Journal of Computer Science and Computational Intelligence
May 2026 | Vol. 2 | No. 1 | Page 19-30
E-ISSN: 3109-7383 P-ISSN: 3110-8393
DOI: https://doi.org/10.30598/algorithm.v2i1.19-30

Deepfake sebagai Ancaman Baru dalam Keamanan Siber: Tinjauan Literatur mengenai Risiko terhadap Identitas Digital, Rekayasa Sosial, dan Kepercayaan Informasi

(Deepfakes as a New Threat in Cybersecurity: A Literature Review on Risks to Digital Identity, Social Engineering, and Information Trust)

Victor Eric Pattiradjawane[1*], Juins Carlo Radjulan[2], Febry de Fretes[3]

[1,2] Program Studi Ilmu Komputer, Universitas Pattimura, Ambon, Indonesia
[3]UPA Teknologi, Informasi dan Komunikasi, Universitas Pattimura, Ambon, Indonesia
[*]Corresponding Author: ***victor.pattiradjawane@lecturer.unpatti.ac.id**

Manuscript submitted:	*Manuscript revision:*	*Accepted for publication:*
15th April 2026	*20th April 2026*	*5th May 2026*

Abstract

The development of Artificial Intelligence (AI) technology, particularly Generative AI, has enabled the creation of increasingly realistic synthetic content, one example of which is deepfake technology. Deepfake is a deep learning-based technique capable of manipulating faces, voices, and videos to mimic real individuals with a high degree of similarity. While it has positive potential in the entertainment, education, and creative industries, this technology also presents various new risks in the cybersecurity domain. This study aims to examine the threat of deepfakes to digital identity, social engineering, and information trust through a literature review approach. The method used is a narrative literature review of various scientific publications, industry reports, and policy documents published between 2022 and 2026. The results of the study indicate that deepfakes have developed into an effective tool for identity forgery, voice and video-based fraud, the spread of disinformation, and the manipulation of public opinion. Furthermore, the increasing quality of synthetic content has given rise to an information trust crisis, a decline in the public's ability to distinguish between genuine and digitally manipulated information. This research also identifies various mitigation strategies, including improving digital literacy, developing deepfake detection technology, digital watermarking, content provenance, and strengthening regulations related to the use of AI. The research findings are expected to contribute to understanding the cybersecurity risks arising from the development of deepfake technology and form the basis for more effective mitigation efforts.

Keywords: *Artificial Intelligence; Cybersecurity; Deepfake; Digital Identity; Social Engineering, Disinformation.*

How to cite this article:
V. E. Pattiradjawane, J. C. Radjulan, and F. de Fretes, "Deepfake sebagai Ancaman Baru dalam Keamanan Siber: Tinjauan Literatur mengenai Risiko terhadap Identitas Digital, Rekayasa Sosial, dan Kepercayaan Informasi", *algorithm*, vol. 2, no. 1, pp.19-30, May 2026.

Journal homepage: https://ojs3.unpatti.ac.id/index.php/algorithm
Research Article Open Access

1. PENDAHULUAN

Perkembangan teknologi kecerdasan buatan (*Artificial Intelligence* atau AI) dalam beberapa tahun terakhir telah membawa perubahan signifikan dalam berbagai sektor kehidupan, termasuk pendidikan, kesehatan, industri kreatif, bisnis, dan keamanan siber. Kemajuan teknologi AI modern, khususnya pada bidang *Generative Artificial Intelligence*, memungkinkan sistem komputer menghasilkan berbagai bentuk konten digital seperti teks, gambar, suara, dan video dengan tingkat realisme yang semakin tinggi. Kemampuan tersebut didukung oleh perkembangan model *deep learning*, termasuk *Generative Adversarial Networks* (GANs), *Diffusion Models*, dan berbagai model generatif lainnya yang mampu mempelajari karakteristik data secara mendalam dan menghasilkan representasi sintetis yang menyerupai data asli [1], [2].

Salah satu implementasi Generative AI yang berkembang sangat pesat adalah teknologi *deepfake*. *Deepfake* merupakan teknik manipulasi digital yang memanfaatkan algoritma pembelajaran mendalam untuk menghasilkan atau memodifikasi wajah, suara, maupun video seseorang sehingga tampak autentik dan sulit dibedakan dari konten asli. Teknologi ini memungkinkan pelaku menciptakan representasi digital seseorang yang dapat berbicara, bergerak, atau melakukan tindakan tertentu tanpa keterlibatan individu yang bersangkutan. Pada awal perkembangannya, *deepfake* banyak digunakan untuk kebutuhan hiburan, industri perfilman, pemasaran digital, serta rekonstruksi visual dalam berbagai kegiatan kreatif. Namun, seiring meningkatnya kualitas dan aksesibilitas teknologi tersebut, *deepfake* juga mulai dimanfaatkan untuk tujuan yang bersifat merugikan dan bahkan kriminal [3], [4].

Dalam perspektif keamanan siber, *deepfake* telah berkembang menjadi ancaman baru yang tidak lagi terbatas pada manipulasi visual semata. Teknologi ini mulai dimanfaatkan sebagai sarana pemalsuan identitas digital (*digital identity impersonation*), penipuan berbasis rekayasa sosial (*social engineering*), penyebaran disinformasi, hingga manipulasi opini publik. Berbagai laporan keamanan menunjukkan peningkatan penggunaan teknologi AI generatif dalam serangan siber modern, termasuk *voice cloning*, pemalsuan video konferensi, dan pembuatan identitas sintetis yang digunakan untuk memperoleh akses terhadap sistem atau layanan tertentu [5], [6]. Kondisi ini menunjukkan bahwa perkembangan *deepfake* tidak hanya menjadi isu teknologi, tetapi juga telah menjadi bagian dari lanskap ancaman keamanan siber global.

Salah satu risiko utama yang ditimbulkan oleh *deepfake* adalah ancaman terhadap identitas digital. Identitas digital merupakan representasi elektronik yang digunakan untuk mengenali dan memverifikasi individu dalam lingkungan digital [7], [8]. Dalam berbagai layanan modern seperti perbankan elektronik, layanan pemerintah digital, sistem pendidikan daring, dan media sosial, identitas digital menjadi komponen penting dalam proses autentikasi dan otorisasi pengguna. Kehadiran teknologi *deepfake* memungkinkan pelaku melakukan impersonasi terhadap individu tertentu melalui manipulasi wajah maupun suara sehingga berpotensi melemahkan mekanisme verifikasi identitas yang selama ini digunakan. Situasi ini menjadi semakin kompleks ketika teknologi *deepfake* dikombinasikan dengan data pribadi yang diperoleh dari media sosial atau hasil kebocoran data (*data breach*).

Selain ancaman terhadap identitas digital, *deepfake* juga memperkuat efektivitas serangan rekayasa sosial. Dalam banyak kasus, keberhasilan serangan siber tidak hanya bergantung pada eksploitasi kelemahan teknis, tetapi juga pada kemampuan pelaku memanipulasi perilaku manusia. Teknologi *deepfake* memberikan dimensi baru dalam serangan rekayasa sosial karena memungkinkan pelaku meniru suara, wajah, dan perilaku seseorang secara realistis. Serangan *voice phishing* (*vishing*), penipuan berbasis panggilan video, serta *Business Email Compromise* (BEC) yang didukung teknologi AI menjadi contoh nyata bagaimana *deepfake* digunakan untuk meningkatkan tingkat keberhasilan serangan terhadap individu maupun organisasi [5], [9].

Di samping ancaman terhadap individu dan organisasi, penyebaran konten *deepfake* secara masif juga menimbulkan dampak sosial yang lebih luas berupa menurunnya tingkat kepercayaan

masyarakat terhadap informasi digital. Fenomena ini sering disebut sebagai information trust crisis, yaitu kondisi ketika masyarakat semakin sulit membedakan antara informasi yang autentik dan informasi yang telah dimanipulasi menggunakan teknologi AI. Apabila tidak diantisipasi dengan baik, kondisi tersebut dapat menyebabkan meningkatnya penyebaran hoaks, disinformasi, polarisasi sosial, serta menurunnya kepercayaan terhadap media, institusi publik, dan berbagai sumber informasi digital lainnya [10], [11]. Dalam konteks demokrasi dan komunikasi publik, fenomena ini menjadi tantangan serius karena dapat memengaruhi persepsi masyarakat terhadap suatu peristiwa, tokoh publik, maupun kebijakan tertentu.

Berbagai penelitian terdahulu umumnya berfokus pada pengembangan algoritma deteksi *deepfake* menggunakan pendekatan *machine learning*, *computer vision*, dan *digital forensics* [4], [6]. Meskipun penelitian tersebut memberikan kontribusi penting dalam upaya identifikasi konten sintetis, kajian yang membahas *deepfake* sebagai ancaman multidimensi dalam perspektif keamanan siber masih relatif terbatas. Padahal, pemahaman mengenai dampak *deepfake* terhadap identitas digital, rekayasa sosial, dan kepercayaan informasi sangat penting untuk mendukung pengembangan strategi mitigasi yang lebih komprehensif.

Berdasarkan kondisi tersebut, penelitian ini bertujuan untuk melakukan tinjauan literatur mengenai perkembangan teknologi *deepfake* serta risiko yang ditimbulkannya terhadap identitas digital, rekayasa sosial, dan kepercayaan informasi. Selain itu, penelitian ini juga mengidentifikasi berbagai strategi mitigasi yang dapat diterapkan untuk mengurangi dampak negatif teknologi *deepfake* dalam ekosistem digital modern. Hasil penelitian diharapkan dapat memberikan pemahaman yang lebih komprehensif mengenai posisi *deepfake* sebagai salah satu ancaman baru dalam keamanan siber sekaligus menjadi referensi bagi akademisi, praktisi, pembuat kebijakan, dan masyarakat dalam menghadapi perkembangan teknologi AI yang semakin kompleks.

2. METODE PENELITIAN

2.1 Pendekatan Penelitian

Penelitian ini menggunakan metode *Narrative Literature Review (NLR)* untuk mengkaji perkembangan teknologi *deepfake* dan implikasinya terhadap keamanan siber. Menurut Snyder[12], literature review dapat digunakan untuk mengidentifikasi, mengevaluasi, dan mensintesis hasil penelitian terdahulu guna membangun pemahaman yang lebih komprehensif terhadap suatu topik penelitian. Pendekatan ini dipilih karena memungkinkan peneliti melakukan analisis, sintesis, dan interpretasi terhadap berbagai hasil penelitian terdahulu guna memperoleh pemahaman yang komprehensif mengenai fenomena yang sedang berkembang [13].

Berbeda dengan *Systematic Literature Review (SLR)* yang menerapkan prosedur seleksi yang sangat ketat dan terstruktur, *Narrative Literature Review* (NLR) lebih menekankan pada sintesis konseptual untuk menjelaskan hubungan antar konsep, tren penelitian, dan perkembangan pengetahuan dalam suatu bidang kajian [14]. Narrative Literature Review juga memungkinkan peneliti mengintegrasikan berbagai hasil penelitian untuk menghasilkan pemahaman konseptual yang lebih luas terhadap suatu fenomena [15].

2.2 Sumber Data

Data penelitian diperoleh dari berbagai sumber literatur ilmiah yang relevan dengan topik *deepfake* dan keamanan siber. Database yang digunakan meliputi: Google Scholar, IEEE Xplore, ACM Digital Library, SpringerLink, ScienceDirect, Scopus Preview, dan laporan resmi lembaga keamanan siber internasional. Kata kunci pencarian yang digunakan meliputi: *Deepfake*, *Artificial Intelligence*, *Generative* AI, *Cybersecurity*, *Digital Identity, Identity Fraud, Social Engineering, Information Trust, Misinformation, Disinformation, Voice Cloning*. Pemilihan rentang publikasi 2022–2026 dilakukan untuk menangkap perkembangan terkini teknologi Generative AI dan *deepfake* yang mengalami pertumbuhan signifikan dalam beberapa tahun terakhir[6], [10].

Dalam penelitian ini, kami menerapkan kriteria inklusi untuk menerapkan literatur yang digunakan. Kriteria inklusi tersebut meliputi jenis literatur yang dipublikasikan pada periode 2022–2026, membahas *deepfake*, Generative AI, atau konten sintetis, berkaitan dengan keamanan siber, identitas digital, rekayasa sosial, atau disinformasi, dipublikasikan pada jurnal, prosiding, laporan industri, atau laporan lembaga resmi yang memiliki kredibilitas tinggi, tersedia dalam bahasa Inggris atau bahasa Indonesia. Sebaliknya, kriteria eksklusi digunakan terhadap literatur yang dianggap tidak memenuhi syarat penelitian ini. Kriteria eksklusi tersebut diantaranya tidak berkaitan dengan topik *deepfake*, hanya membahas implementasi AI tanpa hubungan dengan keamanan siber, merupakan artikel populer tanpa sumber ilmiah yang jelas, dan dipublikasikan sebelum tahun 2022.

Tabel 1. Kriteria Seleksi Literatur

Kriteria Inklusi	Kriteria Eksklusi
Publikasi tahun 2022–2026	Publikasi sebelum 2022
Membahas *deepfake* atau Generative AI	Tidak membahas *deepfake*
Berkaitan dengan cybersecurity	Hanya membahas AI umum
Jurnal, prosiding, laporan resmi	Artikel populer/blog
Bahasa Indonesia atau Inggris	Bahasa lain tanpa terjemahan
Full text tersedia	Hanya abstrak tersedia

Proses pencarian literatur menghasilkan 87 publikasi yang relevan dari berbagai basis data akademik. Setelah dilakukan proses seleksi berdasarkan judul, abstrak, dan kesesuaian dengan kriteria inklusi, sebanyak 34 artikel tersebut dianggap mewakili perkembangan penelitian terkini mengenai *deepfake*, identitas digital, rekayasa sosial, dan kepercayaan informasi sehingga digunakan sebagai dasar analisis dalam penelitian ini.

Tabel 2 Distribusi Literatur Berdasarkan Tema

Tema	Jumlah Artikel
Evolusi *Deepfake*	8
Identitas Digital	7
Rekayasa Sosial	6
Disinformasi dan Kepercayaan Informasi	7
Strategi Mitigasi	6
Total	34

2.3 Tahapan Penelitian

Penelitian ini dilaksanakan melalui tujuh tahapan utama yang dirancang secara sistematis untuk memastikan proses kajian berlangsung komprehensif dan menghasilkan temuan yang valid. Tahap pertama adalah identifikasi literatur, yaitu mengumpulkan berbagai artikel ilmiah, laporan penelitian, dan dokumen relevan dari sejumlah basis data akademik yang memiliki keterkaitan dengan topik *deepfake*, keamanan siber, identitas digital, serta ancaman disinformasi.

Selanjutnya, pada tahap seleksi literatur dilakukan proses penyaringan berdasarkan judul, abstrak, kata kunci, dan tingkat relevansi terhadap tujuan penelitian sehingga hanya sumber yang memenuhi kriteria inklusi yang digunakan dalam analisis. Literatur yang telah terpilih kemudian memasuki tahap klasifikasi, yaitu pengelompokan artikel ke dalam beberapa tema utama yang meliputi evolusi teknologi *deepfake*, identitas digital, rekayasa sosial, disinformasi, dan strategi mitigasi. Pada tahap analisis dan sintesis, setiap artikel ditelaah secara mendalam untuk mengidentifikasi temuan-temuan penting, pola penelitian yang berkembang, kesamaan dan

perbedaan hasil studi, serta hubungan antar tema yang muncul dalam literatur. Hasil proses analisis dan sintesis selanjutnya digunakan untuk mengidentifikasi berbagai risiko keamanan siber yang ditimbulkan oleh teknologi *deepfake*. Berdasarkan risiko yang teridentifikasi, penelitian merumuskan berbagai strategi mitigasi yang direkomendasikan dalam literatur sebelum menyusun kesimpulan akhir penelitian.

Tahap terakhir adalah penyusunan kesimpulan, yang dilakukan dengan merangkum temuan utama penelitian, menjelaskan implikasinya terhadap keamanan siber dan perlindungan identitas digital, serta merumuskan rekomendasi strategi mitigasi yang dapat diterapkan untuk menghadapi ancaman yang ditimbulkan oleh teknologi *deepfake* di era digital.

Gambar 1 Diagram metodologi

3. HASIL DAN PEMBAHASAN

3.1 Evolusi Teknologi *Deepfake* Berbasis Artificial Intelligence

Teknologi *deepfake* berkembang seiring dengan kemajuan kecerdasan buatan, khususnya pada bidang pembelajaran mendalam (*deep learning*) dan teknologi generatif. Dalam beberapa tahun terakhir, kualitas konten yang dihasilkan mengalami peningkatan yang sangat pesat. Jika pada tahap awal video atau suara hasil manipulasi masih relatif mudah dikenali karena adanya ketidaksesuaian visual maupun audio, saat ini hasil yang dihasilkan semakin realistis dan sering kali sulit dibedakan dari konten asli.

Pada dasarnya, *deepfake* dibuat dengan memanfaatkan sejumlah besar data berupa foto, video, atau rekaman suara seseorang. Data tersebut digunakan untuk mempelajari karakteristik tertentu seperti ekspresi wajah, gerakan bibir, pola bicara, dan intonasi suara. Semakin banyak data yang tersedia, semakin tinggi pula tingkat kemiripan hasil yang dapat dihasilkan. Perkembangan berbagai layanan generatif yang mudah diakses turut mempercepat penyebaran teknologi ini. Jika sebelumnya pembuatan *deepfake* memerlukan kemampuan teknis yang cukup tinggi, kini berbagai aplikasi dan layanan daring memungkinkan pengguna menghasilkan video maupun suara sintetis dengan proses yang jauh lebih sederhana. Kondisi ini membuat teknologi *deepfake* tidak lagi terbatas pada kalangan peneliti atau pengembang teknologi.

Dari sisi keamanan siber, kemudahan tersebut membawa konsekuensi baru. Ancaman yang sebelumnya didominasi oleh *malware, phishing*, dan eksploitasi kerentanan sistem kini mulai bergeser ke arah penyalahgunaan identitas digital. *Deepfake* dapat digunakan untuk meniru individu tertentu sehingga pelaku memiliki peluang lebih besar untuk memperoleh kepercayaan korban dan menjalankan berbagai bentuk penipuan.

Perkembangan terbaru juga menunjukkan munculnya kemampuan manipulasi secara langsung dalam komunikasi daring. Wajah dan suara dapat dimodifikasi secara *real-time* selama panggilan video berlangsung. Di tengah meningkatnya penggunaan platform konferensi video dalam dunia kerja, pendidikan, dan pemerintahan, kemampuan tersebut berpotensi menimbulkan risiko baru yang perlu diantisipasi. Perubahan ini menunjukkan bahwa dampak *deepfake* tidak hanya berkaitan dengan aspek teknis, tetapi juga menyentuh aspek sosial, psikologis, dan keamanan informasi. Karena itu, pemahaman mengenai perkembangan teknologi *deepfake* menjadi penting sebagai langkah awal untuk memahami berbagai risiko yang muncul serta strategi mitigasi yang dapat diterapkan.

Untuk memperoleh gambaran yang lebih komprehensif mengenai perkembangan penelitian *deepfake* dalam perspektif keamanan siber, dilakukan sintesis terhadap berbagai literatur yang telah diseleksi. Hasil sintesis menunjukkan bahwa penelitian terkait *deepfake* dapat dikelompokkan ke

dalam beberapa tema utama yang meliputi evolusi teknologi *deepfake*, ancaman terhadap identitas digital, rekayasa sosial, disinformasi, teknik deteksi, hingga tata kelola dan mitigasi risiko. Ringkasan hasil sintesis literatur tersebut disajikan pada Tabel 3.

Tabel 3. Ringkasan Literatur Terkait *Deepfake* dan Keamanan Siber

No	Aspek Kajian	Fokus Penelitian	Temuan Utama	Implikasi Cybersecurity
1	Evolusi *Deepfake*	Perkembangan model generatif	*Deepfake* semakin realistis	Deteksi semakin sulit
2	Voice Cloning	Pemalsuan suara	Risiko penipuan meningkat	Ancaman terhadap autentikasi suara
3	Synthetic Identity	Identitas buatan AI	Sulit dibedakan dari identitas asli	Risiko fraud digital
4	Social Engineering	Pemanfaatan *deepfake* dalam manipulasi korban	Tingkat keberhasilan serangan meningkat	Ancaman organisasi
5	Disinformation	Penyebaran informasi palsu	Opini publik mudah dimanipulasi	Risiko stabilitas sosial
6	Information Trust	Kepercayaan informasi digital	Muncul information trust crisis	Menurunkan kredibilitas media
7	Digital Forensics	Teknik deteksi *deepfake*	Akurasi masih bervariasi	Belum mampu mendeteksi semua kasus
8	Watermarking	Penandaan konten AI	Membantu autentikasi	Mitigasi awal
9	Content Provenance	Pelacakan asal konten	Transparansi meningkat	Memperkuat integritas informasi
10	AI Governance	Regulasi AI	Diperlukan kebijakan baru	Mendukung keamanan nasional

Berdasarkan Tabel 3, terlihat bahwa penelitian mengenai *deepfake* tidak lagi berfokus pada aspek teknis pembentukan dan deteksi konten sintetis semata. Sejumlah penelitian terbaru mulai menyoroti implikasi *deepfake* terhadap keamanan siber, khususnya yang berkaitan dengan pemalsuan identitas, rekayasa sosial, penyebaran disinformasi, dan penurunan kepercayaan terhadap informasi digital. Temuan ini menunjukkan bahwa *deepfake* telah berkembang menjadi ancaman multidimensi yang memerlukan pendekatan mitigasi yang lebih komprehensif.

3.2 Ancaman *Deepfake* terhadap Identitas Digital

Identitas digital saat ini telah menjadi bagian penting dalam kehidupan sehari-hari. Berbagai aktivitas, mulai dari transaksi perbankan, belanja daring, penggunaan media sosial, hingga akses layanan pemerintahan dan pendidikan, bergantung pada kemampuan sistem untuk mengenali dan memverifikasi identitas pengguna. Karena itu, keamanan identitas digital menjadi faktor yang sangat penting dalam menjaga kepercayaan dan keamanan berbagai layanan tersebut. Di tengah perkembangan teknologi *deepfake*, muncul tantangan baru yang tidak dapat diabaikan. Kemampuan teknologi ini untuk meniru wajah dan suara seseorang dengan tingkat kemiripan yang tinggi membuka peluang terjadinya penyalahgunaan identitas. Pelaku dapat memanfaatkan konten hasil manipulasi untuk menyamar sebagai individu tertentu, baik untuk memperoleh keuntungan pribadi maupun untuk melakukan tindakan yang merugikan pihak lain.

Kasus pemalsuan identitas tokoh publik, pimpinan organisasi, maupun figur yang memiliki pengaruh di masyarakat semakin sering ditemukan. Dengan memanfaatkan foto, video, atau rekaman suara yang beredar di internet, pelaku dapat membuat konten yang tampak meyakinkan. Akibatnya, banyak orang kesulitan membedakan mana informasi yang benar-benar berasal dari individu tersebut dan mana yang merupakan hasil manipulasi.

Risiko lainnya adalah munculnya identitas sintetis (*synthetic identity*), yaitu identitas yang dibentuk dari gabungan data asli dan data palsu. Identitas semacam ini dapat digunakan untuk membuat akun fiktif, melakukan penipuan finansial, atau melewati proses verifikasi yang diterapkan oleh berbagai platform digital. Karena tidak sepenuhnya palsu maupun sepenuhnya nyata, identitas sintetis sering kali lebih sulit dideteksi dibandingkan bentuk pemalsuan identitas konvensional. Tantangan ini menjadi semakin kompleks ketika banyak organisasi mulai mengandalkan teknologi biometrik sebagai metode autentikasi. Sistem pengenalan wajah dan suara yang sebelumnya dianggap lebih aman kini menghadapi risiko baru karena karakteristik biometrik tersebut dapat ditiru melalui teknologi *deepfake*. Kondisi ini menunjukkan bahwa penggunaan satu metode autentikasi saja tidak lagi cukup untuk menghadapi ancaman yang terus berkembang.

Oleh sebab itu, perlindungan identitas digital perlu dilakukan melalui beberapa lapisan pengamanan. Penerapan autentikasi multifaktor, pemeriksaan kontekstual terhadap aktivitas pengguna, serta mekanisme pemantauan untuk mendeteksi perilaku yang tidak wajar dapat membantu mengurangi risiko penyalahgunaan identitas. Pendekatan semacam ini menjadi semakin penting seiring meningkatnya kemampuan *deepfake* dalam meniru identitas seseorang secara lebih meyakinkan.

3.3 *Deepfake* sebagai Sarana Rekayasa Sosial (Social Engineering)

Rekayasa sosial merupakan salah satu metode yang paling sering digunakan dalam berbagai serangan siber karena memanfaatkan aspek manusia sebagai target utama. Berbeda dengan serangan yang berfokus pada kelemahan teknis sistem, rekayasa sosial bertujuan memengaruhi cara berpikir dan pengambilan keputusan korban agar bersedia memberikan informasi, akses, atau melakukan tindakan tertentu yang menguntungkan pelaku. Dalam banyak kasus, keberhasilan serangan justru lebih sering disebabkan oleh faktor manusia dibandingkan kegagalan teknologi itu sendiri. Perkembangan teknologi *deepfake* semakin memperluas peluang penyalahgunaan metode ini. Jika sebelumnya pelaku hanya mengandalkan email palsu, pesan singkat, atau panggilan telepon biasa, kini mereka dapat memanfaatkan suara maupun video yang tampak berasal dari orang yang dikenal oleh korban. Kemampuan menghasilkan konten yang menyerupai individu tertentu membuat upaya penipuan menjadi lebih meyakinkan dan lebih sulit dicurigai.

Salah satu bentuk penyalahgunaan yang banyak mendapat perhatian adalah *voice cloning attack*. Melalui teknik ini, pelaku meniru suara seseorang dengan memanfaatkan rekaman yang tersedia di media sosial, video publik, atau sumber lain yang dapat diakses secara terbuka. Suara hasil tiruan tersebut kemudian digunakan untuk menghubungi korban dan menyampaikan permintaan tertentu, seperti transfer dana, pemberian informasi rahasia, atau akses ke sistem organisasi. Karena suara yang digunakan terdengar sangat mirip dengan pemilik aslinya, korban sering kali tidak menyadari bahwa dirinya sedang menjadi target penipuan. Pemanfaatan *deepfake* tidak hanya terbatas pada suara. Dalam beberapa kasus, pelaku juga menggunakan video yang menampilkan wajah dan suara tiruan untuk melakukan komunikasi secara langsung dengan korban. Situasi ini dapat terjadi ketika pelaku menyamar sebagai pimpinan organisasi, rekan kerja, atau bahkan anggota keluarga. Semakin tinggi kualitas hasil manipulasi yang digunakan, semakin besar kemungkinan korban mempercayai identitas yang ditampilkan dalam komunikasi tersebut.

Bagi organisasi, kondisi ini menghadirkan tantangan baru dalam proses verifikasi dan pengambilan keputusan. Berbagai bentuk penipuan korporasi, termasuk *Business Email Compromise* (BEC), berpotensi menjadi lebih efektif ketika didukung oleh audio atau video yang tampak autentik.

Mekanisme verifikasi yang sebelumnya mengandalkan pengenalan suara atau komunikasi tatap muka melalui video tidak lagi dapat dianggap sepenuhnya aman. Perkembangan ini menunjukkan bahwa rekayasa sosial telah mengalami perubahan yang cukup signifikan. Serangan tidak lagi hanya memanfaatkan teks atau email palsu, tetapi juga menggabungkan suara, gambar, dan video dalam satu skenario yang dirancang untuk membangun kepercayaan korban. Oleh sebab itu, organisasi perlu menyesuaikan kebijakan keamanan informasi yang dimiliki, termasuk memperkuat prosedur verifikasi dan meningkatkan kesadaran pengguna terhadap berbagai bentuk penipuan yang memanfaatkan teknologi *deepfake*.

3.4 Dampak *Deepfake* terhadap Kepercayaan Informasi

Salah satu konsekuensi yang paling mengkhawatirkan dari perkembangan *deepfake* adalah berkurangnya kepercayaan masyarakat terhadap informasi yang beredar di ruang digital. Selama bertahun-tahun, foto, rekaman suara, dan video sering dianggap sebagai bukti yang cukup kuat untuk menggambarkan suatu peristiwa. Namun kondisi tersebut mulai berubah ketika teknologi manipulasi digital mampu menghasilkan konten yang sangat menyerupai kenyataan. Kemudahan dalam membuat dan menyebarkan *deepfake* membuka peluang yang lebih besar bagi penyebaran disinformasi dan misinformasi. Konten yang menampilkan tokoh publik, pejabat pemerintah, selebritas, atau figur berpengaruh lainnya dapat dengan cepat menarik perhatian pengguna media sosial. Dalam banyak kasus, informasi tersebut sudah terlanjur menyebar luas sebelum proses verifikasi dilakukan. Situasi ini semakin diperparah oleh mekanisme distribusi pada platform digital yang cenderung mendorong konten yang memicu emosi, kontroversi, atau rasa penasaran pengguna.

Dampak yang ditimbulkan tidak berhenti pada penyebaran informasi palsu semata. *Deepfake* juga dapat dimanfaatkan untuk membangun narasi tertentu, memperkuat propaganda, dan memengaruhi cara masyarakat memandang suatu isu. Dalam konteks politik, misalnya, video atau rekaman suara yang telah dimanipulasi berpotensi membentuk opini publik, memengaruhi preferensi pemilih, bahkan menimbulkan ketegangan sosial apabila tidak segera diklarifikasi. Kondisi tersebut melahirkan tantangan yang lebih mendasar, yaitu menurunnya kepercayaan terhadap informasi digital secara keseluruhan. Masyarakat tidak lagi hanya dihadapkan pada risiko mempercayai informasi yang salah, tetapi juga mulai meragukan informasi yang sebenarnya benar. Ketika seseorang tidak lagi yakin apakah suatu video atau rekaman suara dapat dipercaya, maka nilai informasi sebagai alat komunikasi dan sumber pengambilan keputusan ikut mengalami penurunan.

Bagi organisasi, media, maupun pemerintah, situasi ini menjadi tantangan yang tidak sederhana. Kredibilitas informasi harus dibangun melalui proses verifikasi yang lebih kuat dan lebih transparan dibandingkan sebelumnya. Karena itu, upaya meningkatkan literasi digital masyarakat perlu berjalan beriringan dengan pengembangan mekanisme verifikasi konten yang mampu membantu publik menilai keaslian suatu informasi. Langkah tersebut menjadi semakin penting untuk menjaga integritas informasi di tengah semakin luasnya penggunaan teknologi generatif.

3.5 Strategi Mitigasi Ancaman *Deepfake*

Perkembangan *deepfake* yang semakin pesat menunjukkan bahwa penanganannya tidak dapat bergantung pada satu pendekatan saja. Karakteristik ancaman yang melibatkan aspek teknologi, perilaku manusia, dan penyebaran informasi menuntut adanya strategi mitigasi yang lebih menyeluruh. Oleh sebab itu, upaya menghadapi *deepfake* perlu dilakukan melalui kombinasi pendekatan teknis, kebijakan, dan edukasi masyarakat. Dari sisi teknologi, berbagai penelitian telah menghasilkan beragam metode untuk mendeteksi konten hasil manipulasi. Teknik yang digunakan umumnya berfokus pada identifikasi ketidaksesuaian pada wajah, suara, pola gerakan, maupun karakteristik digital lain yang sulit dikenali oleh manusia secara langsung. Meskipun demikian, perkembangan teknologi pembuatan *deepfake* berlangsung sangat cepat sehingga kemampuan deteksi harus terus diperbarui. Kondisi ini menciptakan situasi yang menyerupai perlombaan

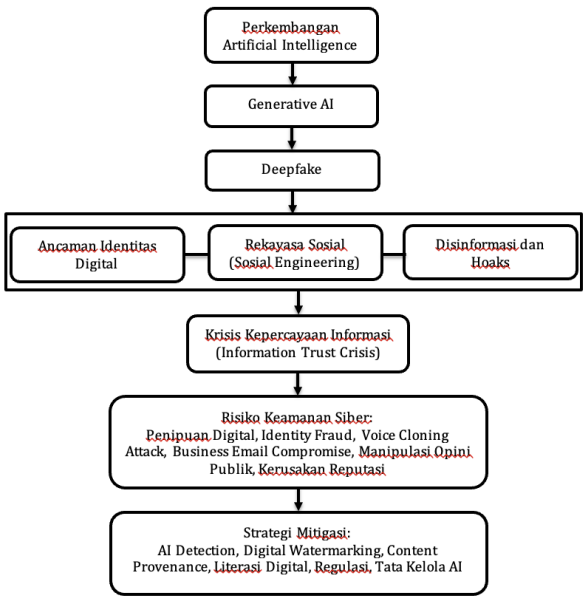
berkelanjutan antara pihak yang mengembangkan teknik manipulasi dan pihak yang berupaya mendeteksinya. Selain deteksi, perhatian juga mulai diarahkan pada upaya memastikan keaslian dan asal-usul konten digital. Konsep seperti *digital watermarking*, *content provenance*, dan *content credentials* dikembangkan untuk membantu pengguna mengetahui dari mana suatu konten berasal serta apakah konten tersebut pernah mengalami perubahan. Pendekatan ini dinilai penting karena memberikan lapisan transparansi tambahan yang dapat meningkatkan kepercayaan terhadap informasi yang beredar.

Namun teknologi saja tidak cukup. Faktor manusia tetap memegang peran yang sangat penting dalam menghadapi ancaman *deepfake*. Masyarakat perlu memiliki kemampuan untuk memeriksa sumber informasi, mengenali indikasi manipulasi digital, dan membiasakan diri melakukan verifikasi sebelum membagikan suatu konten. Semakin baik tingkat literasi digital yang dimiliki, semakin kecil peluang keberhasilan berbagai bentuk penipuan dan rekayasa sosial yang memanfaatkan *deepfake*. Dukungan dari sisi kebijakan juga menjadi bagian yang tidak kalah penting. Pemerintah, penyedia platform digital, dan berbagai lembaga terkait perlu menyusun regulasi yang mampu mengikuti perkembangan teknologi tanpa menghambat inovasi. Regulasi tersebut diharapkan dapat memberikan kejelasan mengenai batasan penggunaan teknologi, tanggung jawab para pihak yang terlibat, serta mekanisme penanganan apabila terjadi penyalahgunaan.

Pada akhirnya, mitigasi ancaman *deepfake* membutuhkan keterlibatan banyak pihak. Pemerintah, akademisi, industri teknologi, media, lembaga keamanan siber, dan masyarakat memiliki peran yang saling melengkapi dalam membangun ekosistem digital yang lebih aman. Kolaborasi tersebut menjadi faktor penting untuk menjaga keamanan identitas digital, mempertahankan integritas informasi, dan meningkatkan kepercayaan publik di tengah pesatnya perkembangan teknologi kecerdasan buatan.

3.6 Model Konseptual Ancaman *Deepfake* dalam Perspektif Keamanan Siber

Berdasarkan hasil kajian terhadap berbagai literatur yang dianalisis, penelitian ini menyusun sebuah model konseptual untuk menggambarkan keterkaitan antara perkembangan teknologi *deepfake* dan berbagai risiko keamanan siber yang muncul sebagai konsekuensinya.



Gambar 2. Diagram konseptual hubungan *deepfake* dan risiko keamanan siber

Model pada Gambar 2, menunjukkan bahwa kemajuan teknologi kecerdasan buatan, khususnya pada bidang *Generative* AI, menjadi faktor utama yang mendorong semakin berkembangnya teknologi *deepfake*. Kemampuan untuk menghasilkan wajah, suara, dan video yang menyerupai individu asli membuat teknologi ini tidak hanya dimanfaatkan untuk tujuan kreatif, tetapi juga berpotensi digunakan untuk berbagai aktivitas yang merugikan.

Hasil sintesis literatur memperlihatkan bahwa ancaman yang muncul dapat dikelompokkan ke dalam tiga area utama. Pertama, ancaman terhadap identitas digital melalui berbagai bentuk pemalsuan dan penyalahgunaan identitas. Kedua, pemanfaatan *deepfake* dalam serangan rekayasa sosial yang bertujuan memengaruhi keputusan atau tindakan korban. Ketiga, penyebaran informasi yang menyesatkan yang berpotensi membentuk persepsi publik berdasarkan informasi yang tidak akurat.

Ketiga aspek tersebut saling berkaitan dan pada akhirnya berkontribusi terhadap menurunnya tingkat kepercayaan masyarakat terhadap informasi digital. Ketika masyarakat semakin sulit membedakan antara konten asli dan hasil manipulasi, maka risiko kesalahan dalam menerima, mempercayai, maupun menyebarkan informasi menjadi semakin besar.

Dalam jangka panjang, kondisi tersebut tidak hanya berdampak pada individu, tetapi juga dapat memengaruhi organisasi, institusi publik, dan masyarakat secara luas. Karena itu, upaya mitigasi perlu dilakukan melalui berbagai pendekatan yang saling melengkapi, mulai dari pengembangan teknologi deteksi, mekanisme verifikasi konten, peningkatan literasi digital, hingga penguatan kebijakan dan tata kelola pemanfaatan kecerdasan buatan.

4. KESIMPULAN DAN SARAN

4.1 Kesimpulan

Perkembangan teknologi kecerdasan buatan dalam beberapa tahun terakhir telah mendorong lahirnya berbagai inovasi yang mampu menghasilkan konten digital dengan tingkat realisme yang semakin tinggi. Salah satu perkembangan yang paling menonjol adalah teknologi *deepfake* yang memungkinkan pembuatan gambar, suara, dan video yang menyerupai individu asli. Di satu sisi, teknologi ini membuka peluang baru pada sektor kreatif, pendidikan, dan industri digital. Di sisi lain, berbagai penelitian menunjukkan bahwa *deepfake* juga menghadirkan tantangan baru dalam bidang keamanan siber.

Hasil kajian literatur menunjukkan bahwa ancaman yang ditimbulkan oleh *deepfake* tidak hanya berkaitan dengan manipulasi konten digital, tetapi juga menyentuh aspek yang lebih luas. Ancaman tersebut mencakup penyalahgunaan identitas digital, pemanfaatan *deepfake* dalam berbagai bentuk rekayasa sosial, serta meningkatnya penyebaran disinformasi yang dapat memengaruhi persepsi masyarakat.

Kajian ini juga menunjukkan bahwa semakin tingginya kualitas konten hasil manipulasi menyebabkan batas antara informasi asli dan informasi palsu menjadi semakin sulit dikenali. Kondisi tersebut berpotensi menurunkan tingkat kepercayaan masyarakat terhadap informasi digital dan memunculkan apa yang dikenal sebagai *information trust crisis*. Apabila tidak diantisipasi dengan baik, fenomena ini dapat berdampak pada meningkatnya risiko penipuan digital, terganggunya integritas informasi, serta menurunnya kepercayaan publik terhadap berbagai institusi.

Berdasarkan temuan tersebut, dapat disimpulkan bahwa *deepfake* tidak lagi sekadar menjadi perkembangan teknologi multimedia, melainkan telah berkembang menjadi salah satu tantangan penting dalam keamanan siber modern. Penanganannya memerlukan perhatian yang serius karena dampaknya mencakup aspek teknologi, sosial, ekonomi, dan keamanan informasi secara bersamaan.

4.2 Rekomendasi

Berdasarkan hasil kajian yang dilakukan, terdapat beberapa langkah yang dapat dipertimbangkan untuk mengurangi risiko yang ditimbulkan oleh teknologi *deepfake*.

1. Peningkatan Literasi Digital
   Kemampuan masyarakat dalam mengenali dan memverifikasi informasi perlu terus ditingkatkan. Pemahaman mengenai cara kerja *deepfake*, ciri-ciri manipulasi digital, serta pentingnya melakukan verifikasi sebelum menyebarkan informasi menjadi bagian penting dalam menghadapi ancaman ini.
2. Pengembangan Teknologi Deteksi *Deepfake*
   Upaya penelitian dan pengembangan teknologi deteksi perlu terus dilakukan agar mampu mengikuti perkembangan kualitas *deepfake* yang semakin realistis. Teknologi deteksi yang lebih akurat akan membantu organisasi maupun masyarakat dalam mengidentifikasi konten yang telah dimanipulasi.
3. Penerapan Content Provenance dan Digital Watermarking
   Pelacakan asal-usul konten digital dapat menjadi salah satu mekanisme untuk meningkatkan transparansi informasi. Penerapan *content provenance* dan *digital watermarking* dapat membantu pengguna menilai keaslian suatu konten sebelum mempercayai atau menyebarkannya.
4. Penguatan Regulasi dan Tata Kelola
   Perkembangan teknologi perlu diimbangi dengan kebijakan yang mampu memberikan perlindungan terhadap penyalahgunaan teknologi tanpa menghambat inovasi. Regulasi yang jelas dapat menjadi dasar dalam penegakan hukum dan pengelolaan risiko yang berkaitan dengan *deepfake*.
5. Kolaborasi Antar Pemangku Kepentingan
   Penanganan ancaman *deepfake* tidak dapat dilakukan oleh satu pihak saja. Kerja sama antara pemerintah, akademisi, industri teknologi, media, lembaga keamanan siber, dan masyarakat diperlukan untuk membangun lingkungan digital yang lebih aman dan terpercaya.

4.3 Saran

Penelitian ini menggunakan pendekatan *Narrative Literature Review*, sehingga hasil yang diperoleh sangat dipengaruhi oleh kualitas dan cakupan literatur yang tersedia. Kajian ini juga tidak melibatkan pengujian empiris terhadap teknologi deteksi *deepfake* maupun analisis kuantitatif mengenai tingkat penyebaran *deepfake* pada platform tertentu. Oleh sebab itu, hasil penelitian perlu dipahami sebagai sintesis konseptual berdasarkan temuan-temuan yang telah dipublikasikan sebelumnya.

Selain itu, perkembangan teknologi generatif berlangsung sangat cepat sehingga karakteristik ancaman yang muncul dapat berubah dalam waktu yang relatif singkat. Kondisi tersebut membuat kajian mengenai *deepfake* perlu terus diperbarui agar tetap relevan dengan perkembangan teknologi dan pola penyalahgunaannya.

Untuk penelitian berikutnya, beberapa topik yang menarik untuk dikaji lebih lanjut antara lain efektivitas teknologi deteksi *deepfake*, tingkat kesiapan masyarakat dalam mengenali konten manipulatif, studi kasus penggunaan *deepfake* dalam insiden keamanan siber, pengembangan model mitigasi yang dapat diterapkan pada organisasi, serta kajian mengenai kebijakan dan tata kelola kecerdasan buatan dalam konteks Indonesia.

REFERENSI

[1] M. Westerlund, “The Emergence of *Deepfake* Technology: A Review,” Nov. 2019.

[2] Stanford University *et al.*, "Artificial Intelligence Index Report 2025," 2025. Accessed: Jun. 06, 2026. [Online]. Available: https://arxiv.org/abs/2504.07139
[3] B. Chesney and D. Citron, "Deep fakes: A looming challenge for privacy, democracy, and national security," *Calif. Law Rev.*, vol. 107, no. 6, pp. 1753–1820, 2019, doi: 10.15779/Z38RV0D15J.
[4] Y. Mirsky and W. Lee, "The Creation and Detection of *Deepfake*s," Jul. 31, 2021, *Association for Computing Machinery*. doi: 10.1145/3425780.
[5] Europol, "Internet Organised Crime Threat Assessment (IOCTA) 2024," Sep. 2024. doi: 10.2813/442713.
[6] ENISA, S. Corbiaux, K. Van Impe, and C. Ardagna, "ENISA THREAT LANDSCAPE 2024," Sep. 2024, doi: 10.2824/0710888.
[7] OECD, "2023 OECD Digital Governance Index," 2024.
[8] NIST and US Department of Commerce, "Artificial Intelligence Risk Management Framework (AI RMF 1.0)," Jan. 2023. doi: 10.6028/NIST.AI.100-1.
[9] D. Team. The Verizon, D. C. Hylender, P. Langlois, A. Pinto, and W. Suzanne, "2024 Data Breach Investigations Report," 2024.
[10] W. E. Forum, M. McLennan, and Group Zurich Insurance, *The Global Risks Report 2024 - 19th edition - insight report*. 2024. [Online]. Available: www.weforum.org
[11] Unesco, *Guidance for generative AI in education and research*. UNESCO, 2023. doi: 10.54675/ewzm9535.
[12] H. Snyder, "Literature review as a research methodology: An overview and guidelines," *J. Bus. Res.*, vol. 104, pp. 333–339, Nov. 2019, doi: 10.1016/j.jbusres.2019.07.039.
[13] A. Booth, A. Sutton, and D. Papaioannou, "Systematic Approaches to a Successful Literature Review," 2016.
[14] Y. Xiao and M. Watson, "Guidance on Conducting a Systematic Literature Review," Mar. 01, 2019, *SAGE Publications Inc.* doi: 10.1177/0739456X17723971.
[15] R. J. Torraco, "Writing Integrative Literature Reviews: Guidelines and Examples," *Human Resource Development Review*, vol. 4, no. 3, pp. 356–367, Sep. 2005, doi: 10.1177/1534484305278283.