulit diwujudkan dalam praktik. Untuk korban deepfake, remediasi yang efektif mencakup setidaknya: penghapusan konten yang cepat dan komprehensif; koreksi publik yang memadai jika reputasi dirugikan; kompensasi finansial untuk kerugian yang dapat dihitung; dukungan psikologis; dan jaminan non-pengulangan.

Realitas yang dihadapi kebanyakan korban sangat jauh dari standar ini. Proses penghapusan konten di platform sering lambat dan tidak transparan. Koreksi publik hampir tidak pernah mencapai skala penyebaran aslinya berita tentang deepfake palsu mungkin viral dengan jutaan tayangan, sementara koreksinya dilihat oleh segelintir orang. Kompensasi finansial melalui jalur hukum memerlukan identifikasi pelaku, proses pengadilan yang panjang, dan kemampuan finansial yang tidak dimiliki semua korban. Dukungan psikologis jarang disediakan oleh negara secara sistematis untuk korban kekerasan digital.

Kesenjangan antara standar remediasi yang dituntut oleh hukum HAM internasional dan realitas yang dihadapi korban adalah salah satu argumen terkuat untuk reformasi sistemik bukan hanya reformasi hukum, tapi reformasi cara platform beroperasi, cara institusi penegak hukum merespons, dan cara sistem layanan dukungan korban didesain.

Mekanisme Akuntabilitas Internasional

Ketika negara gagal memenuhi kewajiban HAM-nya dalam konteks deepfake, mekanisme akuntabilitas internasional apa yang tersedia? Jawabannya bergantung pada instrumen mana yang relevan dan apakah negara bersangkutan telah meratifikasinya dan mengakui yurisdiksi badan-badan terkait.

Universal Periodic Review (UPR) PBB adalah mekanisme yang berlaku untuk semua negara anggota dan yang semakin sering digunakan oleh kelompok masyarakat sipil untuk mengangkat isu kekerasan digital termasuk deepfake. Rekomendasi UPR tidak mengikat secara hukum, tapi tekanan diplomatik yang menyertainya tidak bisa diabaikan begitu saja.

Komite CEDAW dan Komite Hak Anak dapat menerima laporan dari negara dan, dalam beberapa kasus, komunikasi individual dari korban memberikan jalur akuntabilitas yang lebih langsung. Komite CEDAW, khususnya, semakin aktif dalam merespons laporan tentang kekerasan digital terhadap perempuan, termasuk yang melibatkan teknologi deepfake.

Mekanisme-mekanisme ini memiliki keterbatasan yang nyata mereka lambat, tidak memiliki kewenangan eksekusi, dan bergantung pada itikad baik negara untuk mengimplementasikan rekomendasinya. Tapi mereka memainkan fungsi penting dalam mendokumentasikan kegagalan negara, membangun tekanan normatif, dan menyediakan preseden interpretasi yang membentuk standar HAM dari waktu ke waktu.

11.8 IMPLIKASI UNTUK INDONESIA DAN KONTEKS ASIA TENGGARA

Status Ratifikasi dan Kewajiban yang Berlaku

Indonesia telah meratifikasi ICCPR (2005) dan ICESCR (2005), dan sudah meratifikasi CEDAW (1984) serta CRC (1990). Ini berarti kerangka kewajiban yang dibahas dalam bab ini bukan sekadar standar internasional abstrak, namun juga kewajiban hukum yang mengikat Indonesia dan yang dapat dijadikan dasar akuntabilitas di forum internasional.

Yang menjadi pertanyaan bukan apakah kewajiban itu ada, melainkan sejauh mana sistem hukum Indonesia sudah memenuhinya dalam konteks deepfake. Undang-Undang ITE dan berbagai peraturan terkait memberikan beberapa pijakan, tapi kerangka spesifik untuk deepfake masih dalam tahap pembentukan. Analisis kesenjangan (*gap analysis*) antara kewajiban HAM internasional Indonesia dan kerangka hukum domestik yang ada adalah pekerjaan yang sangat diperlukan dan yang sayangnya belum banyak dilakukan.

Konteks Regional: ASEAN dan HAM

ASEAN memiliki instrumen HAM regionalnya sendiri ASEAN Human Rights Declaration (AHRD, 2012) dan Komisi Antarpemerintah ASEAN untuk HAM (AICHR). Tapi kerangka HAM regional ASEAN secara historis lebih lemah dari kerangka universal, dengan prinsip non-intervensi yang kuat dan konsensus yang sering menghambat posisi yang tegas tentang pelanggaran HAM di negara anggota.

Dalam konteks kekerasan digital termasuk deepfake, ASEAN belum mengembangkan posisi atau standar regional yang kohesif. Ini adalah kesenjangan yang seharusnya mendorong negara-negara anggota termasuk Indonesia untuk tidak menunggu konsensus regional tapi bergerak dengan mengacu pada standar HAM universal yang lebih kuat.

Peran Masyarakat Sipil dan Akademisi

Dalam konteks di mana mekanisme negara sering lambat dan tidak memadai, peran masyarakat sipil dan akademisi menjadi sangat penting. Di banyak negara, kemajuan dalam regulasi deepfake didorong bukan oleh inisiatif pemerintah semata, tapi oleh tekanan dari kelompok advokasi korban, penelitian akademis yang mendokumentasikan skala dan dampak masalah, dan litigasi strategis yang membangun preseden hukum.

Komunitas akademik hukum di Indonesia dan kawasan ASEAN memiliki peran yang belum sepenuhnya dimainkan dalam konteks ini: mengembangkan analisis yang mendalam tentang kesenjangan antara kewajiban HAM internasional dan kerangka hukum domestik, mendokumentasikan kasus-kasus dan dampaknya secara sistematis, mengembangkan rekomendasi kebijakan yang berbasis bukti, dan melatih generasi praktisi hukum yang mampu menggunakan kerangka HAM internasional dalam advokasi kasus-kasus konkret.

11.9 SINTESIS: DEEPFAKE SEBAGAI MASALAH HAM YANG TERINTEGRASI

Setelah menelusuri lima lini argumen HAM yang relevan untuk deepfake privasi, martabat dan reputasi, perlindungan perempuan, perlindungan anak, dan keseimbangan kebebasan berekspresi serta menganalisis kewajiban negara yang mengikutinya, apa yang bisa kita simpulkan?

Pertama, deepfake adalah masalah HAM yang nyata dan multidimensional. Deepfake tidak hanya menyentuh satu hak, Deepfake menyentuh jaringan hak yang saling terkait, dan pelanggaran terhadap satu hak sering memicu pelanggaran terhadap yang lain. Analisis yang memadai harus menangkap jaringan ini, bukan hanya satu simpulnya.

Kedua, kerangka HAM internasional yang ada meski tidak dirancang dengan deepfake dalam pikiran menyediakan fondasi yang cukup kuat untuk membangun kewajiban regulasi yang bermakna. Instrumen-instrumen seperti ICCPR, CEDAW, dan CRC tidak perlu direvisi untuk berlaku pada konteks deepfake; yang diperlukan adalah interpretasi yang responsif dan berani dari badan-badan perjanjian yang mengawasinya.

Ketiga, kewajiban negara dalam kerangka ini adalah nyata dan dapat dijadikan dasar akuntabilitas. Due diligence bukan slogan, namun standar hukum yang memiliki konten spesifik: mencegah, menyelidiki, menghukum, dan memulihkan. Negara yang gagal di salah satu dimensi ini tidak memenuhi standar HAM internasional, terlepas dari apakah ia sudah memiliki undang-undang deepfake atau tidak.

Keempat pelajaran yang paling penting dari bab ini regulasi hukum saja tidak cukup. Kerangka HAM menuntut pendekatan yang jauh lebih holistik: reformasi institusional, kapasitas

penegakan yang memadai, akses keadilan yang nyata bagi korban, dan perubahan norma sosial yang menopang semuanya. Bab-bab berikutnya akan mengeksplorasi bagaimana komponen-komponen ini bisa diintegrasikan dalam kebijakan yang koheren dengan Indonesia dan konteks Asia Tenggara sebagai titik rujukan utama.

Satu hal yang layak direnungkan sebagai penutup: kerangka HAM internasional, pada akhirnya, adalah cermin tentang apa yang kita anggap sebagai kehidupan yang layak bagi semua manusia. Deepfake mengancam banyak hal yang ada di cermin itu privasi, martabat, keamanan, partisipasi dalam kehidupan publik. Bahwa teknologi bisa melakukan ini dalam skala dan kecepatan yang belum pernah ada sebelumnya bukan alasan untuk menyerah pada fatalism teknologi, tapi justru alasan untuk memperkuat komitmen terhadap nilai-nilai yang kerangka HAM itu dirancang untuk melindungi.

BAB 12
PEMBUKTIAN DIGITAL: FORENSIK DEEPFAKE DI PERSIDANGAN

"Pengadilan tidak membutuhkan kebenaran mutlak. Ia membutuhkan standar pembuktian yang bisa diandalkan dan teknologi baru selalu menantang apa yang kita anggap sebagai standar yang bisa diandalkan."
— parafrase dari literatur forensik digital kontemporer

Bayangkan skenario berikut: seorang terdakwa menghadapi dakwaan pencemaran nama baik berbasis video sebuah rekaman yang menampilkan dirinya mengucapkan kata-kata yang tidak pernah ia katakan. Tim pengacaranya mengklaim bahwa video itu adalah deepfake. Jaksa penuntut bersikeras bahwa video itu asli. Majelis hakim, yang tidak memiliki latar belakang teknis, harus memutuskan siapa yang benar. Ahli dari mana yang bisa dipercaya? Metode apa yang diakui sah? Standar apa yang digunakan untuk menentukan apakah video itu palsu?

Skenario semacam ini bukan spekulasi akademis. Kasus-kasus yang melibatkan konten media yang dipertanyakan keasliannya sudah mulai masuk ke pengadilan di berbagai negara, dan hakim-hakim di seluruh dunia termasuk Indonesia belum sepenuhnya siap menghadapinya. Ini bukan kritik terhadap kapasitas hakim sebagai individu; ini adalah pengakuan bahwa sistem hukum kita, secara institusional, belum mengembangkan kerangka yang memadai untuk mengevaluasi bukti digital yang semakin kompleks, dan deepfake adalah puncak gunung es dari kompleksitas itu.

Bab ini mendekati masalah pembuktian deepfake dari empat sudut yang saling melengkapi. Pertama, dari sisi hukum acara: bagaimana kerangka alat bukti elektronik dalam sistem hukum Indonesia mengakomodasi atau gagal mengakomodasi bukti berbasis AI. Kedua, dari sisi teknis: apa yang sebenarnya bisa dilakukan oleh forensik deepfake, dan apa batas-batas riilnya yang sering tidak diungkapkan secara jujur. Ketiga, dari sisi prosedural: bagaimana chain of custody digital bekerja dalam praktiknya dan mengapa ia sangat mudah rusak. Keempat, dari sisi kelembagaan: siapa ahli forensik digital yang layak dipercaya di pengadilan, dan seberapa besar kesenjangan antara kapasitas laboratorium forensik Indonesia dengan standar internasional. Keempat sudut ini tidak bisa dipisahkan kegagalan di satu sudut akan merusak keseluruhan upaya pembuktian.

12.1 ALAT BUKTI ELEKTRONIK DALAM HUKUM ACARA INDONESIA

Kerangka Dasar: KUHAP dan Keterbatasannya

Kitab Undang-Undang Hukum Acara Pidana (KUHAP) yang disahkan pada 1981 mendefinisikan lima jenis alat bukti yang sah: keterangan saksi, keterangan ahli, surat, petunjuk, dan keterangan terdakwa. Tidak ada satu pun dari lima kategori ini yang secara eksplisit menyebut

bukti elektronik karena pada 1981, bukti elektronik sebagai konsep praktis belum ada dalam kehidupan hukum sehari-hari Indonesia. Video deepfake, rekaman digital, metadata file, log server semuanya harus dipaksakan masuk ke dalam kategori-kategori yang diciptakan untuk dunia analog.

Pemaksaan ini bukan tanpa konsekuensi. Ketika bukti digital diperlakukan sebagai "surat" dalam pengertian KUHAP, misalnya, prosedur otentikasi yang berlaku adalah prosedur yang dirancang untuk dokumen kertas yang sangat berbeda secara fundamental dari prosedur verifikasi integritas file digital. Ketika saksi ahli komputer bersaksi tentang analisis deepfake, ia masuk dalam kategori "keterangan ahli" tapi tidak ada standar yang jelas tentang kualifikasi apa yang membuat seseorang menjadi "ahli" yang diakui dalam konteks digital forensik.

Undang-Undang Nomor 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik (UU ITE), yang kemudian direvisi melalui UU Nomor 19 Tahun 2016 dan UU Nomor 1 Tahun 2024, memberikan pijakan yang lebih spesifik untuk alat bukti elektronik. Pasal 5 UU ITE menyatakan bahwa informasi elektronik dan dokumen elektronik adalah alat bukti yang sah, dengan syarat bahwa informasi itu diperoleh, disimpan, diolah, dan ditampilkan menggunakan sistem elektronik sesuai ketentuan yang berlaku. Syarat terakhir inilah yang membuka banyak pertanyaan: sistem elektronik seperti apa yang "sesuai ketentuan"? Siapa yang berwenang menentukan kesesuaian itu?

UU ITE dan Standar Admissibility Bukti Digital

Dalam praktik persidangan, admissibility apakah sebuah alat bukti dapat diterima sebagai bukti adalah langkah pertama yang harus dilalui sebelum pertanyaan tentang kekuatan pembuktian (*probative value*) bahkan bisa dibahas. Untuk bukti digital umum, yurisprudensi Indonesia sudah mulai membentuk standar, meski masih dengan banyak ketidakpastian. Untuk deepfake secara spesifik, belum ada yurisprudensi yang jelas.

Dari ketentuan UU ITE dan perkembangan yurisprudensi yang ada, setidaknya ada tiga pertanyaan admissibility yang relevan untuk bukti deepfake. Pertama, apakah bukti diperoleh secara sah apakah video yang diklaim sebagai deepfake diperoleh melalui prosedur yang tidak melanggar hak-hak terdakwa atau pihak lain? Kedua, apakah integritas bukti terjaga apakah ada bukti bahwa file digital tidak dimodifikasi sejak diperoleh? Ketiga, apakah analisis forensik yang dilakukan terhadap bukti itu memenuhi standar metodologi yang dapat dipertanggungjawabkan?

Hal yang membuat kasus deepfake secara khusus rumit adalah bahwa semua tiga pertanyaan ini lebih sulit dijawab dibanding kasus bukti digital biasa. Integritas bukti, misalnya, biasanya diverifikasi melalui hash kriptografis algoritma yang menghasilkan nilai unik dari sebuah file, sehingga modifikasi sekecil apapun pada file akan mengubah nilainya. Ini adalah mekanisme yang bekerja dengan baik untuk dokumen atau foto. Tapi untuk video yang diklaim sebagai deepfake, pertanyaannya bukan hanya apakah file video itu dimodifikasi sejak diperoleh, melainkan apakah konten video itu sendiri merupakan fabrikasi sebuah pertanyaan yang sama sekali berbeda dan yang tidak bisa dijawab hanya dengan verifikasi hash.

Perbandingan dengan Standar Internasional: Daubert dan Frye

Di Amerika Serikat, dua standar utama yang mengatur admissibility keterangan ahli dalam perkara federal adalah standar Frye (lama) dan standar Daubert (yang lebih baru dan lebih dominan). Untuk kepentingan diskusi ini, standar Daubert lebih relevan karena ia secara eksplisit mengharuskan hakim untuk berperan sebagai "gatekeeper" yang mengevaluasi apakah metodologi yang digunakan ahli adalah ilmiah yang sah.

Standar Daubert mencakup beberapa faktor: apakah teori atau teknik yang digunakan telah diuji; apakah ia telah melewati peer review dan publikasi; apakah ada tingkat kesalahan yang diketahui dan standar yang mengontrol operasinya; dan apakah ia diterima secara luas dalam komunitas ilmiah yang relevan. Ketika standar ini diterapkan pada teknik deteksi deepfake, hasilnya tidak selalu menggembirakan bagi para penuntut: banyak alat deteksi deepfake yang tersedia saat ini memiliki tingkat kesalahan yang signifikan, dan komunitas ilmiah belum memiliki konsensus tentang standar metodologi yang diterima secara universal.

Indonesia tidak secara formal mengadopsi standar Daubert atau Frye sistem hukum kita tidak mengoperasikan mekanisme yang ekuivalen. Tapi prinsip dasarnya bahwa keterangan ahli harus didasarkan pada metodologi yang dapat dipertanggungjawabkan adalah prinsip yang berlaku secara universal dalam logika pembuktian. Masalahnya adalah bahwa tanpa standar yang terartikulasi dengan baik, hakim Indonesia tidak memiliki kerangka yang jelas untuk mengevaluasi apakah keterangan ahli forensik digital yang mereka dengar memenuhi standar itu atau tidak.

12.2 TEKNIK DETEKSI DEEPFAKE

Pengantar: Mengapa Deteksi Lebih Sulit dari yang Tampak

Ada kecenderungan dalam liputan media dan bahkan dalam beberapa diskusi kebijakan untuk memperlakukan deteksi deepfake sebagai masalah teknis yang sudah terselesaikan atau segera akan terselesaikan seolah-olah yang diperlukan hanyalah alat yang tepat dan deepfake bisa diidentifikasi dengan kepastian tinggi. Kenyataannya jauh lebih kompleks dan lebih tidak nyaman dari itu.

Deteksi deepfake adalah pertarungan antara sistem yang berusaha membuat deepfake yang semakin tidak terdeteksi dan sistem yang berusaha mendeteksinya. Ini adalah dynamic arms race yang tidak memiliki titik akhir yang stabil: setiap kemajuan dalam deteksi mendorong kemajuan dalam pembuatan, dan sebaliknya. Model deteksi yang dilatih pada dataset deepfake tertentu sering kali gagal ketika berhadapan dengan deepfake yang dibuat menggunakan model generatif yang berbeda atau yang lebih baru. Ini adalah masalah generalisasi yang fundamental, dan ia belum terpecahkan.

Untuk konteks pembuktian di pengadilan, implikasinya serius: tidak ada alat deteksi deepfake yang dapat memberikan kesimpulan dengan kepastian absolut. Yang bisa diberikan adalah probabilitas dan probabilitas, dalam konteks hukum pidana dengan standar *"beyond*

reasonable doubt" atau dalam konteks perdata dengan standar *"preponderance of evidence,"* harus dikomunikasikan dengan sangat hati-hati agar tidak menyesatkan majelis hakim.

Analisis Artefak Visual: Melihat Jejak yang Ditinggalkan AI

Analisis artefak visual adalah pendekatan deteksi deepfake yang paling intuitif: mencari inkonsistensi visual yang ditinggalkan oleh proses generasi AI. Model-model deepfake, meski semakin canggih, masih cenderung meninggalkan jejak-jejak tertentu yang bisa diidentifikasi oleh sistem analisis yang tepat.

Salah satu artefak yang paling sering dibicarakan adalah inkonsistensi di area wajah yang memiliki kompleksitas tinggi: tepi wajah (area di mana wajah yang disisipkan bertemu dengan latar belakang atau rambut), gigi, dan mata. Model-model generatif awal, misalnya, sering menghasilkan gigi yang tampak terlalu sempurna atau terlalu kabur, dan refleksi di kornea mata yang tidak konsisten dengan sumber cahaya di adegan. Model yang lebih baru sudah jauh lebih baik dalam menangani inkonsistensi ini, tapi analisis dengan resolusi tinggi dan teknik magnifikasi tertentu masih bisa mengungkapnya dalam banyak kasus.

Artefak temporal inkonsistensi yang muncul ketika menganalisis video frame per frame adalah area lain yang menjanjikan. Karena model deepfake mengolah setiap frame secara independen atau dalam batch terbatas, transisi antara frame sering tidak sehalus video asli: ada flickering halus di area tertentu, atau gerakan wajah yang tidak persis mengikuti fisika kepala manusia yang sebenarnya. Analisis temporal yang sistematis menggunakan optical flow dan teknik analisis gerak dapat mendeteksi anomali ini.

Artefak visual ini semakin sulit dideteksi seiring kemajuan teknologi. Deepfake yang dibuat dengan model-model terbaru pada tahun 2025 jauh lebih bersih dari yang dibuat pada tahun 2019. Analisis artefak yang berhasil pada deepfake generasi lama mungkin tidak berhasil pada deepfake generasi baru dan ini adalah keterbatasan yang sangat relevan untuk konteks pembuktian, di mana kita sering tidak tahu kapan deepfake yang sedang diperiksa dibuat atau dengan alat apa.

Analisis Metadata: Mempertanyakan Asal-Usul File

Metadata adalah informasi yang tertanam dalam file digital tentang file itu sendiri: kapan dibuat, dengan perangkat apa, di mana (jika ada data GPS), dengan perangkat lunak apa diedit, dan seterusnya. Analisis metadata adalah salah satu teknik forensik yang paling dasar dan yang sering memberikan petunjuk pertama tentang apakah sebuah file mungkin telah dimanipulasi.

Untuk video yang diklaim sebagai rekaman asli tapi sebenarnya adalah deepfake, analisis metadata sering mengungkapkan inkonsistensi yang sulit dijelaskan. Misalnya: metadata kamera yang mengklaim video diambil dengan iPhone tertentu tapi pola noise sensor yang tidak sesuai dengan karakteristik kamera tersebut; timestamp yang diklaim menunjukkan rekaman tahun lalu tapi format file yang hanya tersedia dalam perangkat lunak yang dirilis tahun ini; atau metadata yang menunjukkan bahwa file telah diekspor dari perangkat lunak pengedit video tertentu sebuah langkah yang tidak akan ada jika video itu benar-benar rekaman mentah dari kamera.

Analisis metadata relatif lebih mudah dipahami oleh hakim dan juri dibanding analisis artefak visual yang memerlukan pemahaman teknis mendalam. Ini menjadikannya alat komunikasi yang berharga dalam persidangan tapi dengan satu caveat penting: metadata bisa dimanipulasi. Seseorang yang memiliki pengetahuan teknis yang memadai bisa menghapus, mengubah, atau memalsukan metadata untuk membuat deepfake tampak seperti rekaman asli, atau sebaliknya, untuk membuat rekaman asli tampak seperti deepfake. Kehadiran metadata yang konsisten bukan bukti keaslian; absennya inkonsistensi metadata bukan bukti bahwa manipulasi tidak terjadi.

Watermark dan Content Provenance: Infrastruktur Masa Depan

Watermark digital adalah teknik yang menyematkan penanda yang tidak terlihat ke dalam konten teks, gambar, atau video yang memungkinkan verifikasi asal-usul dan integritas konten tersebut. Ada dua jenis yang relevan untuk konteks deepfake: watermark kriptografis yang membuktikan bahwa konten dihasilkan oleh sistem AI tertentu, dan watermark provenance yang membuktikan bahwa konten tidak dimodifikasi sejak ditandai.

Coalition for Content Provenance and Authenticity (C2PA) adalah standar industri yang sedang dikembangkan secara aktif oleh konsorsium yang mencakup Adobe, Microsoft, Intel, BBC, dan sejumlah perusahaan teknologi besar lainnya. C2PA mendefinisikan format metadata yang menyertai konten media dengan catatan kriptografis tentang asal-usul dan modifikasi yang dilakukan: siapa yang membuatnya, kapan, dengan alat apa, dan apa saja perubahan yang terjadi sejak pembuatan.

Jika diterapkan secara luas, C2PA bisa menjadi fondasi teknis untuk sistem pembuktian keaslian konten media sebuah semacam rantai kustodi digital yang bersifat kriptografis dan karena itu sangat sulit dipalsukan. Beberapa perangkat kamera dan platform media sosial sudah mulai mengimplementasikan elemen-elemen C2PA, dan regulasi China tentang deep synthesis, yang dibahas di Bab 10, sudah mewajibkan standar pelabelan yang konsisten dengan pendekatan ini.

Ada dua keterbatasan besar yang harus diakui. Pertama, C2PA hanya efektif untuk konten yang dibuat setelah sistem ini diimplementasikan dan tidak membantu mengevaluasi keaslian konten yang sudah ada sebelum sistem ini ada. Kedua adalah keterbatasan yang lebih mendasar sistem watermark bisa di-bypass jika seseorang merekam ulang layar yang menampilkan deepfake (*screen recording*), yang efektif menghapus semua metadata asli dan menggantinya dengan metadata baru yang tidak mengandung informasi tentang manipulasi yang terjadi sebelumnya.

Analisis Frekuensi dan Forensik Level Sinyal

Di luar analisis visual dan metadata, ada pendekatan forensik yang beroperasi di level yang lebih dalam: analisis frekuensi dan forensik level sinyal. Pendekatan ini menguji distribusi statistik nilai piksel dalam gambar atau video untuk mendeteksi pola yang tidak konsisten dengan konten yang diambil oleh kamera fisik.

Error Level Analysis (ELA) adalah teknik yang memvisualisasikan perbedaan dalam tingkat kompresi di berbagai area gambar gambar yang dikompresi secara seragam menunjukkan pola

ELA yang konsisten, sementara area yang diedit dan kemudian dikompresi ulang menunjukkan anomali yang terlihat. Untuk deepfake video, varian temporal dari ELA dapat mengungkapkan area di mana generasi AI menyisipkan atau memodifikasi piksel secara berbeda dari konten sekitarnya.

Noise analysis mengeksploitasi fakta bahwa setiap sensor kamera memiliki pola noise yang unik semacam sidik jari sensor yang konsisten di semua gambar yang diambil oleh kamera tersebut. Konten yang dihasilkan oleh AI tidak memiliki pola noise sensor yang konsisten, atau memiliki pola yang tidak sesuai dengan kamera yang diklaim menghasilkannya. Teknik ini, yang dikenal sebagai PRNU (*Photo Response Non-Uniformity*) analysis, sudah cukup mapan dalam forensik gambar, meski penerapannya untuk video deepfake modern masih dalam pengembangan aktif.

Teknik-teknik level sinyal ini lebih sulit untuk dimanipulasi oleh pembuat deepfake dibanding analisis visual atau metadata, karena mereka beroperasi pada properti statistik yang fundamental dari data digital bukan pada atribut yang mudah diedit. Tapi mereka juga lebih sulit untuk dikomunikasikan di pengadilan: menjelaskan kepada majelis hakim mengapa distribusi frekuensi DCT (*Discrete Cosine Transform*) yang anomali membuktikan bahwa video adalah deepfake memerlukan kemampuan komunikasi yang luar biasa dari saksi ahli.

12.3 CHAIN OF CUSTODY DIGITAL: MENJAGA INTEGRITAS BUKTI

Mengapa Chain of Custody Kritis untuk Bukti Digital

Chain of custody adalah rekam jejak yang mendokumentasikan siapa yang memegang sebuah alat bukti, kapan, dan dalam kondisi apa dari saat pertama kali ditemukan atau disita hingga saat disajikan di pengadilan. Untuk bukti fisik, konsep ini relatif intuitif: jika seseorang merusak, menukar, atau memodifikasi barang bukti tanpa dokumentasi yang tepat, integritas bukti itu dipertanyakan dan bisa mengarah pada exclusion (pengecualian bukti) di pengadilan.

Untuk bukti digital, prinsipnya sama tapi implementasinya jauh lebih kompleks karena sifat unik media digital. Berbeda dari barang bukti fisik yang memiliki eksistensi tunggal hanya ada satu pisau, satu dokumen kertas file digital dapat disalin dengan sempurna tanpa meninggalkan jejak perbedaan antara asli dan salinan. Tindakan sekadar membuka sebuah file di komputer bisa mengubah timestamp dan metadata tanpa ada modifikasi konten yang terlihat. Dan berbeda dari barang bukti fisik yang terdegradasi secara fisik dari waktu ke waktu, file digital bisa dimodifikasi tanpa meninggalkan jejak yang kasat mata hanya dengan alat yang tepat.

Ini berarti bahwa prosedur chain of custody untuk bukti digital harus jauh lebih ketat dan lebih teknis dari prosedur untuk bukti fisik konvensional. Kegagalan dalam chain of custody digital tidak selalu berarti bukti dikecualikan hakim memiliki diskresi tapi ia menciptakan kerentanan yang bisa dieksploitasi oleh pihak lawan dan yang bisa, dalam kasus-kasus di mana hakim mengikuti standar yang ketat, mengakibatkan bukti yang secara substansi valid menjadi tidak dapat digunakan.

Prosedur Akuisisi Bukti Digital yang Benar

Akuisisi bukti digital yang benar dimulai dari momen pertama penyidik berinteraksi dengan media yang mengandung bukti dan ini adalah titik di mana kesalahan paling sering terjadi. Prinsip dasar yang sudah mapan dalam forensik digital internasional adalah: jangan pernah bekerja langsung pada media asli.

Forensic imaging pembuatan salinan bit-for-bit dari media penyimpanan asli adalah langkah pertama yang harus dilakukan sebelum analisis apapun. Salinan forensik ini dibuat menggunakan write blocker (perangkat yang mencegah sistem operasi melakukan operasi tulis ke media asli selama proses penyalinan), dan hasilnya diverifikasi menggunakan hash kriptografis biasanya SHA-256 atau MD5 yang membuktikan bahwa salinan identik dengan aslinya. Semua analisis berikutnya dilakukan terhadap salinan ini, bukan terhadap media asli.

Untuk video deepfake yang ditemukan di platform media sosial atau dikirim melalui aplikasi pesan, prosedurnya lebih kompleks. File yang diunduh dari platform sudah melalui kompresi dan transcoding oleh platform tersebut, yang berarti ia bukan lagi representasi persis dari file yang diunggah metadata aslinya mungkin sudah berubah, dan artefak kompresi tambahan sudah ditambahkan. Ini harus didokumentasikan dengan jelas dalam laporan forensik, karena ia bisa mempengaruhi interpretasi hasil analisis berikutnya.

Dalam konteks Indonesia, ada komplikasi tambahan: koordinasi dengan platform asing Meta, TikTok, Telegram untuk mendapatkan data asli (bukan hanya konten yang terlihat oleh pengguna, tapi metadata server, log akses, dan informasi akun) memerlukan proses *Mutual Legal Assistance* (MLA) yang bisa memakan waktu berbulan-bulan, dan platform tidak selalu menyimpan data yang diperlukan selama periode yang cukup lama. Seringkali, yang tersedia untuk dianalisis oleh penyidik Indonesia hanyalah konten yang sudah dikompresi beberapa kali, bukan data aslinya.

Dokumentasi dan Preservasi

Setiap langkah dalam penanganan bukti digital harus didokumentasikan secara rinci: siapa yang melakukan tindakan apa, pada media apa, menggunakan perangkat lunak versi apa, pada waktu apa, dan dengan hasil apa. Dokumentasi ini bukan formalitas birokrasi, melainkan sebuah fondasi yang memungkinkan pihak lawan, ahli independen, atau pengadilan banding untuk mereproduksi atau memverifikasi hasil analisis.

Preservasi jangka panjang adalah tantangan tersendiri yang sering diabaikan. Bukti digital yang diserahkan ke pengadilan harus disimpan dalam kondisi yang memastikan integritas jangka panjang tidak hanya selama proses persidangan, tapi juga untuk kemungkinan banding. Format file berubah, perangkat lunak yang digunakan untuk membaca format tertentu menjadi obsolete, dan media penyimpanan fisik (hard drive, DVD, USB) memiliki umur yang terbatas. Tidak ada standar nasional yang jelas di Indonesia tentang bagaimana bukti digital harus dipreservasi untuk kepentingan jangka panjang.

Chain of Custody dalam Konteks Deepfake: Tantangan Spesifik

Menguji bukti deepfake secara spesifik, ada tantangan chain of custody yang tidak ada dalam kasus bukti digital lain. Ketika penyidik menyita sebuah video yang diduga deepfake, mereka menghadapi pertanyaan yang tidak mudah: haruskah analisis forensik dilakukan sebelum bukti diserahkan secara resmi, atau sesudahnya? Jika dilakukan sebelumnya misalnya oleh tim teknis penyidik apakah prosedur yang digunakan memenuhi standar forensik yang dapat dipertanggungjawabkan di pengadilan? Apakah alat yang digunakan adalah alat yang berlisensi dan yang bisa diverifikasi?

Ada juga pertanyaan tentang siapa yang memiliki akses ke bukti selama proses analisis. Analisis deepfake yang komprehensif mungkin memerlukan pengiriman file ke laboratorium khusus atau ke pihak ketiga yang memiliki keahlian teknis dan setiap transfer ini harus didokumentasikan dengan ketat untuk mempertahankan chain of custody yang valid. Dalam praktik, sering kali ada tekanan waktu yang membuat dokumentasi ini tidak dilakukan dengan sempurna, menciptakan kerentanan yang bisa dieksploitasi di kemudian hari.

12.4 SAKSI AHLI FORENSIK DIGITAL: PERAN, KUALIFIKASI, DAN TANTANGAN DI PENGADILAN

Peran Saksi Ahli dalam Sistem Hukum Indonesia

Dalam sistem hukum Indonesia, saksi ahli (*expert witness*) berperan untuk memberikan keterangan tentang hal-hal yang memerlukan pengetahuan teknis khusus yang tidak dimiliki oleh hakim atau jaksa secara umum. Keterangan ahli masuk dalam kategori alat bukti yang sah berdasarkan Pasal 184 KUHAP, dan hakim tidak terikat untuk menerima atau menolak keterangan ahli, saksi ahli memiliki kebebasan untuk menilai kredibilitas dan relevansinya.

Dalam praktik, peran saksi ahli forensik digital di Indonesia masih dalam tahap pembentukan. Untuk kasus-kasus yang melibatkan kejahatan siber atau bukti digital, penyidik Polri (terutama Direktorat Tindak Pidana Siber Bareskrim) memiliki personel dengan keahlian teknis tertentu. Tapi keahlian dalam kejahatan siber umum belum tentu sama dengan keahlian dalam analisis deepfake spesifik ini adalah subdomain yang sangat spesifik dengan metodologi yang terus berkembang.

Satu masalah praktis yang sering muncul adalah bahwa penyidik yang melakukan analisis teknis dan yang kemudian bersaksi sebagai ahli di pengadilan adalah orang yang sama atau berasal dari lembaga yang sama. Ini menciptakan potensi bias yang, dalam sistem common law, biasanya diatasi dengan memisahkan peran penyidik dan expert witness, atau dengan mengizinkan pihak lawan menghadirkan ahli mereka sendiri untuk menantang keterangan ahli pemerintah.

Kualifikasi yang Diperlukan untuk Ahli Forensik Deepfake

Apa yang membuat seseorang layak disebut ahli dalam analisis forensik deepfake? Ini adalah pertanyaan yang tidak memiliki jawaban yang sepenuhnya standar, karena bidang ini relatif baru dan sertifikasi khusus untuk forensik deepfake belum berkembang sepenuhnya. Tapi beberapa kriteria kualifikasi yang wajar bisa diidentifikasi.

Dari sisi pengetahuan teknis, seorang ahli yang kredibel harus memiliki pemahaman mendalam tentang cara kerja model generatif AI (terutama GAN dan diffusion models), metodologi deteksi deepfake yang berbeda, keterbatasan masing-masing metodologi, dan cara menginterpretasikan hasil analisis dalam konteks yang tepat. Ia juga harus familiar dengan standar forensik digital internasional setidaknya ACPO Good Practice Guide for Digital Evidence (Inggris) atau SWGDE (*Scientific Working Group for Digital Evidence*) guidelines dari AS karena standar-standar ini adalah tolok ukur yang paling umum digunakan secara internasional.

Dari sisi rekam jejak, ahli yang kredibel idealnya memiliki publikasi dalam jurnal peer-reviewed tentang topik yang relevan, atau setidaknya track record dalam kasus-kasus forensik digital yang sudah diputus. Sertifikasi dari lembaga yang diakui seperti EnCE (*EnCase Certified Examiner*), GCFE (*GIAC Certified Forensic Examiner*), atau yang semakin relevan, sertifikasi dari lembaga-lembaga yang secara khusus fokus pada AI forensics menambah kredibilitas tapi bukan pengganti keahlian substantif.

Yang sering diabaikan adalah kemampuan komunikasi. Seorang ahli yang memahami teknisnya dengan sempurna tapi tidak bisa menjelaskannya kepada hakim awam adalah ahli yang tidak efektif di pengadilan. Ini bukan hanya soal simplifikasi simplifikasi yang berlebihan juga berbahaya karena bisa menyesatkan. Yang diperlukan adalah kemampuan untuk menjelaskan kompleksitas teknis dengan tepat tanpa mengasumsikan pengetahuan teknis dari pendengar.

Dilema Certainty: Apa yang Boleh dan Tidak Boleh Diklaim Ahli

Salah satu dilema paling serius yang dihadapi saksi ahli forensik deepfake adalah pertanyaan tentang seberapa pasti klaim yang bisa dibuat. Sistem hukum, terutama dalam konteks pidana, beroperasi dengan ekspektasi tertentu tentang kepastian: apakah video ini deepfake atau bukan? Jawaban "mungkin deepfake dengan probabilitas 78%" tidak mudah diintegrasikan ke dalam struktur pembuktian yang biasa.

Tapi itulah kenyataan teknis yang jujur. Hampir tidak ada teknik deteksi deepfake yang memberikan kepastian absolut dan memberikan probabilitas, atau lebih tepatnya, sekumpulan indikator yang konsisten atau tidak konsisten dengan hipotesis bahwa konten adalah deepfake. Ahli yang mengklaim kepastian mutlak harus dilihat dengan kecurigaan: ia mungkin tidak memahami keterbatasan metodologinya sendiri, atau ia terlalu terpengaruh oleh tekanan dari pihak yang menghadirkannya.

Ahli yang baik akan mengkomunikasikan hal ini dengan jelas di pengadilan: *"Analisis saya menemukan indikator-indikator berikut yang konsisten dengan konten yang dihasilkan oleh AI; tidak ada indikator yang saya temukan yang menyangkal hipotesis ini; tapi saya tidak dapat mengecualikan kemungkinan bahwa konten ini adalah rekaman asli yang memiliki karakteristik yang tidak biasa."* Formulasi seperti ini secara teknis jujur tapi secara hukum tidak memuaskan dan ketegangan antara kejujuran teknis dan utilitas hukum ini adalah salah satu tantangan paling sulit dalam kesaksian ahli forensik deepfake.

Adversarial Expert Testimony: Ketika Ahli Berbeda Pendapat

Dalam sistem common law, adalah hal yang lumrah bahkan diharapkan bahwa penuntut dan terdakwa masing-masing menghadirkan ahli mereka sendiri, dan kedua ahli memberikan kesaksian yang mungkin bertentangan. Hakim atau juri kemudian memutuskan mana yang lebih meyakinkan. Ini adalah mekanisme yang dirancang untuk mengekspos kelemahan dalam metodologi masing-masing pihak melalui cross-examination.

Sistem hukum Indonesia tidak beroperasi persis dengan cara ini KUHAP tidak mengatur adversarial expert testimony dalam format yang sama seperti common law. Tapi dalam praktik, sudah terjadi kasus di mana pihak terdakwa menghadirkan ahli mereka sendiri untuk menantang keterangan ahli pemerintah, dan hakim harus menilai konflik ini. Tanpa kerangka yang jelas untuk mengevaluasi konflik antar ahli dalam konteks teknis yang kompleks seperti forensik deepfake, hakim sering tergantung pada faktor-faktor yang kurang relevan kelancaran komunikasi, tingkat akademis, institusi asal daripada pada kualitas metodologis keterangan yang diberikan.

12.5 LABORATORIUM FORENSIK INDONESIA: KAPASITAS, KESENJANGAN, DAN JALAN MAJU

Kapasitas yang Ada: Sebuah Peta Realistis

Cara menilai kapasitas forensik digital Indonesia dalam konteks deepfake secara jujur, kita perlu memulai dari apa yang memang ada. Indonesia memiliki beberapa lembaga yang memiliki kemampuan forensik digital: Laboratorium Forensik (Labfor) Polri yang tersebar di berbagai wilayah, Direktorat Tindak Pidana Siber (*Dittipidhsiber*) Bareskrim Polri yang memiliki unit teknis khusus, dan Pusat Forensik Digital yang beroperasi di beberapa universitas teknik dan institusi pendidikan teknologi.

Dalam domain forensik digital umum analisis file, pemulihan data terhapus, analisis lalu lintas jaringan, identifikasi malware kapasitas Indonesia sudah berkembang cukup signifikan dalam satu dekade terakhir. Ada personel yang tersertifikasi secara internasional, ada perangkat keras dan perangkat lunak forensik yang diakui secara global (seperti EnCase dan FTK), dan ada pengalaman yang terakumulasi dari penanganan kasus-kasus kejahatan siber.

Tapi analisis deepfake spesifik adalah subdomain yang berbeda secara kualitatif. Ia memerlukan keahlian dalam machine learning dan computer vision yang melampaui forensik digital konvensional, akses ke dataset deepfake yang terus diperbarui untuk melatih dan memvalidasi model deteksi, dan kapasitas komputasi yang signifikan untuk menjalankan analisis mendalam terhadap video. Di area-area ini, kesenjangan dengan standar internasional masih cukup besar.

Kesenjangan Spesifik dengan Standar Global

Standar global untuk laboratorium forensik digital diatur oleh berbagai kerangka internasional, yang paling relevan adalah ISO/IEC 17025 (persyaratan umum untuk kompetensi laboratorium pengujian), standar SWGDE, dan untuk konteks Eropa ENFSI (*European Network of Forensic Science Institutes*) guidelines. Labfor Polri sudah memiliki akreditasi ISO 17025 untuk

beberapa jenis pemeriksaan, yang adalah pencapaian yang tidak kecil. Tapi akreditasi untuk analisis deepfake spesifik belum ada karena standarnya sendiri belum cukup mapan secara internasional.

Kesenjangan pertama adalah pada perangkat dan metodologi. Alat-alat deteksi deepfake yang paling canggih saat ini banyak dikembangkan oleh lembaga penelitian akademis (MIT, Stanford, beberapa universitas Eropa) dan perusahaan teknologi besar dan tidak semuanya tersedia secara komersial atau dengan harga yang terjangkau untuk lembaga pemerintah. Laboratorium forensik di negara-negara yang lebih maju secara teknis memiliki akses ke alat dan metodologi terbaru yang belum tentu tersedia bagi Labfor Polri.

Kesenjangan kedua adalah pada sumber daya manusia. Analisis forensik deepfake memerlukan individu dengan latar belakang yang tidak umum: kombinasi antara keahlian forensik digital, pemahaman mendalam tentang AI generatif, dan kemampuan untuk mengartikulasikan temuan dalam konteks hukum. Profil ini sangat langka bahkan di negara-negara yang sudah lebih maju dalam bidang ini. Di Indonesia, jumlah individu yang memenuhi kualifikasi ini bisa dihitung dengan jari, dan sebagian besar berada di sektor akademis atau swasta, bukan di lembaga penegak hukum.

Kesenjangan ketiga yang mungkin paling sulit diatasi dalam jangka pendek adalah pada akses ke intelligence tentang model-model deepfake terbaru. Untuk mendeteksi deepfake yang dibuat dengan model tertentu, analis forensik idealnya memiliki pengetahuan tentang karakteristik artefak yang khas dari model tersebut. Pengetahuan ini berkembang sangat cepat dan memerlukan keterlibatan aktif dengan komunitas riset global melalui konferensi, publikasi, dan jaringan profesional yang memerlukan investasi waktu dan sumber daya yang signifikan.

Perbandingan dengan Standar Internasional: Beberapa Referensi Kasus

Untuk memberikan gambaran yang lebih konkret, ada baiknya merujuk pada beberapa standar dan praktik internasional yang bisa menjadi referensi. FBI's *Regional Computer Forensics Laboratory* (RCFL) network di Amerika Serikat memiliki protokol khusus untuk analisis konten yang diduga dihasilkan AI, termasuk deepfake, yang diperbarui secara berkala seiring perkembangan teknologi. *Europol's European Cybercrime Centre* (EC3) memiliki unit Forensic Technology yang secara aktif mengembangkan metodologi untuk mendeteksi konten manipulasi AI.

Di Asia, INTERPOL melalui Digital Crime Centre (IGCI) di Singapura menyediakan dukungan teknis kepada anggotanya termasuk Indonesia untuk kasus-kasus yang melibatkan bukti digital kompleks. Kerja sama ini adalah jalur yang sudah ada dan yang belum dimanfaatkan secara optimal oleh Indonesia setidaknya tidak untuk kasus-kasus yang melibatkan deepfake secara spesifik.

Satu model yang menarik untuk dipertimbangkan adalah model konsorsium laboratorium forensik regional: beberapa negara berbagi sumber daya, keahlian, dan infrastruktur untuk membangun kapasitas forensik digital yang tidak ekonomis untuk dibangun sendiri-sendiri.

ASEAN, meski belum memiliki mekanisme semacam ini dalam forensik digital, memiliki preseden kerja sama dalam domain lain yang bisa menjadi model.

Rekomendasi untuk Penguatan Kapasitas

Berdasarkan analisis kesenjangan di atas, setidaknya ada empat area prioritas yang perlu menjadi perhatian dalam pengembangan kapasitas forensik deepfake Indonesia.

Pertama, pengembangan sumber daya manusia yang terstruktur. Ini bukan sekadar mengirim personel ke pelatihan teknis sesekali, hal ini memerlukan program pengembangan jangka panjang yang membangun keahlian yang dalam dan yang menjaga personel tetap terkini dengan perkembangan teknologi yang bergerak cepat. Kemitraan antara Polri dengan universitas teknik terkemuka Indonesia dan dengan lembaga internasional adalah mekanisme yang layak dieksplorasi.

Kedua, investasi dalam infrastruktur komputasi yang memadai. Analisis deepfake yang komprehensif memerlukan GPU yang kuat dan kapasitas komputasi yang jauh melampaui apa yang biasanya tersedia di laboratorium forensik konvensional. Ini bisa diatasi melalui cloud computing, tapi penggunaan cloud untuk data forensik sensitif menghadirkan pertanyaan keamanan data yang harus dijawab terlebih dahulu.

Ketiga, membangun dan merawat jaringan kemitraan internasional yang aktif bukan hanya hubungan formal, tapi pertukaran teknis yang nyata dengan laboratorium forensik di negara-negara yang sudah lebih maju. INTERPOL, Europol, dan jaringan akademis internasional adalah titik masuk yang sudah ada.

Keempat, pengembangan standar prosedur operasional (SOP) yang spesifik untuk analisis deepfake SOP yang tidak hanya mengadaptasi prosedur forensik digital umum tapi yang dikembangkan dengan mempertimbangkan tantangan unik konten deepfake: variabilitas artefak antar model generatif, keterbatasan metodologi deteksi yang ada, dan cara mengkomunikasikan ketidakpastian dalam laporan forensik yang akan digunakan di pengadilan.

12.6 SIMULASI KASUS: BAGAIMANA PEMBUKTIAN DEEPFAKE BEKERJA DALAM PRAKTIK

Skenario A: Deepfake Seksual sebagai Bukti dalam Perkara Pidana

Misalkan seorang terdakwa didakwa menyebarkan konten intim palsu yang menampilkan korban sebuah video yang menurut korban adalah deepfake dari fotonya yang diambil dari media sosial. Terdakwa mengklaim video itu asli. Bagaimana proses pembuktian semestinya berjalan?

Langkah pertama adalah akuisisi forensik yang benar dari video yang menjadi alat bukti dari semua sumbernya dari perangkat terdakwa (jika disita), dari platform tempat video disebarkan (memerlukan koordinasi dengan platform atau perintah pengadilan), dan dari perangkat korban atau saksi yang menerima video tersebut. Setiap sumber harus ditangani sebagai sumber yang independen dengan chain of custody yang terpisah.

Analisis forensik kemudian dilakukan terhadap salinan forensik dari masing-masing sumber, meliputi analisis metadata dari setiap salinan (apakah konsisten satu sama lain? apakah

ada tanda-tanda manipulasi metadata?), analisis artefak visual (apakah ada inkonsistensi di area wajah, tepi, atau dalam analisis temporal frame per frame?), analisis level sinyal (apakah pola noise sensor konsisten dengan kamera yang diklaim?), dan jika ada, verifikasi terhadap foto-foto sumber yang diduga digunakan untuk membuat deepfake (perbandingan geometri wajah, konsistensi pencahayaan, dll.).

Hasil analisis kemudian dituangkan dalam laporan forensik yang menggambarkan temuan secara rinci, metodologi yang digunakan, keterbatasan metodologi tersebut, dan interpretasi temuan. Laporan ini harus cukup rinci untuk memungkinkan ahli independen mereproduksi atau memverifikasi analisisnya ini adalah standar reproducibility yang fundamental dalam forensik ilmiah.

Di persidangan, ahli forensik bersaksi tentang isi laporan dan menjawab pertanyaan dari penuntut dan pembela. Tantangan utama: bagaimana menjelaskan kepada majelis hakim yang mungkin tidak memiliki latar belakang teknis bahwa temuan-temuan teknis yang kompleks mendukung atau menolak klaim bahwa video adalah deepfake tanpa mengklaim kepastian yang melebihi apa yang bisa secara jujur didukung oleh metodologi yang digunakan.

Skenario B: Terdakwa Mengklaim Bukti Video adalah Deepfake

Skenario yang berlawanan juga penting untuk dipertimbangkan terdakwa yang menghadapi bukti video yang merekam tindak pidananya mengklaim bahwa video tersebut adalah deepfake fabrikasi yang sengaja dibuat untuk menjebaknya. Ini adalah skenario yang, seiring kemajuan teknologi, akan semakin sering digunakan sebagai strategi pembelaan.

Dari perspektif pembuktian, klaim ini harus ditangani dengan serius tapi juga tidak boleh secara otomatis melemahkan pembuktian hanya karena terdakwa mengklaimnya. Penuntut dalam situasi ini harus mampu menunjukkan bahwa video yang dijadikan bukti memenuhi standar autentisitas yang memadai: bahwa chain of custody-nya terjaga, bahwa tidak ada tanda-tanda manipulasi yang terdeteksi melalui analisis forensik, dan idealnya, bahwa ada corroborating evidence lain yang konsisten dengan isi video.

Dalam kasus ini yang menjadi masalah adalah bahwa absennya tanda-tanda manipulasi yang terdeteksi tidak sama dengan bukti keaslian deepfake yang dibuat dengan teknologi terbaik mungkin tidak meninggalkan artefak yang bisa dideteksi oleh metodologi yang tersedia saat ini. Dalam sistem yang menerapkan standar beyond reasonable doubt, klaim bahwa video mungkin adalah deepfake bahkan tanpa bukti positif yang mendukungnya bisa menimbulkan keraguan yang cukup untuk mempengaruhi putusan. Ini adalah kerentanan sistemik yang belum ada solusi teknis atau hukum yang memuaskan.

12.7 ANTARA TEKNOLOGI YANG BERLARI DAN HUKUM YANG TERTATIH

Bab ini telah menelusuri empat dimensi pembuktian deepfake hukum acara, teknis, prosedural, dan kelembagaan dan di setiap dimensi, gambaran yang muncul adalah gambaran

sistem yang sedang beradaptasi, dengan kecepatan yang belum cukup untuk mengikuti laju perkembangan teknologi yang harus dihadapinya.

Ini bukan alasan untuk pesimisme, tapi untuk urgensi. Kasus-kasus deepfake sudah masuk atau akan segera masuk ke pengadilan Indonesia. Hakim sudah sekarang harus membuat keputusan tentang bukti digital yang kompleks, sering tanpa kerangka yang memadai untuk mengevaluasinya. Ahli forensik sudah sekarang bersaksi tentang analisis yang metodologinya mungkin tidak dipahami sepenuhnya oleh semua pihak di ruang sidang. Dan sistem keseluruhan ekosistem pembuktian yang mencakup hukum, teknologi, prosedur, dan lembaga sudah beroperasi dalam kondisi yang tidak optimal.

Reformasi yang diperlukan tidak bisa datang hanya dari satu arah. Dibutuhkan revisi atau penyempurnaan kerangka hukum acara yang mengakomodasi secara eksplisit bukti berbasis AI. Dibutuhkan investasi serius dalam kapasitas laboratorium forensik. Dibutuhkan pengembangan standar kualifikasi ahli yang lebih jelas. Dan dibutuhkan mungkin yang paling mendesak program pendidikan berkelanjutan bagi para hakim, jaksa, dan pembela tentang karakteristik unik bukti digital dan deepfake.

Bab-bab berikutnya akan mengeksplorasi bagaimana komponen-komponen ini bisa menjadi bagian dari respons kebijakan yang lebih koheren terhadap tantangan deepfake di Indonesia. Tapi pertama-tama, perlu diapresiasi bahwa pembuktian adalah titik di mana semua kebijakan regulasi akhirnya diuji: undang-undang yang baik tidak ada gunanya jika tidak bisa ditegakkan di pengadilan, dan penegakan tidak bisa efektif tanpa sistem pembuktian yang dapat diandalkan.

BAB 13
TANGGUNG JAWAB PLATFORM DIGITAL DAN EKOSISTEM AI

"Platform bukan sekadar pipa yang mengalirkan informasi. Ia adalah arsitektur yang menentukan apa yang mengalir, seberapa cepat, dan ke manadan arsitektur itu adalah pilihan, bukan takdir."

parafrase dari diskursus tata kelola platform kontemporer

Ketika sebuah deepfake seksual beredar di TikTok dan menghancurkan kehidupan seseorang, siapa yang bertanggung jawab? Pertanyaan ini tampaknya mudah dijawab secara moral pembuat deepfake, tentu saja tapi jauh lebih rumit secara hukum. Si pembuat mungkin anonim, berdomisili di yurisdiksi lain, atau tidak memiliki aset yang bisa dimintai pertanggungjawaban. Sementara TikTok platform yang mengizinkan konten itu diunggah, yang algoritmanya mungkin bahkan mempromosikannya, dan yang memiliki sumber daya yang sangat besar untuk mendeteksi dan mencegahnya berlindung di balik doktrin hukum yang memberikannya kekebalan yang luas atas konten yang diunggah penggunanya.

Ketegangan antara kekebalan platform yang luas dan tanggung jawab yang dirasakan secara moral adalah salah satu debat hukum paling penting di era digital dan deepfake mempertajamnya hingga ke titik yang sulit diabaikan. Bab ini menelusuri debat itu dari beberapa sudut: doktrin safe harbor yang menjadi fondasi kekebalan platform di AS, kewajiban baru yang diimposkan oleh TAKE IT DOWN Act dan DSA Eropa, pertanyaan tentang tanggung jawab penyedia model AI generatif yang menyediakan infrastruktur pembuatan deepfake, dan yang paling relevan untuk pembaca buku ini posisi Indonesia melalui PP 71/2019 dan peraturan Menkominfo yang berlaku.

Satu peringatan awal, ini adalah area hukum yang bergerak sangat cepat. Regulasi baru dikeluarkan, putusan pengadilan membentuk interpretasi baru, dan platform merespons dengan perubahan kebijakan yang terkadang mendahului regulasi formal. Apa yang disajikan dalam bab ini adalah potret kondisi per pertengahan 2025 yang membutuhkan pembaruan berkala seiring perkembangan lapangan.

13.1 FONDASI HISTORIS: MENGAPA PLATFORM MENDAPATKAN KEKEBALAN YANG LUAS

Lahirnya Safe Harbor: Logika Awal

Memahami debat saat ini tentang tanggung jawab platform, perlu memulai dari kondisi yang ada sebelum internet komersial berkembang. Dalam hukum media tradisional, penerbit (*publisher*) bertanggung jawab atas konten yang mereka terbitkan, seorang editor surat kabar bisa dituntut atas artikel yang ia pilih untuk diterbitkan. Sebaliknya, distributor pasif seperti toko buku atau perpustakaan umumnya tidak bertanggung jawab atas konten buku yang mereka distribusikan, karena mereka tidak memilih atau mengedit konten tersebut.

Ketika internet komersial mulai berkembang pesat di pertengahan 1990an, pembuat kebijakan di Amerika Serikat menghadapi pertanyaan mendasar: kategori mana yang berlaku untuk layanan internet baru seperti forum diskusi, email, dan kemudian situs web? Jika platform internet diperlakukan sebagai penerbit dan bertanggung jawab atas semua konten yang ada di platform mereka, konsekuensinya akan sangat besar dan berpotensi menghambat perkembangan internet: tidak ada platform yang bisa bertahan secara finansial jika harus menanggung tanggung jawab untuk miliaran konten yang diunggah penggunanya.

Section 230 Communications Decency Act (1996) adalah respons legislatif Amerika terhadap dilema ini. Pasal yang hanya terdiri dari beberapa kalimat ini menetapkan prinsip yang menjadi fondasi seluruh industri platform digital: *"No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider."* Dalam praktiknya, ini berarti platform seperti Facebook, YouTube, atau TikTok tidak bisa dituntut atas konten yang diunggah penggunanya, kekebalan ini hampir mutlak dan tidak bergantung pada apakah platform mengetahui konten tersebut atau tidak.

Logika di balik Section 230 adalah logika insentif: jika platform tidak mendapatkan kekebalan, mereka tidak akan pernah memoderasi konten karena tindakan moderasi bisa dianggap sebagai bukti bahwa mereka memiliki kontrol editorial dan karena itu harus menanggung tanggung jawab penerbit. Dengan memberikan kekebalan yang terpisah dari keputusan moderasi, Section 230 seharusnya mendorong platform untuk memoderasi konten tanpa takut kehilangan perlindungan hukum.

Kesuksesan dan Kegagalan Logika Safe Harbor

Section 230 berhasil dalam beberapa hal yang tidak bisa diremehkan. Ia memungkinkan tumbuhnya ekosistem platform internet yang menjadi tulang punggung ekonomi digital globaltanpa kekebalan yang diberikannya, sulit membayangkan Facebook, YouTube, atau Twitter bisa berkembang ke skala yang ada sekarang. Ia juga memberikan ruang bagi moderasi konten yang imperfect platform bisa menghapus konten yang mereka anggap melanggar kebijakan tanpa harus khawatir tindakan itu membuat mereka menjadi "penerbit" yang bertanggung jawab.

Tapi dalam beberapa dekade sejak 1996, dunia berubah secara fundamental. Platform yang ada pada 1996 adalah forum diskusi sederhana dengan moderasi minimal. Platform pada 2025 adalah entitas dengan miliaran pengguna, algoritma rekomendasi yang secara aktif mempromosikan konten, sistem monetisasi yang memberikan insentif finansial untuk konten yang memancing reaksi kuat (termasuk konten berbahaya), dan kapasitas teknis yang luar biasa untuk mendeteksi dan merespons konten yang sering kali tidak digunakan secara konsisten.

Kekebalan yang diberikan Section 230 tidak ikut berkembang bersama perubahan ini. Platform yang secara aktif mengkurasi, merekomendasikan, dan memonetisasi konten masih mendapatkan kekebalan yang sama seperti forum diskusi pasif yang ada pada 1996. Dan dalam konteks deepfake, ini berarti platform yang algoritmanya mungkin merekomendasikan deepfake

seksual kepada ribuan pengguna dalam hitungan jam masih terlindungi dari tanggung jawab atas kerusakan yang diakibatkan oleh rekomendasinya itu.

Safe Harbor di Luar AS: Prinsip yang Diadopsi Secara Luas

Prinsip safe harbor tidak unik untuk Amerika Serikat. Directiva tentang eCommerce Uni Eropa (2000/31/EC) memuat ketentuan serupa yang memberikan kekebalan kepada penyedia hosting dari tanggung jawab atas konten yang diunggah pengguna, dengan syarat bahwa platform tidak memiliki pengetahuan aktual tentang konten ilegal dan bertindak cepat untuk menghapus konten begitu mereka mengetahuinya.

Di Indonesia, prinsip serupa tercermin dalam UU ITE dan peraturan pelaksanaannya, platform elektronik yang bersifat pasif tidak menanggung tanggung jawab atas konten pengguna, selama mereka bertindak atas aduan yang masuk. Ini adalah kerangka yang secara fundamentalnya sama dengan safe harbor AS dan Eropa, meski dengan nuansa yang berbeda dalam detail implementasinya.

Yang penting untuk dipahami adalah bahwa safe harbor bukan hak yang diberikan kepada platform tanpa syaratia adalah keseimbangan yang dirancang untuk mengalokasikan insentif dengan cara tertentu. Ketika teknologi dan ekosistem platform berubah sedemikian rupa sehingga keseimbangan itu tidak lagi menghasilkan hasil yang diinginkan ketika platform tidak lagi bertindak sebagai pipa pasif tapi sebagai aktor aktif yang membentuk apa yang beredar argumen untuk mempertahankan safe harbor dalam bentuknya yang asli menjadi semakin lemah.

13.2 TAKE IT DOWN ACT: KEWAJIBAN PLATFORM BARU DI AMERIKA SERIKAT

Mekanisme Kewajiban 48 Jam

TAKE IT DOWN Act (2025) memperkenalkan kewajiban baru yang mengubah secara signifikan cara platform harus merespons laporan tentang deepfake intim nonkonsensual. Undang-undang ini mewajibkan platform untuk menghapus konten deepfake seksual dalam waktu 48 jam setelah menerima laporan yang memenuhi syarat dan ini berlaku tidak hanya untuk konten yang dilaporkan secara spesifik, tapi juga untuk semua salinan dan versi yang diketahui dari konten tersebut di platform yang sama.

Kewajiban 48 jam ini adalah terobosan yang signifikan. Sebelumnya, korban deepfake seksual yang melapor ke platform sering menghadapi proses yang lambat, tidak transparan, dan tidak konsisten: konten mungkin dihapus setelah berhari-hari atau bahkan berminggu-minggu, jika dihapus sama sekali. Dalam periode itu, konten sudah bisa tersebar ke ribuan atau jutaan pengguna, disimpan ulang, dan disebarkan kembali membuat penghapusan akhir yang berhasil menjadi tidak berarti dalam hal pembatasan kerusakan.

Platform yang gagal memenuhi kewajiban 48 jam menghadapi konsekuensi yang nyata: mereka tidak lagi dapat mengandalkan safe harbor Section 230 untuk tuntutan yang terkait dengan konten tersebut. Ini adalah struktur insentif yang dirancang dengan cerdas Section 230

yang biasanya memberikan kekebalan penuh menjadi tersyaratkan pada kepatuhan terhadap kewajiban penghapusan.

Tantangan Implementasi: Apa yang Mudah dan Apa yang Tidak

Kewajiban 48 jam terdengar sederhana, tapi implementasinya jauh lebih kompleks dari yang tampak. Platform seperti TikTok, Instagram, atau X menerima jutaan laporan konten setiap haridari laporan spam yang sepele hingga laporan konten kekerasan yang serius. Memprioritaskan laporan deepfake intim dalam batas 48 jam memerlukan sistem triase yang efektif, yang berarti baik otomatisasi yang canggih maupun tim moderasi manusia yang memadai.

Identifikasi "semua salinan dan versi" adalah tantangan teknis yang bahkan lebih besar. Platform besar memiliki teknologi pencocokan konten YouTube memiliki Content ID, Meta memiliki sistem hashing fotoyang bisa mendeteksi ulangan dari konten yang sudah diidentifikasi. Tapi teknologi ini tidak sempurna, terutama untuk konten yang sudah diedit sedikit (dipotong, diputar, diubah kontrasnya) untuk menghindari deteksi. Dan untuk platform yang lebih kecil yang tidak memiliki infrastruktur teknis yang setara, tantangannya jauh lebih besar.

Ada juga pertanyaan tentang siapa yang berhak melaporkan. TAKE IT DOWN Act mengatur bahwa laporan bisa diajukan oleh korban sendiri atau oleh orang yang bertindak atas namanya. Tapi proses verifikasi identitas pelapor dan verifikasi bahwa konten yang dilaporkan benar-benar deepfake (dan bukan, misalnya, rekaman asli yang seseorang ingin hapus karena alasan lain) adalah langkah yang memerlukan pertimbangan yang cermat. Proses verifikasi yang terlalu mudah membuka potensi penyalahgunaan; proses yang terlalu ketat menghalangi korban yang sesungguhnya.

Jangkauan Ekstrateritorial: Apakah Platform Asing Terikat

Pertanyaan tentang jangkauan ekstrateritorial TAKE IT DOWN Act sangat relevan dalam konteks global. Undang-undang ini berlaku untuk platform yang beroperasi di Amerika Serikat tapi bagaimana dengan platform yang berkantor di luar AS tapi memiliki pengguna Amerika? Atau platform yang tidak memiliki kehadiran hukum di AS sama sekali?

Jawabannya bergantung pada apakah platform tersebut beroperasi secara purposeful dalam pasar AS menerima pengguna Amerika, menjalankan iklan yang menargetkan pengguna Amerika, atau memiliki server di AS. Bagi platform besar seperti TikTok (yang berkantor di Cayman Islands tapi memiliki operasi besar di AS), kewajiban TAKE IT DOWN Act tampaknya berlaku. Bagi platform yang lebih kecil yang beroperasi di tempat lain dan tidak secara aktif melayani pasar AS, situasinya lebih abu-abu.

Yang lebih signifikan secara praktis adalah efek tidak langsung: platform besar yang terikat TAKE IT DOWN Act sering menerapkan kebijakan secara global, bukan hanya untuk pengguna Amerika. Jika TikTok mengembangkan sistem penghapusan deepfake intim dalam 48 jam untuk memenuhi persyaratan AS, sistem itu kemungkinan besar akan berlaku untuk semua pengguna TikTok di seluruh dunia, termasuk di Indonesia. Ini adalah mekanisme di mana regulasi AS secara efektif membentuk standar global tanpa memerlukan perjanjian internasional.

13.3 DIGITAL SERVICES ACT EROPA: KEWAJIBAN BERLAPIS UNTUK PLATFORM BESAR

Arsitektur DSA: Pendekatan Berbasis Risiko

Digital Services Act (DSA), yang mulai berlaku penuh pada awal 2024, adalah upaya Uni Eropa yang paling ambisius untuk meregulasi platform digital secara komprehensif. Berbeda dari pendekatan AS yang berfokus pada kewajiban spesifik untuk jenis konten tertentu, DSA membangun arsitektur kewajiban yang berlapis berdasarkan ukuran dan risiko platform semakin besar dan semakin berisiko sebuah platform, semakin ketat kewajiban yang diimposkan kepadanya.

Dalam hierarki DSA, kategori tertinggi adalah *Very Large Online Platforms* (VLOP) dan Very *Large Online Search Engines* (VLOSE)platform dengan lebih dari 45 juta pengguna aktif bulanan di EU. Platform-platform ini termasuk Meta, Google, TikTok, X, dan sekitar selusin lainnya menghadapi kewajiban yang jauh lebih ketat dari platform yang lebih kecil. Komisi Eropa memiliki kewenangan langsung untuk mengawasi dan menegakkan kewajiban terhadap VLOP, melewati regulator nasional negara-negara anggota.

Ini adalah struktur yang dirancang untuk mengatasi masalah yang nyata: platform terbesar memiliki dampak terbesar terhadap ekosistem informasi publik dan risiko terbesar dari penyalahgunaan, tapi mereka juga memiliki sumber daya terbesar untuk mematuhi kewajiban yang lebih ketat. Pendekatan berbasis risiko ini berbeda secara fundamental dari pendekatan one size fitsall yang tidak membedakan antara platform dengan jutaan pengguna dan forum diskusi kecil dengan ribuan pengguna.

Kewajiban VLOP yang Relevan untuk Deepfake

Perspektif deepfake, beberapa kewajiban VLOP dalam DSA sangat relevan. Pertama, VLOP wajib melakukan penilaian risiko tahunan yang mengidentifikasi risiko sistemik yang ditimbulkan oleh layanan mereka termasuk risiko dari penyebaran konten ilegal dan konten yang merugikan yang tidak selalu ilegal. Deepfake seksual yang memenuhi definisi konten ilegal berdasarkan hukum nasional negara anggota jelas termasuk dalam cakupan ini.

Kedua, VLOP wajib mengambil langkah-langkah mitigasi risiko yang proporsional dan ini jauh melampaui sekadar merespons laporan pengguna. Langkah-langkah ini bisa mencakup penyesuaian sistem rekomendasi untuk mengurangi amplifikasi konten berbahaya, investasi dalam sistem deteksi proaktif, dan kerja sama dengan peneliti independen yang menyelidiki risiko sistemik platform. Ini adalah kewajiban yang jauh lebih substantif dan mahal dari kewajiban notice and take down konvensional.

Ketiga, VLOP wajib memberikan akses data kepada peneliti terakreditasi sebuah kewajiban yang, jika ditegakkan dengan serius, bisa mengubah lanskap penelitian tentang penyebaran deepfake secara dramatis. Selama ini, penelitian independen tentang bagaimana deepfake menyebar di platform besar sangat dibatasi oleh kurangnya akses ke data platform. DSA secara teoritis membuka pintu untuk penelitian yang jauh lebih komprehensif.

Mekanisme Penegakan dan Denda

DSA memberikan Komisi Eropa kewenangan yang sangat kuat: denda hingga 6% dari pendapatan global tahunan untuk pelanggaran, dan hingga 1% untuk ketidakpatuhan terhadap permintaan informasi. Untuk platform dengan pendapatan global ratusan miliar dolar, 6% adalah angka yang tidak bisa diabaikan ini bukan denda simbolis.

Lebih signifikan lagi adalah kemungkinan penangguhan sementara akses ke pasar EU, sebuah sanksi nuklir yang belum pernah diterapkan tapi yang keberadaannya memberikan tekanan yang sangat nyata. Eropa adalah pasar yang terlalu besar dan terlalu penting bagi setiap platform global untuk diabaikan. Ini adalah leverage yang tidak dimiliki oleh sebagian besar yurisdiksi lain di dunia, dan ia menjelaskan mengapa platform besar secara serius mengembangkan kapasitas kepatuhan DSA.

Penegakan awal DSA sudah menunjukkan bahwa Komisi tidak segan-segan bergerak: investigasi sudah diluncurkan terhadap beberapa VLOP terkait berbagai isu kepatuhan, termasuk yang berkaitan dengan konten berbahaya. Bagaimana DSA akan diterapkan secara spesifik dalam konteks deepfake akan menjadi jelas dalam beberapa tahun ke depan seiring yurisprudensi berkembang.

DSA vs. TAKE IT DOWN Act: Filosofi Regulasi yang Berbeda

Membandingkan DSA dan TAKE IT DOWN Act mengungkap perbedaan filosofi regulasi yang lebih dalam dari sekadar perbedaan ketentuan spesifik. TAKE IT DOWN Act adalah regulasi reaktif dan sempit: ia merespons kategori konten tertentu (deepfake intim) dengan kewajiban tertentu (hapus dalam 48 jam). DSA adalah regulasi sistemik dan luas: ia mencoba mengubah cara platform beroperasi secara fundamental, mewajibkan mereka untuk proaktif mengidentifikasi dan mengurangi risiko, bukan hanya merespons laporan.

Pendekatan AS mencerminkan tradisi hukum yang lebih reluctant untuk mengatur kontenia mendefinisikan kategori sempit yang jelas ilegal dan mewajibkan platform untuk merespons. Pendekatan Eropa mencerminkan visi yang lebih luas tentang tanggung jawab platform sebagai entitas yang memiliki dampak sistemik terhadap ruang publik digital dan karena itu harus menanggung kewajiban sistemik yang proporsional dengan dampak tersebut.

Mana yang lebih efektif dalam jangka panjang adalah pertanyaan empiris yang belum bisa dijawab sepenuhnya. Tapi keduanya secara bersama-sama mencerminkan pergeseran global yang jelas: era di mana platform bisa beroperasi dengan kekebalan hampir mutlak dan kewajiban minimal sudah berakhir, setidaknya di yurisdiksi-yurisdiksi yang memiliki kekuatan untuk memaksakan kewajiban kepada platform global.

13.4 TANGGUNG JAWAB PENYEDIA MODEL AI GENERATIF

Ekosistem yang Lebih Kompleks dari yang Terlihat

Ketika kita berbicara tentang tanggung jawab dalam ekosistem deepfake, fokus pada platform distribusi saja tidak cukup. Ada lapisan lain dalam ekosistem ini yang sering luput dari

diskusi: penyedia model AI generatif yang menyediakan infrastruktur teknis yang memungkinkan pembuatan deepfake. OpenAI, Stability AI, Midjourney, dan puluhan perusahaan lain telah mengembangkan model-model yang, dalam berbagai bentuk dan dengan berbagai tingkat kendali, bisa digunakan untuk menghasilkan konten deepfake.

Ekosistem ini berlapis dan tidak linear. Ada model-model yang disediakan hanya melalui API yang dikontrol ketat (seperti DALLE milik OpenAI), di mana pengguna tidak memiliki akses ke model itu sendiri dan harus melalui antarmuka yang bisa difilter. Ada model open source yang dapat diunduh dan dijalankan secara lokal tanpa kendali apapun dari pengembangnya (seperti beberapa varian Stable Diffusion). Dan ada segala sesuatu di antaranya model yang disediakan melalui platform pihak ketiga, model yang dimodifikasi oleh komunitas dan didistribusikan secara informal, dan model yang dikembangkan khusus untuk pembuatan konten yang eksplisit atau manipulatif.

Pertanyaan tentang tanggung jawab tidak bisa dijawab secara seragam untuk seluruh spektrum ini. Tanggung jawab OpenAI atas deepfake yang dibuat menggunakan APInya yang difilter ketat berbeda secara fundamental dari tanggung jawab (jika ada) pengembang yang merilis model open source tanpa kendali apapun, yang kemudian dimodifikasi oleh komunitas untuk tujuan yang tidak dimaksudkan oleh pengembang aslinya.

Teori-Teori Tanggung Jawab yang Mungkin Berlaku

Negligence (kelalaian) adalah teori tanggung jawab perdata yang paling umum dicoba dalam konteks ini. Argumennya: pengembang AI yang gagal mengimplementasikan safe guard yang memadai terhadap penggunaan berbahaya padahal bahaya itu bisa diprediksi dapat dianggap lalai dan bertanggung jawab atas kerusakan yang diakibatkan. Ini adalah argumen yang menarik secara teori tapi yang menghadapi tantangan pembuktian yang serius: haruskah pengembang meramalkan semua kemungkinan penyalahgunaan? Apakah standar kehati-hatian yang berlaku untuk teknologi yang belum ada preseden industrinya?

Products liability (tanggung jawab produk) adalah teori alternatif yang semakin banyak dibahas. Jika model AI diklasifikasikan sebagai "produk" dalam pengertian hukum, pengembang bisa bertanggung jawab atas cacat desain termasuk kegagalan untuk mengintegrasikan safe guard yang memadai terhadap penggunaan berbahaya yang bisa diprediksi. Tapi apakah model AI adalah "produk" atau "layanan" dalam pengertian hukum products liability adalah pertanyaan yang belum memiliki jawaban yang settled, bahkan di yurisdiksi dengan kerangka products liability yang paling berkembang.

Contributory infringement dan aiding and abetting adalah teori yang lebih spesifik bahwa pengembang yang mengetahui atau seharusnya mengetahui bahwa produk mereka digunakan untuk pelanggaran hak, tapi tidak mengambil langkah yang memadai, bisa ikut bertanggung jawab. Ini analogi dari doktrin yang sudah berkembang dalam konteks hak cipta (di mana platform yang memfasilitasi pembajakan bisa bertanggung jawab), yang sekarang coba diterapkan pada konteks deepfake.

Dalam praktik, tanggung jawab penyedia model AI untuk deepfake belum banyak diuji di pengadilan ini adalah area yang masih berkembang, dan hasilnya akan sangat bergantung pada fakta spesifik kasus dan pada yurisdiksi di mana kasus tersebut dibawa. Yang sudah jelas adalah bahwa tekanan regulasi dan reputasional membuat pengembang model besar mengembangkan kebijakan penggunaan yang semakin ketat dan sistem filter yang semakin canggih meski efektivitas filter ini diperdebatkan dan sistem open source yang di luar kendali pengembang asli terus berkembang.

Kebijakan Penggunaan dan Filter Konten: Apa yang Dilakukan Industri

Tanpa menunggu kejelasan regulasi, perusahaan-perusahaan AI besar sudah mengembangkan berbagai mekanisme untuk membatasi penggunaan berbahaya model mereka. OpenAI memiliki usage policy yang secara eksplisit melarang pembuatan konten seksual yang melibatkan tokoh nyata tanpa persetujuan yang jelas, dan sistem filternya berupaya mendeteksi dan menolak permintaan yang melanggar kebijakan ini. Stability AIyang mengembangkan Stable Diffusion menghadapi tantangan yang lebih kompleks karena banyak versi modelnya bersifat opensource dan sudah tidak dalam kendalinya.

Filter konten berbasis AI yang digunakan oleh platform dan pengembang model memiliki keterbatasan yang tidak bisa diabaikan dalam konteks pembuktian dan kebijakan. False positive menolak konten yang sahdan false negative mengizinkan konten berbahaya adalah dua sisi masalah yang sama, dan tidak ada filter yang sempurna dalam menangani keduanya. Sistem filter yang terlalu agresif memblokir penggunaan sah; sistem yang terlalu permisif gagal mencegah penyalahgunaan. Dan pelaku jahat yang berkomitmen selalu menemukan cara untuk melewati filter yang ada melalui teknik adversarial prompting, menggunakan model yang tidak memiliki filter, atau dengan memodifikasi model open source.

Dari perspektif kebijakan, ketergantungan pada kebijakan penggunaan dan filter konten sebagai mekanisme utama regulasi menghadapi masalah struktural: ia mengandalkan perusahaan untuk menegakkan kebijakan mereka sendiri, dengan insentif yang tidak selalu selaras dengan kepentingan korban. Regulasi yang memberikan kewajiban eksternal dengan sanksi untuk ketidakpatuhan adalah pelengkap yang diperlukan, bukan pengganti dari kebijakan internal perusahaan.

Tanggung Jawab dalam Rantai Nilai AI

Ekosistem AI generatif memiliki rantai nilai yang panjang: ada yang mengembangkan model dasar (*foundation model*), ada yang melakukan fine tuning untuk aplikasi tertentu, ada yang menyediakan infrastruktur komputasi untuk melatih dan menjalankan model, ada yang membangun antarmuka pengguna di atas model yang ada, dan ada pengguna akhir yang menggunakan antarmuka itu. Pertanyaan tentang di mana tanggung jawab seharusnya ditempatkan dalam rantai ini belum memiliki jawaban yang konsisten.

Analogi yang berguna meski tidak sempurna adalah tanggung jawab dalam rantai manufaktur. Produsen bahan baku tidak bertanggung jawab atas cara produk jadi yang

menggunakan bahan mereka disalahgunakan, tapi produsen produk jadi memiliki kewajiban untuk memastikan produknya aman untuk penggunaan yang diprediksi. Dalam konteks AI, di mana batas antara "bahan baku" dan "produk jadi" sangat cair, penetapan tanggung jawab yang tepat memerlukan analisis yang jauh lebih nuansir dari analogi manufaktur bisa memberikan.

EU AI Act, yang mulai berlaku bertahap dari 2024, mencoba menjawab pertanyaan ini dengan mendefinisikan kewajiban yang berbeda untuk berbagai peran dalam rantai nilai: provider (yang mengembangkan sistem AI), deployer (yang menggunakannya dalam konteks tertentu), dan pengguna akhir. Untuk sistem AI dengan risiko tinggidan sistem yang bisa digunakan untuk membuat deepfake dari orang nyata sangat mungkin termasuk dalam kategori ini kewajiban yang diimposkan pada provider dan deployer sangat substansial.

13.5 POSISI INDONESIA: PP 71/2019 DAN PERATURAN MENKOMINFO

Kerangka Regulasi yang Ada: Peta dan Kesenjangan

Peraturan Pemerintah Nomor 71 Tahun 2019 tentang Penyelenggaraan Sistem dan Transaksi Elektronik (PP 71/2019) adalah peraturan pelaksana UU ITE yang paling komprehensif dan yang paling relevan untuk pembahasan tanggung jawab platform. PP ini membagi penyelenggara sistem elektronik (PSE) ke dalam dua kategori PSE lingkup publik (instansi negara) dan PSE lingkup privat (entitas swasta)dan menetapkan kewajiban yang berbeda untuk masing-masing.

Dari perspektif tanggung jawab platform terhadap konten pengguna, PP 71/2019 mewajibkan PSE privat untuk menghapus konten yang dilarang dalam waktu yang ditentukan setelah menerima aduan atau perintah dari pemerintah. Kewajiban ini tampak paralel dengan kewajiban notice and takedown yang ada di AS dan Eropa, tapi ada perbedaan penting: di Indonesia, kewajiban ini lebih bersifat administrative kewajiban kepada pemerintah daripada bersifat sipil yang memberikan remediasi langsung kepada korban.

Untuk deepfake secara spesifik, PP 71/2019 tidak memberikan definisi atau ketentuan yang eksplisit. Deepfake yang disebarkan melalui platform digital harus diklasifikasikan di bawah kategori konten yang ada konten yang melanggar kesusilaan, konten yang mencemarkan nama baik, atau konten yang menyebarkan informasi palsu yang masing-masing memiliki kerangka hukum yang berbeda dan yang tidak selalu dirancang untuk menangani karakteristik unik deepfake.

Kewajiban Pendaftaran PSE Asing

Salah satu inovasi paling signifikan dalam regulasi Indonesia terhadap platform digital adalah kewajiban pendaftaran PSE asing. Berdasarkan Peraturan Menkominfo Nomor 10 Tahun 2021 dan peraturan-peraturan yang mendahuluinya, platform asing yang beroperasi di Indonesia termasuk Meta, Google, TikTok, dan lainnya diwajibkan untuk mendaftarkan diri sebagai PSE dan menunjuk perwakilan yang dapat dihubungi oleh pemerintah Indonesia.

Kewajiban ini memiliki beberapa tujuan: memastikan bahwa platform asing dapat dihubungi dan dimintai pertanggungjawaban oleh otoritas Indonesia, memberikan mekanisme untuk menyampaikan permintaan penghapusan konten, dan secara teoritis memberikan dasar hukum untuk tindakan lebih lanjut jika platform tidak mematuhi permintaan tersebut.

Tapi efektivitas mekanisme ini masih menghadapi tantangan nyata. Permintaan penghapusan konten yang dikirim ke perwakilan lokal platform asing tidak selalu ditindaklanjuti dengan kecepatan yang memadai terutama untuk konten yang tidak secara jelas melanggar Terms of Service global platform tersebut. Dan ketika platform tidak merespons, instrumen penegakan yang tersedia ancaman pemblokiran adalah instrumen yang sangat blunt dan yang dampaknya bisa jauh melampaui target yang dimaksud, seperti yang terjadi dalam beberapa kasus pemblokiran platform yang telah ada.

Peraturan tentang Konten Intim dan Deepfake

Indonesia belum memiliki regulasi yang secara spesifik dan komprehensif mengatur deepfake sebagai kategori tersendiri. Konten deepfake seksual yang tersebar di platform digital saat ini paling sering ditangani melalui beberapa jalur hukum yang ada: Pasal 27 UU ITE tentang konten yang melanggar kesusilaan, ketentuan tentang penyebaran konten intim nonkonsensual yang sudah ada dalam UU ITE (khususnya setelah revisi), dan dalam beberapa kasus, KUHP melalui ketentuan pencemaran nama baik.

Tidak ada satu pun dari jalur-jalur hukum ini yang dirancang dengan mempertimbangkan karakteristik spesifik deepfake bahwa konten yang disebarkan adalah fabrikasi, bukan rekaman asli; bahwa pembuatannya memerlukan infrastruktur teknis tertentu yang melibatkan aktor-aktor di luar si penyebar; dan bahwa skala penyebaran dan kecepatan replikasi digital membuat penghapusan setelah fakta sangat tidak efektif. Kesenjangan ini adalah argumen kuat untuk pengembangan regulasi deepfake yang lebih spesifik di Indonesia.

Beberapa diskusi sudah berlangsung di tingkat kementerian tentang kemungkinan peraturan yang lebih spesifik, tapi per pertengahan 2025, belum ada ketentuan yang dikeluarkan secara resmi. Proses legislasi yang berjalan di Indonesia dengan berbagai kepentingan yang harus diakomodasi dan dengan kapasitas teknis yang terbatas dalam drafting regulasi teknologi yang kompleks berarti bahwa regulasi yang komprehensif mungkin masih membutuhkan waktu.

Perbandingan: Di Mana Indonesia Berdiri dalam Spektrum Global

Jika kita menempatkan Indonesia dalam spektrum regulasi platform global yang sudah kita bahas, posisinya tampak berada di tengah: lebih aktif dari banyak negara berkembang yang sama sekali belum memiliki kerangka regulasi platform yang berarti, tapi masih jauh di belakang AS (TAKE IT DOWN Act), Eropa (DSA), dan negara-negara pionir seperti Korea Selatan dalam hal kewajiban spesifik yang diimposkan kepada platform terkait konten berbahaya termasuk deepfake.

Kesenjangan ini bukan hanya soal ketertinggalan, ia mencerminkan perbedaan dalam kapasitas kelembagaan, prioritas politik, dan tekanan dari kelompok kepentingan yang berbeda.

Platform digital besar memiliki kepentingan yang jelas untuk mempertahankan rezim tanggung jawab yang minimal, dan mereka memiliki sumber daya untuk melobi secara efektif. Kelompok advokasi korban dan masyarakat sipil di Indonesia, meski sudah mulai aktif, belum memiliki kekuatan pendorong yang sebanding.

Tapi ada juga faktor yang bisa mempercepat perubahan: tekanan dari luar. Ketika platform global mengembangkan sistem untuk memenuhi TAKE IT DOWN Act AS atau DSA Eropa, sistem itu sering berlaku secara global termasuk untuk pengguna Indonesia. Dan ketika kasus-kasus deepfake yang mengerikan mendapat perhatian media yang luas di Indonesia, tekanan publik untuk respons legislatif bisa bergerak sangat cepat. Dinamika ini kombinasi antara tekanan internal dari korban dan tekanan eksternal dari standar global adalah kondisi yang paling mungkin mendorong reformasi regulasi yang bermakna.

13.6 MODEL TANGGUNG JAWAB YANG BERKELANJUTAN

Keterbatasan Fundamental Model Notice and takedown

Model notice and takedown di mana tanggung jawab platform diaktifkan hanya ketika mereka menerima laporan tentang konten tertentu adalah model yang sudah berusia lebih dari dua dekade dan yang dirancang untuk ekosistem internet yang sangat berbeda dari yang ada sekarang. Keterbatasannya dalam konteks deepfake sangat terasa.

Pertama, model ini berasumsi bahwa korban tahu bahwa konten tentang mereka ada, dan bahwa mereka memiliki kapasitas untuk melaporkannya. Tapi banyak korban deepfake tidak tahu bahwa konten tentang mereka beredar mereka menemukannya secara kebetulan atau melalui orang lain, sering kali sudah terlambat. Dan bahkan ketika mereka tahu, proses pelaporan di banyak platform masih cukup melelahkan dan tidak ramah korban.

Kedua, notice and takedown bereaksi setelah kerusakan sudah terjadi. Deepfake yang sudah disebarkan selama 24 jam sudah bisa ditonton oleh ratusan ribu orang, diunduh dan disimpan secara lokal, dan disebarkan ulang ke platform lain. Penghapusan dari platform asal tidak mengubah fakta bahwa konten sudah ada di luar sanadan dalam era screenshot dan download yang mudah, "ada di luar sana" praktis berarti permanen.

Ketiga, notice and takedown tidak mendorong pencegahan proaktif. Platform yang tidak diwajibkan untuk proaktif mencegah konten berbahaya memiliki sedikit insentif untuk berinvestasi dalam sistem pencegahan yang mahal terutama ketika konten berbahaya (termasuk deepfake yang sensasional) sering menghasilkan engagement yang tinggi, yang pada gilirannya menghasilkan pendapatan iklan.

Menuju Model Tanggung Jawab yang Lebih Proaktif

Beberapa proposal untuk model tanggung jawab yang lebih proaktif sudah mulai mendapatkan traksi dalam diskusi kebijakan global. Salah satu yang paling menarik adalah model tanggung jawab berbasis desain ide bahwa platform harus menanggung tanggung jawab untuk

kerusakan yang bisa diprediksi akibat pilihan desain yang mereka buat secara sadar, termasuk pilihan sistem rekomendasi yang memperkuat penyebaran konten berbahaya.

Model lain adalah tanggung jawab berbasis pengetahuan yang diperluas bukan hanya pengetahuan aktual tentang konten tertentu (seperti dalam notice and takedown), tapi pengetahuan yang seharusnya dimiliki berdasarkan kapasitas teknis platform. Jika sebuah platform memiliki teknologi yang mampu mendeteksi deepfake tapi memilih untuk tidak menggunakannya, apakah ia bisa berargumen bahwa ia tidak "mengetahui" adanya deepfake di platformnya? Beberapa akademisi hukum berargumen bahwa jawaban seharusnya tidak.

Ada juga model tanggung jawab yang didistribusikan di seluruh rantai nilai AIdi mana pengembang model, platform distribusi, dan operator sistem moderasi masing-masing menanggung porsi tanggung jawab yang proporsional dengan kemampuan mereka untuk mencegah atau mengurangi kerusakan. Ini adalah model yang secara konseptual menarik tapi yang menghadapi tantangan praktis yang besar dalam hal alokasi tanggung jawab yang dapat diverifikasi.

Prinsip-Prinsip untuk Regulasi Platform yang Efektif

Dari analisis komparatif yang sudah dilakukan dalam bab ini, beberapa prinsip untuk regulasi platform yang efektif bisa diidentifikasi sebagai titik orientasi bukan sebagai resep yang siap pakai, melainkan sebagai kriteria untuk mengevaluasi proposal regulasi yang ada atau yang akan datang.

Proporsionalitas dengan kapasitas. Kewajiban yang diimposkan kepada platform harus proporsional dengan kapasitas teknis dan finansial mereka. Platform dengan miliaran pengguna dan tim teknis ribuan orang bisa diharapkan memenuhi kewajiban yang jauh lebih ketat dari platform kecil dengan sumber daya terbatas. DSA Eropa dengan kategorisasi VLOP adalah model yang patut dipertimbangkan.

Berorientasi pada korban. Regulasi yang efektif harus memastikan bahwa mekanisme pelaporannya mudah diakses oleh korban yang sesungguhnya bukan hanya secara teknis tersedia, tapi secara praktis dapat digunakan oleh seseorang yang dalam kondisi tertekan dan mungkin tidak memiliki pengetahuan teknis yang mendalam.

Mendorong pencegahan proaktif. Regulasi yang hanya mewajibkan respons terhadap laporan tidak akan pernah cukup. Insentif untuk investasi dalam pencegahan proaktif baik melalui kewajiban positif maupun melalui struktur tanggung jawab yang menghilangkan safe harbor untuk kegagalan pencegahan yang bisa diprediksi adalah komponen yang tidak bisa diabaikan.

Transparansi dan akuntabilitas. Platform harus diwajibkan untuk melaporkan secara berkala tentang volume dan penanganan konten berbahaya, termasuk deepfake bukan hanya kepada regulator tapi juga kepada publik dan peneliti independen. Tanpa data yang dapat diverifikasi secara independen, tidak mungkin mengevaluasi apakah kewajiban yang ada benar-benar dipenuhi.

13.7 IMPLIKASI UNTUK KEBIJAKAN INDONESIA

Reforma Regulasi Yang Diperlukan

Berdasarkan analisis dalam bab ini, setidaknya ada empat area reforma regulasi yang perlu dipertimbangkan oleh pembuat kebijakan Indonesia dalam konteks tanggung jawab platform terhadap deepfake.

- Pertama, penguatan kewajiban penghapusan konten deepfake dengan tenggat waktu yang jelas. Kewajiban umum untuk menghapus konten yang dilarang yang ada dalam PP 71/2019 perlu diperkuat dengan ketentuan khusus untuk deepfake intim atau deepfake yang menimbulkan kerugian nyata, dengan tenggat waktu yang lebih ketat dan mekanisme pelaporan yang lebih mudah diakses oleh korban.
- Kedua, ketentuan yang mengalihkan sebagian beban pembuktian kepada platform. Dalam sistem saat ini, korban harus membuktikan bahwa konten yang mereka laporkan adalah ilegal sebelum platform diwajibkan untuk bertindak. Untuk deepfake intim, ada argumen kuat untuk membalik presumsi ini: platform harus menghapus konten yang dilaporkan sebagai deepfake intim nonkonsensual kecuali mereka dapat memverifikasi keasliannya.
- Ketiga, persyaratan transparansi yang lebih kuat terhadap PSE asing. Kewajiban pendaftaran yang sudah ada perlu dilengkapi dengan kewajiban pelaporan tentang volume konten berbahaya yang ditangani, termasuk deepfake, dan tentang efektivitas sistem moderasi yang digunakan. Ini memberikan data yang diperlukan untuk evaluasi regulasi yang berbasis bukti.
- Keempat, mekanisme koordinasi yang lebih efektif antara Kominfo, Polri, dan platform untuk penanganan kasus-kasus deepfake yang prioritas terutama yang melibatkan korban yang rentan seperti anak-anak atau yang melibatkan ancaman terhadap keamanan publik.

Posisi Indonesia dalam Negosiasi Global tentang Tata Kelola Platform

Indonesia adalah negara dengan populasi pengguna internet yang sangat besar salah satu pasar terbesar di Asia Tenggara dan karena itu memiliki leverage yang tidak diabaikan dalam negosiasi dengan platform global. Leverage ini belum selalu digunakan secara strategis: ancaman pemblokiran yang sesekali dikeluarkan lebih sering menjadi ancaman yang tidak diikuti atau yang terlalu mahal konsekuensi sosialnya untuk benar-benar diterapkan.

Model yang lebih efektif mungkin adalah koordinasi dengan negara-negara ASEAN lain dan dengan mitra global seperti Eropa untuk membangun standar bersama standar yang platform global tidak bisa abaikan karena berlaku untuk pasar yang cukup besar secara kolektif. Indonesia sudah menunjukkan kemampuan untuk memimpin dalam forum regional; kepemimpinan yang serupa dalam konteks tata kelola platform digital adalah langkah yang logis berikutnya.

Rangkuman Bab

Bab ini telah menelusuri lanskap tanggung jawab platform dan ekosistem AI yang bergerak cepat dari fondasi safe harbor yang mulai retak, melalui kewajiban baru yang diimposkan oleh TAKE IT DOWN Act dan DSA, hingga pertanyaan yang belum terjawab tentang tanggung jawab

penyedia model AI generatif, dan berakhir pada posisi Indonesia yang memiliki potensi untuk bergerak lebih jauh dari yang sudah dilakukan.

Yang menjadi benang merahnya adalah bahwa pertanyaan tentang tanggung jawab platform bukan pertanyaan teknis semata ia adalah pertanyaan tentang nilai dan distribusi kekuasaan. Siapa yang menanggung biaya dari penyebaran deepfake berbahaya? Saat ini, sebagian besar biaya itu ditanggung oleh korban melalui kerusakan psikologis, kerusakan reputasi, dan upaya hukum yang melelahkan dan sering tidak berhasil. Sementara platform dan pengembang model menikmati manfaat dari ekosistem yang mereka ciptakan tanpa menanggung biaya penuh dari bahaya yang menyertainya.

Reformasi regulasi yang bermakna adalah reformasi yang menggeser distribusi biaya ini yang memastikan bahwa pihak yang memiliki kapasitas terbesar untuk mencegah kerusakan menanggung insentif yang cukup kuat untuk benar-benar melakukannya. Tidak ada satu formula yang sempurna untuk mencapai ini, dan perdebatan tentang cara yang tepat akan terus berlangsung. Tapi arahnya sudah semakin jelas: era platform tanpa tanggung jawab sudah berakhir, meski kecepatan dan kedalaman perubahan masih akan ditentukan oleh dinamika politik dan hukum yang terus berkembang di berbagai yurisdiksi termasuk Indonesia.

BAB 14
STUDI PERBANDINGAN: APA YANG BISA INDONESIA PELAJARI

"Hukum yang baik tidak diciptakan dalam kekosongan. Ia selalu merupakan dialog antara pengalaman sendiri dan pengalaman orang laindan kebijaksanaan ada pada mereka yang bisa membedakan mana yang bisa dipinjam dan mana yang harus ditempa sendiri."

parafrase dari tradisi studi hukum komparatif

Studi hukum komparatif bukan latihan akademis yang berdiri sendiri. Pada titik terbaiknya, ia adalah instrumen kebijakan yang memungkinkan suatu negara untuk belajar dari pengalaman baik keberhasilan maupun kegagalan yurisdiksi lain tanpa harus mengulangi proses coba-coba yang mahal dan menyakitkan. Dalam konteks regulasi deepfake, pelajaran dari yurisdiksi-yurisdiksi yang sudah bergerak lebih awal sangat berharga, meski tidak bisa begitu saja dipindahkan tanpa adaptasi yang serius terhadap konteks lokal.

Bab-bab sebelumnya telah memetakan lanskap regulasi deepfake di Amerika Serikat, China, Korea Selatan, Inggris, Prancis, dan konteks Eropa yang lebih luas. Bab ini melakukan hal yang berbeda: alih-alih sekadar mendeskripsikan, ia bertanya secara langsung tentang relevan siapa yang bisa dipelajari Indonesia dari masing-masing model, apa yang perlu disesuaikan, dan apa yang sebaiknya tidak diadopsi sama sekali karena tidak sesuai dengan konteks hukum dan sosial Indonesia.

Pendekatan yang digunakan bukan pendekatan yang mencari satu model terbaik untuk diadopsi secara keseluruhan. Tidak ada sistem regulasi deepfake yang sempurna, dan setiap sistem mengandung trade-off yang mencerminkan pilihan nilai yang berbeda-beda. Yang ingin dicapai adalah peta yang lebih jelas tentang pilihan-pilihan yang tersedia, beserta analisis tentang apa yang dipertaruhkan dalam setiap pilihan sehingga pembuat kebijakan, akademisi, dan masyarakat sipil di Indonesia dapat membuat pilihan yang lebih informed.

14.1 METODOLOGI STUDI KOMPARATIF: BAGAIMANA KITA MEMBANDINGKAN

Perangkap Umum dalam Perbandingan Hukum

Sebelum masuk ke substansi perbandingan, ada baiknya mengidentifikasi beberapa perangkap umum dalam studi hukum komparatif yang relevan untuk konteks ini. Perangkap pertama adalah apa yang oleh para ahli hukum komparatif disebut sebagai legal transplant fallacy asumsi bahwa aturan hukum bisa dipindahkan dari satu sistem ke sistem lain tanpa mempertimbangkan lingkungan kelembagaan, budaya hukum, dan kondisi sosial yang membentuk makna dan fungsinya. Undang-undang yang bekerja dengan baik di Korea Selatan dengan sistem peradilan yang efisien dan tingkat literasi hukum yang tinggi mungkin

menghasilkan hasil yang sangat berbeda jika diterapkan verbatim di Indonesia dengan kondisi kelembagaan yang berbeda.

Perangkap kedua adalah bias seleksi: kita cenderung mempelajari sistem yang sudah berjalan baik dan mengabaikan kegagalan-kegagalan yang tidak banyak dipublikasikan. Laporan tentang regulasi deepfake di berbagai negara cenderung berfokus pada legislasi yang berhasil disahkan dan kasus-kasus yang berhasil dituntut bukan pada undang-undang yang tidak pernah ditegakkan, kasus-kasus yang gagal di pengadilan karena masalah pembuktian, atau dampak tidak diinginkan dari regulasi yang terlalu luas.

Perangkap ketiga adalah ekuivalensi fungsional yang diasumsikan tapi tidak diverifikasi. Dua sistem hukum yang memiliki ketentuan yang tampak serupa mungkin menghasilkan hasil yang sangat berbeda karena perbedaan dalam cara ketentuan itu diinterpretasikan, ditegakkan, dan diintegrasikan ke dalam praktik hukum yang lebih luas. Membandingkan teks undang-undang saja tidak cukup kita perlu membandingkan bagaimana teks itu hidup dalam praktik.

Dengan peringatan-peringatan ini dalam pikiran, analisis komparatif yang mengikuti berusaha untuk tidak sekadar membandingkan teks undang-undang, melainkan membandingkan sistem regulasi secara lebih holistik termasuk kelembagaan yang mengimplementasikannya, konteks sosial yang membentuknya, dan bukti efektivitas yang tersedia.

Kriteria Evaluasi

Untuk membuat perbandingan yang bermakna dan operasional, diperlukan kriteria evaluasi yang jelas. Bab ini menggunakan lima kriteria yang diajukan bukan sebagai standar universal yang tidak bisa diperdebatkan, tapi sebagai kerangka kerja yang bisa dikritik dan disempurnakan.

- ☑ Kejelasan definisi: apakah sistem regulasi mendefinisikan deepfake dengan cukup jelas untuk memberikan kepastian hukum, tapi cukup fleksibel untuk mengakomodasi perkembangan teknologi yang cepat? Definisi yang terlalu sempit akan segera usang; definisi yang terlalu luas menciptakan ketidakpastian yang menghambat aktivitas sah.
- ☑ Efektivitas perlindungan korban: apakah sistem memberikan remediasi yang nyata dan dapat diakses oleh korban? Ini mencakup kecepatan respons, kemudahan akses, dan adequacy dari remediasi yang tersedia bukan hanya keberadaan mekanisme secara formal.
- ☑ Proporsionalitas: apakah beban yang diimposkan pada berbagai pihak (platform, pengembang AI, individu) proporsional dengan kemampuan dan kontribusi mereka terhadap masalah? Dan apakah regulasi menghindari dampak berlebihan pada aktivitas yang sah?
- ☑ Kelayakan implementasi: apakah regulasi bisa diimplementasikan dengan kapasitas kelembagaan yang realistis dimiliki Indonesia sekarang dan dalam jangka menengah? Regulasi yang memerlukan kapasitas yang jauh melampaui apa yang tersedia adalah regulasi yang akan tetap menjadi teks di atas kertas.

- ☑ Adaptabilitas: seberapa baik sistem regulasi dapat beradaptasi dengan perkembangan teknologi yang cepat? Regulasi yang sangat spesifik secara teknis bisa menjadi usang dalam hitungan tahun; regulasi berbasis prinsip lebih tahan lama tapi lebih sulit ditegakkan.

14.2 PELAJARAN PERTAMA: MENDEFINISIKAN DEEPFAKE DALAM UNDANG-UNDANG

Mengapa Definisi Bukan Sekadar Formalitas

Setiap sistem regulasi yang efektif dimulai dari definisi yang dapat diandalkan. Ini bukan kepatuhan terhadap formalisme hukum semata ia adalah prasyarat fungsional. Tanpa definisi yang jelas tentang apa yang diatur, penegak hukum tidak tahu apa yang harus ditindak, pengadilan tidak tahu apa yang harus diputuskan, platform tidak tahu apa yang harus dihapus, dan individu tidak tahu perilaku apa yang berbahaya secara hukum. Ketidakjelasan definisi adalah undangan untuk kesewenang-wenangan di satu sisi, dan peluang untuk lolos dari tanggung jawab di sisi lain.

Tantangan mendefinisikan deepfake dalam hukum adalah bahwa ia harus menangkap esensi dari fenomena yang secara teknis bergerak sangat cepat. Definisi yang terlalu terikat pada teknologi tertentu misalnya, "konten yang dihasilkan oleh *Generative Adversarial Network*" akan segera usang ketika teknik generasi yang baru (diffusion models, flow matching, atau teknik yang bahkan belum ada namanya) menjadi dominan. Tapi definisi yang terlalu abstrak misalnya, "konten yang tidak mencerminkan kenyataan" akan terlalu luas dan mencakup fotografi yang diedit, ilustrasi, dan karya fiksi yang tidak seharusnya diregulasi dengan cara yang sama.

Perbandingan Pendekatan Definisi

Amerika Serikat dalam TAKE IT DOWN Act mendefinisikan deepfake dengan mengkombinasikan dua elemen: (1) konten yang dihasilkan atau secara substansial dimodifikasi oleh teknik berbasis AI atau komputer, dan (2) yang menggambarkan seseorang yang dapat diidentifikasi secara realistis dalam cara yang tidak pernah terjadi dalam kenyataan. Kekuatan definisi ini adalah pada elemen "dapat diidentifikasi secara realistis" ia fokus pada dampak terhadap individu nyata, bukan pada teknik yang digunakan untuk menghasilkan konten. Kelemahannya adalah bahwa "secara substansial dimodifikasi" adalah frasa yang memerlukan interpretasi lebih lanjut.

China menggunakan istilah "*deep synthesis*" (深度合成) yang didefinisikan secara lebih luas untuk mencakup seluruh spektrum teknologi yang menggunakan AI untuk menghasilkan atau memanipulasi teks, gambar, audio, video, atau kombinasinya. Kekuatan pendekatan ini adalah komprehensivitasnya ia tidak hanya mencakup deepfake video tradisional tapi juga kloning suara, generasi teks, dan bentuk manipulasi media lain yang mungkin belum diantisipasi. Kelemahannya adalah cakupan yang sangat luas itu membuat regulasi lebih sulit untuk ditegakkan secara selektif dan proporsional.

Korea Selatan dalam undang-undang deepfake seksualnya mendefinisikan konten yang dilarang dengan mengacu pada konten yang "dibuat menggunakan teknologi informasi dan

komunikasi untuk menampilkan seseorang dengan cara seksual yang tidak sesuai dengan kenyataan." Ini adalah definisi yang berorientasi pada dampak dan berfokus pada kategori spesifik (konten seksual)sebuah pilihan strategis yang memungkinkan legislasi lebih mudah disahkan tapi yang juga berarti banyak jenis deepfake berbahaya tidak tercakup.

Uni Eropa melalui AI Act mendefinisikan "AI-generated or manipulated image, audio or video content" dengan cara yang berbasis pada proseskonten yang "*noticeably resembles*" orang, tempat, atau entitas nyata dan "*falsely gives the impression*" bahwa itu nyata. Elemen "*noticeably resembles*" penting karena ia mengecualikan konten yang jelas-jelas fiktif atau yang gayanya sangat berbeda dari kenyataan.

Rekomendasi untuk Definisi dalam Hukum Indonesia

Dari perbandingan pendekatan-pendekatan di atas, beberapa elemen kunci yang sebaiknya masuk dalam definisi deepfake untuk konteks hukum Indonesia bisa diidentifikasi. Definisi yang baik perlu memuat setidaknya tiga komponen: komponen teknologi, komponen representasi, dan komponen dampak.

Komponen teknologi sebaiknya diformulasikan secara fungsional, bukan teknis: "konten yang dihasilkan atau dimanipulasi secara sintetis menggunakan sistem komputasi, termasuk namun tidak terbatas pada kecerdasan buatan dan pembelajaran mesin." Formulasi ini cukup luas untuk mencakup teknik baru yang belum ada tapi tidak mencakup pengeditan manual konvensional yang sudah memiliki kerangka hukum tersendiri.

Komponen representasi sebaiknya berfokus pada kemampuan konten untuk menampilkan orang nyata yang dapat diidentifikasi: "yang menampilkan atau menyuarakan seseorang yang dapat diidentifikasi secara wajar." Ini mengecualikan konten yang sepenuhnya fiktif dan yang tidak menggunakan representasi orang nyata.

Komponen dampak dan ini adalah yang paling kritis sebaiknya menetapkan bahwa konten yang dikategorikan sebagai deepfake adalah konten yang dapat menimbulkan kesan palsu yang menyesatkan pada penonton yang wajar. Ini mengikuti pendekatan EU yang mengakui bahwa beberapa konten AI yang menampilkan orang nyata seperti satire yang jelas-jelas merupakan satire tidak seharusnya diperlakukan sama dengan deepfake yang bertujuan menipu.

Poin ini penting: Indonesia tidak perlu menemukan formulasi definisi dari nol. Ia bisa mengambil elemen-elemen terbaik dari definisi yang sudah ada di berbagai yurisdiksi, menyesuaikannya dengan terminologi hukum Indonesia yang sudah ada (termasuk terminologi dalam UU ITE), dan mengonsultasikannya dengan komunitas akademis dan teknis sebelum difinalkan. Proses konsultasi ini yang sering diabaikan dalam proses legislasi yang terburu-buru adalah kunci untuk menghasilkan definisi yang tahan uji.

14.3 MODEL CHINA: PELABELAN WAJIB DAN PENDAFTARAN KONTEN

Inti dari Pendekatan China

Model China yang dibangun melalui *Administrative Provisions on Deep Synthesis* (2023) berangkat dari premis yang berbeda dari model-model lain: alih-alih hanya melarang penggunaan berbahaya setelah fakta, ia membangun infrastruktur transparansi yang bersifat preventif. Semua konten yang dihasilkan atau dimanipulasi secara sintetis oleh AI harus diberi label yang jelas baik yang dapat dibaca manusia maupun yang tertanam dalam meta data file sehingga setiap orang yang berinteraksi dengan konten itu tahu bahwa ia sedang berhadapan dengan konten yang dihasilkan AI.

Logika di balik pendekatan ini adalah bahwa sebagian besar bahaya deepfake penyebaran disinformasi, manipulasi opini publik, kerusakan reputasi bergantung pada orang percaya bahwa konten yang mereka lihat atau dengar adalah asli. Jika konten selalu berlabel jelas sebagai konten AI, kemampuannya untuk menipu berkurang secara dramatis. Ini adalah solusi yang elegan secara konseptual dan dalam konteks China dengan ekosistem platform yang terkontrol, ia memiliki plausibilitas implementasi yang nyata.

Dari model China, setidaknya dua aspek sangat relevan untuk Indonesia. Pertama adalah konsep pelabelan wajib persyaratan bahwa konten yang secara substansial dihasilkan atau dimanipulasi oleh AI harus diidentifikasi sebagai demikian. Ini bukan ide yang datang hanya dari China; standar industri seperti C2PA yang dikembangkan oleh konsorsium perusahaan teknologi Barat juga bergerak ke arah yang sama. Pelabelan wajib memberikan informasi yang diperlukan oleh konsumen konten untuk membuat penilaian yang informed tentang apa yang mereka konsumsi.

Kedua adalah kewajiban penyedia layanan deep synthesis untuk melakukan verifikasi identitas pengguna. Dalam konteks Indonesia yang sudah memiliki sistem KTP elektronik dan identifikasi digital yang berkembang, persyaratan verifikasi identitas untuk layanan yang secara khusus menyediakan pembuatan konten AI bukan ide yang tidak realistis. Ia tidak mengeliminasi anonimitas untuk semua aktivitas daring sebuah pembatasan yang terlalu luas dan yang bertentangan dengan norma privasi tapi memberikan akuntabilitas yang lebih kuat untuk penggunaan layanan yang memiliki potensi penyalahgunaan tinggi.

Tapi model China tidak bisa diadopsi secara keseluruhan tanpa melihat konteks yang melingkupinya dan konteks itu sangat penting. Regulasi deep synthesis China berjalan dalam ekosistem tata kelola digital yang berbeda secara fundamental dari apa yang ada atau yang seharusnya ada di Indonesia sebagai negara demokrasi. Persyaratan identifikasi pengguna di China terhubung dengan sistem pengawasan yang jauh lebih luas data yang dikumpulkan dalam rangka verifikasi identitas bisa dan dalam praktiknya digunakan untuk tujuan yang melampaui regulasi deepfake.

Klausul tentang "keamanan nasional" dan "kepentingan nasional" dalam regulasi China memberikan otoritas yang sangat luas kepada pemerintah untuk mendefinisikan konten mana

yang boleh dan tidak boleh beredar sebuah kewenangan yang, dalam sistem tanpa check and balance yang independen, rentan terhadap penyalahgunaan politik. Adopsi model China tanpa penyesuaian yang mendalam terhadap konteks demokrasi Indonesia bisa menghasilkan instrumen sensor yang lebih kuat dari yang diperlukan untuk melindungi korban deepfake.

Pelajaran dari China, dengan demikian, adalah pelajaran yang selektif: ambil konsep pelabelan wajib dan verifikasi layanan berisiko tinggi, tapi pastikan ia diimplementasikan dalam kerangka hukum yang memiliki proteksi yang kuat terhadap penyalahgunaan oleh negara termasuk pengawasan yudisial, mekanisme banding yang independen, dan pembatasan yang jelas tentang bagaimana data yang dikumpulkan bisa digunakan.

14.4 MODEL KOREA SELATAN: KRIMINALISASI PEMBUATAN

Logika Kriminalisasi Lebih Awal dalam Rantai Kausalitas

Korea Selatan membuat pilihan regulasi yang paling berani di antara yurisdiksi-yurisdiksi yang kita pelajari: mengkriminalisasi pembuatan deepfake seksual, bukan hanya distribusinya. Ini adalah pergeseran titik intervensi yang sangat signifikan secara doktrin. Sebagian besar hukum pidana dan ini termasuk hukum pidana Indonesia mengkriminalisasi tindakan yang menyebabkan bahaya aktual. Mengkriminalisasi pembuatan sesuatu yang belum disebarkan dan yang mungkin tidak pernah disebarkan adalah langkah yang menantang asumsi-asumsi dasar tentang kapan hukum pidana seharusnya berintervensi.

Argumen Korea Selatan yang sudah dibahas di Bab 10 dan 11pada intinya adalah bahwa bahaya dari pembuatan deepfake seksual seseorang terjadi pada saat pembuatan, bukan hanya pada saat distribusi. Fakta bahwa seseorang telah membuat representasi seksual palsu dari seseorang lain tanpa izin sudah merupakan pelanggaran serius terhadap martabat dan otonomi orang tersebut terlepas dari apa yang kemudian dilakukan dengan konten itu. Ada juga argumen pencegahan: ancaman untuk membuat deepfake seksual sudah digunakan sebagai alat kontrol dan pemerasan bahkan ketika konten itu tidak pernah dibuat atau disebarkan.

Relevansi untuk Indonesia: Kasus yang Kuat

Kasus untuk mengadopsi pendekatan Korea Selatan setidaknya sebagian sangat kuat untuk konteks Indonesia. Pertama, Indonesia sudah memiliki pengalaman dengan kejahatan berbasis gender yang dimediasi teknologi: kasus penyebaran konten intim non-konsensual sudah ditangani di bawah UU ITE, meski dengan hasil yang tidak selalu memuaskan. Memperluas cakupan untuk mencakup pembuatan deepfake intim bahkan tanpa distribusi adalah langkah yang logis dalam kontinum perlindungan ini.

Kedua, profil korban deepfake seksual di Indonesia seperti di Korea Selatan sangat berdimensi gender. Data dari berbagai kasus yang sudah dilaporkan menunjukkan bahwa korban secara konsisten adalah perempuan, sering kali perempuan muda, dan bahwa pelaku menggunakan ancaman pembuatan atau penyebaran deepfake sebagai alat kontrol dalam

hubungan yang sudah ada atau sebagai alat intimidasi terhadap orang asing. Kriminalisasi pembuatan memberikan perlindungan pada tahap lebih awal dari rantai kekerasan ini.

Ketiga dan ini adalah pertimbangan yang lebih nuansir kriminalisasi pembuatan mengirimkan sinyal normatif yang kuat tentang apa yang dianggap oleh masyarakat sebagai perilaku yang tidak dapat diterima. Dalam konteks Indonesia di mana kesadaran tentang deepfake sebagai bentuk kekerasan seksual masih rendah, sinyal normatif dari hukum pidana bisa memainkan peran penting dalam membentuk persepsi publik peran yang tidak bisa dimainkan oleh regulasi administratif yang lebih teknis.

Tantangan Adoptasi: Apa yang Harus Disesuaikan

Tapi adopsi pendekatan Korea Selatan menghadapi tantangan yang nyata dalam konteks Indonesia. Tantangan pertama adalah kapasitas penegakan. Seperti yang dibahas di Bab 12, kapasitas forensik digital Indonesia untuk mendeteksi pembuatan deepfake apalagi pembuatan yang belum disebarkan masih sangat terbatas. Mengkriminalisasi pembuatan tanpa memiliki kapasitas untuk mendeteksinya adalah undangan untuk hukum yang tidak ditegakkan yang pada gilirannya bisa merusak kredibilitas seluruh kerangka hukum.

Tantangan kedua adalah risiko kriminalisasi berlebihan. Definisi "pembuatan deepfake seksual" yang terlalu luas bisa mencakup perilaku yang tidak seharusnya masuk ranah pidana misalnya, pembuatan yang dilakukan oleh individu untuk konsumsi pribadi yang tidak pernah menimbulkan ancaman atau bahaya bagi siapapun. Ini bukan argumen untuk tidak mengkriminalisasi sama sekali, tapi untuk merancang ketentuan pidana dengan cukup cermat sehingga sasarannya tepat.

Tantangan ketiga adalah posisi dalam tradisi hukum Indonesia. KUHP dan UU ITE, dalam tradisinya, lebih berorientasi pada tindakan yang menyebabkan bahaya aktual daripada pada preparatory acts. Memperluas kriminalisasi ke tahap pembuatan akan memerlukan argumentasi yang kuat dalam proses legislasi dan mungkin resistensi dari mereka yang khawatir bahwa ekspansi ini menciptakan preseden yang bisa digunakan untuk mengkriminalisasi aktivitas lain yang belum menyebabkan bahaya.

Rekomendasi yang paling terukur adalah pendekatan bertahap: pertama, memastikan distribusi deepfake seksual dikriminalisasi secara eksplisit dan komprehensif (bukan hanya tersirat di bawah ketentuan kesusilaan yang ada); kemudian, setelah kapasitas penegakan berkembang, mempertimbangkan kriminalisasi pembuatan dengan definisi yang ketat dan dengan pengecualian yang jelas untuk konteks yang tidak menimbulkan bahaya.

14.5 MODEL UNI EROPA: BERBASIS RISIKO DENGAN TRANSPARANSI TERUKUR

Arsitektur Berbasis Risiko: Prinsip dan Praktik

Pendekatan Uni Eropa terhadap regulasi AI secara umum dan deepfake sebagai bagian darinya dibangun di atas prinsip proporsionalitas berbasis risiko: semakin tinggi risiko yang ditimbulkan oleh sistem atau konten tertentu, semakin ketat kewajiban yang diimposkan. Ini

adalah prinsip yang secara intuitif masuk akal dan yang menghindari dua ekstrem yang sama-sama bermasalah: regulasi yang terlalu longgar yang membiarkan semua aktivitas berbahaya, dan regulasi yang terlalu ketat yang membatasi inovasi dan aktivitas sah.

EU AI Act mengklasifikasikan sistem AI ke dalam empat kategori risiko: tidak dapat diterima (dilarang sepenuhnya), risiko tinggi (diizinkan dengan kewajiban ketat), risiko terbatas (dengan kewajiban transparansi), dan risiko minimal (hampir tidak diregulasi). Sistem AI yang digunakan untuk membuat deepfake dari orang nyata termasuk dalam kategori risiko terbatas artinya kewajiban utamanya adalah transparansi: pengguna harus diberi tahu bahwa mereka berinteraksi dengan konten yang dihasilkan AI.

Kewajiban transparansi ini, meski terlihat sederhana, memiliki implikasi yang luas. Platform yang menghosting konten AI-generated wajib memastikan bahwa konten tersebut diberi label yang jelas. Penyedia model AI yang digunakan untuk menghasilkan konten yang menampilkan orang nyata wajib memastikan bahwa outputnya mencakup tanda-tanda yang dapat dideteksi secara teknis. Dan jurnalis, pendidik, peneliti, serta pembuat konten satire mendapat pengecualian yang terartikulasi sebuah keseimbangan yang mencoba mengakomodasi kebebasan berekspresi sah sambil melindungi dari manipulasi berbahaya.

DSA dan Kewajiban Platform Besar

Melengkapi AI Act, DSA membangun kewajiban berlapis untuk platform berdasarkan ukuran. Yang paling relevan untuk Indonesia adalah prinsip yang mendasarinya bahwa platform yang lebih besar dan yang memiliki dampak sistemik yang lebih besar terhadap ekosistem informasi publik harus menanggung kewajiban yang lebih besar bukan detail spesifik kewajiban yang dirancang untuk konteks Eropa.

Prinsip ini sangat relevan untuk Indonesia karena ia memberikan logika untuk menghindari pendekatan *one-size-fits-all* yang menerapkan kewajiban yang sama untuk platform dengan satu juta pengguna dan platform dengan ratusan juta pengguna. Dalam konteks Indonesia dengan ekosistem platform yang sangat beragam dari raksasa global seperti TikTok dan Instagram hingga platform lokal yang lebih kecil diferensiasi kewajiban berdasarkan ukuran dan risiko adalah pendekatan yang jauh lebih proporsional.

Transparansi Terukur: Konsep yang Paling Dapat Diadaptasi

Dari seluruh model EU, konsep "transparansi terukur" (*measurable transparency*) adalah yang paling dapat diadaptasi untuk konteks Indonesia. Ini mencakup beberapa elemen yang bisa dipisah-pisahkan dan diadopsi secara bertahap sesuai kapasitas kelembagaan yang ada.

Pertama, kewajiban pelabelan platform dan penyedia layanan AI wajib memastikan bahwa konten yang secara substansial dihasilkan oleh AI diberi label yang jelas. Ini bisa diimplementasikan secara bertahap: dimulai dengan platform besar yang sudah memiliki infrastruktur teknis, kemudian diperluas ke platform yang lebih kecil seiring standar teknis berkembang.

Kedua, kewajiban pelaporan platform diwajibkan untuk melaporkan secara berkala kepada otoritas tentang volume dan penanganan konten deepfake yang dilaporkan. Tanpa data yang dapat diverifikasi, evaluasi kebijakan menjadi hampir mustahil. Kewajiban pelaporan adalah prasyarat untuk pembuatan kebijakan berbasis bukti.

Ketiga, kewajiban audit independen untuk platform besar memungkinkan peneliti yang sudah diverifikasi untuk mengakses data yang diperlukan untuk mengevaluasi apakah kebijakan platform benar-benar diimplementasikan. Ini adalah mekanisme akuntabilitas yang kuat yang tidak memerlukan aparatur birokrasi yang besaria mengandalkan kapasitas akademis dan masyarakat sipil yang sudah ada.

Keterbatasan Model EU untuk Indonesia

Model EU, untuk semua keunggulannya, memiliki keterbatasan yang harus jujur diakui ketika mempertimbangkan relevansinya untuk Indonesia. Pertama, ia memerlukan kapasitas regulasi yang sangat tinggi. DSA dan AI Act diimplementasikan oleh Komisi Eropa dan regulator nasional yang memiliki sumber daya, keahlian teknis, dan independensi kelembagaan yang signifikan. Membangun kapasitas yang ekuivalen di Indonesia bahkan sebagian kecilnya memerlukan investasi yang serius dan jangka panjang.

Kedua, arsitektur berbasis risiko memerlukan penilaian risiko yang konstan dan berbasis bukti proses yang memerlukan data yang baik, metodologi yang disepakati, dan kapasitas analitik yang tidak selalu tersedia. Tanpa fondasi ini, kategori risiko menjadi kategorisasi yang sewenang-wenang yang tidak mencerminkan risiko aktual.

Ketiga dan ini adalah poin yang lebih politis pendekatan EU yang sangat mendetail dan teknis memerlukan proses legislasi yang panjang, konsultatif, dan berbasis bukti. Dalam konteks sistem legislasi Indonesia yang sering bergerak lambat dan yang sering menghasilkan regulasi yang kurang detail, mengadopsi model EU secara keseluruhan adalah tantangan yang berat. Mengadopsi prinsip-prinsip dasarnya terutama berbasis risiko dan transparansi terukur adalah langkah yang lebih realistis.

14.6 KONTEKS HUKUM INDONESIA: APA YANG MENJADI PEMBEDA

Tradisi Hukum Civil Law Dan Implikasinya

Indonesia mewarisi tradisi hukum civil law dari sistem hukum Belanda sebuah tradisi yang, seperti hukum Prancis yang kita bahas di Bab 10, lebih mengutamakan kodifikasi yang sistematis daripada preseden yurisprudensial. Ini memiliki implikasi langsung untuk strategi regulasi deepfake: alih-alih membiarkan yurisprudensi berkembang secara organik melalui putusan pengadilan (pendekatan common law), Indonesia lebih cocok untuk membangun kerangka legislasi yang komprehensif dan kemudian membiarkan putusan pengadilan mengisi detail implementasinya.

Strategi "biarkan pengadilan memutuskan" yang sering digunakan di negara common law seperti AS bukan strategi yang optimal untuk Indonesia. Korban deepfake di Indonesia tidak bisa

mengandalkan preseden yang berkembang mereka memerlukan undang-undang yang jelas yang memberikan dasar hukum yang tidak ambigu. Ini menempatkan beban yang lebih besar pada proses legislasi untuk menghasilkan teks yang cukup jelas dan komprehensif.

Struktur Plural Hukum Indonesia

Satu faktor yang sering diabaikan dalam diskusi tentang regulasi teknologi di Indonesia adalah kompleksitas struktur hukum Indonesia yang plural: di samping hukum nasional yang berlaku secara umum, ada hukum adat (adat law) yang masih berlaku di banyak komunitas, dan hukum agama (khususnya hukum Islam melalui pengadilan agama) yang berlaku untuk urusan-urusan tertentu.

Untuk regulasi deepfake, pluralisme hukum ini memiliki beberapa implikasi. Pertama, norma-norma tentang kehormatan, martabat, dan reputasi yang sudah ada dalam hukum adat dan hukum agama bisa menjadi sumber legitimasi kultural untuk regulasi deepfake menunjukkan bahwa perlindungan terhadap manipulasi identitas dan kehormatan bukan nilai asing yang diimpor dari Barat, melainkan sesuatu yang sudah ada dalam tradisi hukum Indonesia sendiri. Ini adalah argumen yang berguna dalam proses legislasi dan sosialisasi.

Kedua, pengadilan agama yang menangani kasus-kasus yang melibatkan komunitas Muslim mungkin akan berhadapan dengan kasus deepfake dalam konteks keluarga atau komunitas misalnya, deepfake yang dibuat untuk merusak reputasi seseorang dalam konteks perselisihan keluarga atau komunitas. Memastikan bahwa hakim-hakim pengadilan agama juga memiliki literasi dasar tentang deepfake dan bukti digital adalah kebutuhan yang mungkin tidak langsung terlihat tapi nyata.

Ekosistem Digital Indonesia: Karakteristik Khusus

Indonesia memiliki beberapa karakteristik ekosistem digital yang membedakannya dari yurisdiksi-yurisdiksi yang sudah kita pelajari, dan yang harus membentuk pendekatan regulasi deepfake yang relevan.

Penetrasi smartphone yang sangat tinggi dengan basis komputer desktop yang relatif rendah berarti bahwa produksi dan konsumsi konten digital termasuk deepfake terjadi terutama di perangkat mobile. Aplikasi-aplikasi mobile yang memungkinkan pembuatan deepfake dengan mudah, yang sering didesain dengan antarmuka yang sangat sederhana dan yang menargetkan pengguna non-teknis, adalah tantangan yang sangat nyata. Regulasi yang tidak memperhitungkan karakteristik ekosistem mobile ini misalnya, yang berfokus terutama pada platform desktop akan melewatkan bagian terbesar dari masalah.

Penggunaan WhatsApp dan Telegram yang sangat dominan untuk berbagi konten termasuk konten yang bermasalah menimbulkan tantangan regulasi yang berbeda dari yang dihadapi oleh platform terbuka seperti TikTok atau Instagram. Konten yang dibagikan dalam grup WhatsApp atau Telegram yang terenkripsi jauh lebih sulit dideteksi, dilacak, dan dimoderasi dibanding konten yang diunggah ke platform publik. Ini berarti bahwa regulasi yang hanya

berfokus pada platform terbuka akan melewatkan jalur distribusi yang sangat penting di Indonesia.

Tingkat literasi digital yang masih sangat beragam di berbagai kelompok populasi dan daerah berarti bahwa strategi yang mengandalkan kesadaran pengguna seperti membaca label AI pada konten akan memiliki efektivitas yang sangat berbeda antara pengguna urban yang terdidik dan pengguna pedesaan yang baru mengenal smartphone. Regulasi yang tidak disertai dengan program literasi digital yang masif dan berkelanjutan akan menghasilkan perlindungan yang tidak merata.

Kapasitas Kelembagaan: Realitas yang Tidak Bisa Diabaikan

Setiap rekomendasi kebijakan yang tidak mempertimbangkan kapasitas kelembagaan yang realistis adalah rekomendasi yang tidak berguna secara praktis. Indonesia memiliki sejumlah lembaga yang relevan untuk regulasi dan penegakan hukum deepfake: Kementerian Komunikasi dan Informatika (Kominfo) dengan fungsi pengawasan PSE, Bareskrim Polri dengan Dittipidhsiber, Badan Siber dan Sandi Negara (BSSN) dengan fungsi keamanan siber, dan berbagai komisi yang relevan seperti Komnas Perempuan dan KPAI.

Tapi koordinasi antar lembaga-lembaga ini masih perlu ditingkatkan secara signifikan. Dalam praktiknya, kasus-kasus deepfake yang melibatkan dimensi kekerasan gender, kejahatan siber, dan perlindungan anak secara bersamaan bisa jatuh di antara kursi dari berbagai lembaga yang masing-masing merasa bukan domain utamanya. Membangun mekanisme koordinasi yang efektif bukan hanya secara formal tapi dalam praktik adalah prasyarat untuk regulasi yang efektif.

Di tingkat pengadilan, kapasitas hakim untuk mengevaluasi bukti digital yang kompleks termasuk analisis deepfake masih sangat terbatas. Mahkamah Agung memiliki program pendidikan hakim yang berkelanjutan, tapi kurikulum yang secara spesifik membahas deepfake dan forensik AI belum ada atau sangat terbatas. Ini adalah gap yang konkret yang harus diatasi, tidak hanya melalui perubahan regulasi tapi melalui investasi dalam pendidikan yudisial.

14.7 KERANGKA REKOMENDASI: SINTESIS UNTUK INDONESIA

Strategi Regulasi Bertahap

Berdasarkan analisis komparatif yang sudah dilakukan dan analisis konteks Indonesia yang spesifik, rekomendasi utama adalah strategi regulasi yang bertahap bukan dalam pengertian menunda tindakan, tapi dalam pengertian membangun fondasi yang kuat sebelum membangun struktur yang lebih kompleks di atasnya.

Tahap pertama (jangka pendek, 1–2 tahun): fokus pada fondasi definitional dan kriminalisasi distribusi yang eksplisit. Indonesia perlu memiliki definisi deepfake yang jelas dalam undang-undang baik melalui revisi UU ITE yang sudah ada atau melalui peraturan terpisah dan ketentuan yang secara eksplisit mengkriminalisasi distribusi deepfake intim non-konsensual. Ini adalah langkah yang paling mendesak karena korban deepfake seksual sudah ada sekarang dan memerlukan perlindungan hukum yang jelas. Definisi yang diadopsi sebaiknya mengikuti

pendekatan fungsional yang tidak terlalu terikat pada teknologi spesifik, dengan mengambil elemen terbaik dari definisi AS dan EU.

Tahap kedua (jangka menengah, 2–4 tahun): membangun infrastruktur transparansi dan kewajiban platform yang lebih kuat. Ini mencakup kewajiban pelabelan konten AI untuk PSE besar yang beroperasi di Indonesia, kewajiban pelaporan yang transparan tentang penanganan konten deepfake, dan mekanisme penghapusan konten yang lebih cepat dengan tenggat waktu yang jelas mengambil pelajaran dari TAKE IT DOWN Act AS tapi disesuaikan dengan konteks Indonesia. Tahap ini juga harus mencakup pembangunan kapasitas: pelatihan hakim, penguatan Labfor Polri dalam forensik deepfake, dan pengembangan program literasi digital yang massif.

Tahap ketiga (jangka panjang, 4–7 tahun): mempertimbangkan pendekatan berbasis risiko yang lebih komprehensif mengambil pelajaran dari model EU termasuk kemungkinan kriminalisasi pembuatan deepfake intim mengikuti Korea Selatan, jika kapasitas penegakan sudah berkembang cukup untuk membuat kriminalisasi itu bermakna. Tahap ini juga harus mencakup evaluasi berbasis bukti terhadap efektivitas langkah-langkah dari tahap sebelumnya, dengan kesiapan untuk merevisi pendekatan berdasarkan apa yang berhasil dan apa yang tidak.

Prioritas Lintas-Tahap

Di luar strategi bertahap yang tahap-tahapnya bergantung pada urutan, ada beberapa prioritas yang harus berjalan secara paralel dari awal, bukan menunggu tahap tertentu. Pembangunan kapasitas kelembagaan harus dimulai sekarang, tidak menunggu regulasi yang sempurna. Melatih hakim, membangun kapasitas forensik digital Polri, dan memperkuat mekanisme koordinasi antar lembaga adalah investasi yang manfaatnya akan dirasakan terlepas dari detail regulasi yang akhirnya diadopsi.

Program literasi digital yang secara spesifik mencakup deepfake tentang cara mengenalinya, cara melaporkannya, dan hak-hak yang dimiliki korbanharus menjadi bagian dari strategi pendidikan nasional. Ini bukan hanya tugas Kominfo; kementerian pendidikan, KPAI, dan Komnas Perempuan semuanya memiliki peran yang tidak bisa saling digeser.

Partisipasi aktif dalam forum internasional tentang tata kelola AI dan deepfake termasuk forum ASEAN, forum UN, dan forum multilateral lainnya harus menjadi prioritas diplomatik yang nyata. Indonesia tidak harus hanya mengikuti standar yang dibentuk oleh orang lain; sebagai salah satu negara dengan pengguna internet terbesar di dunia, Indonesia memiliki suara yang seharusnya lebih berpengaruh dalam pembentukan standar global.

Analisis komparatif yang jujur juga harus mengidentifikasi jebakan-jebakan yang sebaiknya dihindari pelajaran negatif yang tidak kalah berharganya dari pelajaran positif.

- Pertama, jangan terburu-buru menghasilkan legislasi yang tidak disiapkan dengan baik hanya untuk menunjukkan responsivitas terhadap tekanan publik. Regulasi yang terburu-buru dengan definisi yang tidak jelas, kewajiban yang tidak realistis, atau pengecualian yang tidak dipikirkan dengan matang lebih buruk dari tidak ada regulasi sama sekali,

karena ia menciptakan ilusi perlindungan sementara tidak memberikan perlindungan nyata, dan karena ia menciptakan preseden hukum yang sulit untuk diperbaiki.

- Kedua, jangan adopsi model yang memerlukan kapasitas kelembagaan yang jauh melampaui apa yang tersedia tanpa terlebih dahulu membangun kapasitas tersebut. Mengadopsi kewajiban platform yang kompleks seperti DSA tanpa memiliki regulator yang mampu mengawasi kepatuhan hanya menghasilkan kewajiban di atas kertas yang tidak dipenuhi oleh siapapun.
- Ketiga, jangan biarkan regulasi deepfake menjadi instrumen sensor konten yang lebih luas. Definisi yang terlalu luas, prosedur penghapusan yang tidak memerlukan verifikasi memadai, atau kewenangan pemerintah yang tidak dibatasi dengan jelas bisa dengan mudah disalahgunakan untuk menyensor konten politik yang tidak disukai. Setiap regulasi harus disertai dengan safe guard yang jelas terhadap penyalahgunaan ini.

Rangkuman Bab

Bab ini telah melakukan perjalanan melalui empat model regulasi deepfake utama China, Korea Selatan, Uni Eropa (melalui AI Act dan DSA), dan Amerika Serikat untuk memetakan apa yang bisa dipelajari Indonesia dan apa yang perlu disesuaikan. Dari setiap model, ada elemen yang menarik dan elemen yang harus diwaspadai.

Dari China: ambil konsep pelabelan wajib dan verifikasi layanan berisiko tinggi, tapi pastikan implementasinya dalam kerangka yang memiliki proteksi kuat terhadap penyalahgunaan oleh negara. Dari Korea Selatan: pertimbangkan kriminalisasi distribusi deepfake intim sebagai langkah segera, dan kriminalisasi pembuatan sebagai langkah jangka menengah setelah kapasitas penegakan berkembang. Dari Uni Eropa: adopsi prinsip berbasis risiko dan transparansi terukur sebagai arsitektur regulasi, dengan diferensiasi kewajiban berdasarkan ukuran dan dampak platform. Dari Amerika Serikat: ambil mekanisme penghapusan konten dengan tenggat waktu yang jelas sebagai kewajiban minimum platform.

Tapi yang tidak boleh dilupakan dan ini adalah pelajaran meta dari studi perbandingan adalah bahwa regulasi terbaik sekalipun hanya berguna jika ia diimplementasikan oleh institusi yang kompeten, ditegakkan oleh sistem yang dapat dipercaya, dan didukung oleh masyarakat yang memiliki literasi yang cukup untuk menggunakannya. Membangun fondasi kelembagaan dan sosial itu secara paralel dengan pembangunan kerangka hukumnya adalah pekerjaan yang tidak bisa ditunda dan yang tidak bisa diselesaikan oleh undang-undang sendirian.

Perjalanan regulatif Indonesia dalam menghadapi deepfake baru dimulai. Bab-bab berikutnya akan mengeksplorasi aspek-aspek spesifik lain dari perjalanan itu termasuk peran pendidikan, peran masyarakat sipil, dan arsitektur kelembagaan yang diperlukan untuk membuat regulasi apapun yang diadopsi menjadi bermakna dalam kehidupan nyata korban dan masyarakat yang lebih luas.

BAB 15
USULAN KERANGKA REGULASI DEEPFAKE UNTUK INDONESIA

"Regulasi yang baik bukan yang paling ketat, bukan pula yang paling longgar. Ia adalah yang paling tepat sasaran yang tahu dengan jelas apa yang ingin dilindunginya, dari ancaman apa, dan dengan instrumen apa."

parafrase dari prinsip kebijakan publik berbasis bukti

Bab ini adalah kulminasi dari seluruh perjalanan analisis yang sudah dilakukan dalam buku ini. Setelah memahami cara kerja teknologi deepfake, memetakan dampaknya terhadap individu dan demokrasi, menelusuri perkembangan regulasi di berbagai yurisdiksi, menganalisis kerangka HAM yang berlaku, memeriksa tantangan pembuktian dan forensik, membahas tanggung jawab platform, dan melakukan studi perbandingan yang jujur kini saatnya mengajukan pertanyaan yang paling langsung dan paling sulit dijawab: apa yang konkret harus dilakukan Indonesia?

Bab ini tidak bermaksud menyajikan draft undang-undang yang final dan siap diketuk palu pekerjaan itu memerlukan proses konsultasi yang jauh lebih luas, melibatkan ahli hukum, teknologi, masyarakat sipil, dan yang paling penting perwakilan komunitas yang paling terdampak oleh deepfake. Yang ingin dicapai di sini adalah sesuatu yang berbeda: sebuah kerangka konseptual dan teknis yang cukup konkret untuk menjadi titik awal yang berguna bagi proses penyusunan regulasi, beserta argumentasi yang menjelaskan mengapa masing-masing komponen kerangka itu diperlukan dan bagaimana ia dirancang untuk menghindari jebakan-jebakan yang sudah diidentifikasi dalam analisis komparatif di bab sebelumnya.

Kerangka yang diusulkan terdiri dari lima pilar yang saling memperkuat definisi yang jelas, kriminalisasi yang tepat sasaran, kewajiban platform yang dapat ditegakkan, lembaga pengawas yang independen, dan peta jalan menuju RUU Kecerdasan Buatan. Tidak ada satu pilar yang bisa berdiri sendiri efektivitas kerangka ini bergantung pada seberapa baik kelima pilar itu bekerja bersama.

15.1 MENGAPA KERANGKA REGULASI BARU DIPERLUKAN

Argumen Yang Masih Terdengar: "Hukum Yang Ada Sudah Cukup"

Sebelum masuk ke substansi usulan, ada baiknya merespons satu argumen yang masih cukup sering terdengar dalam diskusi kebijakan: bahwa hukum yang ada sudah cukup untuk menangani masalah deepfake, dan yang diperlukan hanyalah penegakan yang lebih baik, bukan regulasi baru. Argumen ini biasanya menunjuk pada UU ITE, KUHP, dan berbagai peraturan terkait sebagai instrumen yang sudah ada.

Argumen ini tidak sepenuhnya salah ada kasus-kasus tertentu yang bisa ditangani dengan instrumen yang ada. Tapi ia melewatkan beberapa hal fundamental yang sudah dibahas secara mendalam dalam bab-bab sebelumnya. Pertama, instrumen yang ada tidak mendefinisikan deepfake harus dipaksakan masuk ke kategori seperti "konten yang melanggar kesusilaan" atau "informasi palsu" yang tidak sepenuhnya menangkap sifat unik manipulasi AI. Ketidakjelasan ini merugikan korban yang tidak mendapat perlindungan yang diprediksi, dan menguntungkan

pelaku yang bisa berargumen di pengadilan bahwa tindakannya tidak tercakup oleh ketentuan yang ada.

Kedua, instrumen yang ada tidak mengatur tanggung jawab platform secara spesifik untuk konten deepfake platform bisa berdalih bahwa kewajiban penghapusan umum tidak mensyaratkan langkah yang lebih aktif. Ketiga, instrumen yang ada tidak mengatur penyedia model AI generatif sama sekali sebuah celah yang semakin signifikan seiring teknologi deepfake semakin mudah diakses. Keempat dan mungkin yang paling mendasar instrumen yang ada tidak menyediakan remediasi yang memadai bagi korban, proses hukum yang panjang dan mahal untuk sesuatu yang memerlukan respons cepat.

Argumen "hukum yang ada sudah cukup" juga mengabaikan fungsi ekspresif hukum fakta bahwa undang-undang tidak hanya mengatur perilaku melalui sanksi, tapi juga mengomunikasikan nilai-nilai kolektif masyarakat. Tidak adanya ketentuan yang secara eksplisit melarang dan mengkriminalisasi deepfake berbahaya mengirimkan sinyal bahwa masyarakat tidak menganggap perilaku itu cukup serius untuk diatur secara khusus sinyal yang bertentangan dengan realitas kerusakan yang dialami korban.

Prinsip-Prinsip yang Memandu Kerangka Usulan

Kerangka regulasi yang diusulkan dalam bab ini dipandu oleh beberapa prinsip yang secara eksplisit harus diartikulasikan, karena mereka akan menentukan pilihan-pilihan yang dibuat dalam setiap pilar.

- Prinsip orientasi pada korban: setiap komponen regulasi dievaluasi pertama-tama dari perspektif apakah ia memberikan perlindungan dan remediasi yang nyata bagi korban. Regulasi yang terlihat komprehensif di atas kertas tapi yang tidak memberikan jalan yang dapat diakses bagi korban untuk mendapatkan keadilan adalah regulasi yang gagal.
- Prinsip proporsionalitas: beban regulasi harus proporsional dengan kapasitas dan kontribusi pihak yang dibebani terhadap masalah yang diatur. Platform besar dengan miliaran pengguna dan infrastruktur teknis yang canggih menanggung kewajiban yang lebih besar dari platform kecil. Pengembang model AI yang komersinya bergantung pada teknologi generatif menanggung kewajiban yang lebih besar dari pengguna individual.
- Prinsip kepastian hukum: definisi dan ketentuan harus cukup jelas untuk memberikan kepastian pihak yang tunduk pada regulasi harus bisa memahami apa yang dilarang dan apa yang diizinkan, dan pengadilan harus memiliki standar yang cukup jelas untuk menerapkan regulasi secara konsisten.
- Prinsip adaptabilitas: mengingat kecepatan perkembangan teknologi AI, regulasi harus dirancang untuk tidak menjadi usang dengan cepat. Ini berarti mendefinisikan hal-hal secara fungsional (apa yang dilakukan teknologi) daripada secara teknis (teknologi spesifik apa yang digunakan), dan membangun mekanisme yang memungkinkan regulasi diperbarui tanpa harus melalui proses legislasi penuh setiap kali teknologi berubah.
- Prinsip perlindungan kebebasan berekspresi: regulasi harus secara eksplisit mengecualikan aktivitas yang sah satire, jurnalisme, pendidikan, seni, penelitian ilmiah dan memastikan bahwa mekanisme penegakannya tidak bisa digunakan untuk membungkam ekspresi yang dilindungi.

15.2 PILAR PERTAMA: DEFINISI DEEPFAKE DALAM PERATURAN PERUNDANG-UNDANGAN

Strategi Definitional: Amandemen UU ITE versus Peraturan Baru

Pertanyaan pertama yang harus dijawab dalam menyusun definisi adalah: di mana definisi itu ditempatkan? Ada dua opsi utama: mengamandemen UU ITE yang sudah ada untuk menambahkan definisi dan ketentuan deepfake, atau membuat peraturan baru yang secara khusus mengatur deepfake (baik setingkat undang-undang maupun peraturan pemerintah atau peraturan menteri).

Masing-masing opsi memiliki kelebihan dan kekurangan. Amandemen UU ITE memberikan kekuatan hukum yang langsung ketentuan yang dimasukkan ke dalam UU ITE akan memiliki kedudukan yang sama dengan ketentuan lain dalam undang-undang itu dan akan menjadi bagian dari sistem yang sudah dikenal oleh hakim, jaksa, dan penyidik. Tapi proses amandemen UU ITE secara politis tidak sederhana revisi UU ITE selalu mengundang perdebatan yang luas dan kadang menghasilkan hasil yang tidak terduga.

Peraturan pemerintah (PP) atau peraturan presiden yang lebih spesifik bisa bergerak lebih cepat dari proses amandemen undang-undang, dan bisa lebih mudah diperbarui seiring perkembangan teknologi. Hal tersebut memiliki kedudukan hukum yang lebih rendah, tetapi tidak bisa menciptakan ketentuan pidana baru (yang memerlukan undang-undang), dan ia lebih rentan terhadap perubahan kebijakan yang tidak mengikuti proses legislatif.

Rekomendasi yang paling realistis adalah pendekatan dua jalur yang berjalan paralel: dalam jangka pendek, keluarkan peraturan pemerintah yang mendefinisikan deepfake dan menetapkan kewajiban platform (yang tidak memerlukan pembuatan ketentuan pidana baru); sambil bersamaan memulai proses amandemen UU ITE atau penyusunan undang-undang baru yang memasukkan ketentuan pidana spesifik untuk deepfake berbahaya.

Usulan Formulasi Definisi

Berdasarkan analisis komparatif di Bab 14 dan prinsip-prinsip yang diidentifikasi di awal bab ini, berikut adalah formulasi definisi deepfake yang diusulkan untuk konteks hukum Indonesia. Formulasi ini dimaksudkan sebagai titik awal untuk konsultasi, bukan sebagai teks final.

Usulan Definisi — Pasal [X] Ayat (1):

"Konten Sintetis Digital" adalah konten berupa gambar, video, audio, atau kombinasinya yang dihasilkan atau dimanipulasi secara substansial oleh sistem komputasi berbasis kecerdasan buatan atau pembelajaran mesin, sehingga menampilkan atau menyuarakan seseorang yang dapat diidentifikasi secara wajar dalam situasi, pernyataan, atau tindakan yang tidak mencerminkan kenyataan, dan yang dapat menimbulkan kesan menyesatkan pada penonton yang beritikad baik.

Usulan Definisi — Pasal [X] Ayat (2):

"Deepfake Berbahaya" adalah Konten Sintetis Digital sebagaimana dimaksud pada ayat (1) yang dibuat atau disebarkan tanpa persetujuan yang sah dari orang yang ditampilkan, dan yang bertujuan atau berakibat pada: (a) merugikan kehormatan, martabat, atau reputasi orang tersebut; (b) menimbulkan kerugian finansial; (c) memfasilitasi tindak pidana lain; atau (d) mengganggu proses demokrasi atau pemilihan umum.

Usulan Definisi — Pasal [X] Ayat (3):

Ketentuan sebagaimana dimaksud pada ayat (1) dan ayat (2) tidak berlaku terhadap: (a) konten satire atau parodi yang dengan jelas dapat dikenali sebagai karya satire atau parodi oleh penonton yang wajar; (b) konten yang dihasilkan untuk kepentingan jurnalistik, pendidikan, penelitian ilmiah, atau seni yang diberi keterangan yang jelas tentang sifat sintetisnya; (c) konten yang dibuat atas persetujuan eksplisit dari orang yang ditampilkan; (d) rekaman atau konten yang dihasilkan oleh proses produksi kreatif yang lazim yang tidak bertujuan menampilkan kenyataan.

Beberapa aspek dari formulasi ini perlu penjelasan. Frasa "dapat diidentifikasi secara wajar" dipilih untuk menghindari dua ekstrem: definisi yang terlalu sempit yang hanya mencakup identifikasi sempurna, dan definisi yang terlalu luas yang mencakup representasi yang sangat abstrak. "Penonton yang beritikad baik" dimasukkan untuk mengecualikan kasus di mana kesan palsu hanya terjadi pada penonton yang sudah memiliki predisposisi untuk percaya, bukan karena kualitas manipulasinya.

Pengecualian untuk satire, jurnalisme, pendidikan, dan seni adalah komponen yang tidak boleh dihapus, perlindungan terhadap penyalahgunaan regulasi untuk membungkam ekspresi sah. Tapi pengecualian ini harus cukup ketat: satire yang tidak jelas sebagai satire, atau yang tidak memberi keterangan bahwa kontennya dimanipulasi AI, tidak mendapat perlindungan pengecualian ini.

Mekanisme Pembaruan Definisi

Mengingat cepatnya perkembangan teknologi AI, definisi yang ditetapkan dalam undang-undang hari ini berisiko menjadi tidak memadai dalam lima tahun ke depan. Untuk mengantisipasi ini, perlu ada mekanisme pembaruan yang tidak memerlukan proses amandemen undang-undang penuh setiap kali teknik generatif baru muncul.

Mekanisme yang direkomendasikan adalah pemberian kewenangan kepada lembaga pengawas AI yang diusulkan dalam Pilar Keempat untuk menetapkan pedoman teknis tentang apa yang masuk dalam kategori "sistem komputasi berbasis kecerdasan buatan atau pembelajaran mesin" sebagaimana disebutkan dalam definisi. Pedoman teknis ini bisa diperbarui tanpa proses legislatif, selama ia tetap dalam koridor definisi induk yang ditetapkan oleh undang-undang. Ini adalah model yang digunakan oleh banyak sistem regulasi yang menghadapi tantangan teknologi yang bergerak cepat undang-undang menetapkan prinsip, regulasi teknis mengisi detailnya.

15.3 PILAR KEDUA: KRIMINALISASI YANG TEPAT SASARAN

Apa yang Harus Dikriminalisasi dan Mengapa

Kriminalisasi adalah instrumen hukum yang paling kuat sekaligus yang paling mahal-mahal dalam pengertian bahwa ia menggunakan sumber daya negara yang signifikan (penyidik, jaksa, pengadilan) dan bahwa konsekuensinya terhadap individu yang dihukum sangat berat. Karena itu, kriminalisasi harus digunakan secara selektif dan tepat sasaran hanya untuk perilaku yang bahayanya cukup serius untuk membenarkan sanksi pidana, dan yang tidak bisa secara memadai ditangani oleh instrumen hukum perdata atau administratif.

Dalam konteks deepfake, setidaknya tiga kategori perilaku memenuhi ambang ini: distribusi deepfake intim non-konsensual (yang sudah menimbulkan kerusakan psikologis dan reputasional yang terdokumentasi pada korban), penggunaan deepfake untuk memfasilitasi tindak pidana lain (penipuan, pemerasan, pencurian identitas), dan pembuatan atau distribusi deepfake yang mengganggu proses pemilihan umum (yang menyerang fondasi demokrasi).

Pembuatan deepfake intim tanpa distribusi adalah kategori yang lebih diperdebatkan—rekomendasi di sini adalah untuk mempertimbangkannya sebagai prioritas jangka menengah setelah kapasitas penegakan berkembang.

Usulan Ketentuan Pidana

Usulan Pasal Kriminalisasi Distribusi — Pasal [Y] Ayat (1):
"Setiap orang yang dengan sengaja mendistribusikan, menyebarluaskan, atau mentransmisikan Deepfake Berbahaya sebagaimana dimaksud dalam Pasal [X] ayat (2) huruf (a) yang menampilkan seseorang dalam konteks seksual atau intim tanpa persetujuan orang yang bersangkutan, dipidana dengan pidana penjara paling lama 6 (enam) tahun dan/atau denda paling banyak Rp1.000.000.000,00 (satu miliar rupiah)."

Usulan Pasal Kriminalisasi Pembuatan — Pasal [Y] Ayat (2):
"Setiap orang yang dengan sengaja membuat Konten Sintetis Digital sebagaimana dimaksud dalam Pasal [X] ayat (1) yang menampilkan seseorang dalam konteks seksual atau intim tanpa persetujuan orang yang bersangkutan, dipidana dengan pidana penjara paling lama 4 (empat) tahun dan/atau denda paling banyak Rp750.000.000,00 (tujuh ratus lima puluh juta rupiah)."

Usulan Pasal Kriminalisasi Penggunaan untuk Kejahatan — Pasal [Y] Ayat (3):
"Setiap orang yang menggunakan Konten Sintetis Digital sebagaimana dimaksud dalam Pasal [X] ayat (1) sebagai sarana untuk melakukan tindak pidana penipuan, pemerasan, pencemaran nama baik, atau ancaman, dipidana dengan pidana yang berlaku untuk tindak pidana yang difasilitasi ditambah sepertiga dari ancaman pidana maksimum tindak pidana dimaksud."

Usulan Pasal Kriminalisasi Deepfake Pemilu — Pasal [Y] Ayat (4):
"Setiap orang yang membuat atau mendistribusikan Konten Sintetis Digital yang menampilkan penyelenggara pemilihan umum, peserta pemilihan umum, atau pemilih dalam pernyataan atau tindakan yang tidak pernah terjadi, dengan tujuan memengaruhi hasil pemilihan umum atau merusak kepercayaan publik terhadap proses pemilihan umum, dipidana dengan pidana penjara paling lama 8 (delapan) tahun dan/atau denda paling banyak Rp2.000.000.000,00 (dua miliar rupiah)."

Beberapa catatan penting tentang usulan ketentuan pidana ini. Pertama, ancaman pidana yang diusulkan mencerminkan graduasi berdasarkan tingkat bahaya: distribusi deepfake intim (6 tahun) lebih berat dari pembuatan tanpa distribusi (4 tahun), dan deepfake pemilu mendapat ancaman tertinggi (8 tahun) karena dampaknya yang melampaui korban individual dan menyerang institusi demokrasi.

Kedua, untuk Ayat (3) tentang penggunaan deepfake sebagai sarana kejahatan, pendekatan yang digunakan adalah penambahan sepertiga dari ancaman tindak pidana yang difasilitasi bukan penciptaan tindak pidana baru yang terpisah. Ini konsisten dengan pendekatan KUHP Indonesia yang sudah ada untuk pemberatan dalam kasus penggunaan teknologi dalam kejahatan.

Ketiga, semua ketentuan ini mensyaratkan kesengajaan (dolus). Pembuatan atau distribusi yang tidak disengaja misalnya, seseorang yang meneruskan konten tanpa mengetahui bahwa itu

adalah deepfake tidak memenuhi unsur pidana. Ini adalah prinsip dasar hukum pidana yang harus dipertahankan untuk menghindari kriminalisasi yang berlebihan.

Ketentuan Pembuktian Khusus

Mengingat tantangan pembuktian yang sudah dibahas secara mendalam di Bab 12, ketentuan kriminalisasi deepfake harus disertai dengan beberapa ketentuan pembuktian khusus yang mengakomodasi sifat unik bukti digital.

Pertama, perlu ketentuan yang secara eksplisit menetapkan bahwa keterangan ahli forensik digital adalah alat bukti yang sah dalam perkara deepfake, dan yang menetapkan standar minimum kualifikasi ahli yang bisa diterima sebagai saksi ahli dalam perkara ini. Tanpa standar kualifikasi yang jelas, persidangan deepfake berisiko menghadapi "*battle of experts*" yang tidak terkontrol di mana kualitas keterangan tidak dinilai berdasarkan metodologi yang dapat dipertanggungjawabkan.

Kedua, perlu ketentuan yang menetapkan bahwa chain of custody digital prosedur perolehan, penyimpanan, dan analisis bukti elektronik memenuhi standar yang ditentukan oleh Peraturan Kepala Bareskrim tentang prosedur digital forensik. Referensi ke standar yang sudah ada ini menghindari kebutuhan untuk mendefinisikan ulang seluruh prosedur dalam undang-undang.

Ketiga, perlu dipertimbangkan ketentuan tentang beban pembuktian dalam konteks tertentu. Ketika seseorang yang diduga sebagai pembuat deepfake menguasai alat atau sistem yang digunakan untuk membuat konten yang bersangkutan, ada argumen untuk menempatkan beban untuk menjelaskan penggunaan alat tersebut pada pihak yang bersangkutan tanpa sepenuhnya membalik presumsi tidak bersalah yang adalah prinsip fundamental hukum pidana.

15.4 PILAR KETIGA: KEWAJIBAN PLATFORM YANG DAPAT DITEGAKKAN

Diferensiasi Berdasarkan Ukuran Dan Risiko

Mengikuti prinsip proporsionalitas dan pelajaran dari model EU yang dibahas di Bab 14, kewajiban platform sebaiknya dibedakan berdasarkan ukuran dan tingkat risiko platform. Usulan ini mengklasifikasikan platform ke dalam tiga kategori: Platform Digital Besar (PDB), Platform Digital Menengah (PDM), dan Platform Digital Kecil (PDK), dengan kewajiban yang secara signifikan berbeda untuk masing-masing kategori.

Kriteria Klasifikasi Platform:

“Platform Digital Besar: lebih dari 10.000.000 (sepuluh juta) pengguna aktif bulanan di Indonesia. Platform Digital Menengah: antara 1.000.000 (satu juta) sampai dengan 10.000.000 (sepuluh juta) pengguna aktif bulanan. Platform Digital Kecil: kurang dari 1.000.000 (satu juta) pengguna aktif bulanan.”

Threshold ini memang agak sewenang-wenang angka pastinya bisa diperdebatkan tapi prinsip diferensiasinya tidak. Platform yang melayani puluhan juta pengguna Indonesia memiliki dampak sistemik yang jauh lebih besar dan sumber daya yang jauh lebih besar untuk memenuhi kewajiban, dibanding platform yang melayani ratusan ribu pengguna. Regulasi yang memberikan kewajiban yang sama untuk keduanya tidak proporsional dalam dua arah sekaligus: terlalu berat untuk yang kecil, dan terlalu ringan untuk yang besar.

Kewajiban Penghapusan dan Respons Laporan

Usulan Kewajiban PDB — Pasal [Z] Ayat (1):

"Platform Digital Besar wajib: (a) menyediakan mekanisme pelaporan Deepfake Berbahaya yang mudah diakses oleh pengguna, tersedia dalam bahasa Indonesia, dan dapat digunakan tanpa biaya; (b) menindaklanjuti laporan Deepfake Berbahaya dalam jangka waktu 24 (dua puluh empat) jam sejak laporan diterima untuk konten seksual atau intim, dan 72 (tujuh puluh dua) jam untuk kategori lainnya; (c) memastikan penghapusan seluruh salinan konten yang dilaporkan yang dapat diidentifikasi dalam platform yang sama; (d) menyampaikan pemberitahuan kepada pelapor tentang tindakan yang telah diambil dalam jangka waktu yang sama."

Usulan Kewajiban PDM — Pasal [Z] Ayat (2):

"Platform Digital Menengah wajib: (a) menyediakan mekanisme pelaporan konten yang dapat digunakan untuk melaporkan Deepfake Berbahaya; (b) menindaklanjuti laporan Deepfake Berbahaya dalam jangka waktu 72 (tujuh puluh dua) jam sejak laporan diterima; (c) menunjuk penanggung jawab konten yang dapat dihubungi oleh pengguna dan otoritas yang berwenang."

Untuk Platform Digital Kecil, kewajiban minimum adalah memiliki mekanisme pelaporan yang fungsional dan merespons perintah penghapusan dari otoritas yang berwenang dalam waktu yang ditentukan. Beban administrasi yang berlebihan pada platform kecil akan mencegah berkembangnya inovasi digital lokal sesuatu yang harus dihindari.

Kewajiban Pelabelan Konten AI

Kewajiban pelabelan adalah komponen yang mengambil pelajaran paling langsung dari model China dan dari standar industri seperti C2PA yang berkembang secara global. Tapi implementasinya di Indonesia perlu disesuaikan dengan realitas teknis dan kapasitas yang ada.

Usulan Kewajiban Pelabelan Pasal [Z] Ayat (3):

"Platform Digital Besar dan Platform Digital Menengah yang menyediakan fitur untuk menghasilkan atau memanipulasi konten menggunakan kecerdasan buatan wajib: (a) memastikan bahwa konten yang dihasilkan melalui fitur tersebut diberi keterangan yang jelas dan mudah terlihat bahwa konten tersebut dihasilkan atau dimanipulasi menggunakan kecerdasan buatan; (b) menyematkan metadata yang dapat dibaca mesin pada konten yang dihasilkan yang mengidentifikasi konten sebagai konten berbasis AI; (c) mempertahankan pelabelan tersebut ketika konten diunduh, dibagikan, atau ditransmisikan melalui platform yang sama. Ketentuan teknis tentang standar pelabelan ditetapkan lebih lanjut oleh Lembaga Pengawas Kecerdasan Buatan sebagaimana dimaksud dalam Pasal [W]."

Perlu dicatat bahwa kewajiban pelabelan ini hanya berlaku untuk konten yang dihasilkan melalui fitur yang disediakan platform bukan untuk konten yang diunggah oleh pengguna dari luar platform. Ini bukan kelonggaran yang tidak disengaja; ia mencerminkan batas kemampuan teknis yang realistis.

Platform tidak bisa secara andal mengidentifikasi semua konten AI yang diunggah pengguna teknologi deteksi yang ada belum cukup akurat untuk dijadikan dasar kewajiban hukum. Yang bisa dan harus dilakukan platform adalah memastikan bahwa konten yang dihasilkan oleh alat mereka sendiri selalu berlabel.

Kewajiban Pelaporan dan Transparansi

Usulan Kewajiban Transparansi — Pasal [Z] Ayat (4):

"Platform Digital Besar wajib menyampaikan laporan transparansi kepada Lembaga Pengawas Kecerdasan Buatan setiap 6 (enam) bulan, yang memuat sekurang-kurangnya: (a) jumlah laporan Deepfake Berbahaya yang diterima; (b) jumlah laporan yang ditindaklanjuti dengan penghapusan konten beserta rata-rata waktu penghapusan; (c) jumlah laporan yang ditolak beserta kategori alasan penolakan; (d) jumlah akun yang dihentikan atau dibatasi akibat pelanggaran ketentuan terkait Deepfake Berbahaya; (e) langkah-langkah teknis yang diimplementasikan untuk mendeteksi dan mencegah penyebaran Deepfake Berbahaya. Laporan tersebut dipublikasikan oleh Lembaga Pengawas Kecerdasan Buatan dalam format yang dapat diakses oleh publik."

Kewajiban transparansi ini adalah komponen yang sangat penting tapi yang sering diabaikan dalam desain regulasi. Tanpa data yang dapat diverifikasi tentang bagaimana platform menangani deepfake, tidak ada cara untuk mengevaluasi apakah regulasi yang ada efektif—dan tidak ada dasar yang kokoh untuk memutuskan apakah regulasi perlu diperkuat. Laporan transparansi yang dipublikasikan juga memberi masyarakat sipil dan akademisi data yang diperlukan untuk melakukan pengawasan independen.

15.5 PILAR KEEMPAT: LEMBAGA PENGAWAS AI INDEPENDEN

Mengapa Dibutuhkan Lembaga Baru

Salah satu pertanyaan yang selalu muncul dalam diskusi tentang regulasi AI adalah apakah dibutuhkan lembaga pengawas baru atau apakah fungsi pengawasan bisa diintegrasikan ke dalam lembaga yang sudah ada Kominfo, BSSN, atau Komisi Penyiaran Indonesia. Argumen untuk menggunakan lembaga yang ada biasanya berkisar pada efisiensi dan menghindari duplikasi birokrasi. Argumen untuk lembaga baru berkisar pada kebutuhan akan keahlian spesifik, independensi dari tekanan politik dan komersial, dan fokus yang tidak terpecah.

Rekomendasi yang diajukan di sini adalah bahwa lembaga pengawas AI yang independen diperlukan, dengan catatan penting: "independen" di sini berarti independen dari kepentingan komersial industri AI dan dari tekanan politik yang bisa mendistorsi pengawasan, bukan independen dari akuntabilitas publik. Lembaga ini harus memiliki akuntabilitas yang jelas kepada lembaga legislatif dan publik.

Alasan utamanya adalah bahwa regulasi AI termasuk deepfake memerlukan keahlian teknis yang sangat spesifik yang tidak dimiliki oleh birokrasi generalis. Kominfo memiliki mandat yang sangat luas dan sumber daya yang terbatas; meminta Kominfo untuk juga menjadi regulator AI yang kompeten dalam konteks teknis yang cepat berkembang adalah permintaan yang tidak realistis tanpa penguatan kelembagaan yang sangat signifikan. BSSN memiliki fokus pada keamanan siber yang berbeda dari regulasi konten AI. Dan tidak ada lembaga yang ada yang memiliki mandat dan kapasitas untuk mengawasi kewajiban platform dalam skala yang diperlukan.

Usulan Arsitektur Kelembagaan

Usulan Nomenklatur dan Status:

"Lembaga Pengawas Kecerdasan Buatan Nasional (LPKBN) ditetapkan sebagai lembaga pemerintah nonkementerian yang bertanggung jawab langsung kepada Presiden. LPKBN

menjalankan fungsi pengaturan, pengawasan, dan penegakan regulasi terkait kecerdasan buatan di Indonesia, termasuk namun tidak terbatas pada regulasi konten sintetis digital dan deepfake."

Penetapan sebagai lembaga pemerintah nonkementerian mengikuti model seperti BSSN atau Komnas HAM memberikan lembaga ini independensi operasional dari kementerian tertentu sambil tetap dalam struktur pemerintahan yang memiliki akuntabilitas yang jelas.
Usulan Komposisi Pimpinan:

"LPKBN dipimpin oleh Dewan Pengawas yang terdiri dari 7 (tujuh) anggota yang diangkat oleh Presiden atas persetujuan Dewan Perwakilan Rakyat dengan masa jabatan 5 (lima) tahun yang dapat diperpanjang satu kali. Anggota Dewan Pengawas berasal dari unsur: (a) akademisi dengan keahlian dalam kecerdasan buatan atau ilmu komputer (2 orang); (b) akademisi dengan keahlian dalam hukum teknologi atau hukum siber (2 orang); (c) perwakilan masyarakat sipil yang memiliki rekam jejak dalam advokasi hak digital (1 orang); (d) praktisi industri teknologi informasi yang tidak memiliki konflik kepentingan dengan entitas yang diawasi (1 orang); (e) unsur pemerintah dari kementerian atau lembaga yang relevan (1 orang). Anggota Dewan Pengawas dilarang memiliki kepentingan finansial dalam entitas yang diawasi selama masa jabatan."

Komposisi yang beragam ini dirancang untuk memastikan bahwa lembaga memiliki keahlian teknis yang cukup (akademisi AI dan hukum), legitimasi dari masyarakat sipil, perspektif industri yang tidak memiliki konflik kepentingan, dan koordinasi dengan pemerintah. Persyaratan persetujuan DPR untuk pengangkatan anggota adalah mekanisme akuntabilitas demokratis yang penting.

Kewenangan dan Fungsi LPKBN

LPKBN perlu memiliki kewenangan yang memadai untuk menjalankan fungsinya secara efektif termasuk kewenangan yang sering diabaikan dalam desain lembaga regulasi: kewenangan untuk mendapatkan informasi yang diperlukan dari entitas yang diawasi.

- Kewenangan pengaturan: menetapkan pedoman teknis tentang standar pelabelan konten AI, standar kualifikasi ahli forensik digital untuk keperluan persidangan, dan pedoman prosedur operasional standar untuk penanganan laporan deepfake oleh platform.
- Kewenangan pengawasan: menerima dan menindaklanjuti laporan transparansi dari platform, melakukan audit kepatuhan terhadap kewajiban platform baik secara berkala maupun atas dasar laporan pelanggaran, dan meminta informasi dari platform dan penyedia layanan AI yang diperlukan untuk fungsi pengawasan.
- Kewenangan penegakan administratif: menjatuhkan sanksi administratif kepada platform yang tidak memenuhi kewajiban termasuk peringatan, denda administratif, dan dalam kasus pelanggaran berulang atau serius, rekomendasi pencabutan izin operasional kepada kementerian yang berwenang. LPKBN tidak memiliki kewenangan pidana penegakan pidana tetap menjadi domain Polri dan kejaksaan.
- Kewenangan edukasi dan penelitian: mengembangkan dan mempublikasikan panduan bagi masyarakat tentang cara mengenali dan melaporkan deepfake, mendanai atau memfasilitasi penelitian tentang dampak deepfake dan efektivitas regulasi, dan memberikan rekomendasi kepada DPR dan pemerintah tentang penyesuaian regulasi berdasarkan evaluasi yang berbasis bukti.

Koordinasi dengan Lembaga yang Ada

LPKBN tidak beroperasi dalam isolasi harus memiliki mekanisme koordinasi yang jelas dengan lembaga-lembaga yang sudah ada. Koordinasi dengan Kominfo diperlukan untuk kasus-kasus yang memerlukan tindakan pemblokiran atau pemutusan akses. Koordinasi dengan Polri khususnya Dittipidhsiber Bareskrim diperlukan untuk kasus-kasus yang masuk ranah pidana dan yang memerlukan bantuan teknis forensik. Koordinasi dengan Komnas Perempuan dan KPAI diperlukan untuk kasus-kasus yang melibatkan korban perempuan dan anak.

Mekanisme koordinasi yang diusulkan adalah Gugus Tugas Lintas Lembaga yang bertemu secara berkala (setidaknya triwulanan) untuk membahas perkembangan regulasi, kasus-kasus prioritas, dan kebutuhan kapasitas bersama. Gugus tugas ini tidak memiliki kewenangan operasional tersendiri, forum koordinasi, bukan lembaga baru tetapi keberadaannya secara formal mencegah silo kelembagaan yang selama ini menjadi hambatan nyata dalam penanganan kasus-kasus yang lintas yurisdiksi lembaga.

15.6 PILAR KELIMA: PETA JALAN MENUJU RUU KECERDASAN BUATAN INDONESIA

Deepfake Dalam Konteks Yang Lebih Luas: AI Act Indonesia

Regulasi deepfake yang diusulkan dalam bab ini bisa berdiri sendiri sebagai undang-undang atau peraturan yang spesifik dan dalam jangka pendek, mungkin itulah yang paling realistis untuk segera dilaksanakan. Tapi dalam perspektif yang lebih panjang, regulasi deepfake yang terisolasi akan kurang efektif dibanding regulasi yang menjadi bagian dari kerangka regulasi AI yang lebih komprehensif. Ini adalah pelajaran yang bisa ditarik dari pengalaman EU: AI Act yang komprehensif memberikan fondasi yang lebih kokoh untuk regulasi spesifik seperti deepfake dibanding undang-undang sektoral yang terpisah.

Indonesia sudah memiliki beberapa inisiatif yang bergerak ke arah regulasi AI yang lebih komprehensif. Pada 2020, pemerintah sudah menerbitkan Strategi Nasional Kecerdasan Buatan Indonesia 2020–2045 yang menetapkan visi dan prioritas pengembangan AI, meski bukan regulasi yang mengikat. Berbagai diskusi tentang perlunya RUU AI sudah berlangsung di berbagai forum akademis dan kebijakan. Yang belum ada adalah momentum politik yang cukup untuk mengubah diskusi itu menjadi proses legislasi yang konkret.

Mengapa RUU AI Berbasis Risiko Lebih Tepat dari Regulasi Sektoral

Argumen untuk RUU AI yang komprehensif berbasis risiko mengikuti prinsip (meski bukan detail) EU AI Act adalah argumen tentang koherensi sistem. Deepfake adalah salah satu aplikasi AI yang berbahaya, tapi bukan satu-satunya. Diskriminasi algoritmik dalam layanan kredit, manipulasi konten dalam sistem rekomendasi, penggunaan AI dalam pengambilan keputusan yang mempengaruhi hak individu semuanya menimbulkan kekhawatiran HAM dan keamanan yang memerlukan respons regulasi. Menangani masing-masing secara terpisah menghasilkan kerangka regulasi yang tidak konsisten dan yang memiliki celah di antara undang-undang yang berbeda.

Regulasi AI berbasis risiko memberikan kerangka tunggal yang menentukan bagaimana sistem AI diklasifikasikan berdasarkan risiko yang ditimbulkannya, kewajiban apa yang berlaku untuk masing-masing kategori risiko, siapa yang bertanggung jawab di berbagai titik dalam rantai nilai AI, dan bagaimana lembaga pengawas menjalankan fungsinya. Deepfake terutama yang digunakan untuk tujuan berbahaya masuk dalam kategori risiko tinggi dalam kerangka ini dan mendapat kewajiban yang sesuai.

Peta Jalan yang Diusulkan

Berikut adalah peta jalan yang diusulkan untuk perjalanan dari kondisi regulasi saat ini menuju kerangka regulasi AI yang lebih komprehensif, dengan penanda waktu yang realistis untuk konteks proses legislasi Indonesia.

- ☑ 2025–2026 (Fondasi): pengesahan regulasi deepfake spesifik melalui amandemen UU ITE atau peraturan pemerintah yang memasukkan definisi, kewajiban platform dasar, dan ketentuan pidana untuk distribusi deepfake intim dan deepfake pemilu. Bersamaan, pembentukan LPKBN melalui Peraturan Presiden sambil menunggu legislasi yang lebih permanen. Peluncuran program literasi digital tentang deepfake yang terintegrasi dengan program-program yang sudah ada di Kemendikbud dan Kominfo.
- ☑ 2026–2028 (Konsolidasi): evaluasi berbasis bukti terhadap efektivitas regulasi yang sudah ada menggunakan data laporan transparansi dari platform dan data penegakan hukum. Pengembangan standar teknis oleh LPKBN, termasuk standar pelabelan konten AI yang kompatibel dengan standar internasional. Penguatan kapasitas forensik digital Polri melalui kerja sama dengan lembaga internasional. Dimulainya proses konsultasi publik yang luas untuk RUU AI melibatkan akademisi, industri, masyarakat sipil, dan perwakilan komunitas yang terdampak.
- ☑ 2028–2030 (Pematangan): pengesahan RUU AI yang komprehensif berbasis risiko, yang mengintegrasikan ketentuan deepfake yang sudah ada ke dalam kerangka yang lebih luas. Ratifikasi atau aksesi terhadap instrumen internasional tentang tata kelola AI jika sudah tersedia. Evaluasi berkelanjutan dan penyesuaian regulasi berdasarkan perkembangan teknologi dan bukti efektivitas.

Peta jalan ini bukan jadwal yang kaku, namun orientasi tentang urutan prioritas. Jika ada momentum politik yang kuat untuk bergerak lebih cepat, peta jalan ini harus menyesuaikan diri. Yang penting adalah urutan fondasi sebelum struktur yang lebih kompleks: definisi yang jelas sebelum kriminalisasi yang komprehensif, kriminalisasi distribusi sebelum kriminalisasi pembuatan, kapasitas lembaga sebelum kewajiban yang mensyaratkan kapasitas tersebut.

15.7 KOMPONEN PENDUKUNG: YANG TIDAK BISA DILUPAKAN

Remediasi Non-Pidana untuk Korban

Kerangka yang berfokus pada kriminalisasi berisiko melewatkan kebutuhan yang paling mendesak dari sebagian besar korban deepfake: remediasi cepat yang tidak memerlukan menunggu proses pidana yang panjang. Jalur perdata harus diperkuat secara paralel dengan jalur pidana.

Hal ini mencakup setidaknya dua komponen. Pertama, mekanisme penghapusan konten berbasis perintah pengadilan yang cepat (*injunction*) korban harus bisa mendapatkan perintah pengadilan untuk penghapusan konten dalam hitungan hari, bukan bulan. Beberapa sistem hukum memiliki mekanisme semacam ini melalui ketentuan tentang provisional measures atau interim orders; Indonesia perlu memastikan bahwa mekanisme yang ada bisa diterapkan secara efektif untuk kasus deepfake.

Kedua, hak atas kompensasi finansial melalui jalur perdata yang lebih mudah diakses. Tuntutan perdata konvensional terlalu panjang dan mahal untuk kebanyakan korban deepfake. Perlu dipertimbangkan sistem small claims atau mekanisme mediasi yang dipercepat untuk kasus-kasus deepfake di mana kerugian dapat dikuantifikasi.

Pendidikan dan Literasi Digital

Regulasi yang paling baik sekalipun akan memiliki efektivitas yang terbatas jika masyarakat tidak memiliki literasi yang cukup untuk menggunakannya. Literasi deepfake mencakup tiga hal yang berbeda: kemampuan untuk mengenali deepfake (tidak selalu mungkin tapi setidaknya menyadari kemungkinannya), pengetahuan tentang hak-hak yang dimiliki korban dan mekanisme pelaporan yang tersedia, dan pemahaman tentang risiko hukum bagi mereka yang membuat atau menyebarkan deepfake berbahaya.

Mengintegrasikan literasi deepfake ke dalam kurikulum pendidikan dari tingkat sekolah menengah hingga perguruan tinggi adalah langkah yang memiliki dampak jangka panjang yang melampaui dampak regulasi jangka pendek apapun. Program pendidikan untuk guru yang mengajar mata pelajaran yang relevan (TIK, PKn, bahasa Indonesia dalam konteks media) perlu dikembangkan dan program literasi digital untuk masyarakat umum yang disampaikan melalui berbagai saluran media sosial, program televisi, komunitas, organisasi kemasyarakatan perlu menjadi komponen permanen dari strategi nasional, bukan kampanye sesaat.

Penelitian dan Pemantauan Berkelanjutan

Regulasi yang tidak dievaluasi secara berkala adalah regulasi yang tidak bisa diperbaiki. Indonesia perlu membangun kapasitas penelitian tentang dampak deepfake dan efektivitas regulasi yang independen dari kepentingan pemerintah dan industri. Ini berarti mendanai penelitian akademis melalui hibah BRIN atau skema pendanaan lain yang secara berkala mendokumentasikan: seberapa luas masalah deepfake di Indonesia, populasi mana yang paling terdampak, seberapa efektif regulasi yang ada, dan teknologi atau tren baru apa yang muncul dan yang memerlukan respons regulasi.

Tanpa penelitian yang berkelanjutan, proses evaluasi dan pembaruan regulasi yang dibangun dalam peta jalan di atas tidak akan memiliki bahan bakar yang diperlukan untuk bergerak. Komitmen untuk mendanai dan menggunakan penelitian independen adalah komponen dari kerangka regulasi yang efektif, bukan sekadar aksesori akademis.

Rangkuman Bab

Bab ini telah mengajukan kerangka regulasi deepfake yang terdiri dari lima pilar definisi yang jelas, kriminalisasi yang tepat sasaran, kewajiban platform yang dapat ditegakkan, lembaga pengawas yang independen, dan peta jalan menuju RUU AI beserta komponen pendukung yang tidak kalah pentingnya. Kerangka ini bukan produk akhir; ia adalah titik awal untuk diskusi yang harus melibatkan lebih banyak suara dari yang bisa diakomodasi dalam sebuah buku.

Yang ingin ditekankan sebagai penutup bab ini adalah bahwa kerangka regulasi, sebaik apapun rancangannya, hanya akan berdampak jika ada komitmen politik yang nyata untuk mengimplementasikannya. Komitmen itu tidak lahir secara spontan dibentuk oleh tekanan dari korban yang berani berbicara, dari advokasi masyarakat sipil yang konsisten, dari penelitian akademis yang mendokumentasikan bahaya secara kredibel, dan dari perdebatan publik yang jujur tentang apa yang dipertaruhkan.

Komunitas akademis termasuk mahasiswa dan dosen yang membaca buku ini memiliki peran yang tidak bisa diremehkan dalam proses pembentukan komitmen itu. Penelitian yang kredibel, analisis kebijakan yang tajam, advokasi yang berbasis bukti, dan partisipasi aktif dalam konsultasi publik adalah kontribusi yang konkret dan yang dibutuhkan. Hukum tidak berubah sendiri; ia berubah karena ada orang-orang yang cukup peduli dan cukup tahu untuk mendorong perubahannya ke arah yang benar.

Deepfake adalah tantangan yang nyata dan yang sudah di depan pintu. Indonesia belum terlambat untuk merespons dengan kerangka regulasi yang serius tapi jendela untuk membangun fondasi yang kuat sebelum masalah menjadi jauh lebih besar tidak akan terbuka selamanya. Waktu untuk memulai adalah sekarang, dan tempat untuk memulai adalah percakapan yang buku ini berharap bisa membantu memulai.

BAB 16
LITERASI DIGITAL, EDUKASI HUKUM, DAN PERLINDUNGAN MASYARAKAT

"Teknologi tidak berbahaya atau aman dengan sendirinya. Yang menentukan adalah siapa yang menggunakannya, dengan pengetahuan apa, dan dalam kerangka nilai apa."

parafrase dari tradisi etika teknologi

Empat belas bab sebelumnya bergerak terutama dalam register akademis dan kebijakan menganalisis teknologi, memetakan regulasi, menguraikan kerangka HAM, dan mengusulkan reformasi kelembagaan. Bab penutup ini sengaja bergerak ke register yang berbeda: yang lebih dekat ke tanah, lebih dekat ke individu yang hidup dalam dunia di mana deepfake sudah bukan ancaman hipotetis melainkan realitas yang sudah dirasakan oleh sejumlah orang Indonesia.

Tapi pergeseran register ini bukan berarti penurunan kedalaman. Justru sebaliknya: literasi digital, edukasi hukum, dan perlindungan masyarakat adalah dimensi dari respons terhadap deepfake yang paling sulit dibangun dan yang paling tahan lama dampaknya. Regulasi bisa diubah dalam hitungan bulan oleh parlemen; teknologi deteksi bisa usang dalam hitungan tahun; tapi pemahaman yang tertanam dalam individu dan komunitas tentang cara mengenali manipulasi, cara melindungi diri, dan cara meminta pertanggungjawaban itu adalah kapasitas yang bertahan jauh lebih lama.

Bab ini menyatukan dua jenis konten yang karakternya berbeda tapi yang saling melengkapi: analisis tentang peran pendidikan, masyarakat sipil, dan kerja sama internasional (yang mengikuti tradisi akademis bab-bab sebelumnya), dan panduan praktis yang bisa langsung digunakan oleh siapa pun yang terdampak deepfake di Indonesia (yang sengaja ditulis dengan bahasa yang lebih aksesibel). Keduanya perlu ada dalam sebuah buku yang serius tentang deepfake karena masalah ini terlalu nyata untuk hanya menjadi objek analisis tanpa memberikan respons yang bisa digunakan.

16.1 CARA MENDETEKSI DEEPFAKE: KEMAMPUAN DAN BATAS LITERASI VISUAL

Mengapa Deteksi Manual Semakin Sulit

Ada ekspektasi yang cukup umum bahwa dengan cukup latihan, seseorang bisa belajar mendeteksi deepfake secara visual dengan melihat dengan cermat dan mengetahui apa yang harus dicari. Ekspektasi ini ada benarnya untuk deepfake generasi lama, tapi semakin tidak akurat untuk deepfake yang dihasilkan oleh model-model terbaru. Bab 12 sudah mendiskusikan ini dari perspektif forensik; di sini yang relevan adalah implikasinya untuk literasi publik.

Penelitian tentang kemampuan manusia untuk mendeteksi deepfake konsisten menunjukkan hasil yang mengkhawatirkan: akurasi rata-rata manusia dalam membedakan video asli dari deepfake hanya sedikit di atas peluang acak sekitar 50 hingga 60 persen dalam kondisi eksperimental yang terkontrol, dan lebih rendah dalam kondisi nyata ketika konten dikonsumsi dengan cepat di feed media sosial. Lebih mengkhawatirkan lagi, kepercayaan diri seseorang terhadap penilaiannya tidak berkorelasi kuat dengan akurasinya orang yang paling yakin bahwa mereka bisa mendeteksi deepfake tidak selalu yang paling akurat.

Ini bukan alasan untuk menyerah pada literasi deteksi. Ia adalah alasan untuk bersikap realistis: tujuan literasi visual bukan membuat semua orang menjadi detektor deepfake yang sempurna itu tidak mungkin. Tujuannya adalah membangun kebiasaan skeptisisme yang sehat, kemampuan untuk mengenali tanda-tanda peringatan yang menunjukkan bahwa konten perlu diverifikasi lebih lanjut, dan pengetahuan tentang alat apa yang tersedia untuk verifikasi.

Indikator Visual yang Masih Relevan

Meski tidak ada indikator yang sempurna dan tidak ada yang berlaku untuk semua jenis deepfake, beberapa tanda peringatan visual masih cukup konsisten untuk menjadi bagian dari literasi dasar. Yang penting untuk disampaikan dalam konteks pendidikan adalah bahwa indikator-indikator ini bersifat indikatif, bukan konklusif kehadirannya menyarankan pemeriksaan lebih lanjut, bukan vonis.

Area transisi wajah dan latar belakang adalah titik di mana model deepfake sering paling kesulitan: batas antara wajah yang disisipkan dengan rambut, leher, atau latar belakang kadang tampak blur, berwarna tidak konsisten, atau memiliki artifak visual yang tampak seperti "halo" tipis di sekeliling wajah. Ini terutama terlihat ketika orang bergerak atau ketika ada cahaya yang tidak seragam.

Inkonsistensi kedipan mata dan ekspresi wajah bagian bawah adalah artefak yang cukup konsisten pada deepfake generasi lama. Model awal jarang menghasilkan kedipan yang natural karena dataset pelatihan lebih banyak mengandung foto dengan mata terbuka. Model yang lebih baru sudah jauh lebih baik dalam hal ini, tapi dalam kualitas video yang rendah, inkonsistensi kecil masih kadang terlihat.

Sinkronisasi bibir dan audio yang tidak sempurna adalah indikator yang lebih mudah dideteksi telinga daripada mata: ketika gerakan bibir tidak sepenuhnya sinkron dengan suara, atau ketika kualitas audio tampak berbeda dari kualitas video, ini adalah tanda peringatan. Deepfake audio-video yang dibuat secara terpisah kemudian digabungkan sering memiliki inkonsistensi semacam ini.

Gerakan tidak natural pada rambut, telinga, dan leher adalah area yang masih menjadi tantangan bagi banyak model generatif: rambut yang tampak seperti dirender secara terpisah dari kepala, telinga yang sedikit tidak simetris atau yang tampak "ditempelkan", dan gerakan leher yang tidak sepenuhnya mengikuti fisika kepala manusia.

Satu kebiasaan yang paling berguna untuk dikembangkan bukan kemampuan mendeteksi artefak spesifik tapi kebiasaan untuk memperlambat konsumsi konten yang menimbulkan reaksi emosional yang kuat. Konten yang dirancang untuk menimbulkan kemarahan, ketakutan, atau sensasi mendalam adalah konten yang paling mungkin disebarkan tanpa verifikasi dan yang paling mungkin adalah manipulasi. Jeda sebelum berbagi adalah kebiasaan yang sederhana tapi sangat efektif.

Alat Bantu Teknologi untuk Verifikasi

Di luar deteksi visual, ada sejumlah alat berbasis teknologi yang bisa digunakan oleh masyarakat umum bukan hanya oleh para ahli forensik untuk memverifikasi konten yang meragukan. Penting untuk menyampaikan ini dengan ekspektasi yang realistis: tidak ada alat yang memberikan kepastian absolut, dan banyak alat ini memerlukan koneksi internet dan kemampuan teknis dasar.

Pencarian gambar terbalik (*reverse image search*) menggunakan Google Images, TinEye, atau Yandex Images adalah langkah pertama yang paling mudah untuk gambar yang meragukan. Jika gambar yang diklaim sebagai foto terbaru dari seseorang ternyata sudah muncul di internet bertahun-tahun lalu dalam konteks yang berbeda, itu adalah tanda peringatan yang jelas. Untuk video, mengambil screenshot dari frame-frame tertentu kemudian melakukan pencarian gambar terbalik dari screenshot itu adalah pendekatan yang bisa memberikan petunjuk.

Platform deteksi deepfake yang tersedia publik seperti Deepware Scanner, Microsoft Video Authenticator (dalam versi yang tersedia publik), dan beberapa alat berbasis browser lainnya memungkinkan pengguna mengunggah video untuk dianalisis. Hasilnya berupa skor probabilitas bukan kepastian dan interpretasinya memerlukan pemahaman tentang keterbatasan yang sudah dibahas di Bab 12. Tapi sebagai alat penyaringan awal, mereka memberikan nilai tambah yang nyata dibanding tidak ada pemeriksaan sama sekali.

Pemeriksaan metadata menggunakan alat seperti Jeffrey's Exif Viewer atau ExifTool memungkinkan siapa pun melihat metadata yang tertanam dalam file gambar atau video termasuk informasi tentang kapan file dibuat, dengan perangkat apa, dan apakah sudah diedit dengan perangkat lunak tertentu. Ini memerlukan mengunduh file aslinya (bukan hanya melihatnya di platform), dan interpretasi hasilnya memerlukan sedikit pengetahuan teknis.

Yang perlu disampaikan dengan jujur dalam program literasi digital adalah bahwa tidak ada alat yang sempurna, bahwa deepfake yang canggih bisa melewati semua alat yang ada, dan bahwa yang paling penting bukan kemampuan untuk membuktikan bahwa konten adalah deepfake melainkan kemampuan untuk menahan diri dari menyebarkan konten yang belum diverifikasi. Skeptisisme yang produktif adalah aset yang lebih berharga dari kemampuan deteksi yang tidak sempurna.

Deepfake versus Hoax: Membedakan di Mata Hukum Indonesia

Sebelum membahas langkah-langkah praktis bagi korban, ada satu konsep yang perlu diperjelas karena sering menimbulkan kebingungan dalam diskusi publik dan bahkan dalam penanganan kasus: perbedaan antara deepfake dan hoax di mata hukum Indonesia.

Hoax adalah istilah yang merujuk pada konten berita bohong, disinformasi, atau narasi yang menyesatkan tanpa spesifikasi tentang cara pembuatannya. Medianya bisa berupa teks, foto yang dikontekstualisasikan secara salah, video yang dipotong, atau kombinasi keduanya. Deepfake, sebaliknya, adalah teknik pembuatan konten menggunakan AI untuk menghasilkan representasi sintetis yang hiper-realistis. Dengan kata lain: deepfake adalah alat atau metode, sementara hoax adalah hasil. Semua deepfake yang menyesatkan bisa dikategorikan sebagai hoax, tapi tidak semua hoax dibuat dengan deepfake.

Perbedaan ini penting secara hukum karena pasal yang dikenakan berbeda. Untuk hoax, instrumen utamanya adalah Pasal 28 ayat (3) UU ITE No. 1/2024 tentang penyebaran informasi palsu yang menimbulkan keresahan atau kerugian—ancaman maksimum enam tahun penjara dan denda Rp1 miliar. Untuk deepfake yang melibatkan penggunaan wajah atau suara seseorang tanpa izin, ada lapisan tambahan dari UU PDP (Undang-Undang Perlindungan Data Pribadi) No. 27/2022, khususnya Pasal 58 tentang pemrosesan data biometrik tanpa izin—ancaman enam tahun penjara dan denda Rp6 miliar. Ini berarti pelaku deepfake yang merugikan orang lain bisa menghadapi dua lapisan tuntutan secara bersamaan.

Perlu dicatat dengan jujur bahwa per 2025–2026, UU ITE dan KUHP yang berlaku belum memiliki pasal yang secara spesifik menyebut "deepfake" sebagai kategori tersendiri. Penegak hukum menggunakan kombinasi pasal-pasal yang ada UU ITE, UU PDP, UU TPKS untuk kasus seksual, dan KUHP untuk menjangkau tindakan yang melibatkan deepfake. Ini adalah celah yang diidentifikasi dalam Bab 15 sebagai salah satu alasan mengapa regulasi spesifik deepfake diperlukan.

16.2 PANDUAN PRAKTIS BAGI KORBAN DEEPFAKE DI INDONESIA

Langkah Pertama yang Kritis: Jangan Hapus, Arsipkan

Ketika seseorang menemukan bahwa dirinya menjadi korban deepfake baik itu video seksual yang menggunakan wajahnya, rekaman suara palsu yang mengatasnamakan dirinya, atau konten yang merusak reputasinya respons instingtif pertama sering kali adalah ingin segera menghapus semua bukti dari pandangan. Ini adalah respons yang sangat manusiawi tapi yang bisa merusak proses hukum berikutnya.

Prinsip yang paling penting untuk dipahami: jangan hapus bukti sebelum mengarsipkannya secara lengkap. Dalam sistem hukum yang mengandalkan bukti digital, konten yang sudah dihapus sangat sulit dipulihkan dan bisa membuat laporan tidak bisa diproses. Langkah pertama adalah mendokumentasikan semua yang bisa didokumentasikan: URL lengkap dari halaman tempat

konten berada (bukan hanya nama platformnya), nama akun atau profil yang menyebarkan konten, tanggal dan waktu penemuan, dan jumlah tayangan atau interaksi jika terlihat.

Cara terbaik untuk mengarsipkan adalah dengan merekam layar (*screen recording*) yang secara bersamaan menampilkan URL di address bar browser, konten yang bersangkutan, dan tanggal/waktu dari sistem. Rekaman ini lebih kuat sebagai bukti daripada screenshot biasa karena memperlihatkan konteks yang lebih lengkap. Jika memungkinkan, gunakan lebih dari satu perangkat untuk merekam rekaman dari perangkat berbeda memberikan redundansi bukti yang berharga.

Daftar Bukti yang Harus Disiapkan sebelum Melapor:

1. Rekaman layar (*screen recording*) yang menampilkan URL + konten + tanggal/waktu
2. Screenshot dengan metadata timestamp yang terlihat
3. Link/URL lengkap dari setiap platform tempat konten beredar
4. Nama akun, username, atau nomor telepon penyebar yang dapat diidentifikasi
5. Kronologi singkat: kapan pertama kali menemukan, dari siapa mendapat informasi
6. Surat pernyataan bermaterai yang menyatakan bahwa konten tersebut bukan dokumentasi nyata dan bahwa Anda tidak memberikan persetujuan
7. Fotokopi KTP atau identitas resmi
8. Jika ada: bukti percakapan dengan penyebar atau ancaman yang menyertai konten

Jalur Pelaporan Pertama: Bareskrim Siber

Jalur hukum pidana di Indonesia untuk kasus deepfake dimulai dari pelaporan kepada Kepolisian. Untuk kasus yang melibatkan konten digital, unit yang paling relevan adalah Direktorat Tindak Pidana Siber (Dittipidhsiber) Bareskrim Polri atau unit Kriminal Khusus (Krimsus) di tingkat Polda untuk daerah yang tidak berada di Jakarta.

Ada dua jalur pelaporan yang tersedia. Untuk jalur daring, Polri menyediakan portal Patrolisiber di alamat *https://patrolisiber.id* yang memungkinkan pelaporan awal secara online, serta alamat email *cybercrime@polri.go.id* untuk laporan yang disertai lampiran bukti. Pelaporan daring biasanya digunakan sebagai langkah awal, dan pelapor kemudian akan diminta untuk datang secara fisik untuk melengkapi proses pelaporan formal.

Untuk jalur langsung, korban dapat datang ke kantor Dittipidhsiber Bareskrim Polri di Jakarta atau ke unit Krimsus di Polda wilayah masing-masing. Datang dengan membawa seluruh bukti yang sudah diarsipkan, identitas diri, dan sebaiknya ditemani oleh pendamping—baik anggota keluarga, teman, atau pendamping dari lembaga masyarakat sipil yang familiar dengan kasus kekerasan digital.

Dasar hukum yang biasa digunakan dalam pelaporan kasus deepfake adalah UU ITE Pasal 27 ayat (1) untuk konten yang melanggar kesusilaan, Pasal 27 ayat (3) untuk pencemaran nama baik, Pasal 28 ayat (3) untuk penyebaran informasi palsu, dan UU PDP Pasal 58 untuk pemrosesan data biometrik tanpa izin. Untuk kasus deepfake seksual, tambahkan UU TPKS (Tindak Pidana

Kekerasan Seksual) No. 12/2022. Menyebutkan dasar hukum yang relevan saat melapor membantu penyidik mengklasifikasikan kasus dengan tepat dari awal.

Satu hal yang penting untuk diingat: minta surat tanda terima laporan (STPL) setelah pelaporan diterima. STPL adalah dokumen resmi yang membuktikan bahwa laporan sudah diterima dan yang memberikan nomor referensi untuk menindaklanjuti perkembangan kasus. Tanpa STPL, tidak ada rekam jejak formal bahwa laporan sudah masuk.

Jalur Pelaporan Kedua: Komdigi untuk Penghapusan Konten

Proses hukum pidana membutuhkan waktu—kadang berbulan-bulan sebelum ada tindakan konkret. Untuk mendapatkan penghapusan konten yang lebih cepat, jalur yang lebih langsung adalah melalui Kementerian Komunikasi dan Digital (Komdigi, yang sebelumnya bernama Kominfo) melalui kanal aduan konten negatif.

Komdigi menyediakan beberapa kanal aduan: portal web di *https://aduankonten.id* , dan layanan WhatsApp di *0811-9224-545*. Dalam pengaduan, sertakan URL lengkap konten yang ingin dihapus atau diblokir, penjelasan singkat tentang mengapa konten itu merugikan dan melanggar ketentuan yang berlaku, serta bukti bahwa konten tersebut menyangkut diri pelapor. Untuk konten yang dikategorikan sebagai mendesak terutama konten seksual yang melibatkan wajah nyata proses penanganan bisa berlangsung dalam 1x24 jam, meski tidak semua kasus mendapatkan respons secepat itu. Perlu dicatat bahwa penghapusan atau pemblokiran oleh Komdigi tidak menggantikan proses hukum pidana, ini adalah tindakan administratif yang mencegah konten terus beredar sementara proses hukum berjalan. Keduanya perlu dijalankan secara paralel, bukan berurutan.

Jalur Pelaporan Ketiga: Platform Tempat Konten Beredar

Secara paralel dengan pelaporan ke polisi dan Komdigi, korban harus segera melaporkan konten ke platform tempat deepfake beredar menggunakan mekanisme pelaporan yang disediakan platform. Semua platform besar yang beroperasi di Indonesia TikTok, Instagram, Facebook, X (Twitter), YouTube memiliki fitur pelaporan konten yang bisa diakses dari konten yang bersangkutan.

Untuk platform-platform ini, pilih kategori pelaporan yang paling sesuai dengan sifat konten: "Peniruan identitas" atau "Impersonation" untuk deepfake yang mengatasnamakan seseorang, "Konten seksual non-konsensual" atau "Non-consensual intimate images" untuk deepfake seksual, "Pencemaran nama baik" untuk konten yang merusak reputasi, atau "Konten palsu" untuk deepfake yang menyebarkan informasi menyesatkan. Menggunakan kategori yang tepat meningkatkan kemungkinan laporan ditangani oleh tim yang tepat.

Untuk Telegram dan WhatsApp yang sering digunakan untuk menyebarkan konten dalam grup tertutup, mekanismenya sedikit berbeda. Di Telegram, laporan bisa dikirim ke @notoscam atau langsung melalui tombol lapor dalam aplikasi; untuk kasus yang lebih serius, email ke abuse@telegram.org dengan menyertakan detail kasus. Untuk WhatsApp, laporan bisa diajukan melalui fitur lapor dalam aplikasi atau melalui email ke *support@whatsapp.com*.

Jangan bergantung hanya pada satu platform untuk menghapus konten—deepfake yang sudah beredar sering sudah diunduh dan diunggah ulang ke berbagai platform sekaligus. Pantau secara aktif apakah konten muncul di platform lain, dan ulangi proses pelaporan untuk setiap kemunculan baru.

Jalur Khusus untuk Deepfake Seksual

Kasus deepfake yang bersifat pornografi atau seksual memerlukan penanganan yang berbeda dari kasus deepfake umum, karena ia masuk dalam kategori kekerasan seksual yang dilindungi oleh UU TPKS No. 12/2022—sebuah undang-undang yang memberikan perlindungan tambahan bagi korban, termasuk hak atas pendampingan psikologis, hak atas kerahasiaan identitas, dan mekanisme perlindungan terhadap reviktimisasi.

Untuk kasus jenis ini, selain melapor ke Dittipidhsiber, korban juga sebaiknya menghubungi Unit Pelayanan Perempuan dan Anak (Unit PPA) di Polres atau Polda setempat—unit ini memiliki personel yang terlatih untuk menangani korban kekerasan seksual dan yang familiar dengan ketentuan UU TPKS. Proses pelaporan di Unit PPA umumnya lebih ramah korban dibanding pelaporan di unit umum.

Kontak Lembaga Pendukung Korban Kekerasan Digital:
Komnas Perempuan: pengaduan@komnasperempuan.go.id | Telepon: 021-3903963
KPAI (untuk korban anak): pengaduan@kpai.go.id | Telepon: 021-3190-5456
LBH APIK Jakarta: apik@cbn.net.id | untuk pendampingan hukum perempuan
Yayasan Pulih: yayasanpulih@gmail.com | untuk dukungan psikologis
IJRS (Indonesia Judicial Research Society): untuk riset dan advokasi hukum digital

16.3 PERAN PERGURUAN TINGGI DALAM RISET DAN ADVOKASI KEBIJAKAN AI

Perguruan Tinggi sebagai Aktor Kebijakan

Dalam diskursus tentang regulasi teknologi, perguruan tinggi sering ditempatkan hanya sebagai produsen pengetahuan akademis menghasilkan penelitian yang kemudian dibaca oleh pembuat kebijakan yang kemudian mengambil keputusan. Model linier ini tidak akurat secara empiris dan tidak cukup ambisius secara normatif. Perguruan tinggi, terutama di negara berkembang di mana lembaga think tank independen masih relatif lemah, memiliki kapasitas dan legitimasi untuk berperan jauh lebih aktif dalam pembentukan kebijakan.

Dalam konteks regulasi deepfake, peran perguruan tinggi yang efektif mencakup setidaknya empat dimensi: produksi pengetahuan yang relevan dan aksesibel bagi pembuat kebijakan, pembentukan keahlian praktis yang dibutuhkan sistem hukum (termasuk ahli forensik digital yang bisa bersaksi di pengadilan), advokasi berbasis bukti yang membawa temuan penelitian ke dalam proses kebijakan, dan pendidikan publik yang mengubah pengetahuan akademis menjadi literasi yang bisa digunakan oleh masyarakat luas.

Agenda Riset yang Mendesak

Ada beberapa area riset tentang deepfake yang sangat mendesak untuk dikembangkan dalam konteks Indonesia mendesak karena tanpa data dan analisis yang memadai, pembuat kebijakan terpaksa membuat keputusan berdasarkan asumsi atau berdasarkan data dari konteks yang sangat berbeda.

Pemetaan skala dan dampak deepfake di Indonesia adalah kebutuhan riset yang paling mendasar. Seberapa luas masalah deepfake di Indonesia? Kelompok mana yang paling terdampak? Jenis deepfake apa yang paling umum seksual, politik, penipuan finansial, atau perundungan? Pertanyaan-pertanyaan ini belum memiliki jawaban yang berbasis data yang komprehensif, dan tanpa jawaban itu, respons kebijakan yang proporsioanl sulit dirancang.

Evaluasi efektivitas hukum yang ada adalah riset yang tidak kurang pentingnya. Kasus-kasus deepfake yang sudah dilaporkan ke polisi: berapa yang berhasil dituntut? Berapa yang gagal di tahap penyidikan? Apa hambatan utamanya pembuktian, identifikasi pelaku, atau ketidakjelasan pasal? Data tentang ini sangat diperlukan untuk mengevaluasi apakah reformasi regulasi yang diusulkan di Bab 15 menyasar hambatan yang tepat.

Riset tentang dampak psikologis dan sosial pada korban deepfake di konteks Indonesia mempertimbangkan dinamika budaya spesifik yang bisa mempengaruhi bagaimana korban merespons, seberapa besar hambatan mereka untuk melapor, dan dampak jangka panjang terhadap kesehatan mental dan partisipasi sosial. Riset semacam ini belum ada dalam jumlah yang memadai untuk Indonesia.

Riset tentang persepsi publik dan efektivitas literasi digital tentang deepfake apakah orang Indonesia sudah mengetahui apa itu deepfake, seberapa mampu mereka mengenalinya, dan bagaimana intervensi literasi yang berbeda mempengaruhi kemampuan itu. Tanpa riset ini, program literasi digital dirancang berdasarkan asumsi tentang apa yang tidak diketahui masyarakat asumsi yang mungkin tidak akurat.

Klinik Hukum dan Bantuan Hukum Digital

Salah satu kontribusi paling konkret yang bisa diberikan perguruan tinggi hukum adalah melalui klinik hukum yang secara spesifik menangani kasus-kasus kekerasan digital, termasuk deepfake. Model klinik hukum di mana mahasiswa hukum, di bawah supervisi dosen, memberikan bantuan hukum nyata kepada klien nyata sudah ada di beberapa fakultas hukum Indonesia. Tapi klinik yang secara spesifik berfokus pada kekerasan digital dan yang memiliki kapasitas teknis untuk menangani bukti digital masih sangat jarang.

Mengembangkan klinik hukum digital yang terintegrasi menggabungkan keahlian hukum dengan keahlian teknis dari fakultas ilmu komputer atau teknik informatika adalah model yang menjanjikan dan yang sudah diujicobakan di beberapa universitas di luar negeri. Mahasiswa dari dua disiplin yang berbeda bekerja bersama dalam kasus nyata, belajar dari situasi yang tidak sempurna dan dari kompleksitas yang tidak bisa diajarkan hanya melalui kuliah.

16.4 KERJA SAMA INTERNASIONAL: ASEAN, INTERPOL, DAN UNESCO DALAM TATA KELOLA AI

Mengapa Regulasi Nasional Tidak Cukup Sendiri

Deepfake adalah masalah yang secara fundamental bersifat lintas batas. Konten bisa dibuat di satu negara, disebarkan melalui platform yang berkantor di negara lain, menargetkan korban di negara ketiga, dan menggunakan model AI yang dikembangkan di negara keempat. Tidak ada satu yurisdiksi nasional yang bisa, sendiri, mengatasi masalah ini secara komprehensif. Kerja sama internasional bukan pilihan tambahan—ia adalah prasyarat untuk efektivitas.

Tapi kerja sama internasional memiliki banyak bentuk dan tingkatan, dan tidak semuanya sama efektifnya. Ada perbedaan yang penting antara kerja sama yang bersifat deklaratif (negara-negara mengeluarkan pernyataan bersama tentang pentingnya regulasi AI yang bertanggung jawab) dan kerja sama yang bersifat operasional (negara-negara berbagi kapasitas teknis, data, dan mekanisme penegakan hukum lintas batas). Yang pertama lebih mudah dicapai tapi dampaknya terbatas; yang kedua lebih sulit tapi jauh lebih bermakna.

ASEAN: Potensi yang Belum Dioptimalkan

ASEAN sebagai forum regional memiliki beberapa mekanisme yang relevan untuk tata kelola deepfake. ASEAN Digital Masterplan 2025 menetapkan visi untuk transformasi digital kawasan, dan beberapa prinsip di dalamnya relevan untuk regulasi AI. ASEAN *Working Committee on E-Commerce* (WCE) menangani isu-isu yang bersinggungan dengan regulasi platform digital. Dan AICHR (Komisi Antarpemerintah ASEAN untuk HAM) secara teoritis bisa mengangkat isu kekerasan digital berbasis gender termasuk deepfake.

Tapi dalam praktik, ASEAN dalam konteks regulasi teknologi masih lebih sering menghasilkan pernyataan prinsip daripada komitmen operasional yang mengikat. Prinsip konsensus dan non-intervensi yang menjadi fondasi ASEAN membuat standardisasi regulasi yang substantif sangat sulit dicapai karena setiap negara anggota bisa menolak standar yang dianggap mencampuri kedaulatan regulasinya.

Peluang yang lebih menjanjikan mungkin ada di tingkat sub-regional atau bilateral. Kerja sama teknis antara Indonesia, Malaysia, Filipina, dan Singapura negara-negara dengan ekosistem digital yang lebih berkembang di kawasan dalam hal forensik digital, pelatihan penegak hukum, dan berbagi data intelijen tentang jaringan deepfake berbahaya adalah langkah yang lebih konkret dan lebih mudah dicapai dari standardisasi regulasi ASEAN secara keseluruhan.

INTERPOL: Kapasitas Operasional yang Sudah Ada

INTERPOL melalui Digital Crime Centre (IGCI) di Singapura sudah memiliki kapasitas operasional yang relevan untuk kasus deepfake yang melibatkan dimensi lintas batas termasuk identifikasi jaringan pelaku yang beroperasi dari berbagai yurisdiksi, koordinasi penyelidikan lintas negara, dan dukungan teknis untuk laboratorium forensik digital di negara-negara anggota.

Indonesia sebagai anggota INTERPOL sudah memiliki akses ke sumber daya ini, tapi pemanfaatannya untuk kasus-kasus deepfake spesifik masih belum optimal. Salah satu

hambatannya adalah kapasitas internal Polri untuk mengidentifikasi kasus-kasus yang memerlukan koordinasi INTERPOL dan untuk menyiapkan permintaan kerja sama yang memenuhi standar yang diperlukan. Memperkuat kapasitas ini melalui pelatihan personel dan pengembangan prosedur standar untuk kasus lintas batas adalah investasi yang menghasilkan manfaat yang melampaui kasus deepfake saja.

UNESCO dan Kerangka Etika AI Global

Recommendation on the Ethics of Artificial Intelligence yang diadopsi oleh UNESCO pada November 2021 adalah dokumen pertama yang memberikan kerangka etika AI yang disetujui secara global disetujui oleh 193 negara anggota, termasuk Indonesia. Dokumen ini menetapkan nilai-nilai dan prinsip-prinsip yang seharusnya memandu pengembangan dan penggunaan AI: proporsionalitas, keamanan, keadilan, keberlanjutan, privasi, transparansi, dan lain-lain.

Dalam konteks deepfake, Rekomendasi UNESCO ini memiliki beberapa relevansi langsung. Pertama, ia memberikan legitimasi normatif internasional bagi regulasi deepfake yang berbasis nilai menunjukkan bahwa perlindungan terhadap manipulasi identitas digital bukan hanya kepentingan nasional tapi komitmen nilai yang sudah disetujui secara global. Kedua, ia mewajibkan negara-negara anggota untuk mengembangkan kebijakan AI yang konsisten dengan prinsip-prinsipnya sebuah kewajiban yang, meski tidak mengikat secara hukum keras, memberikan dasar untuk akuntabilitas dalam forum internasional.

Yang perlu dicatat adalah bahwa Rekomendasi UNESCO bersifat soft law, dan tidak memiliki mekanisme penegakan yang mengikat. Nilainya ada pada kekuatan normatif dan pada kerangka referensi yang diberikannya untuk dialog kebijakan, bukan pada sanksi atas ketidakpatuhan. Bagi Indonesia, memanfaatkan Rekomendasi UNESCO sebagai referensi dalam penyusunan regulasi AI nasional adalah langkah yang memberikan legitimasi internasional tanpa memerlukan proses ratifikasi yang panjang.

16.5 ETIKA AI DAN TANGGUNG JAWAB MORAL DI ERA SINTETIS

Mengapa Etika Tidak Bisa Diserahkan Sepenuhnya kepada Hukum

Sepanjang buku ini, fokusnya telah banyak tertuju pada regulasi hukum definisi, kriminalisasi, kewajiban platform, kelembagaan. Ini adalah fokus yang tepat untuk tujuan kebijakan yang ingin dicapai. Tapi ada dimensi yang tidak bisa sepenuhnya ditangkap oleh hukum, dan yang tanpanya hukum sendiri tidak akan cukup efektif: dimensi etika.

Hukum menetapkan batas minimum apa yang dilarang dan apa konsekuensinya. Etika menetapkan standar yang lebih tinggi apa yang seharusnya dilakukan bahkan ketika tidak ada yang melarang melakukan sebaliknya. Seseorang bisa membuat deepfake satire yang secara hukum dilindungi tapi yang secara etis masih bisa dipertanyakan. Platform bisa mematuhi semua kewajiban hukum yang berlaku tapi masih gagal dalam tanggung jawab moral yang lebih luas terhadap penggunanya. Dan pengembang AI bisa mematuhi semua regulasi yang ada sambil tetap

memilih untuk tidak mengimplementasikan safe guard yang mereka tahu akan mengurangi risiko penyalahgunaan.

Di era di mana teknologi bergerak lebih cepat dari regulasi kondisi yang tampaknya struktural, bukan sementara etika adalah yang mengisi celah antara apa yang diizinkan secara hukum dan apa yang seharusnya dilakukan. Pengembang yang secara etis bertanggung jawab tidak menunggu regulasi untuk memasang safe guard; peneliti yang bertanggung jawab tidak mempublikasikan model yang mereka tahu akan disalahgunakan hanya karena tidak ada larangan hukum; platform yang bertanggung jawab tidak memaksimalkan engagement di atas keselamatan pengguna hanya karena bisa.

Tanggung Jawab Berlapis: Pengembang, Platform, Pengguna

Etika AI dalam konteks deepfake adalah tanggung jawab yang berlapis dan yang tidak bisa sepenuhnya dibebankan ke satu pihak saja. Pengembang model AI baik yang bekerja di perusahaan komersial maupun di institusi akademis memiliki tanggung jawab untuk melakukan apa yang dalam diskursus etika AI disebut sebagai anticipatory ethics: memikirkan kemungkinan penyalahgunaan sebelum teknologi dirilis, dan mengambil langkah-langkah proaktif untuk memitigasinya.

Ini bukan tanggung jawab yang mudah ditunaikan. Pengembang sering tidak bisa memprediksi semua kemungkinan penyalahgunaan, dan ada tekanan kompetitif yang kuat untuk merilis teknologi secepat mungkin. Tapi ini tidak berarti tanggung jawab itu tidak ada. Paralel dengan industri farmasi atau penerbangan yang juga menghadapi tekanan untuk bergerak cepat tapi yang memiliki standar keamanan yang ketat menunjukkan bahwa industri dengan potensi dampak besar dapat dan seharusnya mengembangkan budaya keselamatan yang lebih kuat.

Platform memiliki tanggung jawab yang berbeda tapi tidak kalah signifikan: tanggung jawab untuk tidak mendesain sistem yang secara struktural menguntungkan penyebaran konten berbahaya karena konten itu menghasilkan engagement yang tinggi. Algoritma rekomendasi yang memperkuat konten kontroversial karena konten itu memancing reaksi emosional kuat dan yang karena itu sering memperkuat deepfake yang provokatif adalah pilihan desain, bukan takdir teknologis. Memilih algoritma yang berbeda adalah pilihan etis yang bisa dilakukan tanpa menunggu regulasi.

Pengguna individual juga memiliki tanggung jawab moral tanggung jawab untuk tidak menjadi agen penyebaran konten yang belum diverifikasi, untuk tidak mengonsumsi atau menikmati konten deepfake yang jelas-jelas dibuat untuk merugikan orang lain, dan untuk menggunakan suara mereka sebagai konsumen untuk menekan platform dan pengembang menuju praktik yang lebih bertanggung jawab. Tanggung jawab individual ini lebih kecil dari tanggung jawab pengembang dan platform, tapi ia tidak nol terutama dalam skala agregat di mana pilihan jutaan pengguna membentuk insentif bagi seluruh ekosistem.

Deepfake dan Kebenaran: Krisis Epistemik yang Lebih Dalam

Di balik semua diskusi tentang regulasi, forensik, dan tanggung jawab, ada krisis yang lebih dalam yang perlu diakui: ancaman deepfake terhadap kemampuan kita untuk membedakan yang nyata dari yang palsu dan konsekuensinya terhadap kepercayaan sosial yang menjadi fondasi kehidupan bersama.

Dalam jangka menengah, deepfake yang semakin canggih bisa menghasilkan apa yang kadang disebut sebagai "liar's dividend" manfaat yang tidak adil bagi orang-orang yang berbohong. Jika bukti video tidak lagi bisa dipercaya karena bisa saja deepfake, maka seseorang yang tertangkap kamera melakukan sesuatu yang memalukan atau ilegal bisa berargumen bahwa rekaman itu adalah deepfake dan argumen itu, tanpa alat forensik yang memadai dan tanpa literasi yang cukup di pengadilan, bisa berhasil. Dalam situasi semacam ini, siapa yang kena rekaman video mengalami penambahan beban pembuktian yang tidak adil, sementara pelaku mendapat ruang penolakan yang baru.

Merespons krisis epistemik ini memerlukan lebih dari regulasi dan memerlukan investasi yang serius dalam infrastruktur kepercayaan: standar keaslian konten yang diadopsi secara luas, kapasitas forensik yang bisa diakses oleh sistem hukum, dan literasi yang cukup di masyarakat untuk memahami bahwa tidak semua klaim "ini deepfake" adalah valid. Membangun infrastruktur kepercayaan ini adalah proyek jangka panjang yang tidak bisa diselesaikan oleh satu bab atau satu buku, tapi yang harus dimulai dengan mengakui skala dan urgensi tantangannya.

Rangkuman Bab

Bab ini dan buku ini berakhir dengan pengakuan yang mungkin terasa tidak nyaman: tidak ada solusi yang komplet dan memuaskan untuk tantangan deepfake. Tidak ada regulasi yang bisa sepenuhnya mencegah pembuatan konten berbahaya. Tidak ada teknologi yang bisa mendeteksi semua deepfake. Tidak ada sistem pendidikan yang bisa membuat semua orang kebal terhadap manipulasi. Dan tidak ada kerja sama internasional yang bisa sepenuhnya menghilangkan celah yurisdiksi yang dieksploitasi oleh pelaku kejahatan.

Tapi pengakuan ketidaklengkapan solusi ini bukan alasan untuk nihilisme kebijakan pandangan bahwa karena tidak ada solusi sempurna, tidak ada gunanya mencoba. Sebaliknya, ia adalah undangan untuk bersikap realistis tentang apa yang bisa dicapai: regulasi yang lebih jelas dan lebih tepat sasaran akan mengurangi celah penyalahgunaan tanpa mengeliminasinya sepenuhnya; kapasitas forensik yang lebih kuat akan meningkatkan kemungkinan pelaku dimintai pertanggungjawaban tanpa menjamin setiap kasus berhasil dituntut; literasi yang lebih luas akan mengurangi penyebaran deepfake berbahaya tanpa menghentikannya sepenuhnya.

Pengurangan bertahap dari kerusakan yang nyata bukan eliminasi sempurna dari risiko yang tidak mungkin dihilangkan adalah standar yang tepat untuk mengevaluasi keberhasilan respons terhadap deepfake. Dan dalam kerangka standar itu, ada banyak yang bisa dan harus dilakukan: oleh pembuat kebijakan yang bersedia belajar dari pengalaman yurisdiksi lain dan yang berkomitmen pada proses konsultasi yang inklusif; oleh peneliti yang menghasilkan pengetahuan

yang relevan dan yang bersedia terlibat dalam advokasi kebijakan; oleh praktisi hukum yang mengembangkan keahlian yang diperlukan dan yang menjadikan bantuan hukum bagi korban kekerasan digital sebagai bagian dari misi mereka; oleh pendidik yang mengintegrasikan literasi digital dan etika teknologi ke dalam kurikulum; dan oleh setiap individu yang memilih, sehari-hari, untuk tidak menjadi bagian dari rantai penyebaran konten yang belum diverifikasi. Era sintetis sudah tiba. Yang menentukan bukan apakah kita akan menghadapinya kita tidak punya pilihan untuk tidak menghadapinya melainkan dengan kejernihan apa, dengan kapasitas apa, dan dengan komitmen nilai apa kita melakukannya.

DAFTAR PUSTAKA

Al-Mihimdi, S. F. K. (2025). The Associated Challenges with the Legal Regulation of Deepfake" A comparative analytical legal study". *Twejer Journal*, *8*(4), 447-465.

Alshamsi, M. S. (2023). *The Future of Deep Fakes: Analyzing the Potential Future Consequences of the Widespread Use of Deepfakes on the Policing Sector*. Rochester Institute of Technology.

Apolo, Y., & Michael, K. (2024). Beyond a reasonable doubt? Audiovisual evidence, AI manipulation, deepfakes, and the law. *IEEE Transactions on Technology and Society*, *5*(2), 156-168.

Asadi, O. (2025). Exploring Current and Potential Solutions: The Rise of Deepfakes in Legislative, Legal, and Technological Arenas. *Berkeley Undergraduate Journal*, *39*(1).

Ayata, O. (2024). Artificial Realities: Mitigations against Deepfakes. *The Centre for International Governance Innovation (CIGI)*.

Ben-David, A. & Murie, H. (2020). Cheapfakes, Shallowfakes, and the Proliferation of Synthetic Media. Journal of Information Technology and Politics, 17(4), 1–15.

Bermudez, Y. A. B. (2025). Deepfakes and artificial intelligence in social engineering: Emerging threats in 21st-century cyberfraud. *IUSTA*, (63), 54-71.

Birrer, A., & Just, N. (2025). What we know and don't know about deepfakes: An investigation into the state of the research and regulatory landscape. *New media & society*, *27*(12), 6819-6838.

Busacca, A., & Monaca, M. A. (2023). Deepfake: Creation, purpose, risks. In *Innovations and economic and social changes due to artificial intelligence: the state of the art* (pp. 55-68). Cham: Springer Nature Switzerland.

Chawki, M. (2024). Navigating legal challenges of deepfakes in the American context: a call to action. *Cogent Engineering*, *11*(1), 2320971.

Chen, R., et al. (2021). Synthetic Data in Machine Learning for Medicine and Healthcare. Nature Biomedical Engineering, 5(6), 493–497.

Chesney, R. & Citron, D.K. (2019). Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. California Law Review, 107(6), 1753–1820.

Citron, D.K. (2022). The Fight for Privacy: Protecting Dignity, Identity, and Love in the Digital Age. New York: W.W. Norton & Company.

de Souza, R. R. M. (2026). Legal Remedies and Regulatory Frameworks to Combat AI-Driven Deepfakes. In *Mitigating the Risks of AI Deepfakes* (pp. 94-116). CRC Press.

Delfino, R. A. (2022). Deepfakes on trial: a call to expand the trial judge's gatekeeping role to protect legal proceedings from technological fakery. *Hastings LJ*, *74*, 293.

Fernandez, A. (2021). 'Deep fakes': disentangling terms in the proposed EU Artificial Intelligence Act. *UFITA Archiv für Medienrecht und Medienwissenschaft*, *85*(2), 392-433.

Gambín, Á. F., Yazidi, A., Vasilakos, A., Haugerud, H., & Djenouri, Y. (2024). Deepfakes: current and future trends. *Artificial Intelligence Review*, *57*(3), 64.

Goodfellow, I., et al. (2014). Generative Adversarial Nets. Advances in Neural Information Processing Systems, 27.

Grillo, G. (2025). *Deepfake and Generative AI: Legal Challenges and Technical Strategies for Detection and Prevention* (Doctoral dissertation, Politecnico di Torino).

Gunawan, I. J., & Janisriwati, S. (2023). Legal analysis on the use of deepfake technology: Threats to Indonesian banking institutions. *Law and Justice*, *8*(2), 192-210.

Haworth, K. (2022). The Ethics of Voice Cloning in Assistive Technology. AI and Society, 37(4), 1623–1631.

Ingimundarson, J. I. (2025). The silent threat: Technology facilitated sexual violence, AI deepfakes and the European Union's Artificial Intelligence act.

Judijanto, L., Utama, A. S., & Setiyawan, H. (2025). Implementation of ethical artificial intelligence law to prevent the use of AI in spreading false information (deepfake) in Indonesia. *The Easta Journal Law and Human Rights*, *3*(02), 101-109.

Kadri, T. E., & West, S. R. (2025). Deepfake Torts: Emerging Tort Frameworks in US Deepfake Regulation. *Journal of Tort Law*, *18*(2), 515-552.

Kalpokas, I., & Kalpokiene, J. (2022). *Deepfakes: a realistic assessment of potentials, risks, and policy regulation*. Springer Nature.

Kishwar, S. D., Tripathi, A., Khatoon, S., Poddar, D., & Khurana, B. (2025). Regulating deep fakes and synthetic media: Privacy, policy and global regulatory challenges. *Journal of Data Protection & Privacy*, *8*(1), 78-96.

Łabuz, M. (2023). Regulating deep fakes in the artificial intelligence act. *Applied Cybersecurity & Internet Governance*, *2*(1), 1-42.

Langa, J. (2021). Deepfakes, real consequences: Crafting legislation to combat threats posed by deepfakes. *BUL Rev.*, *101*, 761.

Lin, L. S. (2025). Examining the role of deepfake technology in organized fraud: Legal, security, and governance challenges. *Frontiers in Law*, *4*, 6-17.

Mekkawi, M. H. (2023). The challenges of digital evidence usage in deepfake crimes era. *Journal of Law and Emerging Technologies*, *3*(2), 176-232.

Meskys, E., Kalpokiene, J., Jurcys, P., & Liaudanskas, A. (2020). Regulating deep fakes: legal and ethical considerations. *Journal of Intellectual Property Law & Practice*, *15*(1), 24-31.

Micheal, D. (2025). Detecting digital threats in the age of ai: Deep learning approaches for deepfakes and intrusion detection in decentralized systems.

Muhammad, O. (2025). AI-Generated Deepfakes and the Crisis of Legal Authenticity: Reconstructing Evidentiary Standards in the Digital Age. *Sarhad Journal of Legal Studies*, *1*(1), 01-15.

Nandal, M. (2025, June). Mitigating Deepfake Threats to Privacy: Legal Frameworks and Technological Safeguards. In *Proceedings of the National Seminar on Enhancing Privacy Protection in the Digital Age: Legal Challenges & Innovations (NSEPPDA 2025)* (p. 225). Springer Nature.

Nnamdi, N., Oniyinde, O. A., & Abegunde, B. (2023). An appraisal of the implications of deep fakes: The need for urgent international legislations. *American Journal of Leadership and Governance*, *8*(1), 43-70.

Panwar, K. S., & Roy, N. D. (2024). Rising menace of deepfakes with the help of AI: Legal implications in India. *Indian Journal of Integrated Research in Law Volume IV Issue III| ISSN*, *2583*, 0538.

Paris, B. & Donovan, J. (2019). Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence. Data & Society Research Institute.

Parsan, S. J., & Kim, M. (2024). Unmasking DeepFakes A review of technology, regulation, challenges and policy implications. *Regulation, Challenges and Policy Implications (May 29, 2024)*.

Piers, M., & Osaer, H. C. (2025). When Seeing is No Longer Believing: Evidentiary Challenges of Deepfakes in Arbitration. *Dispute Resolution International*, (100).

Ramluckan, T. (2024, March). Deepfakes: The legal implications. In *International Conference on Cyber Warfare and Security* (Vol. 19, No. 1, pp. 282-288). Academic Conferences International Limited.

Ray, A. (2021). Disinformation, deepfakes and democracies: The need for legislative reform. *TheUNIVERSITY OF NEW SOUTH WALES LAW JOURNAL*, *44*(3), 983-1013.

Romero-Moreno, F. (2024). Deepfake Fraud Detection: Safeguarding Trust in Generative AI. *Available at SSRN 5031627*.

Roy, A., Chattopadhyay, S., & Chakrabarty, S. P. (2025, February). Deepfakes: Navigating Ethical Concerns and Legal Frameworks. In *International Ethical Hacking Conference* (pp. 337-355). Singapore: Springer Nature Singapore.

Schick, N. (2020). Synthetic: How AI Is Creating Narratives, Identity, and Reality. London: Hachette Books.

Sharma, A. K., & Sharma, R. (2024). Generative artificial intelligence and legal frameworks: Identifying challenges and proposing regulatory reforms. *Kutafin Law Review*, *11*(3), 415-451.

Singh, A., & Shanker, N. (2024). Redefining cybercrimes in light of artificial intelligence: Emerging threats and challenges. *International Journal of Innovations in Science, Engineering And Management*, 192-201.

Suwajanakorn, S., Seitz, S.M. & Kemelmacher-Shlizerman, I. (2017). Synthesizing Obama: Learning Lip Sync from Audio. ACM Transactions on Graphics (TOG), 36(4), 1–13.

Van der Sloot, B., & Wagensveld, Y. (2022). Deepfakes: regulatory challenges for the synthetic society. *Computer Law & Security Review*, *46*, 105716.

Zumarno, G. C., Lesmana, S. J., & Indayatun, R. (2026). Bridging Legal Gaps in AI-Generated Deepfake Pornography: A Comparative Approach to Privacy and Digital Ethics. *Jurnal Ius Constituendum*, *11*(2), 237-263.

Deepfake:

Rekayasa Konten Palsu, Hasil produk AI

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

Bio Data Penulis

Penulis lahir di Semarang pada tanggal 1 Maret 1983. Penulis menempuh pendidikan Sarjana Teknik Elektro di Universitas Kristen Satya Wacana (UKSW), lulus tahun 2004, kemudian tahun 2005 melanjutkan studi pada Magister Desain di Fakultas Seni Rupa dan Desain, Institut Teknologi Bandung (ITB), dan kemudian melanjutkan studi pada program studi Teknologi Multimedia di Swinburne University of Technology Australia. Penulis sejak tahun 2010, menjadi dosen pada program studi Desain Grafis Universitas Sains dan Teknologi Komputer (Universitas STEKOM), memiliki Jabatan Akademik Lektor Kepala 700. Penulis juga seorang wirausaha di bidang toko online yang berhasil di kota Semarang dan juga aktif sebagai freelancer dalam bidang fotografi, web design dan multimedia.

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 978-634-7695-23-9 (PDF)
9 786347 695239