BAB II
TINJAUAN PUSTAKA

2.1. Studi Sebelumnya

2.1.1. Pengalaman Generasi Z

Berdasarkan penelitian Shahid dkk. [14], melakukan penelitian tentang persepsi dan sikap pengguna media sosial terhadap video palsu termasuk konten *Deepfake*, dengan melibatkan 36 pengguna media sosial di India yang diperlihatkan tiga video *Deepfake* dan satu video asli menggunakan pendekatan kualitatif melalui wawancara semi terstruktur. Hasil penelitian menunjukkan bahwa sebagian besar pengguna tidak menyadari keberadaan video yang telah dimanipulasi secara digital, dan mayoritas pengguna hanya menganggap sebuah video sebagai palsu apabila informasi yang disampaikan terasa tidak akurat bagi mereka.

Generasi Z yang menjadi peserta dalam penelitian ini adalah pengguna muda berusia 18-27 tahun yang menunjukkan pengalaman beragam terhadap *Deepfake.* Meskipun sebagian besar pengguna muda media sosial menggunakan smartphone lebih dari lima tahun, mereka tetap kesulitan mengenali video *Deepfake* yang canggih. Beberapa peserta muda mengetahui bahwa *Deepfake* digunakan untuk membuat konten pornografi, namun tetap tidak mampu mendeteksinya secara langsung. Peserta muda seperti Anushka 27 tahun pernah menjumpai video dengan figur manusia buatan AI di Instagram yang sangat sulit dibedakan dari manusia nyata. Bahkan peserta muda seperti Vikram 22 tahun dan Abdul 19 tahun mengakui bahwa mereka lebih fokus pada kesesuaian konten dengan pandangan mereka daripada memverifikasi keaslian video tersebut. Hal ini menunjukkan bahwa meskipun Gen Z tumbuh di era digital, mereka tetap rentan terhadap *Deepfake* dan belum memiliki kemampuan yang memadai untuk mendeteksi manipulasi video yang semakin canggih.

2.1.2. Kemampuan Identifikasi *Deepfake*

Berdasarkan penelitian Hoe dkk. [17], melakukan penelitian kualitatif menggunakan wawancara semi-terstruktur secara individual kepada 38 peserta untuk menggali strategi yang digunakan dalam mengidentifikasi *Deepfake*. Profil peserta menunjukkan bahwa mereka merupakan pengguna aktif platform video online, di mana seluruh peserta menggunakan YouTube, 30 dari 38 peserta menggunakan Facebook, dan 28 peserta menggunakan Instagram, sehingga mereka dinilai relevan sebagai subjek penelitian karena terbiasa menonton video online dan berpotensi terpapar konten *Deepfake* dalam keseharian mereka. Pengumpulan data dilakukan dalam dua tahap, di mana wawancara individu dilakukan t untuk menggali latar belakang dan kebiasaan peserta, kemudian pada disuguhkan 4 video yang terdiri dari 2 video asli dan 2 video *Deepfake*, dengan urutan video yang diacak untuk setiap peserta guna meminimalkan efek urutan. Peserta diminta untuk menentukan mana video yang asli dan mana yang *Deepfake*, kemudian melalui wawancara semi-terstruktur mereka diminta menjelaskan bagaimana mereka sampai pada kesimpulan tersebut, sehingga peneliti dapat menggali strategi identifikasi apa yang mereka gunakan secara mendalam.

Kemampuan manusia dalam mengidentifikasi *Deepfake* dengan membandingkan performa manusia biasa terhadap model deteksi *Deepfake* terdepan berbasis kecerdasan buatan. Berbeda dengan anggapan umum bahwa manusia memiliki kemampuan yang sangat terbatas, hasil penelitian menunjukkan bahwa kemampuan deteksi manusia sebanding dengan model mesin terbaik, meskipun keduanya membuat jenis kesalahan yang berbeda. Dalam eksperimen pertama, sebanyak 82% peserta mampu melampaui akurasi model mesin terdepan. Kemampuan identifikasi *Deepfake* pada manusia dipengaruhi oleh beberapa faktor, antara lain kemampuan pemrosesan wajah secara holistik yang dimiliki manusia, kondisi emosional seperti kemarahan yang terbukti menurunkan akurasi deteksi, serta kualitas video *Deepfake* itu sendiri. Selain itu, kolaborasi antara manusia dan AI terbukti meningkatkan akurasi

identifikasi, di mana akurasi peserta meningkat dari 66% menjadi 73% setelah mendapat prediksi dari model AI.

2.1.3. Dampak *Deepfake* terhadap Kepercayaan Informasi

Pada penelitian Chesney dkk. [9], di mana mereka tidak memiliki cukup informasi untuk menilai apakah suatu konten benar atau salah. Ketidakpastian ini terbukti menurunkan kepercayaan pengguna terhadap informasi di media sosial secara menyeluruh, bahkan terhadap konten yang sebenarnya autentik sekalipun tercermin dari temuan bahwa pengguna jauh lebih sering melabeli media nyata sebagai *Deepfake* dibandingkan sebaliknya, dan banyak yang menyatakan tidak dapat mempercayai footage apapun dari konflik tersebut.

Kondisi ini dijelaskan melalui konsep *liar's dividend*, yakni situasi di mana meningkatnya kesadaran publik tentang keberadaan *Deepfake* justru mempermudah penolakan terhadap bukti video yang autentik, sehingga bahaya terbesar *Deepfake* bukan terletak pada kemampuannya menipu, melainkan pada kemampuannya menabur keraguan secara meluas. Dengan demikian, ketika Generasi Z berulang kali terpapar *Deepfake* di Facebook tanpa disertai mekanisme klarifikasi yang memadai, akumulasi ketidakpastian tersebut dapat berujung pada penurunan kepercayaan yang menyeluruh terhadap informasi, perubahan perilaku konsumsi konten, hingga kelelahan kognitif akibat beban verifikasi yang terus-menerus.

2.1.4. Bagaimana Publik Memahami *Deepfake*

Pada penelitian Caci dkk. [19], melakukan penelitian tentang representasi mental publik terhadap teknologi *Deepfake* menggunakan pendekatan kualitatif melalui analisis teks dari forum diskusi daring Quora. Penelitian ini mengidentifikasi delapan tema dominan dalam cara publik memahami *Deepfake*, antara lain kekhawatiran terhadap manipulasi informasi, hilangnya kepercayaan terhadap konten digital, serta ketidakpastian dalam membedakan konten asli dari yang palsu. Penelitian ini relevan karena menggunakan pendekatan kualitatif dalam menggali persepsi kognitif pengguna terhadap *Deepfake.*

Pada penelitian tentang representasi mental publik terhadap teknologi *Deepfake* menggunakan pendekatan kualitatif melalui analisis teks dari forum diskusi daring Quora. Dengan menganalisis 166 jawaban dari 17 pertanyaan menggunakan perangkat lunak T-Lab 10, penelitian ini mengidentifikasi delapan tema dominan dalam cara publik memahami *Deepfake*. Tema-tema tersebut mencakup teknik pembuatan *Deepfake*, pendeteksian anomali visual, aplikasi pembuatan *Deepfake*, refleksi etis, inovasi hiburan, ancaman terhadap integritas informasi, pemetaan wajah, serta gerakan wajah yang tidak wajar. Selain itu, penelitian ini menemukan adanya kekhawatiran publik yang semakin meningkat terhadap manipulasi informasi, kurangnya kepercayaan terhadap konten digital, serta ketidakmampuan masyarakat dalam membedakan konten asli dari yang palsu. Penelitian ini juga menunjukkan bahwa diskusi terkait implikasi etis *Deepfake* meningkat signifikan dari waktu ke waktu, mencerminkan kesadaran publik yang terus berkembang. Penelitian ini relevan karena menggunakan pendekatan kualitatif berbasis analisis tematik untuk menggali secara mendalam persepsi dan representasi kognitif pengguna terhadap fenomena *Deepfake* dalam konteks media sosial.

2.1.5. Facebook sebagai Platform Penyebaran *Deepfake*

Berdasarkan penelitian Twomey dkk. [18], bahwa platform media sosial dengan arsitektur berbagi konten yang terbuka merupakan ekosistem yang paling kondusif bagi penyebaran konten *Deepfake*. Hal ini disebabkan oleh kemampuan platform tersebut dalam mendisssssstribusikan konten secara masif dan viral kepada jutaan pengguna dalam waktu yang sangat singkat, jauh sebelum proses verifikasi sempat dilakukan. Dalam konteks ini, Facebook menjadi salah satu platform yang paling rentan, mengingat hoaks dan rumor palsu terbukti menyebar sepuluh kali lebih cepat dibandingkan berita akurat, sementara kebohongan 70 persen lebih mungkin disebarkan ulang dibanding informasi yang benar.

Kondisi ini diperparah oleh tingginya tingkat kepercayaan pengguna terhadap konten yang dibagikan oleh orang-orang dalam jaringan pertemanan mereka, di mana setelah melihat rekomendasi teman secara online, individu cenderung meneruskan rekomendasi tersebut ke jaringan mereka sendiri. Selain itu, sistem algoritma platform seperti Facebook cenderung memprioritaskan konten dengan tingkat interaksi tinggi tanpa mempertimbangkan keasliannya. sehingga semakin mempercepat penyebaran konten *Deepfake* yang bersifat negatif dan sensasional.

Tabel 2.1. Perbandingan Penelitian Sebelumnya

No.	Penulis	Tahun	Tujuan	Metode	Hasil	Posisi Peneliti	Gap Penelitian
1.	Shahid dkk. [14]	2022	Mengkaji persepsi dan sikap pengguna media sosial terhadap video palsu termasuk *Deepfake*	Kualitatif wawancara semi terstruktur	Mayoritas pengguna tidak menyadari video dimanipulasi kemampuan deteksi sangat rendah	Landasan awal Gen Z rentan terhadap deepfake, namun hanya pada konteks India dan platform campuran (bukan khusus Facebook)	Tidak mengkaji Gen Z secara eksklusif tidak spesifik pada satu platform (Facebook) tidak mengeksplorasi dampak terhadap kepercayaan informasi konteks terbatas pada India.
2.	Hoe dkk. [17]	2024	Mengungkap strategi yang digunakan orang dalam mengidentifikasi video *Deepfake* serta tantangan yang dihadapi	Wawancara semi terstruktur individual dengan 38 peserta yang diminta menonton dan mengidentifikasi 8	Ditemukan 11 strategi identifikasi *Deepfake* yang terbagi dalam 3 kategori yaitu *Media-based* (cues visual & audio), *Knowledge-based* (pengetahuan isi pesan & tokoh), dan *Search-based*	Paling dekat secara metode dengan penelitian in; menggali strategi identifikasi deepfake secara kualitatif, namun tidak	Tidak berfokus pada Gen Z tidak spesifik pada platform Facebook tidak meneliti dimensi paparan awal maupun dampak

No.	Penulis	Tahun	Tujuan	Metode	Hasil	Posisi Peneliti	Gap Penelitian
				video (4 asli + 4 *Deepfake)*	(pencarian di internet & bertanya kepada orang lain), namun sebagian besar peserta gagal mengidentifikasi dengan benar karena prasangka awal mereka tentang ciri-ciri *Deepfake*	spesifik Gen Z/Indonesia/Facebook	kepercayaan informasi konteks bukan Indonesia.
3.	Twomey dkk . [18]	2023	Menganalisis bagaimana orang bereaksi terhadap konten *Deepfake* di Twitter selama invasi Rusia ke Ukraina, dan mencari bukti nyata dampak epistemik *Deepfake* terhadap kepercayaan publik.	*Thematic analysis* pada 1.231 tweet berbahasa Inggris, Jerman, dan Prancis yang dikumpulkan via Twitter API menggunakan kata kunci terkait	Ditemukan tiga tema utama (1) *Deepfake* dan misinformasi, (2) skeptisisme yang dipicu *Deepfake,* dan (3) Temuan penting justru *media nyata lebih sering dilabeli sebagai Deepfake* daripada sebaliknya, dan *Deepfake* mendorong teori konspirasi	Menjadi landasan teoretis dimensi ketiga penelitian ini (dampak kepercayaan informasi) menunjukkan bahwa deepfake menurunkan kepercayaan menyeluruh bahkan terhadap konten asli,	Konteks spesifik konflik geopolitik (Ukraina), bukan penggunaan Facebook sehari-hari tidak menyasar Gen Z secara eksklusif tidak mengeksplorasi pengalaman identifikasi konten deepfake secara

No.	Penulis	Tahun	Tujuan	Metode	Hasil	Posisi Peneliti	Gap Penelitian
				Deepfake dan perang Ukraina.	serta ketidakpercayaan umum terhadap media.	yang menjadi fokus pertanyaan penelitian PP3	langsung oleh pengguna platform yang diteliti adalah Twitter, bukan Facebook.
4.	Caci dkk. [19]	2024	Menggali representasi mental publik terhadap teknologi *Deepfake*	Kualitatif analisis teks forum Quora	Publik didominasi kekhawatiran manipulasi informasi dan hilangnya kepercayaan konten digital	Memetakan persepsi kognitif publik umum tentang deepfake, bukan pengalaman subjektif generasi tertentu pada platform tertentu	Tidak menyasar kelompok usia/generasi tertentu; data berasal dari forum terbuka (bukan wawancara) tidak spesifik pada platform Facebook maupun konteks Indonesia.
5.	Chesney dkk. [9]	2019	Memberikan penilaian komprehensif pertama tentang dampak teknologi	Kajian literatur dan analisis hukum (non-empiris) mencakup	Tidak ada solusi tunggal yang efektif. Teknologi deteksi masih tertinggal, hukum	Fondasi teoretis & kebijakan paling awal memberikan kerangka	Tidak ada pengamatan, tidak ada data primer dari pengguna nyata; tidak

No.	Penulis	Tahun	Tujuan	Metode	Hasil	Posisi Peneliti	Gap Penelitian
			deep fake terhadap privasi, demokrasi, dan keamanan nasional, serta mengeksplorasi respons hukum dan kebijakan yang tersedia.	telaah teknologi, dampak sosial, dan survei berbagai solusi (teknologi, hukum pidana/perdata, regulasi, militer, dan pasar).	pidana/perdata terbatas jangkauannya, regulasi agensi federal sangat terbatas, dan solusi pasar (seperti *life logging*) berpotensi merusak privasi. Penulis merekomendasikan kombinasi reformasi Section 230, kebijakan platform yang transparan, dan kewaspadaan publik.	besar ancaman deepfake di platform media sosial terbuka termasuk Facebook	mengkaji kelompok usia tertentu tidak meneliti pengalaman identifikasi atau dampak kepercayaan secara langsung.
6.	Usulan Penulis Peneliti	2026	Mengeksplorasi pengalaman subjektif Generasi Z Indonesia dalam mengidentifikasi konten *Deepfake* di Facebook	Kualitatif *Reflexive Thematic Analysis* (RTA) wawancara mendalam, semi	Hasil yang diharapkan Deskripsi mendalam pengalaman Gen Z Indonesia dalam mengidentifikasi *Deepfake* di Facebook	Satu-satunya penelitian yang menggabungkan 3 dimensi sekaligus, konteks Indonesia, platform Facebook	-

No.	Penulis	Tahun	Tujuan	Metode	Hasil	Posisi Peneliti	Gap Penelitian
			menggunakan pendekatan kualitatif, mencakup pengalaman terpapar, kemampuan identifikasi berbasis kriteria teknis, dampak kepercayaan, dan proses identifikasi	terstruktur dengan, *member checking*	berdasarkan kriteria teknis visual, serta pemahaman dampaknya terhadap kepercayaan informasi	spesifik, dan Gen Z sebagai subjek utama	

2.2. Dasar Teori

2.2.1. *Artificial Intelligence*

Momeni dkk. [20], menjelaskan kecerdasan buatan (*artificial intelligence*) AI merupakan teknologi yang memungkinkan mesin untuk mempelajari pola dari data yang telah dilatihkan sebelumnya. Kemudian menggunakan pemahaman tersebut untuk menghasilkan konten baru secara mandiri, termasuk teks, gambar, video, dan audio yang tampak seperti dibuat oleh manusia dalam perkembangannya, AI tidak hanya mampu mengenali informasi, tetapi juga mampu menciptakan informasi baru yang sebelumnya tidak pernah ada inilah yang dikenal sebagai AI generatif (*generative AI*). Salah satu terobosan paling signifikan dalam AI generatif adalah kemunculan algoritma *Generative Adversarial Networks*, yaitu sistem pembelajaran mesin yang terdiri dari dua jaringan neural yang bekerja secara kompetitif dan bersamaan.

Jaringan pertama disebut *generator*, bertugas menciptakan konten sintetis yang semirip mungkin dengan data asli, sementara jaringan kedua disebut *discriminator*, bertugas menilai apakah konten yang dihasilkan cukup meyakinkan atau masih dapat dibedakan dari konten asli. Proses kompetisi antara kedua jaringan ini berlangsung secara terus-menerus hingga konten yang dihasilkan mencapai tingkat kerealistisan yang sangat tinggi dan sulit dideteksi oleh mata manusia biasa. Teknologi GAN inilah yang kemudian menjadi fondasi utama dari *Deepfake*, yaitu konten video dan audio sintetis berbasis AI yang mampu memanipulasi wajah, suara, dan ekspresi seseorang secara sangat realistis seolah-olah orang tersebut benar-benar melakukan atau mengucapkan sesuatu yang tidak pernah terjadi. Lebih jauh, seiring semakin mudahnya akses terhadap aplikasi berbasis AI seperti DeepFaceLive, Reface, dan Sora, pembuatan *Deepfake* kini tidak lagi memerlukan keahlian teknis tinggi maupun peralatan canggih karena siapa pun dapat membuatnya hanya dengan beberapa klik. Kondisi inilah yang menjadikan penyebaran konten *Deepfake* di platform media sosial seperti Facebook semakin masif, cepat, dan semakin sulit untuk

dikendalikan, khususnya di kalangan pengguna yang belum memiliki literasi digital yang memadai untuk membedakan konten asli dari konten hasil rekayasa AI.

2.2.2. *Deepfake*

Menurut E. P. [21], *Deepfake* adalah teknologi manipulasi media yang menggunakan teknik pembelajaran mesin untuk menciptakan atau memodifikasi konten visual dan audio sehingga menyerupai individu nyata. Teknologi ini mampu meniru wajah, suara, serta ekspresi seseorang dengan tingkat kemiripan yang tinggi, sehingga sulit dibedakan dari konten asli. Seiring perkembangan AI generatif, pembuatan *Deepfake* menjadi semakin mudah dan dapat dilakukan oleh masyarakat umum dengan perangkat lunak yang tersedia secara luas.

Keberadaan *Deepfake* menimbulkan tantangan serius dalam dunia komunikasi digital, terutama terkait keaslian informasi dan perlindungan identitas individu. Penelitian di Indonesia menyoroti bahwa *Deepfake* berpotensi digunakan untuk penipuan, pencemaran nama baik, dan penyebaran disinformasi. Selain merugikan individu yang menjadi objek manipulasi, *Deepfake* juga berdampak pada menurunnya kepercayaan publik terhadap konten digital secara keseluruhan, karena pengguna menjadi semakin sulit membedakan antara fakta dan rekayasa.

2.2.3. Literasi Digital

Menurut Alkalai [22], menjelaskan bahwa literasi digital bukan sekadar kemampuan teknis mengoperasikan perangkat atau software. Konsep ini jauh lebih luas, mencakup keterampilan kognitif, motorik, sosial, sampai emosional yang dibutuhkan seseorang supaya bisa berfungsi dengan baik di lingkungan digital. Dengan kata lain, orang yang literat secara digital bukan cuma bisa "memakai" teknologi, tapi juga mampu membaca situasi, menilai informasi, dan memahami aturan main yang berlaku di dunia maya, meskipun aturan itu tidak pernah tertulis secara resmi.

Untuk memetakan kemampuan tersebut, menurut Alkalai [22], membagi literasi digital ke dalam lima jenis keterampilan *photo-visual literacy* kemampuan membaca tampilan visual dan antarmuka grafis, *reproduction literacy* kemampuan menciptakan sesuatu yang baru dan bermakna dari materi yang sudah ada, *branching literacy* kemampuan menyusun pemahaman dari informasi yang tersebar dan tidak berurutan, seperti saat menjelajah hyperlink, *information literacy*, dan *socio-emotional literacy*. Dua jenis yang terakhir paling relevan dengan konteks penelitian ini. *Information literacy* digambarkan sebagai "seni bersikap skeptis", yaitu kemampuan berpikir kritis dan tidak mudah percaya begitu saja pada informasi yang terlihat meyakinkan, sekalipun sumbernya tampak otoritatif. Sementara itu, *socio-emotional literacy* berkaitan dengan kemampuan seseorang mengenali jebakan atau manipulasi di ruang digital, termasuk memilah mana informasi yang tulus dan mana yang menipu. Kedua jenis literasi ini menjadi krusial ketika pengguna dihadapkan pada konten yang sengaja dirancang untuk terlihat asli, seperti video hasil manipulasi Deepfake, karena di situlah kemampuan untuk curiga, mengecek ulang, dan tidak langsung percaya benar-benar diuji.