IDENTIFIKASI ANCAMAN KEAMANAN DIGITAL PADA TREN MANIPULASI FOTO BERBASIS AI: A SYSTEMATIC LITERATURE REVIEW

IDENTIFICATION OF DIGITAL SECURITY THREATS IN THE TREND OF AI-BASED PHOTO MANIPULATION: A SYSTEMATIC LITERATURE REVIEW

Nurfudiniyah[1]
[1]Magister Ilmu Komunikasi, Universitas Gadjah Mada
[1]Jl. Socio Yustisia No. 1 Bulaksumur, Yogyakarta, Indonesia
[1]nurfudiniyah@mail.ugm.ac.id

Diterima tgl. 10 Februari 2026 Direvisi tgl. 7 Juni 2026 Disetujui tgl. 30 Juni 2026

ABSTRACT

Advances in AI-based technologies have given rise to a new trend in the creation of manipulated images by combining facial photographs with specific prompts. However, behind what is often perceived as a form of entertainment, this trend presents potential digital security threats to privacy, identity, and personal data that remain largely unrecognized due to limited public literacy and persistent digital divides. Therefore, this study aims to map the digital security threats associated with the growing trend of AI-generated photo manipulation using a Systematic Literature Review (SLR) approach. Employing a qualitative methodology, data were collected through searches of the Scopus database and Publish or Perish, covering publications from 2020 to 2025. Following the identification, screening, and inclusion stages adapted from the PRISMA guidelines, 239 studies were initially identified, of which 18 were selected for further analysis based on their relevance and credibility. The findings reveal that digital identity threats associated with AI-based photo manipulation include privacy violations, non-consensual image manipulation, biometric data theft, identity misuse, and the exploitation of digital footprints. These threats can facilitate various forms of digital fraud, distort representations of past events, and enable sexual harassment targeting vulnerable groups, particularly women and children. Moreover, their impacts may extend beyond individual and social environments to the national level. This study therefore highlights the urgency of strengthening digital security literacy in response to the increasing prevalence of AI-driven visual manipulation. Furthermore, the findings provide a valuable reference for the development of more adaptive regulatory frameworks aimed at mitigating digital security risks and emerging threats.

Keywords: *Digital security threats, Systematic Literature Review, AI photo trends*

ABSTRAK

Kemajuan teknologi berbasis AI melahirkan tren baru berupa penciptaan gambar manipulasi dengan memasukkan foto wajah dan *prompt* tertentu. Namun, di balik tren yang dianggap sebagai sebuah hiburan ini, memungkinkan ancaman keamanan digital pada privasi, identitas, dan data pribadi yang belum banyak disadari karena kurangnya literasi masyarakat dan kesenjangan digital. Oleh karena itu, penelitian ini berupaya memetakan ancaman keamanan digital di balik tren foto AI tersebut dengan metode Systematic Literature Review. Menggunakan pendekatan kualitatif, data dikumpulkan berdasarkan pencarian Scopus serta Publish or Perish dari tahun 2020-2025. Melalui tahap identifikasi, penyaringan sebagai proses screening, serta tahapan inklusi yang diadaptasi dari pedoman PRISMA, 239 literatur ditemukan dan 18 literatur terpilih yang dianalisis berdasarkan relevansi serta kredibilitasnya. Hasil penelitian ini menunjukkan bahwa ancaman identitas digital berupa pelanggaran privasi, manipulasi foto non-konsensual, pencurian data biometrik, penyalahgunaan identitas, serta rekam jejak digital bisa disalahgunakan untuk melakukan penipuan digital, distorsi peristiwa masa lalu, hingga pelecehan seksual

yang menyasar kelompok rentan seperti perempuan dan anak-anak. Bahkan, dampaknya bisa terjadi dari level lingkungan sosial hingga nasional. Sehingga, penelitian ini menegaskan urgensi literasi keamanan digital sebagai respon dari tingginya manipulasi visual dengan AI. Selain itu, temuan penelitian ini dapat menjadi referensi perumusan regulasi yang lebih adaptif sebagai mitigasi risiko ancaman keamanan digital.

Kata Kunci: Ancaman keamanan digital, *Systematic Literature Review*, Tren foto AI

1. PENDAHULUAN

Seiring dengan perkembangan teknologi, AI (*artificial intelligence)* atau akal imitasi hadir menawarkan beragam fungsi untuk penggunanya, seperti mengoptimisasi tugas, memberikan preferensi baru, meningkatkan akurasi dan efisiensi tugas manusia, serta memungkinkan personalisasi layanan. Fungsi-fungsi tersebut biasa diterapkan dalam berbagai bidang, baik pendidikan, keuangan, kesehatan, seni, hingga fotografi (Sepriano, 2023).

Kepopuleran AI semakin meningkat ketika kemampuannya dimanfaatkan untuk menciptakan gambar berdasarkan perintah teks yang dimasukkan ke dalam AI (*prompt*), bahkan aplikasi AI juga bisa membuat foto yang dimasukkan dengan *prompt* tertentu menjadi foto artistik, seperti tren foto pribadi yang dijadikan karakter Disney, karakter Pixar, karakter studio Ghibli, Anime, miniatur dan berbagai macam tren foto AI lainnya (Enjellina et al., 2023). Foto-foto yang dihasilkan tersebut berasal dari sistem *AI-Generated Image* yang bisa menghasilkan foto tiruan berdasarkan foto asli yang diunggah. *AI-Generated Image* adalah sistem yang dihidupkan kembali oleh OpenAI sebagai perusahaan pengembang ChatGPT yang menggunakan model Transformer di GPT-2 dan GPT-3. Sistem ini memungkinkan seseorang menggunakan model AI teks-ke-gambar yang bisa menghasilkan gambar berkualitas tinggi berdasarkan perintah bahasa yang dimasukkan (Kartika, 2025).

Salah satu tren foto dengan memanfaatkan aplikasi berbasis AI adalah pembuatan foto manipulasi bersama figur publik atau tokoh idola. Melalui teknologi *AI-Generated Image*, pengguna dapat membuat foto yang seolah mencitrakan mereka yang sedang berfoto dengan selebriti, aktor, musisi, atau *influencer* favorit dengan hasil yang sangat realistis. Hal tersebut sejalan dengan survei dari Katadata Insight Center yang mengungkapkan bahwa 44,8% masyarakat Indonesia menggunakan AI untuk mengedit foto atau video (Maheswara, 2025). Menurut Detta Rahmawan, faktor masyarakat mengikuti tren tersebut adalah FOMO *(Fear of Missing Out)* atau ketakutan akan ketinggalan tren yang sedang berlangsung, khususnya di media sosial. FOMO tersebut mencerminkan adanya kalkulasi rasional antara manfaat yang diperoleh dengan risiko yang ditanggung, culnan dan amstrong menyebutnya sebagai *Privacy Calculus Theory.* Meskipun, hal tersebut juga bisa dikaitkan dengan perilaku berlebihan dalam memuja selebritis serta figur publik (Infipop, 2025). Perilaku tersebut menurut (Wohl & Horton, 1956) adalah sebuah interaksi parasosial, yaitu hubungan satu arah yang dirasakan pengguna media terhadap figur publik seolah terdapat kedekatan interpersonal. Hal ini ini juga dapat mendorong seseorang membuat foto bersama sebagai pemenuhan kebutuhan afiliasi.

Namun, kemampuan literasi digital masyarakat Indonesia masih belum mumpuni dan belum siap menghadapi cepatnya perkembangan teknologi. Hal tersebut ditujukan dari menurunnya Indeks Literasi Digital masyarakat Indonesia ke level 49,28 dibanding tahun sebelumnya yang mencapai 58,25 poin (Fajriadi, 2025). Sehingga, rendahnya literasi digital tersebut menjadi tantangan tersendiri bagi perkembangan teknologi informasi dan komunikasi, khususnya dalam melindungi data pribadi dan privasi di media digital. Selain itu, jumlah pengguna media sosial yang sangat besar di Indonesia turut menjadi saluran utama tersebarnya *deepfake.* Hal tersebut

membuat masyarakat lebih terekspos pada konten palsu yang sulit diverifikasi (Kumar & Singh, 2025; Lumingkewas & Halim, 2026).

Kosongnya undang-undang khusus yang mengatur *deepfake* di Indonesia juga membuat pelaku kejahatan lebih leluasa memanfaatkan teknologi ini (Purwadi et al., 2022). Di sisi lain, sistem algoritmik seperti *platform* AI menurut Diakopoulos & Johnson (2021) juga harusnya bertanggung jawab terhadap dampak yang ditimbulkan oleh proses otomatisnya. Hal tersebut dinyatakan sebagai *Algorithmic Accountability,* yang juga menegaskan bahwa ancaman keamanan digital bukan hanya bersumber dari pengguna jahat, tapi juga dari desain algoritmik yang tidak bertanggung jawab dan menjadi dasar rekomendasi regulasi *platform* dalam penelitian ini.

Dibanding empat pilar literasi digital lainnya, isu terkait keamanan digital memiliki skor terendah pada tahun 2021, yang berada pada poin 3,10 (Adi, 2022). Skor tersebut menunjukkan bahwa kesadaran dan literasi digital terhadap keamanan digital masyarakat Indonesia masih tergolong rendah. Menurut Sanjaya et al. (2024), rendahnya literasi keamanan digital dipengaruhi oleh pengetahuan digital, kepercayaan digital, dan kewaspadaan digital.

Kurangnya pengetahuan digital tersebut membuat maraknya *deepfake* berbasis AI dengan tren foto manipulasi. *Deepfake* merupakan hasil pembuatan konten audio-visual yang sengaja dibuat-buat dan sangat realistis, dapat digunakan sebagai alat untuk menyebarkan disinformasi, manipulasi opini publik, hingga menimbulkan ketidakpastian sistemik masyarakat (Nurdin & Nugraha, 2025). Pembuatan *deepfake* di Indonesia kian meningkat seiring dengan peningkatan penetrasi internet dan perkembangan AI, Riset Sensity AI menunjukkan bahwa adanya peningkatan hingga 550% produksi konten *deepfake* dalam lima tahun terakhir (Firyalfatin, 2025). Hal tersebut menunjukkan bahwa perkembangan AI lebih cepat daripada kemampuan pemahaman masyarakat terhadap konsekuensinya. Selain itu, berdasarkan *Protection Motivation Theory* rendahnya literasi digital bisa menghasilkan *threat appraisal* yang rendah dan membuat masyarakat tidak menyadari keparahan ancaman. Sehingga, motivasi protektif juga rendah (Rogers, 1975).

Tren pembuatan foto manipulasi bersama tokoh publik telah menjadi *deepfake* yang digunakan untuk menyebut teknik pemalsuan konten digital, seperti video, gambar, dan suara menggunakan teknologi AI. Teknologi tersebut memungkinakan pembuatan video yang terlihat sangat nyata, namun penuh manipulasi (Sulianta, 2025a). *Deepfake* merupakan sebuah teknik untuk sintesis citra manusia berdasarkan AI. *Deepfake* dibuat dengan menggabungkan serta menempatkan gambar atau video dengan menggunakan sistem *machine learning* yang disebut dengan *generative adversarial network (GAN)* (Khusna, 2019).

Deepfake telah membuat beberapa atlet Timnas Sepakbola Indonesia resah dan melarang fotonya dijadikan konten serupa. Dalam unggahan *Instagram story,* mereka menunjukkan foto hasil *deepfake* buatan fans mereka yang dinilai telah melanggar etika dan privasi, foto-foto hasil AI tersebut menunjukkan para atlet Timnas Sepakbola Indonesia sedang berfoto bersama, memeluk, hingga mencium fans mereka. Bahkan foto Rizki Ridho dinilai seolah telah melakukan pelecehan karena tangannya berada di area sensitif seorang fans perempuannya ketika memeluk, Ridho menuliskan *"Teman teman minta tolong lebih sopan lagi yaa, tidak perlu edit yg kaya gini."* Sandi Wals, pemain Timnas lainnya juga mengatakan *"I would like to kindly ask people to stop making AI photo's of me without my consent as this can lead to problem/misunderstandings in the future Terima kasih"* (Majid, 2025).

Dampak psikologis yang dominan dialami oleh korban *deepfake* adalah munculnya rasa tidak nyaman dan malu, yang dapat mempengaruhi kehidupan sosial serta emosional mereka (Herdian & Sumarwan, 2025). Hal yang sama juga diungkapkan oleh Ali et al bahwa efek

deepfake terhadap seseorang bisa menjadi sebuah trauma psikologis, kerusakan reputasi, dan viktimisasi berulang (Ali et al., 2025). Sedangkan, menurut Khusna dan Pangestu, *deepfake* bisa dimanfaatkan untuk menyebarkan kebencian dan menjadi alat propaganda politik. Sehingga, *deepfake* bisa berpotensi untuk menyebarkan kejahatan, khususnya terhadap pribadi korban dan keamanannya di media digital, seperti penipuan identitas, pelanggaran privasi, hingga akses illegal terhadap informasi sensitif (Khusna, 2019).

Oleh karena itu, masih rendahnya literasi digital masyarakat, khususnya dalam pilar keamanan digital serta keterbatasan kemampuan masyarakat dalam memahami konsekuensi pada penggunaan AI yang semakin berkembang pesat, membuat identifikasi ancaman keamanan digital di balik tren AI tersebut diperlukan, khususnya untuk menjembatani kesenjangan digital antara adopsi teknologi dan kesiapan pengguna.

Keamanan menjadi hal yang diperhatikan secara khusus karena pengendalian terhadap data berada di tangan pengguna sendiri dalam menggunakan internet yang memiliki sifat menghubungkan pengguna secara global, kendali keamanan data ada pada masing-masing individu pengguna internet. Oleh karena itu, keamanan digital bukan hanya untuk mengamankan data yang kita miliki saja, namun juga melindungi data pribadi yang sifatnya rahasia. Sehingga, hal tersebut membuka peluang bagi pihak-pihak dengan niat jahat untuk melakukan penipuan atau peretasan. Oleh karena itu keamanan digital juga menjadi area kompetensi literasi digital yang juga dijadikan kerangka berfikir dalam riset-riset keamanan digital (Adikara et al., 2021).

Berdasarkan *systematic literature review* (SLR), kajian Jada dan Mayayise mengemukakan bahwa AI mempengaruhi peningkatan pertahanan siber. Namun, disisi lain juga menghadirkan kebutuhan data yang sangat tinggi yang menyebabkan inefisiensi AI (Jada & Mayayise, 2024). Hal yang serupa juga menjadi temuan kajian dari Pongoh et al, yang juga menggunakan SLR dengan hasil bahwa AI telah berhasil mendeteksi serangan siber, hingga mengoptimalkan respon keamanan (Pongoh et al., 2024). Sedangkan, menurut (Nurdin & Nugraha, 2025), perkembangan AI bisa memunculkan masalah politis, psikologis, hingga struktural. Sehingga, perlunya regulasi dan literasi digital untuk mengatasi hal tersebut.

Meskipun penelitian mengenai AI dan deepfake telah banyak dilakukan, studi yang secara khusus memetakan ancaman keamanan digital dalam konteks tren manipulasi foto hiburan melalui pendekatan sistematis masih terbatas. Kajian sebelumnya cenderung berfokus pada aspek teknis deteksi deepfake atau level institusional (Jada & Mayayise, 2024; Pongoh et al., 2024), sehingga belum menyentuh perspektif pengguna awam dengan literasi digital rendah. Selain itu, saat pengguna mengunggah foto pribadi ke aplikasi AI untuk dimanipulasi, terdapat risiko data biometrik wajah yang tersimpan di server pihak ketiga yang rentan terhadap serangan siber (Chua et al., 2025). Hal ini menjadi pintu masuk ancaman keamanan digital yang belum banyak dikaji secara sistematis. Oleh karena itu, penelitian ini bertujuan untuk memetakan secara sistematis berbagai ancaman keamanan digital yang muncul dari tren manipulasi foto berbasis AI dengan menggunakan metode Systematic Literature Review (SLR), guna memberikan gambaran komprehensif serta rekomendasi kebijakan bagi perlindungan data pribadi masyarakat di era kecerdasan buatan. Harapannya, penelitian ini bisa menambah kajian tentang literasi digital dan teknologi AI, meningkatkan kesadaran publik tentang ancaman keamanan digital di balik tren foto AI dan memperkuat pemahaman etika dan keamanan dalam penggunaan teknologi, serta menjadi dasar konseptual bagi pengembangan kebijakan terkait.

2. METODE PENELITIAN

Penelitian ini mengadopsi *systematic literature review* (SLR) sebagai metode menggunakan pendekatan kualitatif pada kajian terkait AI dan keamanan digital. SLR adalah metode untuk

mengidentifikasi, mengevaluasi, dan menafsirkan semua riset yang tersedia dan relevan dengan tujuan penelitian tertentu, bidang topik, atau isu yang menjadi fokus (Kouchaksaraei & Karl, 2004).

Penelitian ini dilandasi oleh dua pertanyaan penelitian (Research Questions) berikut: RQ1: Apa saja jenis ancaman keamanan digital yang ditimbulkan oleh tren manipulasi foto berbasis AI? RQ2: Siapa kelompok yang paling rentan terhadap ancaman tersebut dan bagaimana implikasinya dari level individu hingga nasional? Kedua pertanyaan ini menjadi panduan dalam proses seleksi, ekstraksi, dan sintesis literatur yang dilakukan. Sehingga, penelitian ini mengkaji data sekunder dari artikel-artikel yang telah dipublikasikan di jurnal internasional terindeks Scopus. Sedangkan, penggunaan pendekatan kualitatif pada penelitian ini untuk mengumpulkan hasil penelitian yang didapatkan dan bersifat non-numerikal (Creswell, 2018). SLR dalam kajian ini menggunakan kerangka kerja Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 dengan berbagai penyesuaian untuk mengekstrak informasi dalam artikel jurnal (PRISMA, 2020). Tahap pertama yaitu identifikasi dengan melakukan pencarian literatur dari basis data Google Scholar dan Scopus berdasarkan kriteria tahun terbit mulai dari 2020-2025, artikel menggunakan bahasa Inggris, dalam rumpun keilmuan *social science,* serta dipublikasikan dalam jurnal akademik. Pencarian literatur didasarkan pada kata kunci *"AI-Generated Image Impact"* DAN *"Digital Security"; "Digital Security Risk"* DAN *"Cyber Risk",* menghasilkan 239 jurnal ilmiah dari Scopus dan 100 jurnal ilmiah dari Google Scholar memanfaatkan aplikasi Publish or Perish. String pencarian lengkap yang digunakan adalah: ("AI-generated image" OR "deepfake" OR "AI photo manipulation") AND ("digital security" OR "cyber risk" OR "digital threat"). Penggunaan operator Boolean AND/OR bertujuan untuk memperluas cakupan hasil pencarian sekaligus memastikan relevansi topik.

Tahap kedua sebagai *screening* awal dengan penyaringan berdasarkan relevansi judul artikel, menghasilkan 142 artikel yang dihimpun dari basis data Google Scholar dan Scopus. *Screening* lanjutan dilakukan melalui penyaringan berdasarkan relevansi abstrak dengan tujuan penelitian, menghasilkan 58 artikel. Selanjutnya, *screening* dilakukan dengan penyaringan artikel yang disempurnakan lebih lanjut menggunakan kriteria pengecualian tambahan, yaitu kredibilitas artikel. Sehingga, hanya artikel yang terindeks Scopus Q1-Q2 saja yang akan masuk dalam tahap selanjutnya. Pada penyaringan ini, menghasilkan 37 artikel.

Tahap terakhir adalah *included* dengan meninjau relevansi keseluruhan isi artikel. Pada tahap ini, hanya artikel yang memiliki relevansi tinggi berdasarkan kriteria inklusi yaitu selaras dengan tujuan penelitian, berbentuk artikel publikasi dan naskah lengkap, serta bisa diakses *(open access).* Tahap ini juga menghasilkan artikel berjumlah 18 yang kemudian disintesis untuk mendapatkan temuan tentang ancaman keamanan digital dalam tren foto manipulasi AI. Data dari 18 artikel tersebut diekstraksi ke dalam matriks ekstraksi yang memuat komponen: penulis, tahun terbit, metode, jenis ancaman yang dibahas, dan temuan kunci.

Sintesis dilakukan menggunakan pendekatan *thematic synthesis* yang dikembangkan oleh Thomas & Harden (2008). Teknik ini terdiri dari tiga tahap: (1) *free line-by-line coding* terhadap temuan setiap artikel terpilih; (2) pengembangan descriptive themes berdasarkan kesamaan pola koding; dan (3) mengonstruksi *analytical themes* yang melampaui deskripsi individual menuju sintesis interpretatif lintas artikel. Tema-tema analitik yang dihasilkan kemudian dipetakan ke dalam indikator keamanan digital Japelidi. Selain itu, penilaian kualitas konten *(quality assessment)* juga dilakukan terhadap metodologi masing-masing artikel untuk memastikan kredibilitas temuan yang disintesis.

3. HASIL DAN PEMBAHASAN

Berdasarkan SLR yang telah dilakukan sebelumnya, 18 artikel dipilih untuk mengidentifikasi dan memetakan ancaman-ancaman keamanan digital di balik tren foto AI dengan tokoh publik. Terdapat 1 artikel yang terbit pada tahun 2022, 2 artikel pada tahun 2023, 7 artikel pada tahun 2024, 7 artikel pada tahun 2025, dan 1 artikel pada tahun 2026. Selain itu, terdapat 11 artikel yang terindeks Scopus Q1 dan 7 artikel dengan indeks Scopus Q2.

Tabel 1. Matriks Ringkasan 18 Literatur Terpilih

No.	Penulis & Tahun	Fokus Kajian	Metode	Temuan Utama	Relevansi	Indeks
1	Furizal et al. (2025)	Implikasi sosial, hukum, etika deepfake pornografi AI	SLR (PRISMA)	Konten deepfake non-konsensual melanggar otonomi individu; memicu kerusakan reputasi & dampak psikologis	Memperkuat ancaman identitas digital & keamanan anak	**Q1**
2	Nightingale & Wade (2022)	Dampak dan mitigasi media visual palsu	Review eksperimental	Face morphing memudahkan pencurian identitas; sulit dideteksi publik awam	Dasar konseptual ancaman identitas biometrik	**Q1**
3	Roe & Perkins (2024)	Deepfake dalam konteks pendidikan tinggi	Scoping Review	Deepfake vocal membuka celah pencurian data sensitif; konten non-konsensual meningkat	Relevan untuk ancaman penipuan & rekam jejak digital	**Q2**
4	Vasist & Krishnan (2023)	Keterlibatan pengguna dengan deepfake via SST Theory	Meta-sintesis	Platform amplifikasi deepfake; perlindungan anak lemah di ranah digital	Relevan untuk ancaman keamanan digital bagi anak	**Q1**
5	Kharvi (2024)	Dampak deepfake pada opini publik, wacana politik, keamanan pribadi	Kajian teori-empiris	Deepfake merusak sistem biometrik; ancaman keamanan nasional dari pemalsuan identitas	Relevan untuk ancaman identitas digital & keamanan nasional	**Q1**
6	K. Verma (2024)	Dampak deepfake terhadap hak privasi	Kajian hukum-etis	Konten AI non-konsensual sebagai bentuk kekerasan baru; kelompok rentan paling terdampak	Menguatkan ancaman identitas & ancaman bagi anak	**Q2**
7	Ali et al. (2025)	Deepfake dan viktimologi	Kajian yuridis-kriminologi	Deepfake menimbulkan trauma psikologis, kerusakan reputasi, viktimisasi berulang	Memperkuat analisis dampak psikologis ancaman identitas digital	**Q2**
8	Lowenstein et al. (2025)	Implikasi etis deepfake prosocial pada IPV	Analisis komentar media sosial (kualitatif)	Penggunaan deepfake pada korban meninggal memunculkan isu persetujuan digital posthumous	Relevan untuk ancaman rekam jejak digital	**Q1**
9	Momeni (2025)	Deepfake politik dan persepsi warga	Kajian teoritis-komunikasi	Deepfake mendistorsi ingatan kolektif; platform memperkuat mis/disinformasi pada skala masif	Relevan untuk ancaman rekam jejak digital & opini publik	**Q1**
10	Okolie (2023)	AI, pelecehan seksual berbasis gambar, privasi data	Kajian hukum-gender	Deepfake seksual menarget perempuan dan anak; ancaman privasi meningkat di media sosial	Relevan untuk ancaman identitas digital & keamanan anak	**Q2**
11	Maniyal & Kumar (2024)	Klasifikasi dan trajektori deepfake	Review teknis	Deepfake difasilitasi GAN; klasifikasi ancaman	Memberikan kerangka teknis pemetaan ancaman	**Q1**

				mencakup identitas, penipuan, dan propaganda		
12	Li & Zhao (2024)	Intensi perilaku terkait deepfake	Survei kuantitatif	Konsumen deepfake menunjukkan normalisasi konten manipulatif; risiko rekam jejak digital meningkat	Relevan untuk rekam jejak digital & perilaku pengguna	**Q1**
13	Jada & Mayayise (2024)	Dampak AI pada keamanan siber organisasi	SLR	AI meningkatkan kapasitas pertahanan siber sekaligus menciptakan kerentanan baru	Konteks makro ancaman keamanan digital berbasis AI	**Q1**
14	Pongoh et al. (2024)	Pemanfaatan AI untuk meningkatkan keamanan siber	SLR	AI efektif mendeteksi serangan siber; penyalahgunaan AI membuka vektor ancaman baru	Konteks kebijakan dan mitigasi teknis	**Q2**
15	Zheng et al. (2025)	Regulasi deepfake komparatif AS, UE, China	Kajian hukum komparatif	Deepfake digunakan untuk penipuan keuangan, penyamaran identitas, perusakan reputasi	Relevan untuk ancaman penipuan digital & rekam jejak digital	**Q1**
16	A. Verma (2025)	Krisis keaslian digital dan tantangan etika media sintetis	Kajian etis-komunikasi	Deepfake merendahkan tokoh publik, terutama perempuan dalam politik	Relevan untuk ancaman rekam jejak digital & penipuan	**Q1**
17	Yang & Jiang (2025)	Persepsi risiko dan perilaku positif pengguna AIGC (Model CAC)	Kuantitatif-survei	Persepsi risiko diferensial memengaruhi niat protektif; literasi digital sebagai moderator kunci	Mendukung argumen PMT dan perilaku pengguna dalam menghadapi ancaman AI	**Q1**
18	Nurdin & Nugraha (2025)	Ancaman deepfake dan disinformasi AI terhadap keamanan nasional	Kajian teoritis-kebijakan	Deepfake memunculkan masalah politis, psikologis, struktural; urgensi regulasi & literasi digital	Relevan untuk semua indikator ancaman, khususnya level sosial-nasional	**Q2**

Sumber: Olahan Peneliti

Ancaman keamanan digital penggunaan AI dalam penelitian ini dianalisis berdasarkan indikator keamanan digital dari Japelidi. Indikator tersebut dipilih karena memiliki relevansi dengan pilar kompetensi keamanan digital sesuai dengan konteks penelitian ini, serta berhubungan dengan perlindungan *hardware*, perlindungan identitas digital serta data pribadi di media digital, penipuan daring, rekam jejak media digital, hingga *catfishing* (Adikara et al., 2021).

3.1. Ancaman Keamanan Identitas Digital

Dalam menjaga keamanan digital, seseorang harus mengerti maksud dari identitas digital dan risiko yang terjadi ketika tidak mampu menjaga identitas tersebut. Identitas digital merujuk pada identitas digital merupakan jati diri seseorang sebagai pengguna media digital, yang identitasnya tidak selalu sama dengan identitasnya di ruang publik atau kehidupan nyata karena terdiri dari gabungan beragam identitas parsial yang ingin ditunjukan oleh pengguna (Adikara et al., 2021).

Sayangnya, kurangnya pemahaman seseorang dalam menggunakan aplikasi digital seperti AI, khususnya yang digunakan untuk memanipulasi foto seseorang, menimbulkan permasalahan tentang etika di ruang digital, khususnya mengenai identitas digital dan hak pribadi. Penggunaan kemiripan seseorang secara non-konsensual seperti pada kasus atlet sepakbola yang fotonya dimanipulasi dengan AI seolah sedang memeluk penggemarnya, telah melanggar prinsip persetujuan, kebenaran, transparansi, dan otonomi individu di ruang digital (Furizal et al., 2025). Berdasarkan *Privacy Calculus,* pengguna yang sukarela mengunggah foto wajah mereka ke

platform AI telah melakukan kalkulasi yang tidak optimal, mereka bersedia menukar manfaat kesenangan, hiburan, dan pengakuan sosial dengan risiko privasi jangka panjang yang jauh lebih besar (Culnan & Armstrong, 1999). Hal tersebut diperparah dengan adanya *parasocial interaction* yang membuat adanya rasa kedekatan pengguna dengan figur publik di ruang digital dengan idolanya (Wohl & Horton, 1956). Akibatnya, mereka terdorong untuk memvisualisasikan kedekatan itu dengan bantuan AI tanpa menyadari konsekuensinya.

Nightingale & Wade (2022) mendefinisikan manipulasi foto tersebut sebagai *face morphing,* yaitu penggabungan dua atau lebih gambar individu secara digital untuk membuat gambar baru yang menyerupai masing-masing identitas aslinya. *Morphing* ini memberikan peluang aplikasi pengeditan digital untuk melakukan pencurian identitas berupa foto wajah seseorang. Bahkan, saat ini, foto wajah pada paspor bisa dipalsukan dengan menggunakan *deepfake* yang sulit dideteksi dan juga berpeluang digunakan untuk memfasilitasi kejahatan lain seperti pencurian identitas, perdagangan manusia, dan imigrasi ilegal. Potensi pencurian identitas lainnya juga dibahas oleh Roe & Perkins (2024), sebagai akibat dari adanya konten *deepfake* non-konsensual yang melibatkan foto seseorang atau beberapa orang.

Identitas digital juga diartikan sebagai karakter seseorang di *platform* digital. Karakter tersebut juga bisa dimanipulasi dengan *deepfake* yang disalahgunakan untuk menciptakan citra lainnya, seperti seksual non-konsensual, melakukan penipuan keuangan dan pencurian identitas, dan memicu kampanye informasi yang salah. Pencitraan negatif seseorang dengan menggunakan AI tersebut berpotensi menyebabkan kerusakan reputasi (Furizal et al., 2025; Maniyal & Kumar, 2024). Khususnya, apabila penerapan teknologi AI menyebabkan penyebar luasan gambar dan video palsu untuk tujuan kriminal terhadap identitas tokoh publik.

Sedangkan, identitas digital dalam cakupan yang lebih luas mencakup data pribadi, yang meliputi identitas, simbol, kode, hingga penanda personal seseorang yang bersifat rahasia. Selain itu, data pribadi juga dapat dipahami sebagai informasi individu yang disimpan dan dijaga keamanannya karena bersifat privat (Adikara et al., 2021). Oleh karena itu, tren penggunaan AI untuk membuat foto manipulasi juga menciptakan ancaman keamanan digital yang berhubungan dengan data pribadi seseorang.

Media yang dibuat secara artifisial ini dapat dengan mudah memanipulasi persepsi individu dan kolektif, sehingga membahayakan kepercayaan terhadap proses demokrasi, privasi pribadi, dan keaslian media. Selain itu, teknologi *deepfake* dapat digunakan untuk memproduksi media non-konsensual yang menampilan kemiripan visual dengan seseorang tanpa adanya persetujuan sebelumnya (Roe & Perkins, 2024). Oleh karena itu, kondisi ini bisa melanggar privasi, merugikan individu dan keluarga, serta digunakan untuk tindakan-tindakan manipulatif seperti pencurian identitas. Hal tersebut menunjukkan bahwa ancaman keamanan digital terhadap identitas digital individu semakin kompleks karena adanya konten digital yang tampak realistis dan meyakinkan. Sehingga, batas antara konten dengan identitas asli dan manipulasi menjadi kabur di ruang digital.

Ancaman keamanan pada identitas digital tersebut semakin mengkhawatirkan bagi kelompok berisiko yang terdiri dari anak-anak dan perempuan karena lebih berpotensi dieksploitasi melalui penyebaran konten-konten manipulatif buatan AI (K. Verma, 2024). Manipulasi identitas tersebut dapat merusak reputasi seseorang yang berdampak jangka panjang bagi reputasi individu. Selain itu, penyebaran konten manipulatif bisa memicu kampanye informasi yang salah dan bisa mempengaruhi persepsi masyarakat terhadap isu tertentu.

Penggunaan kemiripan seseorang secara nonkonsensual untuk tujuan jahat semakin menimbulkan pertanyaan tentang etika yang signifikan mengenai identitas digital dan hak pribadi (K. Verma, 2024). Konten dewasa non-konsensual, penyalahgunaan foto wajah seseorang, serta

pemalsuan identitas melalui *AI-Generated Image* berupa *deepfake* menciptakan bentuk kekerasan baru yang tidak hanya tampak secara fisik, namun juga menimbulkan dampak psikologis. Rendahnya *threat appraisal* pada masyarakat dengan literasi digital rendah yang membuat mereka tidak menganggap serius konsekuensi tindakan mengunggah fotonya tersebut ke AI. Sehingga, motivasi protektif (*coping appraisal*) juga tidak terbentuk. Dalam kerangka *Protection Motivation Theory* (PMT) hal ini bisa menciptakan lingkaran kerentanan yang berulang terus menerus (Rogers, 1975).

Ancaman serius lainnya terjadi dalam sistem pengenalan wajah yang digunakan dalam mekanisme keamanan. Identitas digital individu yang berupa wajah merupakan data biometrik yang terekam sebagai pengenalan identitas dalam berbagai sistem keamanan (Kharvi, 2024). Namun, jika *deepfake* atau foto manipulasi visual digunakan untuk mengelabui sistem keamanan biometrik tersebut, risiko keamanan digital berupa akses ilegal terhadap data pribadi seseorang bisa semakin meningkat. Oleh karena itu, saat ini ancaman keamanan digital karena manipulasi konten bukan hanya dalam bentuk pelanggaran privasi, namun juga merambah pada aspek infrastruktur keamanan (Wati et al., 2024). Berdasarkan kerangka Cybersecurity Framework (CSF), pengguna harus mampu mengidentifikasi foto wajahnya sebagai aset digital yang sensitif dan rentan. Selain itu, adanya fungsi *protect* dann *detect* berperan dalam menjaga sistem keamanan agar tidak dieksploitasi sebagai konten *deepfake (Pratomo et al., 2018).*

3.2. Ancaman Penipuan Digital

Penipuan digital merupakan tindakan penyalahgunaan media digital oleh orang yang tidak bertanggung jawab dan menyasar masyarakat dengan tingkat kompetensi literasi digital rendah (Riti et al., 2024). Penipuan ini menjadi modus kejahatan yang memanfaatkan jaringan internet melalui *platform* digital untuk mengelabui korban. Sehingga, pelaku mendapatkan keuntungan tertentu. Penipuan digital termasuk dalam kejahatan siber (*cybercrime)* yang kian marak seiring perkembangan teknologi. Japelidi juga mengungkapkan, bahwa kasus penipuan menjadi ancaman keamanan digital yang sering ditemui, baik melalui surel maupun SMS. Bahkan, saat ini modus penipuan berbasis daring marak terjadi pada media sosial hingga aplikasi *e-commers (Adikara et al., 2021).*

Kecerdasan buatan atau AI saat ini juga menjadi teknologi yang bisa digunakan oleh masyarakat secara bebas yang menimbulkan potensi penyalahgunaan AI, khususnya manipulasi foto orang lain, seperti pada kasus penipuan viral di media sosial yang menampilkan panggilan video menggunakan wajah artis terkenal dengan modus memberikan hadiah (*give away*) (Vidi, 2024). Oleh karena itu, hal tersebut menciptakan kekhawatiran tersendiri bagai masyarakat, karena dapat menjadi media yang menyesatkan atau digunakan sebagai alat penipuan yang bisa merugikan pihak lain (Lowenstein et al., 2025).

Modus penipuan yang memanfaatkan teknologi seperti AI dengan bentuk *deepfake* banyak digunakan untuk mencari keuntungan berupa uang. Sehingga, *deepfake* menjadi model baru penipuan keuangan, khususnya melalui panggilan video palsu maupun pemerasan dengan identitas korban yang telah diedit menggunakan *AI-Generated Image,* menciptakan materi visual yang melecehkan dan bisa mengakibatkan kerugian keuangan (Nightingale & Wade, 2022).

Bukan hanya penggunaan identitas atau wajah dalam *deepfake* buatan *AI-Generated Image* saja yang dijadikan alat penipuan digital, suara seseorang juga bisa dimanfaatkan untuk modus yang sama. Menurut Roe & Perkins (2024), dengan munculnya *deepfake vocal,* autentikasi suara dari seseorang bisa dijadikan informasi untuk mencuri data pribadi seperti rekening bank. Bahkan, modus penipuan digital untuk mendapatkan informasi pribadi dan sensitif seperti *phising* bisa lebih meyakinkan pada korban yang tingkat literasi digitalnya rendah dan menyertakan video atau

suara, yang dimanipulasi dari hasil *deepfake* serta dibagikan oleh teman terpercaya (Xiaobil, 2025). Penipuan suara tersebut juga sempat terjadi di Indonesia dengan memanfaatkan teknologi AI yang disebut dengan AI *Voice Spoofing.* Tindakan tersebut dilakukan dengan menelepon korban dan merekam suaranya untuk dijadikan bahan manipulasi dan menipu orang lain (Wpj, 2025). Berdasarkan perspektif *Algorithmic Accountability, platform* AI yang tidak memiliki mekanisme moderasi konten yang memadai, secara tidak langsung turut bertanggung jawab atas terfasilitasinya penipuan-penipuan tersebut. Tanggung jawab algoritmik tidak dapat dipisahkan dari dampak sosial yang ditimbulkan (Diakopoulos & Johnson, 2021).

Dalam skala yang lebih luas, *deepfake* dapat dieksploitasi oleh penjahat untuk menyamar sebagai individu yang dikenal korban di platform media sosial, sehingga memfasilitasi aktivitas penipuan (Zheng et al., 2025). Adanya *deepfake* yang menciptakan citra gambar realistis dan sangat meyakinkan, mampu membuat pelaku kejahatan membangun kepercayaan palsu, memancing korban untuk mengirimkan dana, hingga mendapatkan akses pada akun dan data sensitif lainnya. Sehingga, fenomena tersebut memperlihatkan bahwa ancaman *deepfake* bukan hanya menyasar pada individu, namun juga bisa merusak ekosistem kepercayaan digital yang dibangun di media sosial.

Selain itu, penyebaran gambar palsu merupakan ancaman serius bagi keamanan nasional, karena gambar-gambar ini dapat dimanfaatkan untuk pemalsuan identitas pada berbagai konteks, seperti keuangan illegal hingga sistem pemerintahan (Kharvi, 2024). Keamanan nasional juga bisa terancam jika identitas yang dipalsukan adalah identitas dari tokoh publik yang memiliki peran krusial di sebuah negara, seperti *deepfake* video pidato Menteri Keuangan Sri Mulyani tentang gaji pengajar di Indonesia, kasus tersebut membuat para pengajar, baik guru atau dosen melakukan aksi protes melalui media sosial (Yanwardahana, 2025).

Oleh karena itu, *deepfake* atau foto hasil *generate* AI yang memasukan foto wajah tokoh publik juga berpotensi mengancam keamanan digital dan ruang publik dengan jangkauan yang luas. Melalui penyebaran propaganda berupa visual yang menggiring opini publik tersebut, stabilitas sosial bisa terganggu. Meskipun CSF telah merekomendasikan fungsi *respond* dan *recover* sebagai respon institusional terhadap insiden penipuan digital, jika tidak diimbangi dengan literasi digital, maka respon individu tetap lemah dan kejadian serupa bisa terjadi. Sedangkan, kunci literasi digital yang efektif menurut *Protection Motivation Theory* Rogers (1975*)* adalah meningkatnya *coping appraisal* masyarakat, yaitu keyakinan bahwa mereka mampu mengenali dan menghindari penipuan *deepfake.*

3.3. Ancaman Rekam Jejak Digital

Teknologi digital telah menjadi bagian yang tidak terpisahkan dari masyarakat Indonesia, yang memudahkan aktivitas sehari-hari. Bahkan, laporan digital dari Global Overview pada tahun 2025 menyebutkan bahwa negara Indonesia menempati posisi teratas di dunia dalam penggunaan ponsel untuk mengakses internet, dengan rata-rata waktu online harian masyarakat adalah 6 jam 38 menit (Kemp, 2025).

Tingginya penggunaan internet tersebut mengancam keamanan digital penggunanya, karena saat seseorang menggunakan internet, maka mereka telah memberikan akses pada pihak lain untuk mengetahui kebiasaan yang dilakukannya setiap hari melalui rekam jejak yang ditinggalkan pengguna. Hal tersebut karena media digital memerlukan data aktivitas pengguna untuk mengembangkan teknologinya. Sehingga, setiap kali seseorang mengunggah konten, berinteraksi di media sosial, maupun sekedar menjelajahi internet, mereka telah meninggalkan jejak digital yang bisa dianalisis oleh berbagai pihak (Sulianta, 2025).

Platform media sosial memainkan peran penting dalam sirkulasi deepfake dalam skala besar dan bisa meningkatkan bahayanya. *Deepfake* mengintensifkan masalah misinformasi dengan menantang kemampuan publik untuk membedakan realitas dari fiksi. Oleh karena itu, konten audio-visual yang *hiperrealistis* namun hasil dari manipulasi dapat menyesatkan dan menipu audiens dalam skala yang tidak terduga dan belum pernah terjadi sebelumnya (Momeni, 2025).

Dalam konteks kampanye kekerasan seksual terhadap perempuan di Israel, AI digunakan untuk membuat visual kesaksian dari korban perempuan yang sudah meninggal. Fenomena tersebut memunculkan diskusi di antara para penontonnya, khususnya mengenai pro kontra karena para korban tersebut tidak bisa memberikan persetujuan atas visual dirinya yang dimanipulasi dengan AI dalam proyek *deepfake* apapun. Keprihatinan penonton mereka ditunjukan dalam komentar media sosialnya, khususnya menyoroti tentang moral dan etika penyelenggara. Selain itu, implikasi emosional bagi keluarga terdampak juga menjadi keprihatinan mereka (Lowenstein-Barkai et al., 2025).

Kampanye digital tersebut juga menyiratkan risiko keamanan digital pada identitas korban dan rekam jejak digital yang tidak bisa dihapus. Sehingga, Meese menyebutnya sebagai keabadian digital. Angela Wijayanto juga mengungkapkan bahwa seluruh kegiatan seseorang di ruang digital selalu tercatat dan bukan menjadi hal yang mudah untuk menghapus jejak tersebut (Adikara et al., 2021b; Lowenstein-Barkai et al., 2025). Dalam perspektif *Privacy Calculus* ancaman rekam jejak digital ini bersifat asimetri, biaya yang ditanggung korban seperti kerusakan reputasi jangka panjang tidak proporsional dengan manfaat yang diperoleh pelaku (Culnan & Armstrong, 1999). *Platform* yang tidak menerapkan *Algorithmic Accountability* semakin menambah amplifikasi ancaman ini tanpa bisa diminta pertanggungjawaban (Diakopoulos & Johnson, 2021).

Selain itu, perkembangan teknologi *deepfake* juga digunakan untuk kepentingan komersil, seperti pertukaran wajah yang digunakan untuk meningkatkan strategi pemasaran online. Dengan memanfaatkan visual *deepfake* yang menarik, strategi tersebut dianggap bisa meningkatkan keterlibatan pengguna dan penjualan di media digital (Zheng et al., 2025). Namun, ketika wajah seseorang dipasang sebagai teknik pemasaran, tanpa disadari, jejak digitalnya bisa dimanfaatkan sebagai bentuk asosiasi yang berbeda tanpa mereka rencanakan sebelumnya. Sehingga, ekspansi *deepfake* dalam media digital turut menggerus kendali individu terhadap identitas dan citra dirinya. Oleh karena itu, teknologi *deepfake* bisa mengikis kepercayaan hingga mengakibatkan tekanan psikologis pada korbannya (Wahab, 2025).

Deepfake atau manipulasi terhadap acara publik terkemuka dapat mengubah ingatan kolektif tentang peristiwa tersebut. Visual yang dipalsukan dalam sebuah peristiwa mampu membentuk ulang persepsi masyarakat (Momeni, 2025). Ketika d*eepfake* disalahgunakan untuk menciptakan citra lainnya, seperti seksual non-konsensual terhadap seseorang, menimbulkan risiko terhadap rekam jejak digital korban dan bisa menimbulkan citra baru atau merusak reputasi dalam jangka panjang (Li & Zhao, 2024).

Di ranah politik, ancaman keamanan digital berupa rekam jejak digital ini semakin nyata ketika *deepfake* digunakan sebagai senjata untuk menjelekkan dan menggambarkan tokoh politik secara negatif untuk meningkatkan peluang lawan mereka, serta menyerang reputasi tokoh tertentu (Zheng et al., 2025). Hal tersebut meninggalkan jejak digital berupa konten disinformasi yang sulit dihapus. *Deepfake* yang juga digunakan untuk menyebarkan dan memicu emosi negatif guna memengaruhi atau memprovokasi pendukung dalam kampanye distorsi menciptakan gambaran palsu dalam benak publik (Okolie, 2023)

Di sisi lain, visual memiliki kekuatan pengaturan agenda yang memengaruhi konstruksi citra kandidat dalam politik tradisional, menciptakan identifikasi, dan membantu daya tarik emosional.

Namun, kehadiran *deepfake* membuat adanya risiko manipulasi rekam jejak digital untuk merendahkan tokoh politik tertentu, terutama perempuan. *Deepfake* dapat digunakan untuk menciptakan gambar eksplisit tanpa persetujuan, yang mungkin secara khusus digunakan untuk mengobjektifikasi dan merendahkan, alat untuk melakukan perundungan siber atau eksploitasi. Konten *deepfake* tersebut menjadi rekam jejak digital manipulative yang bisa tersebar tanpa kontrol, membentuk narasi digital, serta persepsi baru yang bisa mempengaruhi opini publik (A. Verma, 2025).

Kemampuan *deepfake* untuk memproduksi konten yang realistis dan eksplisit telah memicu maraknya pornografi non-konsensual dan pemerasan seksual, di mana video deepfake digunakan untuk mengeksploitasi dan melecehkan korban. Hal tersebut secara langsung telah merusak rekam jejak digital korban. Ketika *deepfake* dimanfaatkan untuk mengintimidasi seseorang, dampaknya bukan hanya ketika konten tersebut beredar. Namun, berdampak jangka panjang pada jejak digital korban. Sehingga, korban kesulitan mengontrol narasi tentang dirinya di media digital. Fungsi *recovery* ini menjadi problematic karena pemulihan reputasi digital hampir tidak mungkin dilakukan sepenuhnya karena konten yang tersebar di media digital bersifat presisten (Pratomo et al., 2018).

3.4. Ancaman Keamanan Digital Bagi Anak

Media digital dengan bebas bisa digunakan oleh siapa saja. Sehingga, pengguna media digital bersifat tidak terbatas, termasuk kelompok rentan seperti penyandang disabilitas, perempuan, anak-anak, dan lansia dari berbagai wilayah, termasuk daerah 3T (Tertinggal, Terluar, Terdepan). Oleh sebab itu, dalam pilar literasi digital berupa keamanan digital bagi anak, pengguna media sosial harus memperhatikan bahwa apa yang dilakukannya di media sosial seperti memproduksi pesan dan membagikannya bisa berdampak pada kelompok-kelompok rentan tersebut, khususnya anak-anak (Vasist & Krishnan, 2023)

Anak-anak menjadi *digital native* yang akrab dengan perangkat digital. Oleh karena itu, tidak dapat dipungkiri bahwa konektivitas antara dunia digital dan dunia anak saat ini semakin erat. Namun, berkembangnya teknologi AI dalam media digital memungkinkan terjadinya pelecehan seksual berbasis gambar yang muncul sebagai bentuk baru pelecehan di media sosial. Kemampuan teknologi AI dalam menciptakan visual yang sangat realistis memungkinkan tersebarnya misinformasi dan berita palsu yang bisa mengarahkan pada tindakan hoaks dan perundungan (Yang & Jiang, 2025).

Selain itu, muatan konten yang mengandung pelecehan seksual bisa menyasar pada pengguna media sosial dari kelompok anak-anak karena mereka tidak terlalu memahami bahaya yang bisa mengancamnya di media sosial. Hal tersebut karena anak-anak memiliki kapasitas *threat appraisal* dan *coping appraisal* yang jauh lebih rendah dari pada orang dewasa (Rogers, 1975). Sehingga, fenomena maraknya *deepfake* bukan hanya mengaburkan batas antara konten nyata dan rekayasa. Namun, juga mempersulit upaya perlindungan anak dalam media digital yang belum memiliki kapasitas literasi digital yang mumpuni (K. Verma, 2024). Oleh karena itu, diperlukan tanggung jawab yang sekaligus menjadi inti dari tuntutan *Algorithmic Accountability* kepada *platform* untuk tidak melepaskan diri dari kewajiban melindungi pengguna anak-anak dari eksposur konten *deepfake* berbahaya. Fungsi *protect* dalam CSF harus diimplementasikan secara aktif, termasuk verifikasi usia dan penyaringan konten yang menyasar kelompok rentan (Diakopoulos & Johnson, 2021).

Dari fenomena tren foto polaroid AI dengan tokoh idola, menyiratkan bahwa di tengah berbagai macam ancaman dan bentuk *deepfake*, risiko terhadap hak privasi individu semakin meningkat. Hal tersebut karena konten manipulasi buatan AI bisa dengan mudah dibuat dan

diedarkan tanpa kendali, sehingga bisa memicu dampak berantai. Oleh karena itu, kondisi tersebut mengkhawatirkan bagi anak-anak dan perempuan karena mereka rentan menjadi target eksploitasi digital (Okolie, 2023). Secara keseluruhan, keempat kategori ancaman yang telah dibahas (ancaman identitas digital, penipuan digital, rekam jejak digital, dan ancaman terhadap anak) bukan merupakan ancaman yang berdiri sendiri, melainkan saling berkaitan dalam satu rantai risiko. Pencurian identitas melalui deepfake membuka celah bagi penipuan digital, yang kemudian meninggalkan rekam jejak digital permanen yang sulit dihapus, dan pada akhirnya berdampak paling berat pada kelompok paling rentan seperti anak-anak dan perempuan. Pemahaman atas keterkaitan ini penting untuk merumuskan pendekatan perlindungan yang holistik, bukan hanya menangani satu jenis ancaman secara terisolasi.

4. PENUTUP

Berdasarkan *systematic literature review* terhadap 18 artikel kredibel menggunakan indikator keamanan digital dari Japelidi, semua artikel terkait ancaman keamanan digital dan manipulasi visual dengan AI membahas kasus-kasus penyalahgunaan *deepfake* atau foto hasil buatan *AI-Generated Image*. Penyalahgunaan tersebut menyasar pada identitas digital seseorang, baik data pribadi, maupun manipulasi citra atau privasi melalui konten non-konsensual. Data pribadi yang bisa diambil secara bebas dari hasil manipulasi foto dan suara dengan AI bisa dimanfaatkan sebagai alat penipuan digital yang memberikan keuntungan tertentu bagi pelaku. Sayangnya, rekam jejak digital yang sulit dihapus dan dengan mudah tersebar tanpa kendali pemilik konten menjadi ancaman keamanan digital tersendiri, khususnya pada *deepfake* hasil *AI-Generated Image* yang menyesatkan dan pornografi non-konsensual. Anak-anak sebagai pengguna media digital juga rentan terhadap ancaman keamanan digital dengan *deepfake* tersebut, mereka bisa menjadi sasaran dan korban kekerasan seksual. Selain ancaman pada individu, *deepfake* juga bisa menimbulkan ancaman keamanan digital berdasarkan cakupan dampak yang ditimbulkan, seperti pada level sosial dan politik melalui misinformasi serta distorsi peristiwa sejarah tertentu yang bisa memanipulasi opini publik. *Deepfake* juga bisa mengancam keamanan nasional dengan konten yang dimanipulasi berdasarkan figur publik tertentu, serta penyalahgunaan data biometrik pada infrastruktur keamanan publik.

Berdasarkan temuan penelitian tersebut, terdapat tiga rekomendasi, yaitu; memperkuat regulasi adaptif terkait penggunaan AI dan konten *deepfake,* kerangka Cybersecurity Framework dapat diadopsi sebagai standar nasional pengelolaan ancaman digital berbasis AI oleh pemerintah. Selain itu, pemerintah juga perlu mengintegrasikan materi keamanan digital dan risiko AI dalam kurikulum pendidikan formal. *Platform* digital perlu mengimplementasikan mekanisme *Algorithmic Accountability* yang transparan, termasuk audit berkala, sistem deteksi konten *deepfake,* dan kebijakan ketat terkait verifikasi usia serta perlindungan data biometrik. Masyarakat perlu meningkatkan kesadaran dalam menggunakan *platform* digital seperti AI dengan mengintegrasikan peningkatan *threat appraisal* dan *coping appraisal* secara bersamaan.

Penelitian ini memiliki beberapa limitasi. Pertama, jumlah literatur yang dianalisis terbatas pada 18 artikel dengan fokus pada database Scopus Q1-Q2, sehingga mungkin belum menjangkau seluruh perspektif, khususnya dari jurnal nasional terakreditasi SINTA yang membahas isu lokal Indonesia secara lebih mendalam. Kedua, perkembangan teknologi AI yang sangat cepat membuat temuan penelitian ini perlu diperbarui secara berkala. Oleh karena itu, peneliti selanjutnya diharapkan dapat melakukan studi empiris mengenai efektivitas modul literasi digital keamanan berbasis AI bagi kelompok rentan seperti anak-anak dan perempuan di Indonesia, serta mengeksplorasi regulasi yang lebih adaptif sesuai dinamika perkembangan teknologi terkini.

DAFTAR PUSTAKA

Adi. (2022, January 20). *Meski Indeks Literasi Digital Membaik, Keamanan Digital Punya Nilai Buruk*. PasaRDana. https://pasardana.id/news/2022/1/20/meski-indeks-literasi-digital-membaik-keamanan-digital-punya-nilai-buruk/

Adikara, G. J., Kurnia, N., Adhrianti, L., Astuty, S., Wijayanto, X. A., Desiana, F., & Astuti, S. I. (2021a). *Aman Bermedia Digital*. Kementrian Komunikasi dan Informatik Republik Indonesia, Jepilidi, SIberkreasi.

Adikara, G. J., Kurnia, N., Adhrianti, L., Astuty, S., Wijayanto, X. A., Desiana, F., & Astuti, S. I. (2021b). *Aman Bermedia Digital*. Kementrian Komunikasi dan Informatik Republik Indonesia, Jepilidi, SIberkreasi.

Ali, M., Fernando, Z. J., Huda, C., & Mahmutarom, M. (2025). Deepfakes and Victimology: Exploring the Impact of Digital Manipulation on Victims. *Substantive Justice International Journal of Law*, *8*(1), 1–12. https://doi.org/10.56087/substantivejustice.v8i1.306

Alif Ilham Fajriadi. (2025). *Mengapa Skor Literasi Digital Anjlok | tempo.co*. Tempo. https://www.tempo.co/ekonomi/kenapa-literasi-digital-indonesia-anjlok-2077770

Angel Kartika. (2025, August 19). *Apakah Kita Tidak Bisa Percaya Lagi pada Apapun yang Kita Lihat?: AI-Generated Images Nowadays - Program Studi Desain Komunikasi Visual Universitas Ciputra Surabaya*. Ciputra.Ac.Id. https://www.ciputra.ac.id/vcd/apakah-kita-tidak-bisa-percaya-lagi-pada-apapun-yang-kita-lihat-ai-generated-images-nowadays/

Chua, M., Yeow, C. Y. T., Chwee, C. X. Y., Leong, S. M., & Phan, R. C. W. (2025). Securing Face ID: Privacy Preservation for Non-Retentive Face Recognition System. *IEEE Region 10 Annual International Conference, Proceedings/TENCON*, 1028–1032. https://doi.org/10.1109/TENCON66050.2025.11374999

Culnan, M. J., & Armstrong, P. K. (1999). Information Privacy Concerns, Procedural Fairness and Impersonal Trust: An Empirical Investigation. *Organization Science*, *10*(1), 104–115.

Diakopoulos, N., & Johnson, D. (2021). Anticipating and addressing the ethical implications of deepfakes in the context of elections. *New Media and Society*, *23*(7), 2072–2098. https://doi.org/10.1177/1461444820925811

Enjellina, Beyan, E. V. P., & Anastasya Gisela Cinintya Rossy. (2023). Review of AI Image Generator: Influences, Challenges, and Future Prospects for Architectural Field. *Journal of Artificial Intelligence in Architecture*, *2*(1), 53–65. https://doi.org/10.24002/jarina.v2i1.6662

Firyalfatin. (2025, September 12). *Deepfake Kian Mengkhawatirkan, Pemerintah Minta Platform Global Hadirkan Fitur Deteksi Konten AI*. Hukumonline.Com. https://www.hukumonline.com/berita/a/deepfake-kian-mengkhawatirkan--pemerintah-minta-platform-global-hadirkan-fitur-deteksi-konten-ai-lt68c3c6073feff/

Furizal, F., Ma'arif, A., Maghfiroh, H., Suwarno, I., Prayogi, D., Lonang, S., & Sharkawy, A.-N. (2025). Social, legal, and ethical implications of AI-Generated deepfake pornography on digital platforms: A systematic literature review. *Social Sciences and Humanities Open*, *12*. https://doi.org/10.1016/j.ssaho.2025.101882

Herdian, A., & Sumarwan, U. (2025). Analisis Kriminologi Deepfake Melalui Media Sosial Berdasarkan Teori Rational Choice. *IKRA-ITH HUMANIORA: Jurnal Sosial Dan Humaniora*, *9*(1), 323–331.

Wohl, R., & Horton, D. (1956). Mass Communication and Para-Social Interaction: Observations on Intimacy at a Distance. Psychiatry. *Psychiatry*, *3*(1), 215-229.

Infipop. (2025). *Ikut-ikutan tren itu gapapa, cuma, ya, jangan sampai merugikan orang lain dan diri kita. Lebih bijak lagi ya #Whatspop*. Infipop Berkolaborasi Dengan Girls Beyond. https://www.instagram.com/infipop.id/p/DPeG7mnE9WD/

Itsna Hidayatul Khusna, S. P. (2019). Deepfake, Tantangan Baru Untuk Netizen Deepfake, a New Challenge for Netizen. *Agustus 1945 Jakarta 1 Promedia*, *5*(2), 1–24.

Jada, I., & Mayayise, T. O. (2024). The impact of artificial intelligence on organisational cyber security: An outcome of a systematic literature review. *Data and Information Management*, *8*(2), 100063. https://doi.org/10.1016/J.DIM.2023.100063

John W. Creswell, J. D. C. (2018). Research Design: Qualitative, Suantitaive, and Mixed Method Approaches. In *Writing Center Talk over Time: A Mixed-Method Study* (fith). SAGE Publication Inc. https://doi.org/10.4324/9780429469237

Kharvi, P. L. (2024). Understanding the Impact of AI-Generated Deepfakes on Public Opinion, Political Discourse, and Personal Security in Social Media. *IEEE Security and Privacy*, *22*(4), 115–122.

https://doi.org/10.1109/MSEC.2024.3405963;PAGE:STRING:ARTICLE/CHAPTER
Kouchaksaraei, H. R., & Karl, H. (2004). Procedures for Performing Systematic Reviews. In *Keele University Technical Report TR/SE-0401*. https://doi.org/10.1145/3328905.3332505
Kumar, P., & Singh, D. (2025). A Study on Deepfake Dangers in the Digital Age: Social Media Misinformation and the Youth at Risk. In *Innovations in Cryptocrime and Financial Fraud* (pp. 341–362). IGI Global. https://doi.org/10.4018/979-8-3373-0675-9.ch010
Li, W., & Zhao, H. (2024). "It's Up to Me Whether I Do—Or Don't—Watch Deepfakes": Deepfakes and Behavioral Intention. *SAGE Open*, *14*(4). https://doi.org/10.1177/21582440241302282
Lowenstein, H., Steinfeld, N., & Rosenberg, H. (2025). The Ethical Implications of Prosocial Synthetic Resuscitation: Analysing User Comments to a Deepfake Campaign Addressing Intimate Partner Violence. *Journal of Creative Communications*, *20*(1), 23–40. https://doi.org/10.1177/09732586241276984
Lowenstein-Barkai, H., Steinfeld, N., & Rosenberg, H. (2025). The Ethical Implications of Prosocial Synthetic Resuscitation: Analysing User Comments to a Deepfake Campaign Addressing Intimate Partner Violence. *Journal of Creative Communications*, *20*(1), 23–40. https://doi.org/10.1177/09732586241276984
Lumingkewas, E. V., & Halim, E. (2026). Deepfake Fraud Increased: Study of Indonesian Gen Z's Trust and Acceptance toward Deepfake. *2026 6th International Conference on Advanced Research in Computing: Responsible AGI: Balancing Intelligence, Responsibility and Sustainability, ICARC 2026 - Conference Proceedings*, 1–6. https://doi.org/10.1109/ICARC68737.2026.11454120
Maniyal, V., & Kumar, V. (2024). Unveiling the Deepfake Dilemma: Framework, Classification, and Future Trajectories. *IT Professional*, *26*(2), 32–38. https://doi.org/10.1109/MITP.2024.3369948
Momeni, M. (2025). Artificial Intelligence and Political Deepfakes: Shaping Citizen Perceptions Through Misinformation. *Journal of Creative Communications*, *20*(1), 41–56. https://doi.org/10.1177/09732586241277335
Naufal Majid. (2025, September 21). *Ancaman Keamanan Data Pribadi di Balik Tren Foto AI*. Tirto.Id. https://tirto.id/ancaman-keamanan-data-pribadi-di-balik-tren-foto-ai-hh5B
Nightingale, S. J., & Wade, K. A. (2022). Identifying and Minimising the Impact of Fake Visual Media: Current and Future Directions. *Memory, Mind & Media*, *1*, e15. https://doi.org/10.1017/MEM.2022.8
Nurdin, S. W., & Nugraha, I. F. (2025). Ancaman Deepfake Dan Disinformasi Berbasis Ai: Implikasi Terhadap Keamanan Siber Dan Stabilitas Nasional Indonesia. *JIMR: Journal Of International Multidisciplinary Research*, *4*(1), 73. http://azramedia-indonesia.azramediaindonesia.com/index.php/JIMR/indexDOI:https://doi.org/10.62668/jimr.v4i01.1551
Okolie, C. (2023). Artificial Intelligence-Altered Videos (Deepfakes), Image-Based Sexual Abuse, and Data Privacy Concerns. *Journal of International Women's Studies*, *25*(2). https://vc.bridgew.edu/jiws/vol25/iss2/11
Pongoh, A. G., Fahreza, R. A., Al Kindi, B., Pribadi, F. S., & Aprilianto, R. A. (2024). Systematic Literature Review (SLR): Dampak Pemanfaatan Artificial Intelligence untuk Meningkatkan Cyber Security. *Cyber Security Dan Forensik Digital*, *7*(1), 34–41. https://doi.org/10.14421/CSECURITY.2024.7.1.4486
Pratomo, B. A., Marwan, A., Wibowo, S., Kariadi, M. T., & Faridah, S. (2018). Kerangka Kerja untuk Meningkatkan Keamanan Siber Infrastruktur Kritis. *National Institute of Standards and Technology*, *1.1*(April), 1-51Pratomo, B. A., Marwan, A., Wibowo, S., Kariad.
PRISMA. (2020). *PRISMA 2020 flow diagram*. Prisma Statement. https://www.prisma-statement.org/prisma-2020-flow-diagram
Purwadi, A., Serfiyani, C. Y., & Serfiyani, C. R. (2022). Legal Landscape on National Cybersecurity Capacity in Combating Cyberterrorism Using Deep Fake Technology in Indonesia. *International Journal of Cyber Criminology*, *16*(1), 123–140. https://doi.org/10.5281/zenodo.4766560
Raka Maheswara. (2025, February 15). *Intensitas Penggunaan AI di Indonesia Berdasarkan Jenis Aktivitas – Dataloka.id*. Dataloka. https://dataloka.id/humaniora/2691/intensitas-penggunaan-ai-di-indonesia-berdasarkan-jenis-aktivitas/
Riti, Y. F., Teknik, F., Katolik, U., & Cendika, D. (2024). Penyuluhan dan Edukasi Identifikasi Modus Penipuan Mellaui Media Sosial Bagi Masyarakat. *Jurnal CSDS*, *3*(1), 245–252.
Roe, J., & Perkins, M. (2024). Deepfakes and Higher Education: A Research Agenda and Scoping

Review of Synthetic Media. *Journal of University Teaching and Learning Practice* , *21*(10). https://doi.org/10.53761/2y2np178

Rogers, R. W. (1975). A Protection Motivation Theory of Fear Appeals and Attitude Change. *The Journal of Psychology*, *91*(1), 93–114. https://doi.org/10.1080/00223980.1975.9915803

Sanjaya, S., Fitriati, L. R., Hakim, M. A., Yasin, M. Y., & Maesaroh, S. S. (2024). Analisis Literasi Keamanan Digital Bagi Mahasiswa Universitas Pendidikan Indonesia Kampus Tasikmalaya: Tingkat Pengetahuan, Kepercayaan, dan Kewaspadaan. *Innovative: Journal Of Social Science Research*, *4*(3), 8205–8216. http://j-innovative.org/index.php/Innovative/article/view/10648%0Ahttps://j-innovative.org/index.php/Innovative/article/download/10648/7791

Sepriano, L. P. I. K. R. S. Y. Z. K. H. P. D. P. W. A. M. T. I. G. I. Su. H. P. G. C. S. (2023). *Fenomena Artificial Intelligence (AI)*. PT. Sonpedia Publishing Indonesia.

Simon Kemp. (2025, February 5). *Digital 2025: Global Overview Report* . Data Reportal. https://datareportal.com/reports/digital-2025-global-overview-report

Sulianta, F. (2025a). *Deepfake: Teknologi, Dampak, dan Potensinya*.

Sulianta, F. (2025b). *Jejak Digital: Memahami dan Mengelola Reputasi di Era Digital*. Universitas Widyatama. https://www.researchgate.net/publication/389505333_JEJAK_DIGITAL_MEMAHAMI_DAN_MENGELOLA_REPUTASI_DI_ERA_DIGITAL

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*, *8*, 1–10. https://doi.org/10.1186/1471-2288-8-45

Vasist, P. N., & Krishnan, S. (2023). Engaging with Deepfakes: A Meta-synthesis from the Perspective of Social Shaping of Technology Theory. *Internet Research*, *33*(5), 1670–1726. https://doi.org/10.1108/INTR-06-2022-0465

Verma, A. (2025). Deepfakes and the Crisis of Digital Authenticity: Ethical Challenges in the Age of Synthetic Media. *Journal of Information, Communication and Ethics in Society*. https://doi.org/10.1108/JICES-04-2025-0083/1271845

Verma, K. (2024). Digital Deception: The Impact of Deepfakes on Privacy Rights. *Lex Scientia Law Review*, *8*(2), 859–896. https://doi.org/10.15294/LSLR.V8I2.13749

Vidi, A. (2024, January 14). *Cek Fakta: Hoaks Baim Wong Bagikan Giveaway Rp 50 Juta dengan Cara Daftar Via Whatsapp*. Liputan6.Com. https://www.liputan6.com/cek-fakta/read/5504720/cek-fakta-hoaks-baim-wong-bagikan-giveaway-rp-50-juta-dengan-cara-daftar-via-whatsapp

Wahab, A. (2025). Futures of Deepfake and Society: Myths, Metaphors, and Future Implications for a Trustworthy Digital Future. *Futures*, *173*, 103672. https://doi.org/10.1016/J.FUTURES.2025.103672

Wati, D. S., Nurhaliza, S., Sari, M. W., & Amallia, R. (2024). Dampak Cyber Crime Terhadap Keamanan Nasional dan Strategi Penanggulangannya : Ditinjau Dari Penegakan Hukum. *Jurnal Bevinding*, *02*(01), 44–55.

Wpj. (2025, November 20). *Waspada Scam Metode Voice Spoofing, Bisa Tiru Suara Keluarga*. Cnnindonesia.Com. https://www.cnnindonesia.com/teknologi/20251119164130-185-1297187/waspada-scam-metode-voice-spoofing-bisa-tiru-suara-keluarga

Xiaobil. (2025, October 2). *Waspada Deepfake & Voice Scam: Ketika Teknologi Menjadi Alat Penipuan Canggih – Politeknik Penerbangan Palembang*. Poltekbangplg.Ac.Id. https://poltekbangplg.ac.id/waspada-deepfake-voice-scam-ketika-teknologi-menjadi-alat-penipuan-canggih/#google_vignette

Yang, Z., & Jiang, Y. (2025). How Differential Risk Perception Influences Users' Positive Information Behaviour Toward AIGC Technologies: An Analysis Based on the Cognitive-Affective-Conative (CAC) Model. *International Journal of Human-Computer Interaction*. https://doi.org/10.1080/10447318.2025.2526582

Yanwardahana, E. (2025, August 8). *Besaran Gaji Guru Viral di Medsos, Sri Mulyani Buka Suara*. Cnbcindonesia.Com. https://www.cnbcindonesia.com/news/20250808091949-4-656281/besaran-gaji-guru-viral-di-medsos-sri-mulyani-buka-suara

Zheng, G., Shu, J., & Li, K. (2025). Regulating Deepfakes between Lex Lata and Lex ferenda—a comparative analysis of regulatory approaches in the U.S., the EU and China. *Crime, Law and Social Change*, *83*(1). https://doi.org/10.1007/s10611-024-10197-z