

TANTANGAN PSIKOLOGIS DALAM MENGHADAPI FENOMENA DEEFAKE DI ERA POST-TRUTH: SCOPING REVIEW

Tizty Annur Triananda¹, Bonar Hutapea^{1*}, Ghaisyani Rizky Dwiagustin Handoko¹, Tasya Viviana¹, Lourdes Clayrence¹

¹Faculty of Psychology, Universitas Tarumanagara,
Letjen S. Parman No. 1, Tomang, Grogol petamburan, Kota Jakarta Barat,
Daerah Khusus Ibukota Jakarta, Indonesia 11440

*bonarh@fpsi.untar.ac.id

Abstrak

Fenomena *deepfake* menandai pergeseran besar dalam lanskap post-truth, di mana batas antara realitas dan rekayasa digital menjadi semakin kabur. Penelitian ini bertujuan memetakan tantangan psikologis yang muncul dalam menghadapi *deepfake*, dengan meninjau faktor kognitif, afektif, sosial, dan teknis yang memengaruhi persepsi dan perilaku individu. Menggunakan pendekatan *scoping review*, studi ini menganalisis 16 artikel internasional terbitan 2020 sampai 2025 dari basis data PubMed, Scopus, Web of Science, PLOS ONE, ScienceDirect, dan SAGE Journals. Hasil menunjukkan bahwa keterbatasan kognitif seperti *illusory truth effect* dan *overconfidence bias* membuat individu rentan terhadap manipulasi *deepfake*. Di sisi afektif, persepsi keaslian memodulasi respons emosional dan neural terhadap ekspresi wajah atau suara sintetis. Sementara pada tingkat sosial, kepercayaan terhadap institusi dan interaksi antar kelompok turut melemah akibat penyebaran konten palsu yang meyakinkan. Namun, beberapa studi juga menemukan bahwa *deepfake* tidak selalu lebih manipulatif dibanding misinformasi non-*deepfake* bila kontrol eksperimental diperketat. Studi ini menegaskan perlunya integrasi antara teori psikologi dan strategi literasi digital berbasis *prebunking* untuk mengurangi dampak psikologis *deepfake* dalam era post-truth.

Kata kunci: bias kognitif; *deepfake*; kepercayaan; *post-truth*; tantangan psikologis

Abstract

The rise of deepfakes marks a significant shift in the post-truth era, where the boundary between reality and digital fabrication becomes increasingly blurred. This study aims to map the psychological challenges in confronting deepfakes by examining cognitive, affective, social, and technical factors that influence individual perception and behavior. Using a scoping review approach, this study analyzed 16 international articles published between 2020 and 2025 from databases including PubMed, Scopus, Web of Science, PLOS ONE, ScienceDirect, and SAGE Journals. The findings reveal that cognitive limitations such as the illusory truth effect and overconfidence bias make individuals vulnerable to deepfake manipulation. Affectively, perceived authenticity modulates emotional and neural responses to synthetic faces and voices. At the social level, public trust in institutions and interpersonal relationships are weakened by the spread of highly convincing false content. However, several studies also indicate that deepfakes are not necessarily more manipulative than non-deepfake misinformation when strict experimental controls are applied. This study highlights the importance of integrating psychological theory with digital literacy and prebunking strategies to mitigate the psychological impacts of deepfakes in the post-truth era.

Keywords: cognitive bias; deepfake; trust; post-truth; psychological challenges

PENDAHULUAN

Perkembangan teknologi *deepfake* berkontribusi pada pergeseran cara masyarakat menilai kebenaran, di mana konten audio-visual realistis namun palsu mengaburkan batas antara fakta dan rekayasa serta mengikis kepercayaan publik terhadap bukti visual sebagai dasar validitas informasi (Köbis et al., 2021). Teknologi ini berbasis kecerdasan buatan generatif yang memungkinkan manipulasi suara, wajah, dan ekspresi emosional dengan tingkat realisme tinggi (Weikmann et al., 2024). Berbeda dengan manipulasi digital tradisional, *deepfake*

memiliki kualitas hiper-realistis yang menantang batas antara fakta dan ilusi. Dalam lanskap *post-truth*, emosi dan keyakinan personal lebih berpengaruh daripada data objektif (Lewandowsky et al., 2021). Hal ini membuat masyarakat semakin sulit membedakan antara bukti autentik dan rekayasa digital. Sifat audio-visual *deepfake* membuatnya lebih meyakinkan dibanding misinformasi berbasis teks atau foto (Weikmann et al., 2024), sehingga meningkatkan risiko internalisasi *belief* secara cepat.

Keterbatasan kognitif manusia menjadi titik lemah utama dalam menghadapi *deepfake*, terutama ketika konten disajikan realistis dan konsisten dengan orientasi politik atau nilai pribadi (Köbis et al., 2021; Lewandowsky & van der Linden, 2021). Studi eksperimental menunjukkan bahwa kemampuan publik membedakan *deepfake* dari konten asli sangat terbatas, bahkan tidak lebih baik daripada tebakan acak (Josephs et al., 2023; Köbis et al., 2021). Individu seringkali yakin dapat mengenali rekayasa visual, menimbulkan kesenjangan antara persepsi diri dan kenyataan objektif. Fenomena *overconfidence bias* ini menegaskan keterbatasan metakognitif manusia dalam menilai kapabilitas sendiri. Efek pada memori juga muncul melalui *mis-attribution*, di mana individu sulit membedakan informasi asli dan manipulasi (Josephs et al., 2023). Dengan demikian, kerentanan kognitif ini membuka jalur bagi *deepfake* untuk mempengaruhi keyakinan dan perilaku secara signifikan.

Dampak kognitif *deepfake* kemudian menjalar ke ranah sosial, mempengaruhi trust kolektif terhadap institusi dan partisipasi publik. Studi lintas negara menunjukkan bahwa konten politik berbasis *deepfake* menurunkan kepercayaan publik terhadap lembaga demokrasi, serta memperbesar keraguan terhadap validitas informasi politik (Dan., 2025). Eksperimen lain memperlihatkan bahwa *deepfake* yang menampilkan kegagalan infrastruktur publik menurunkan kepercayaan terhadap pemerintah (Ahmed et al., 2024). Dampak ini terlihat pula pada perilaku sosial, seperti niat voting atau keterlibatan warga dalam partisipasi politik. Dengan demikian, pengaruh *deepfake* tidak hanya pada persepsi individu, tetapi juga pada stabilitas sosial dan kontrak sosial antara warga dan institusi.

Dalam konteks psikologi sosial, kepercayaan terhadap sumber informasi memainkan peran penting dalam pembentukan *belief*. Semakin realistis konten *deepfake*, semakin besar kemungkinan diinternalisasi sebagai keyakinan, bahkan tanpa proses verifikasi kritis (Martel et al., 2020). Fenomena ini diperkuat oleh *illusory truth effect*, dimana klaim palsu yang dipaparkan berulang tetap dipercaya, sekalipun telah terbantahkan. Bias konfirmasi dan *motivated reasoning* juga mempermudah penerimaan konten yang sesuai orientasi politik atau nilai pribadi (Ahmed et al., 2024; Dan., 2025). Dengan demikian, *deepfake* dapat memperkuat realitas paralel yang diyakini masing-masing kelompok, sehingga meningkatkan polarisasi sosial.

Kerangka *dual-system thinking* dari Kahneman (2011) menjelaskan kerentanan ini lebih lanjut. Konsumsi informasi di media sosial cenderung mengaktifkan *System 1* yang intuitif dan heuristik, membuat individu lebih rentan menerima klaim visual yang salah tanpa evaluasi kritis (Ahmed & Chua, 2023). Individu yang berpikir reflektif (*System 2*) lebih mampu menolak misinformasi (Bago et al., 2020; Martel et al., 2020). Faktor psikologis lain yang memperkuat dampak *deepfake* termasuk *overconfidence bias* dan dominasi *System 1*. Selain itu, faktor sosial seperti tingkat kepercayaan terhadap sumber informasi dan polarisasi kelompok turut menentukan sejauh mana *belief* dan perilaku dipengaruhi. Faktor teknis seperti kualitas realisme ekspresi wajah, sinkronisasi audio-visual, dan kesesuaian konteks juga mempengaruhi efektivitas *deepfake* (Ahmed & Chua, 2023; Eiserbeck et al., 2023).

Teknologi *deepfake* yang digunakan dalam studi empiris beragam, mulai dari *Generative Adversarial Network* (GAN) untuk manipulasi wajah, *voice cloning* untuk imitasi suara, hingga *full-body reenactment* untuk simulasi gerak (Weikmann et al., 2024). Kualitas dan realisme teknologi terbukti mempengaruhi kemampuan deteksi partisipan serta internalisasi *belief*. Semakin tinggi kualitas *deepfake*, semakin sulit partisipan membedakan konten asli dan manipulasi (Weikmann et al., 2024). Temuan ini menunjukkan bahwa tidak hanya kognisi yang terpengaruh, tetapi juga bagaimana individu merespons konten secara emosional. Dengan kata lain, kualitas *deepfake* dapat memperkuat pengaruhnya terhadap persepsi dan reaksi afektif, sehingga interaksi antara faktor psikologis, sosial, dan teknis menjadi semakin kompleks. Hal ini menjelaskan mengapa *belief*, memori, dan perilaku individu dapat berubah secara signifikan akibat paparan *deepfake*.

Selain kognisi, aspek emosional juga memainkan peran penting dalam dampak *deepfake*. Persepsi terhadap keaslian ekspresi wajah memicu respons afektif yang berbeda. Senyuman yang diyakini palsu menimbulkan reaksi lebih lemah dibanding senyuman nyata. Sementara ekspresi marah tetap memicu respons kuat meskipun diketahui palsu (Eiserbeck et al., 2023). Temuan ini menegaskan bahwa *deepfake* memengaruhi kualitas pengalaman emosional dalam interaksi sosial, bukan hanya penilaian kognitif. Dengan demikian, pemahaman terhadap pengaruh emosional *deepfake* diperkirakan penting sebagai dasar perancangan intervensi terhadap informasi palsu, serta mendukung pengembangan strategi mitigasi oleh pembuat kebijakan, peneliti, dan praktisi media sosial yang mempertimbangkan aspek kognitif dan afektif. Pendekatan ini berpotensi membantu mengurangi dampak negatif terhadap *belief*, memori, dan perilaku sosial.

Meskipun penelitian mengenai dampak psikologis *deepfake* terus berkembang, masih terdapat sejumlah kesenjangan penting yang perlu mendapat perhatian serius. Sebagian besar studi yang ada bersifat eksperimental jangka pendek, sehingga belum mampu menggambarkan konsekuensi psikologis jangka panjang dari paparan berulang terhadap konten manipulatif (Weikmann et al., 2024). Paparan semacam itu berpotensi menimbulkan *information fatigue*, penurunan empati, dan melemahnya kepercayaan terhadap bukti visual yang sah (Weikmann et al., 2024). Selain itu, literatur yang tersedia masih terpusat pada konteks budaya Barat, sementara pengaruh faktor demografis, sosial, dan budaya terhadap kerentanan psikologis terhadap *deepfake* belum banyak dieksplorasi (Ahmed et al., 2024; Dan., 2025). Padahal, persepsi terhadap keaslian dan tingkat kepercayaan terhadap media sangat dipengaruhi oleh norma sosial serta ekspektasi budaya yang khas di setiap masyarakat.

Integrasi antara aspek teknologi dan psikologi masih belum sepenuhnya terjalin secara mendalam. Faktor teknis seperti tingkat realisme visual, sinkronisasi audio, dan ekspresi wajah terbukti memengaruhi persepsi manusia, namun hubungan antara kompleksitas teknologis *deepfake* dengan respons kognitif dan emosional individu belum dikaji secara sistematis (Ahmed & Chua, 2023; Weikmann et al., 2024). Hal ini menunjukkan perlunya pendekatan lintas disiplin yang mampu menjembatani analisis teknis dengan pemahaman psikologis manusia.

Selain itu, kesenjangan antara teori dan praktik juga masih nyata. Banyak temuan teoritis yang menjelaskan mekanisme psikologis di balik penerimaan *deepfake*, namun belum diterjemahkan ke dalam intervensi yang aplikatif, seperti literasi digital berbasis prebunking atau edukasi yang berfokus pada koreksi bias kognitif (Lewandowsky & van der Linden, 2021; Pramod et al., 2025). Dengan memperhatikan berbagai kesenjangan tersebut, penelitian ini berupaya memberikan pemetaan konseptual yang lebih komprehensif mengenai tantangan psikologis

dalam menghadapi fenomena *deepfake* terkait pola psikologis dari literatur penelitian terkait kognitif, emosional, sosial, dan neural di era *post-truth*. Meskipun penelitian mengenai dampak *deepfake* telah berkembang pesat, temuan yang ada masih tersebar di berbagai pendekatan psikologis dengan fokus variabel dan desain yang beragam, sehingga belum tersedia pemetaan konseptual yang terintegrasi. Kondisi ini menyulitkan penarikan pola mekanisme psikologis utama serta pengembangan intervensi berbasis bukti yang terukur. Oleh karena itu, *scoping review* diperlukan untuk memetakan lanskap temuan empiris, mengelompokkan mekanisme psikologis yang terlibat, serta mengidentifikasi kesenjangan penelitian sebagai dasar bagi pengembangan riset dan strategi mitigasi di masa depan.

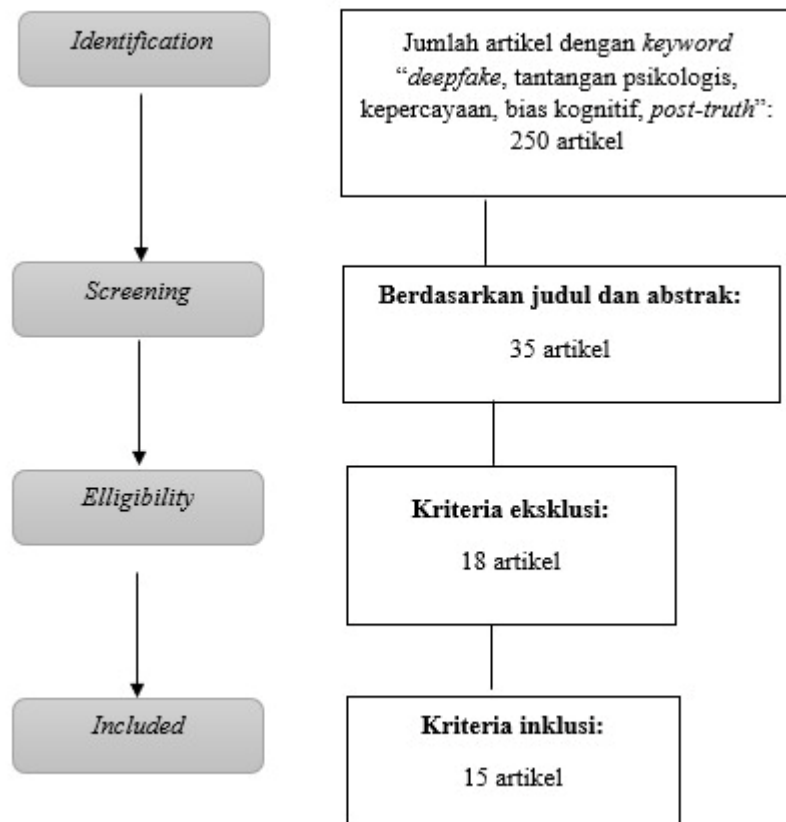
METODE

Tinjauan ini menggunakan pendekatan *scoping review* untuk memetakan dan mensintesis penelitian terkini mengenai tantangan psikologis dalam menghadapi *deepfake* pada era *post-truth*. Pendekatan ini dipilih karena sesuai untuk mengidentifikasi pola konseptual, memperjelas hubungan antara faktor kognitif, afektif, sosial, dan teknis, serta menemukan kesenjangan pengetahuan dalam literatur yang membahas pengaruh *deepfake* terhadap persepsi, kepercayaan, memori, dan perilaku (Peters et al., 2022).

Proses pencarian literatur dilakukan pada basis data internasional bereputasi, yaitu PubMed, Scopus, Web of Science, PLOS ONE, ScienceDirect, dan SAGE Journals, untuk memastikan cakupan literatur lintas disiplin yang relevan dengan topik *deepfake* dari perspektif psikologi, komunikasi, dan teknologi. Scopus dan Web of Science menyediakan indeksasi jurnal multidisipliner bereputasi dengan standar seleksi ketat, PubMed menjangkau studi kognitif dan neuropsikologis, sementara ScienceDirect, SAGE Journals, dan PLOS ONE memperluas cakupan pada bidang teknologi, media, dan ilmu perilaku. Penggunaan beberapa basis data bertujuan mengurangi bias seleksi sumber dan meningkatkan kelengkapan identifikasi studi.

Cakupan publikasi dengan rentang tahun 2020 hingga 2025 didasari dari hasil analisis bibliometrik bahwa penelitian *deepfake* mulai meningkat pesat dengan jumlah publikasi yang mengalami pertumbuhan cepat pada rentang tahun 2014-2024, lalu kemudian pada tahun 2020, *deepfake* mulai tersebar semakin luas dalam bidang yang lebih spesifik seperti hiburan, politik, dan penipuan keuangan (Acim et al., 2025). Untuk memastikan relevansi terhadap perkembangan terkini teknologi *deepfake* dan kajian psikologis yang menyertainya. Pencarian dilakukan menggunakan kombinasi kata kunci utama "*deepfake*", "*synthetic media*", dan "*AI-generated content*", yang kemudian dipadukan dengan istilah yang mencerminkan aspek psikologis seperti "*cognitive bias*", "*trust*", "*false memory*", "*emotion*", dan "*dual-system thinking*", serta istilah sosial seperti "*polarization*" dan "*social trust*". Kombinasi kata kunci berbasis konsep dengan operator Boolean AND dan OR, yang di mana sinonim digabungkan dengan OR dan antar konsep digabungkan dengan AND. Format dari pencarian kemudian disesuaikan dengan fitur masing-masing basis data.

Proses seleksi artikel dilakukan secara bertahap dengan mengikuti pedoman PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*) untuk menjamin transparansi dan kualitas analisis. Dari pencarian awal diperoleh 250 artikel, kemudian setelah penyaringan berdasarkan judul dan abstrak tersisa 35 artikel. Selanjutnya, 18 artikel diperiksa dalam bentuk teks penuh untuk menilai relevansi dan kualitas metodologisnya, dan akhirnya 15 artikel memenuhi kriteria inklusi untuk dimasukkan ke dalam sintesis akhir, sebagaimana tertera dalam Gambar 1.



Gambar 1. Proses Seleksi Literatur dengan *Prisma Flow Diagram*

RESULT AND DISCUSSION

Tabel 1.

Hasil Penyaringan Artikel

Nama peneliti, Tahun dan Judul Penelitian	Desain, Sampel dan Instrumen Penelitian	Hasil	Kesimpulan
Ahmed et al. (2024). <i>Social media news use amplifies the illusory truth effects of viral deepfake.</i>	Desain eksperimen lintas negara dengan pendekatan kuantitatif. Sampel terdiri dari partisipan dewasa di delapan negara. Instrumen berupa skala kepercayaan informasi dan niat berbagi konten.	Penggunaan berita media sosial secara signifikan meningkatkan <i>illusory truth effect</i> terhadap konten <i>deepfake</i> di seluruh konteks negara.	Paparan berulang melalui media sosial memperkuat penerimaan <i>deepfake</i> sebagai kebenaran lintas budaya.
Ahmed & Chua (2023). <i>Perception and deception: Exploring individual responses to</i>	Desain eksperimen kuantitatif. Sampel dewasa. Instrumen penilaian akurasi, kepercayaan, dan niat	<i>Deepfake</i> berbentuk video dinilai paling kredibel dan paling mendorong niat berbagi dibanding <i>modality</i> lain.	Modalitas visual meningkatkan kerentanan psikologis terhadap <i>deepfake</i> .

Nama peneliti, Tahun dan Judul Penelitian	Desain, Sampel dan Instrumen Penelitian	Hasil	Kesimpulan
<i>deepfakes across different modalities</i>	berbagi pada stimulus video, audio, dan teks.		
Ahmed et al. (2025). <i>False failures, real distrust</i>	Eksperimen kuantitatif. Sampel warga Amerika Serikat. Instrumen kepercayaan terhadap pemerintah.	Paparan <i>deepfake</i> kegagalan infrastruktur menurunkan kepercayaan publik terhadap pemerintah.	<i>Deepfake</i> berpotensi menggerus kepercayaan institusional.
Bago et al. (2020). <i>Fake news, fast and slow</i>	Eksperimen laboratorium. Sampel dewasa. Instrumen penilaian kebenaran headline.	Pemrosesan deliberatif menurunkan kepercayaan pada berita palsu namun tidak pada berita benar.	Pemikiran cepat meningkatkan kerentanan terhadap misinformasi.
Dan (2025). <i>Deepfakes as a democratic threat</i>	Eksperimen kuantitatif. Sampel dewasa. Instrumen kepercayaan politik dan persepsi ancaman.	Efek negatif <i>deepfake</i> dapat direduksi melalui <i>fact-check</i> jurnalistik.	Intervensi informasi efektif menekan dampak psikologis <i>deepfake</i> .
Eiserbeck et al. (2023). <i>Deepfake smiles matter less</i>	Eksperimen dengan EEG. Sampel dewasa. Instrumen ERP (P1, N170, LPP).	Persepsi stimulus sebagai sintesis menurunkan respons afektif awal dan meningkatkan evaluasi kognitif.	Keyakinan autentisitas memodulasi pemrosesan afektif dan neural.
Josephs et al. (2023). <i>Artifact magnification on deepfake videos</i>	Eksperimen kuantitatif. Sampel dewasa. Instrumen akurasi deteksi dan kepercayaan diri.	Pembesaran artefak visual meningkatkan akurasi deteksi <i>deepfake</i> .	Bantuan visual teknis dapat meningkatkan kemampuan deteksi manusia.
Kahneman (2011). <i>Thinking, Fast and Slow</i>	Kajian teoretis psikologi kognitif.	Pemrosesan cepat (<i>System 1</i>) rentan bias dibanding pemrosesan lambat (<i>System 2</i>).	Bias kognitif menjelaskan kerentanan terhadap misinformasi.
Köbis et al. (2021). <i>Foiled twice</i>	Eksperimen kuantitatif. Sampel dewasa. Instrumen akurasi dan kepercayaan diri deteksi.	Partisipan gagal mendeteksi <i>deepfake</i> namun menunjukkan <i>overconfidence</i> .	Manusia tidak andal mendeteksi <i>deepfake</i> secara intuitif.
Lewandowsky & van der Linden (2021). <i>Countering misinformation through prebunking</i>	Review teoretis.	<i>Prebunking</i> efektif meningkatkan resistensi psikologis terhadap misinformasi.	Intervensi preventif lebih efektif dibanding koreksi pasca paparan.

Nama peneliti, Tahun dan Judul Penelitian	Desain, Sampel dan Instrumen Penelitian	Hasil	Kesimpulan
Lovato et al. (2024). <i>Diverse misinformation</i>	Simulasi jaringan dan eksperimen.	<i>Homophily</i> memperkuat bias, jejaring beragam meningkatkan peluang koreksi kolektif.	Struktur jejaring sosial memengaruhi penyebaran <i>deepfake</i> .
Martel et al. (2020). <i>Reliance on emotion promotes belief in fake news</i>	Eksperimen kuantitatif. Sampel dewasa. Instrumen ketergantungan emosi.	Ketergantungan emosi meningkatkan kepercayaan pada berita palsu.	Faktor afektif memperbesar kerentanan misinformasi.
Pramod et al. (2025). <i>Two-theory approach on deepfakes</i>	Eksperimen kuantitatif. Sampel konsumen. Instrumen PMT dan TPB.	Sikap dan norma sosial lebih kuat memprediksi niat protektif.	Intervensi perlu menargetkan sikap dan norma berbagi.
Roswandowitz et al. (2024). <i>Cortical-striatal network distinguishes deepfake voices</i>	Studi neuroimaging. Sampel dewasa.	Jaringan <i>cortico-striatal</i> membedakan suara asli dan <i>deepfake</i> .	Pemrosesan suara <i>deepfake</i> memiliki penanda neural spesifik.
Weikmann et al. (2024). <i>After Deception</i>	Eksperimen longitudinal. Sampel dewasa.	Paparan <i>deepfake</i> menurunkan kepercayaan terhadap media.	Dampaknya bersifat jangka panjang terhadap persepsi media.

Hasil lintas studi memperlihatkan pola yang konsisten. *Social media news use* meningkatkan peluang terjadinya *illusory truth* sehingga klaim dalam *deepfakes* terasa lebih benar bagi audiens dan pola ini tampak konsisten di delapan negara dengan konteks beragam (Ahmed et al., 2024). Perbedaan modality juga berpengaruh. Video *deepfake* cenderung dinilai lebih kredibel dan lebih ingin dibagikan dibanding *cheapfake*. Audio memberi dampak akurasi yang lebih subtil namun tetap dapat mendorong *intention to share*. Kapasitas kognitif yang lebih tinggi berkaitan dengan penurunan *misbelief* dan penurunan niat berbagi konten keliru (Ahmed & Chua, 2023). Pada saat yang sama manusia rata-rata lemah dalam mendeteksi *deepfakes* dan sering menunjukkan *overconfidence* sedangkan *awareness* singkat atau insentif kecil umumnya tidak banyak memperbaiki akurasi deteksi (Köbis et al., 2021). Pada ranah memori dan keyakinan paparan *deepfake* dapat menimbulkan *false memory* dan *false belief*.

Dasar dari efek tersebut dapat dilihat melalui sisi mekanisme kognitif dua proses dasar tampak dominan yaitu *processing fluency* dan *repetition*. Paparan berulang membuat klaim terasa akrab dan mudah diproses sehingga penilaian kebenaran meningkat khususnya ketika konsumsi berita berlangsung di platform sosial yang dirancang untuk distribusi cepat dan paparan berulang (Ahmed et al., 2024). Arsitektur *platform* serta *network structure* ikut mengarahkan lintasan penyebaran. Jejaring yang homogen memperkuat kluster keyakinan sedangkan jejaring yang lebih beragam membuka peluang *collective correction* meski efek tersebut sering kalah oleh dinamika *virality* yang cepat (Lovato et al., 2024).

Transisi dari proses kognitif ke reaksi afektif dan neural menjelaskan peran sentral keyakinan atas autentisitas. Ketika orang percaya bahwa stimulus *AI-generated* respons positif awal

terhadap senyum melemah dan pada tahap evaluasi muncul pemrosesan yang lebih *effortful*. Pola ini tercermin pada dinamika ERP melalui komponen P1 N170 EPN yang menurun serta LPP yang meningkat ketika autentisitas diragukan (Eiserbeck et al., 2023). menegaskan bahwa keyakinan atas autentisitas berfungsi sebagai *gatekeeper* afektif yang memodulasi baik atensi awal maupun evaluasi lambat. Pada domain suara jaringan *cortico-striatal* menampilkan penanda yang memudahkan pembedaan identitas *deepfake* dari suara asli sehingga bias penilaian dapat ditelusuri pada konfigurasi jaringan tersebut (Roswadowitz et al., 2024).

Bias tersebut kemudian berhubungan dengan dampak sosial yang terlihat pada konteks layanan publik melalui peningkatan *distrust* terhadap institusi pada kondisi tertentu dengan besaran efek yang termoderasi oleh tingkat pendidikan pada sampel Amerika Serikat (Ahmed et al., 2024). Selain itu kerentanan tidak hanya ditentukan oleh karakter individu tetapi juga struktur jejaring sosial. *Homophily* memperkuat bias konfirmasi dan mempersulit koreksi bersama sedangkan keragaman koneksi memperbesar peluang *collective correction*. Di sisi lain *liar's dividend* memungkinkan aktor mendelegitimasi bukti sah hanya dengan menunjuk pada keberadaan teknologi manipulatif sehingga otoritas bukti menjadi rapuh dalam sengketa informasi (Lovato et al., 2024; Simon, 2025).

Pertanyaan mengenai keunikan daya manipulasi menghasilkan jawaban yang bernuansa. Pada sejumlah domain seperti memori dan keyakinan *deepfakes* tidak selalu lebih manipulatif dibanding misinformasi *non deepfake* ketika komparator metodologis disusun secara ketat. Implikasi kebijakan karenanya lebih baik bertumpu pada estimasi dampak yang terukur dan dapat direplikasi agar perumusan intervensi terhindar dari *moral panic* dan tetap berorientasi pada bukti (Ching et al., 2025). Konsistensi temuan justru berada pada tiga pilar yang paling kuat yaitu kesulitan deteksi oleh manusia pengaruh *social media news use* sebagai penguat *illusory truth* serta modulasi afektif ketika autentisitas dipersepsi sebagai hasil *AI-generated* (Ahmed et al., 2024; Eiserbeck et al., 2023; Köbis et al., 2021).

Implikasi teoretis bagi intervensi memberi arah yang jelas. *Theory of Planned Behavior* dan *Protection Motivation Theory* menjelaskan *protective intention* konsumen ketika berhadapan dengan konten sintesis. *Attitude* dan *subjective norms* cenderung lebih stabil dalam memprediksi niat protektif dibanding *perceived behavioral control*. Karena itu strategi *prebunking* perlu menumbuhkan sikap dan norma kebiasaan *check before share* sekaligus meningkatkan *response efficacy* dan *self-efficacy*. Pendekatan *tips spotting* yang singkat biasanya tidak cukup menaikkan akurasi sehingga diperlukan desain lingkungan informasi yang mengurangi paparan berulang dan menambah friksi berbagi pada konteks berisiko (Köbis et al., 2021; Pramod et al., 2025).

Keterbatasan korpus dan standar pelaporan masih menonjol. Banyak studi memakai stimulus arsip atau materi buatan sendiri namun pelaporan kualitas stimulus dan tingkat keterdeteksian belum seragam sehingga menyulitkan perbandingan lintas studi. Rekomendasi utama meliputi pilot test kredibilitas pelaporan mutu stimulus yang terstandar serta penyediaan *open materials* untuk replikasi dan meta analisis. Selain itu sebagian besar *outcome* diukur segera setelah paparan sehingga bukti jangka panjang masih terbatas dan memerlukan studi longitudinal yang kuat untuk menilai ketahanan efek dalam kehidupan sehari-hari yang sarat paparan (Ching et al., 2025). Keterbatasan ini juga mengungkap kesenjangan penelitian yang lebih luas, seperti perlunya penelitian lintas budaya serta integrasi antara aspek teknis dan psikologis secara lebih sistematis guna memahami dampak psikologis dari *deepfakes* secara menyeluruh (Dan et al., 2025; Weikmann et al., 2024).

Kajian ini menunjukkan bahwa tantangan psikologis dalam menghadapi *deepfake* di era *post-truth* bersifat multidimensi, melibatkan aspek kognitif, afektif, sosial, dan teknis. Individu rentan terhadap *illusory truth effect*, *processing fluency*, dan *overconfidence bias* yang meningkatkan penerimaan terhadap konten palsu, sementara keyakinan atas autentisitas berperan dalam memodulasi respons emosional dan pemrosesan neural. Pada tingkat sosial, paparan *deepfake* berpotensi melemahkan kepercayaan terhadap institusi serta memperkuat polarisasi. Namun, temuan *scoping review* ini menunjukkan bahwa *deepfake* tidak selalu lebih manipulatif dibandingkan misinformasi *non-deepfake* ketika kontrol metodologis diterapkan secara ketat. Secara teoretis, hal ini menegaskan bahwa dampak psikologis *deepfake* terutama dimediasi oleh mekanisme kognitif umum dalam pemrosesan informasi, bukan oleh karakteristik teknologi itu sendiri. Secara praktis, temuan ini mengindikasikan bahwa strategi mitigasi perlu difokuskan pada koreksi bias kognitif, pengelolaan paparan informasi, dan pendekatan *prebunking*, alih-alih bertumpu pada deteksi teknis semata. Oleh karena itu, penelitian selanjutnya perlu mengembangkan studi longitudinal, memperluas konteks lintas budaya, serta mengintegrasikan pendekatan teknis dan psikologis untuk memperkuat resiliensi masyarakat terhadap disinformasi digital.

KESIMPULAN

Kajian ini menunjukkan bahwa tantangan psikologis dalam menghadapi *deepfake* di era *post-truth* bersifat multidimensi, melibatkan aspek kognitif, afektif, sosial, dan teknis. Individu rentan terhadap *illusory truth effect*, *processing fluency*, dan *overconfidence bias* yang meningkatkan penerimaan terhadap konten palsu, sementara keyakinan atas autentisitas berperan penting dalam memodulasi reaksi emosional dan pemrosesan neural. Pada tingkat sosial, paparan *deepfake* dapat menurunkan kepercayaan terhadap institusi dan memperkuat polarisasi. Namun, temuan *scoping review* ini menunjukkan bahwa *deepfake* tidak selalu lebih manipulatif dibandingkan misinformasi *non-deepfake* ketika kontrol metodologis diterapkan secara ketat. Secara teoretis, hal ini menegaskan bahwa dampak psikologis *deepfake* terutama dimediasi oleh mekanisme kognitif umum, bukan oleh karakteristik teknologi itu sendiri. Secara praktis, temuan ini mengarahkan strategi mitigasi pada koreksi bias kognitif, pengelolaan paparan informasi, dan *prebunking*, alih-alih fokus eksklusif pada deteksi teknis. Oleh karena itu, penelitian selanjutnya perlu memperluas konteks lintas budaya, mengembangkan studi longitudinal, serta mengintegrasikan pendekatan teknis dan psikologis dalam rancangan intervensi berbasis literasi digital dan *prebunking* guna memperkuat resiliensi masyarakat terhadap disinformasi digital.

REFERENCES

- Acim, B., Boukhelif, M., Ouhni, H., Kharmoum, N., & Ziti, S. (2025). A decade of deepfake research in the generative AI era, 2014–2024: A Bibliometric analysis. *Publications*, 13(4), 50. <https://doi.org/10.3390/publications13040050>
- Ahmed, S., Bee, A. W. T., Ng, S. W. T., & Masood, M. (2024). Social media news use amplifies the illusory truth effects of viral deepfakes: a Cross-National study of eight countries. *Journal of Broadcasting & Electronic Media*, 68(5), 778–805. <https://doi.org/10.1080/08838151.2024.2410783>
- Ahmed, S., & Chua, H. W. (2023). Perception and deception: Exploring individual responses to deepfakes across different modalities. *Heliyon*, 9(10), e20383. <https://doi.org/10.1016/j.heliyon.2023.e20383>

- Ahmed, S., Masood, M., Bee, A. W. T., & Ichikawa, K. (2025). False failures, real distrust: the impact of an infrastructure failure deepfake on government trust. *Frontiers in Psychology, 16*, 1574840. <https://doi.org/10.3389/fpsyg.2025.1574840>
- Bago, B., Rand, D. G., & Pennycook, G. (2020). Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. *Journal of Experimental Psychology General, 149*(8), 1608–1613. <https://doi.org/10.1037/xge0000729>
- Dan, V. (2025). Deepfakes as a Democratic threat: Experimental evidence shows noxious effects that are reducible through journalistic fact checks. *The International Journal of Press/Politics*. <https://doi.org/10.1177/19401612251317766>
- Eiserbeck, M., Szyck, G. R., Müller, K., & Hasson, U. (2023). Deepfake smiles matter less: The psychological and neural effects of believing a face is synthetic. *Scientific Reports, 13*(1), 20391. <https://doi.org/10.1038/s41598-023-42802-x>
- Josephs, E., Fosco, C., & Oliva, A. (2023). Artifact magnification on deepfake videos increases human detection and subjective confidence. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2304.04733>
- Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux. [https://mlsu.ac.in/econtents/2950_Daniel%20Kahneman%20-%20Thinking,%20Fast%20and%20Slow%20\(2013\).pdf](https://mlsu.ac.in/econtents/2950_Daniel%20Kahneman%20-%20Thinking,%20Fast%20and%20Slow%20(2013).pdf)
- Köbis, N. C., Doležalová, B., & Soraperra, I. (2021). Fooled twice: People cannot detect deepfakes but think they can. *iScience, 24*(11), 103364. <https://doi.org/10.1016/j.isci.2021.103364>
- Lewandowsky, S., & Van Der Linden, S. (2021). Countering misinformation and fake news through inoculation and prebunking. *European Review of Social Psychology, 32*(2), 348–384. <https://doi.org/10.1080/10463283.2021.1876983>
- Lovato, J., St-Onge, J., Harp, R., Lopez, G. S., Rogers, S. P., Haq, I. U., Hébert-Dufresne, L., & Onaolapo, J. (2024). Diverse misinformation: impacts of human biases on detection of deepfakes on networks. *Npj Complexity, 1*(1). <https://doi.org/10.1038/s44260-024-00006-y>
- Martel, C., Pennycook, G., & Rand, D. G. (2020). Reliance on emotion promotes belief in fake news. *Cognitive Research Principles and Implications, 5*(1), 47. <https://doi.org/10.1186/s41235-020-00252-3>
- Peters, M. D., Godfrey, C. M., Khalil, H., McInerney, P., Parker, D., & Soares, C. B. (2015). Guidance for conducting systematic scoping reviews. *International Journal of Evidence-Based Healthcare, 13*(3), 141–146. <https://doi.org/10.1097/xeb.0000000000000050>
- Pramod, D., Patil, K. P., & S, V. B. (2025). Is it really unreal? A two-theory approach on the impact of deepfakes technology on the protection motivation of consumers. *Cogent Business & Management, 12*(1). <https://doi.org/10.1080/23311975.2025.2461239>
- Roswandowitz, C., Kathiresan, T., Pellegrino, E., Dellwo, V., & Frühholz, S. (2024). Cortical-striatal brain network distinguishes deepfake from real speaker identity. *Communications Biology, 7*(1), 711. <https://doi.org/10.1038/s42003-024-06372-6>
- Simon, J. (2025). Generative ai, quadruple deception, & trust. *Social Epistemology, 40*(1), 101–115. <https://doi.org/10.1080/02691728.2025.2491087>
- Weikmann, T., Greber, H., & Nikolaou, A. (2024). After deception: How falling for a deepfake affects the way we see, hear, and experience media. *The International Journal of Press/Politics, 30*(1), 187–210. <https://doi.org/10.1177/19401612241233539>