BAB I

PENDAHULUAN

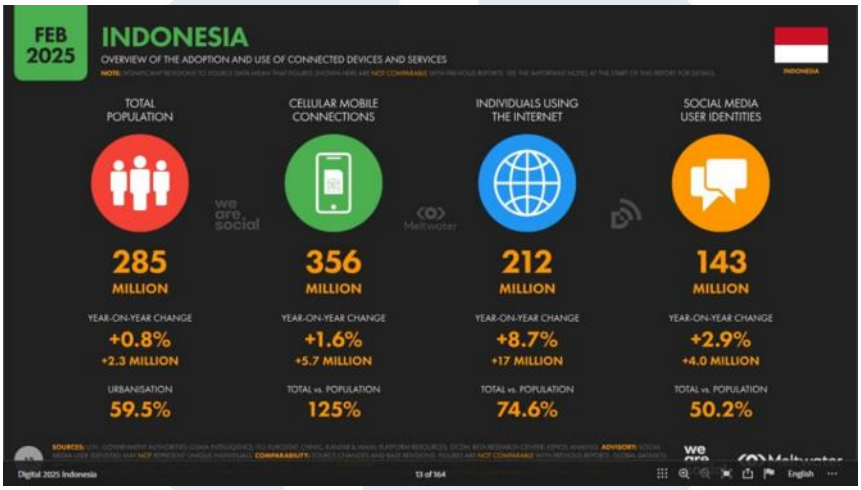
1.1 Latar Belakang Penelitian

Teknologi kecerdasan buatan (*artificial intelligence*) telah mengubah cara masyarakat memproduksi, mendistribusikan, dan memaknai informasi. Di satu sisi, teknologi ini mempermudah berbagai aktivitas mulai dari pencarian informasi hingga produksi konten digital. Di sisi lain, kemudahan yang sama membuka ruang penyalahgunaan yang semakin sulit dikenali. Salah satu bentuk penyalahgunaan yang kini menjadi perhatian global adalah *deepfake*, yaitu teknologi yang mampu menghasilkan atau memanipulasi foto, audio, maupun video sehingga tampak menyerupai individu tertentu secara realistis (Chesney & Citron, 2019; Verdoliva, 2020). Kemampuan teknologi ini menghasilkan kemiripan yang nyaris sempurna membuat batas antara representasi dan rekayasa menjadi semakin kabur, bahkan bagi mata yang terlatih sekalipun.

Persoalan menjadi lebih kompleks ketika manipulasi tersebut secara spesifik digunakan untuk meniru, memerankan, atau mengatasnamakan identitas seseorang yang sudah dikenal baik tokoh publik maupun individu biasa tanpa sepengetahuan dan persetujuan yang bersangkutan. Fenomena ini dikenal sebagai *deepfake* berbasis peniruan identitas (Chesney & Citron, 2019). Berbeda dengan *deepfake* generatif yang menciptakan wajah atau suara sepenuhnya sintetis dan tidak merujuk pada siapa pun, peniruan identitas memanfaatkan identitas individu yang benar-benar ada. Konsekuensinya bersifat personal dan melekat: yang dirugikan bukan entitas fiktif, melainkan orang yang wajah, suara, dan reputasinya dicatut. Bentuk manipulasi ini tidak hanya berisiko menyesatkan publik, tetapi juga merugikan individu yang identitasnya digunakan, baik dari sisi reputasi, privasi, maupun keamanan digital.

Kerentanan tersebut tumbuh di atas populasi digital Indonesia yang sangat besar. Jumlah pengguna internet di Indonesia terus meningkat dari tahun ke tahun, sebagaimana ditunjukkan pada Gambar 1.1. Semakin banyak individu yang

menyimpan jejak visual berupa foto dan video di ruang digital, semakin luas pula bahan mentah yang tersedia bagi pembuatan konten manipulatif. Dalam konteks ini, keberadaan di ruang digital tidak lagi netral; ia sekaligus menjadi bentuk keterpaparan.



Gambar 1.1 Jumlah Pengguna Internet di Indonesia

Sumber: diolah dari We Are Social (2025)

Fenomena *deepfake* mengalami peningkatan signifikan dalam beberapa tahun terakhir, baik secara global maupun di Indonesia. Berdasarkan laporan Sumsub (2023), jumlah kasus *deepfake* yang terdeteksi secara global meningkat lebih dari sepuluh kali lipat dibandingkan tahun sebelumnya, dengan kawasan Asia Pasifik mengalami peningkatan hingga 1.530 persen dalam kurun waktu satu tahun. Tren tersebut berlanjut hingga Identity Fraud Report 2025–2026 yang mencatat kategori sophisticated fraud kombinasi *deepfake*, identitas sintetis, dan rekayasa sosial meningkat 180 persen, dengan *deepfake* menyumbang sekitar 11 persen dari seluruh kasus penipuan identitas tingkat pertama (*first-party fraud*) sepanjang 2025. Di Indonesia, data Otoritas Jasa Keuangan bersama Indonesia Anti-Scam Centre menunjukkan lebih dari 274.000 laporan penipuan keuangan sepanjang akhir 2024 hingga 2025 dengan estimasi kerugian lebih dari Rp6 triliun, sebagian di antaranya memanfaatkan rekayasa wajah dan suara sintetis untuk meniru identitas individu tertentu.

Angka-angka tersebut memotret satu wajah dari penyalahgunaan *deepfake*, yaitu kerugian finansial yang dapat dihitung. Namun terdapat wajah lain yang bekerja di ranah jauh lebih personal dan jarang terwakili dalam statistik kerugian ekonomi, yaitu ketika wajah individu khususnya perempuan dimanipulasi ke dalam konten bermuatan seksual tanpa persetujuannya. Bentuk penyalahgunaan ini termasuk dalam kategori Kekerasan Berbasis Gender Online (KBGO), dan skalanya di Indonesia tidak kecil. Komnas Perempuan (2026) melalui Catatan Tahunan 2025 mencatat 376.529 kasus kekerasan berbasis gender terhadap perempuan sepanjang 2025, meningkat 14,07 persen dibandingkan tahun sebelumnya dan menjadi angka pengaduan tertinggi dalam satu dekade terakhir. Dari keseluruhan kasus di ranah publik, KBGO menempati posisi bentuk kekerasan yang paling dominan dengan 1.091 kasus, yang terdiri atas 977 kasus KBGO bermuatan seksual dan 114 kasus non-seksual. Data pendampingan dari SAFEnet (2025) memperkuat gambaran tersebut, dengan mencatat 1.698 kasus KBGO sepanjang sembilan bulan pertama 2025, atau rata-rata lebih dari enam kasus setiap hari.

Yang menjadikan data tersebut semakin relevan bagi penelitian ini adalah sebaran usia korbannya. Komnas Perempuan (2026) mencatat kelompok usia 18–24 tahun sebagai kelompok yang paling banyak mengalami kekerasan, sementara pemantauan SAFEnet (2025) juga menunjukkan mayoritas korban KBGO berada pada rentang usia 18–25 tahun. Rentang usia tersebut beririsan langsung dengan kelompok yang paling aktif memproduksi dan membagikan konten visual di media sosial, sebagaimana tergambar pada komposisi pengguna internet Indonesia menurut kelompok generasi pada Gambar 1.2. Dengan kata lain, kelompok yang paling banyak hadir di ruang digital adalah juga kelompok yang paling rentan mengalami kekerasan di dalamnya.

Gambar 1.2 Pengguna Internet Indonesia Berdasarkan Kelompok Generasi

Sumber: Indonesiabaik.id, diolah dari APJII (2024)

Ringkasan perbandingan kedua wajah penyalahgunaan *deepfake* tersebut disajikan pada Tabel 1.1 untuk memperjelas posisi fenomena yang menjadi fokus penelitian ini.

Tabel 1.1 Dua Bentuk Penyalahgunaan *Deepfake* Berbasis Peniruan Identitas

Aspek Pembeda	**Peniruan Identitas untuk Penipuan**	**Peniruan Identitas untuk Konten Seksual**
Sasaran utama	Tokoh publik, pejabat, figur otoritas, pimpinan perusahaan	Perempuan, termasuk individu biasa dan pembuat konten
Motif pelaku	Keuntungan finansial melalui penipuan	Pelecehan, penghinaan, dominasi, dan komodifikasi tubuh
Pihak yang dirugikan	Pihak ketiga yang tertipu; identitas yang dicatut sebagai alat	Individu yang identitasnya dicatut sebagai sasaran langsung
Bentuk kerugian	Kerugian materiil yang terukur dalam nominal	Kerugian psikologis, sosial, reputasional, dan profesional
Keterlihatan dalam data	Terekam dalam statistik kejahatan keuangan	Banyak yang tidak dilaporkan karena stigma
Sumber data rujukan	OJK dan IASC; Identity Fraud Report 2025–2026	Komnas Perempuan (2026); SAFEnet (2025)

Sumber: diolah peneliti dari berbagai sumber (2026)

Selain berdampak secara finansial dan institusional, penyalahgunaan *deepfake* berbasis peniruan identitas dirasakan pada level personal oleh individu yang identitasnya disalahgunakan. Bentuknya dapat berupa pencatutan identitas dalam konten menyesatkan, penyebaran konten tidak senonoh berbasis manipulasi wajah, hingga pemalsuan identitas untuk tujuan penipuan sosial. Meskipun tidak selalu menimbulkan kerugian finansial yang terukur, dampaknya bersifat psikologis dan sosial khususnya berkaitan dengan bagaimana korban memaknai, mengelola, dan mengomunikasikan pengalaman tersebut dalam interaksi sosialnya sehari-hari.

Kondisi ini semakin rumit ketika korban adalah individu yang kehadirannya di ruang digital merupakan bagian dari pekerjaannya, seperti content creator. Bagi mereka, wajah dan citra diri bukan sekadar atribut personal melainkan modal profesional. Ketersediaan foto wajah dalam jumlah besar di akun publik menjadikan mereka sasaran yang secara teknis paling mudah dijangkau, sebab teknologi face-swap berbasis kecerdasan buatan hanya memerlukan sejumlah kecil citra wajah untuk menghasilkan manipulasi yang meyakinkan. Ketika wajah tersebut dimanipulasi menjadi konten tidak pantas, korban menghadapi dilema ganda: menarik diri dari ruang digital berarti kehilangan sumber penghidupan, sementara tetap hadir berarti mempertahankan visibilitas yang justru menjadi pintu masuk kerentanan. Dilema inilah yang menjadikan keputusan untuk bercerita atau diam bukan sekadar persoalan psikologis, melainkan persoalan komunikasi yang penuh perhitungan.

Gambaran tersebut tercermin dalam observasi awal yang peneliti lakukan melalui wawancara pendahuluan terhadap berbagai individu yang mengalami *deepfake* berbasis peniruan identitas. Hasil wawancara pendahuluan menunjukkan bahwa selain memunculkan respons emosional berupa keterkejutan, rasa malu, dan kecemasan, korban juga menghadapi dilema dalam memutuskan kepada siapa, kapan, dan sejauh mana pengalaman tersebut dapat diceritakan. Korban cenderung membatasi diri dan hanya bersedia menceritakan pengalamannya kepada pihak-pihak tertentu yang dianggap dapat dipercaya, sementara terhadap lingkungan yang lebih luas ia memilih menahan informasi karena khawatir terhadap stigma dan

penilaian negatif. Kondisi ini menunjukkan bahwa menjadi korban *deepfake* bukan semata-mata persoalan teknologi, melainkan juga persoalan komunikasi khususnya terkait bagaimana individu mengelola batas privasi dan memutuskan proses pengungkapan diri (*self-disclosure*) atas pengalaman yang dialaminya.

Meskipun demikian, sudut pandang tersebut belum banyak diangkat dalam kajian akademik di Indonesia. Kajian mengenai *deepfake* di Indonesia sejauh ini didominasi perspektif hukum dan kebijakan, terutama yang membahas kekosongan pengaturan dalam Undang-Undang Informasi dan Transaksi Elektronik serta Undang-Undang Perlindungan Data Pribadi dalam melindungi korban. Kajian semacam ini penting, namun berhenti pada pertanyaan mengenai bagaimana negara seharusnya merespons, bukan bagaimana korban sendiri mengalami dan mengomunikasikan peristiwa yang menimpanya. Sementara itu, kajian komunikasi yang menggunakan teori Manajemen Privasi Komunikasi (Communication Privacy Management) di Indonesia telah cukup berkembang, tetapi konteks yang diteliti umumnya berupa pengungkapan status kesehatan (Bate & Amrullah, 2022), pengalaman pelecehan seksual siber (Saptyasari dkk., 2023), pengalaman kekerasan seksual konvensional (Lewi & Raras, 2024), maupun pengelolaan privasi dalam penggunaan fitur media sosial (Rahman & Wijaya, 2025). Belum ditemukan penelitian komunikasi di Indonesia yang secara khusus menempatkan korban *deepfake* sebagai subjek utama untuk memahami proses pengungkapan dirinya.

Pada tataran internasional, penelitian mengenai korban *deepfake* seksual dan penyebaran citra intim non-konsensual memang telah dilakukan secara kualitatif (Mclocklin dkk., 2024; Flynn dkk., 2025). Namun kajian-kajian tersebut sebagian besar berangkat dari disiplin kriminologi, viktimologi, dan psikologi, sehingga penekanannya jatuh pada dampak psikologis, hambatan mencari bantuan, serta rekomendasi kebijakan. Proses pengungkapan diri korban belum dibaca sebagai tindakan komunikasi yang memiliki aturan, dimensi, dan konsekuensi relasionalnya sendiri. Selain itu, konteks sosial-budaya Indonesia yang ditandai kuatnya stigma terhadap korban dan kecenderungan menyalahkan korban (*victim blaming*) belum terwakili dalam literatur tersebut. Dengan demikian terdapat celah yang jelas:

belum ada penelitian yang membaca pengungkapan diri korban *deepfake* sebagai proses manajemen privasi komunikasi dalam konteks Indonesia.

Celah tersebut penting untuk diisi karena keputusan korban untuk berbicara atau diam memiliki konsekuensi nyata yang melampaui ranah personal. Diamnya korban bukan hanya berarti seorang memendam luka, melainkan juga berarti kasus tidak terlaporkan, dukungan tidak diberikan, dan pola kejahatan tidak terbaca oleh publik maupun pembuat kebijakan. Selisih besar antara angka pengaduan yang tercatat dan perkiraan kasus yang sesungguhnya terjadi mengindikasikan bahwa persoalan utama bukan semata ketiadaan aturan, melainkan juga hambatan komunikasi yang membuat korban memilih tidak bersuara. Memahami logika di balik keputusan tersebut karenanya menjadi prasyarat bagi perancangan pendampingan korban yang tepat sasaran.

Berangkat dari persoalan tersebut, penelitian ini menggunakan teori Manajemen Privasi Komunikasi dari Sandra Petronio sebagai landasan teori. Pemilihan ini didasarkan pada kesesuaian antara inti persoalan yang diteliti dengan inti penjelasan teori. Di balik respons emosional yang muncul, korban senantiasa dihadapkan pada proses pengambilan keputusan mengenai kepada siapa, kapan, dan sejauh mana pengalaman tersebut boleh diceritakan sebuah proses pengelolaan batas privasi yang bersifat personal sekaligus situasional. Manajemen Privasi Komunikasi secara khusus menjelaskan bagaimana individu mengelola keputusan untuk membuka maupun menyembunyikan informasi privat melalui mekanisme aturan dan batas privasi (*privacy rules and boundaries*), kepemilikan bersama atas informasi privat (*co-ownership*), serta turbulensi batas (*boundary turbulence*) ketika informasi tersebut bergerak di luar kendali pemiliknya.

Sebagai landasan konsep, penelitian ini menggunakan konsep pengungkapan diri (*self-disclosure*) dari Joseph A. DeVito untuk mendeskripsikan dimensi dan kualitas keterbukaan korban. Keduanya berada pada tingkat abstraksi yang berbeda dan karena itu saling melengkapi alih-alih tumpang tindih: Manajemen Privasi Komunikasi menjelaskan mengapa dan bagaimana batas privasi dibentuk dan dikelola, sedangkan konsep DeVito membantu memerikan seperti apa wujud pengungkapan yang akhirnya dilakukan seberapa banyak, seberapa dalam, dengan

nada seperti apa, dan sejujur apa. Pembagian peran ini diuraikan lebih rinci pada Subbab 2.4.

Adapun dari sisi metode, mengingat pengalaman menjadi korban *deepfake* merupakan peristiwa yang sensitif, personal, dan sulit diakses secara luas, penelitian ini memerlukan pendekatan yang mampu menggali proses subjektif tersebut secara mendalam dan kontekstual, bukan sekadar mengukurnya secara umum. Penelitian ini menggunakan paradigma post-positivisme dengan pendekatan kualitatif deskriptif melalui metode studi kasus sebagaimana dikemukakan Robert K. Yin. Penelitian melibatkan satu partisipan yang dipilih secara purposif sebagai kasus tunggal (single case) yang bersifat unik dan mengungkap (revelatory), mengingat terbatasnya akses terhadap individu yang bersedia membagikan pengalamannya sebagai korban *deepfake* berbasis peniruan identitas. Pilihan desain kasus tunggal ini bukan semata konsekuensi keterbatasan akses, melainkan justru sesuai dengan karakteristik kasus yang diteliti, sebagaimana dijelaskan pada Subbab 3.3.

Berdasarkan seluruh uraian di atas, peneliti memandang perlu dilakukan penelitian berjudul "Manajemen Privasi Komunikasi dan Pengungkapan Diri Korban *Deepfake* Berbasis Peniruan Identitas (Studi Kasus pada Seorang Content Creator Perempuan di Media Sosial)".

1.2 Rumusan Masalah

Teknologi *deepfake* memungkinkan wajah seseorang dipindahkan ke dalam konten yang tidak pernah ia buat dan tidak pernah ia setujui. Ketika hal itu terjadi, korban kehilangan sesuatu yang dalam kondisi normal sepenuhnya berada dalam kendalinya, yaitu hak untuk menentukan bagaimana dirinya ditampilkan kepada orang lain. Terdapat ketidaksesuaian mendasar antara kondisi ideal dan kondisi aktual di sini: secara ideal, individu adalah pemilik tunggal atas informasi mengenai dirinya dan berhak menentukan siapa saja yang boleh mengetahuinya; secara aktual, korban *deepfake* justru harus menyusun aturan privasi atas informasi yang batasnya telah dilanggar terlebih dahulu oleh pihak lain. Persoalan komunikasi yang muncul kemudian bukanlah bagaimana korban menjaga rahasia, melainkan

bagaimana korban mengelola sesuatu yang sudah terlanjur bocor, sembari memutuskan kepada siapa ia bersedia menceritakan versinya sendiri.

Keputusan tersebut tidak diambil dalam ruang hampa. Korban berhadapan dengan stigma sosial, kemungkinan disalahkan, serta beban tambahan berupa keharusan meyakinkan orang lain bahwa konten tersebut palsu beban yang tidak dihadapi korban kekerasan berbasis citra dalam bentuk konvensional. Ketika korban sekaligus berprofesi sebagai content creator, perhitungan tersebut menjadi semakin berlapis karena menyangkut kelangsungan pekerjaannya. Sementara itu, literatur yang tersedia belum menjelaskan bagaimana proses pengelolaan batas privasi dan pengungkapan diri berlangsung pada situasi ketika pelanggaran batas terjadi lebih dahulu daripada pembentukan aturan. Berdasarkan uraian tersebut, rumusan masalah dalam penelitian ini adalah bagaimana pengalaman korban *deepfake* berbasis peniruan identitas dalam melakukan pengungkapan diri atas peristiwa yang dialaminya, ditinjau dari perspektif Manajemen Privasi Komunikasi dengan konsep pengungkapan diri DeVito sebagai landasan konsep.

1.3 Pertanyaan Penelitian

Rumusan masalah tersebut dielaborasi menjadi tiga pertanyaan penelitian berikut.

1. Bagaimana korban *deepfake* berbasis peniruan identitas membentuk dan menerapkan aturan serta batas privasi (*privacy rules and boundaries*) atas informasi mengenai peristiwa yang dialaminya?
2. Bagaimana proses pengambilan keputusan korban dalam melakukan pengungkapan diri kepada orang lain, ditinjau dari dimensi *self-disclosure* yang dikemukakan Joseph A. DeVito?
3. Bagaimana korban menghadapi dan mengelola turbulensi batas privasi (*boundary turbulence*) ketika informasi mengenai peristiwa tersebut tersebar kepada pihak lain di luar kendalinya?

1.4 Tujuan Penelitian

Berdasarkan rumusan dan pertanyaan penelitian tersebut, penelitian ini bertujuan untuk:

1. Memahami pembentukan dan penerapan aturan serta batas privasi korban *deepfake* berbasis peniruan identitas atas informasi mengenai peristiwa yang dialaminya, berdasarkan perspektif Manajemen Privasi Komunikasi;
2. Menganalisis proses pengambilan keputusan korban dalam melakukan pengungkapan diri kepada orang lain, ditinjau dari dimensi self-disclosure Joseph A. DeVito; dan
3. Menganalisis cara korban menghadapi dan mengelola turbulensi batas privasi ketika informasi mengenai peristiwa tersebut tersebar di luar kendalinya.

1.5 Kegunaan Penelitian

1.5.1 Kegunaan Akademis

Penelitian ini diharapkan memberikan kontribusi pada pengembangan kajian ilmu komunikasi, khususnya komunikasi antarpribadi dan privasi digital, dengan memperluas penerapan teori Manajemen Privasi Komunikasi ke dalam konteks yang belum banyak dijangkau, yaitu korban kejahatan digital berbasis kecerdasan buatan. Secara lebih spesifik, penelitian ini berpotensi menguji satu asumsi urutan dalam teori tersebut. Dalam penjelasan baku Manajemen Privasi Komunikasi, turbulensi batas dipahami sebagai akibat dari kegagalan koordinasi setelah informasi dibagikan. Pada kasus *deepfake*, urutan tersebut berpotensi terbalik karena pelanggaran batas terjadi sebelum korban sempat menyusun aturan apa pun. Pengujian atas kemungkinan pembalikan urutan inilah yang menjadi sumbangan teoretis yang ditawarkan penelitian ini.

Selain itu, penelitian ini menawarkan pembacaan atas fenomena *deepfake* dari sudut pandang pengalaman subjektif korban sudut pandang yang selama ini kurang

mendapat tempat dibandingkan pendekatan hukum dan teknologi. Hasil penelitian juga dapat menjadi rujukan metodologis bagi peneliti selanjutnya yang hendak menggunakan desain studi kasus tunggal untuk topik-topik sensitif yang aksesnya terbatas.

1.5.2 Kegunaan Praktis

Bagi lembaga pendamping korban, konselor, dan organisasi masyarakat sipil yang bergerak di bidang keamanan digital, penelitian ini diharapkan memberikan gambaran mengenai pertimbangan yang bekerja di balik keputusan korban untuk bercerita atau diam. Pemahaman tersebut berguna dalam merancang pendekatan pendampingan yang tidak memaksa korban berbicara sebelum ia siap, sekaligus dalam membangun saluran pelaporan yang menghargai kendali korban atas informasi pribadinya. Bagi content creator dan pengguna media sosial secara umum, temuan penelitian dapat menjadi bahan refleksi mengenai pengelolaan jejak visual di ruang digital serta langkah yang dapat ditempuh apabila menghadapi situasi serupa.

1.5.3 Kegunaan Sosial

Penelitian ini diharapkan turut berkontribusi pada upaya mengurangi stigma terhadap korban kejahatan digital. Dengan menghadirkan pengalaman korban secara utuh dan bermartabat, penelitian ini berupaya menggeser cara pandang publik dari mempertanyakan korban menuju memahami posisi korban. Selain itu, penelitian ini dapat memperkuat kesadaran masyarakat bahwa manipulasi wajah dan identitas melalui kecerdasan buatan merupakan bentuk kekerasan yang nyata, bukan sekadar lelucon atau rekayasa digital yang dianggap tidak berkonsekuensi.

1.5.4 Batasan Penelitian

Penelitian ini berfokus pada proses komunikasi dan pengungkapan diri korban, sehingga tidak mencakup analisis teknis mengenai proses pembuatan, deteksi, maupun aspek teknologi *deepfake* itu sendiri.