BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Tinjauan pustaka ini mengulas penelitian-penelitian relevan yang menjadi landasan teoritis dan empiris bagi penelitian mengenai risiko *Deepfake AI* dan kemampuan masyarakat mendeteksi konten palsu.

Penelitian oleh Altuncu, Franqueira, dan Li (2024) berjudul *"Deepfake: Definitions, Performance Metrics and Standards, Datasets, and a Meta-Review"* memberikan kerangka komprehensif tentang definisi *Deepfake*, metrik kinerja sistem deteksi, standar evaluasi, serta kumpulan dataset yang tersedia. Penelitian ini menyimpulkan bahwa akurasi model deteksi *Deepfake* sangat bervariasi tergantung pada kualitas dataset dan metode yang digunakan, dan masih terdapat celah besar dalam standarisasi evaluasi yang menyebabkan sulitnya perbandingan antar sistem.

Darmawan, Junaidi, dan Khaerudin (2025) dalam penelitian berjudul "Penegakan Hukum terhadap Penyalahgunaan *Deepfake* pada Pornografi Anak di Era *Artificial Intelligence* di Indonesia" mengkaji dimensi hukum dari penyalahgunaan *Deepfake*, khususnya dalam konten yang merugikan anak-anak. Penelitian ini mengungkapkan bahwa regulasi di Indonesia masih belum memadai untuk menangani kasus-kasus *Deepfake* secara spesifik, dan diperlukan pembaruan peraturan perundang-undangan yang mengakomodasi perkembangan teknologi *AI.*

Temuan ini relevan karena menunjukkan bahwa risiko *Deepfake* tidak hanya bersifat teknis, tetapi juga memiliki implikasi hukum dan sosial yang serius.

Fadhilah dan Retnoningsih (2024) dalam penelitian "Perancangan Kampanye Digital Melawan Disinformasi melalui *Artificial Intelligence* dan *Deepfake* di Kalangan Pra Lansia Usia *Deepfake.* Hasil 45-55 Tahun" mengkaji kerentanan kelompok usia pra-lansia terhadap konten penelitian menunjukkan bahwa kelompok usia ini memiliki tingkat literasi digital yang lebih rendah dibandingkan kelompok usia muda, sehingga lebih rentan terhadap manipulasi konten *Deepfake.* Penelitian ini merekomendasikan pendekatan kampanye digital yang disesuaikan dengan karakteristik demografis sasaran sebagai strategi efektif peningkatan literasi digital.

Seveney, Wicaksono, dan Soetijono (2025) dalam "Urgensi Regulasi terhadap Penyalahgunaan *Deepfake* Berbasis *AI* pada Konten Pornografi" menekankan urgensi regulasi yang komprehensif terhadap penyalahgunaan *Deepfake*. Penelitian ini menemukan bahwa ketiadaan regulasi yang spesifik menciptakan ruang abu-abu hukum yang dimanfaatkan oleh pelaku kejahatan *siber*. Temuan ini memperkuat pentingnya memahami risiko sosial *Deepfake* dari perspektif kebijakan publik, yang relevan dengan tujuan penelitian ini dalam mengidentifikasi faktor-faktor risiko yang dihadapi masyarakat.

Payne (2026) dalam entri ensik lopedia Britannica mendefinisikan *Deepfake* sebagai representasi digital sintetis yang dihasilkan melalui algoritma pembelajaran mendalam, dan menyoroti dampaknya terhadap kepercayaan publik terhadap media. Definisi ini menjadi landasan konseptual dalam penelitian ini untuk

memahami karakteristik dan batasan teknologi Deepfake. Selain itu, tinjauan ini juga mencatat bahwa literasi media digital merupakan faktor kritis yang menentukan kemampuan seseorang dalam mengidentifikasi konten Deepfake secara akurat.

Penelitian mengenai perkembangan teknologi pada era industri 4.0 dan society 5.0 menunjukkan bahwa Artificial intelligence (AI) menjadi salah satu inovasi utama yang mendorong percepatan otomatisasi dan pengambilan keputusan cerdas. perkembangan ini memperlihatkan bagaimana AI terus memperluas perannya dalam kehidupan modern. Studi terdahulu juga menyoroti bahwa AI digunakan pada berbagai aplikasi seperti peramalan cuaca, sistem penerjemahan, hingga sistem rekomendasi di berbagai sektor;bahwa pemanfaatan tersebut semakin meluas dari tahun ke tahun. Selain itu, AI semakin banyak di manfaatkan dalam pembuatan konten digital, termasuk gambar, audio, dan video, karena kemampuannya menghasilkan output secara cepat dan efisien. Temuan-temuan ini menegaskan peran signifikan AI dalam mendukung transformasi digital di berbagai bidang. (Lestari, P. A et al., 2026)

2.2 Landasan Teori

2.2.1 Kecerdasan Buatan (Artificial Intelligence)

Artificial Intelligence (AI) merupakan bidang ilmu komputer yang bertujuan untuk menciptakan mesin atau program komputer yang dapat melakukan tugas yang memerlukan kecerdasan manusia, seperti pengenalan pola, pengambilan keputusan, dan pemrosesan bahasa alami (Juditha, 2025). AI telah diterapkan secara luas pada berbagai bidang,

mulai dari peramalan cuaca, sistem penerjemahan, sistem rekomendasi, hingga pembuatan konten digital.

Verdoliva (2023) menjelaskan bahwa perkembangan *AI* generatif menyebabkan konten digital hasil manipulasi semakin realistis sehingga sulit dikenali hanya melalui pengamatan manusia. Oleh karena itu, bidang *media forensics* berkembang sebagai disiplin ilmu yang berfokus pada identifikasi keaslian gambar, video, dan audio menggunakan analisis karakteristik digital.

Penelitian ini menguraikan berbagai pendekatan deteksi, mulai dari analisis artefak visual, analisis frekuensi citra, hingga penggunaan model *deep learning*. Selain itu, penulis menegaskan bahwa perlombaan antara teknologi pembuat *Deepfake* dan teknologi pendeteksi akan terus berlangsung sehingga pengembangan metode deteksi harus dilakukan secara berkelanjutan.

2.2.2 *Deepfake*

Deepfake merupakan media sintetis yang dihasilkan menggunakan teknologi *AI* untuk menampilkan sesuatu yang sebenarnya tidak pernah terjadi. Istilah *Deepfake* berasal dari gabungan kata *Deep learning*, yang merujuk pada teknologi pembelajaran mesin yang mendalam, dan *fake* yang berarti palsu. Kemunculan *Deepfake* pertama kali digunakan untuk memberikan kemudahan bagi manusia dalam menciptakan hasil foto atau video yang lebih nyata dalam konten hiburan seperti film atau gim,

namun seiring berjalannya waktu teknologi ini mulai disalah gunakan untuk berbagai kepentingan negatif (Seveney, Wicaksono, & Soetijono, 2025).

Dalam penelitian ini, istilah *Deepfake* digunakan dalam pengertian yang luas (broad sense), yaitu mencakup segala bentuk konten visual yang dimanipulasi atau dihasilkan menggunakan teknologi *AI* generatif, tidak terbatas pada teknik *face-swap* berbasis *Generative Adversarial Network* (*GAN*) saja, tetapi juga mencakup hasil *AI image editing generatif* seperti yang tersedia pada *Gemini* (*Google*), sepanjang hasilnya menampilkan representasi visual yang menyerupai kenyataan namun sesungguhnya telah direkayasa.

2.2.3 *Generative Adversarial Network* (*GAN*)

GAN merupakan metode utama yang digunakan dalam pembuatan konten *Deepfake,* yang bekerja dengan menggunakan dua model jaringan saraf tiruan yang saling bersaing, yaitu generator yang menghasilkan konten palsu dan discriminator yang berupaya membedakan konten palsu dengan konten asli. Proses ini berlangsung berulang kali sehingga menghasilkan konten yang semakin realistis dan sulit dibedakan dari konten asli (Altuncu, Franqueira, & Li, 2024).

Mirsky dan Lee (2021) menjelaskan bahwa *Deepfake* merupakan media sintetis yang dihasilkan menggunakan teknik *deep learning* sehingga mampu memanipulasi wajah, suara, maupun gerakan seseorang dengan tingkat kemiripan yang sangat tinggi. Teknologi ini berkembang

pesat karena didukung oleh model pembelajaran mendalam, terutama *Generative Adversarial Network* (*GAN*), yang mampu menghasilkan konten digital menyerupai kondisi nyata. Selain membahas proses pembuatan *Deepfake*, penelitian ini juga mengulas berbagai metode deteksi, seperti analisis inkonsistensi pencahayaan, gerakan wajah, kedipan mata, tekstur kulit, hingga pendekatan berbasis *Convolutional Neural Network* (CNN). Penelitian tersebut menyimpulkan bahwa seiring meningkatnya kualitas teknologi generatif, proses pendeteksian menjadi semakin sulit sehingga diperlukan kombinasi teknologi deteksi otomatis dan peningkatan literasi masyarakat.

Wang dkk. (2023) mengulas secara komprehensif perkembangan teknologi pembangkitan (*generation*) dan deteksi (*detection*) *Deepfake*. Penelitian ini menjelaskan bahwa kualitas *Deepfake* meningkat sangat cepat karena penggunaan model *AI* generatif modern yang mampu menghasilkan gambar dan video dengan detail tinggi. Penulis mengelompokkan metode deteksi menjadi beberapa kategori, seperti deteksi berbasis fitur visual, deteksi berbasis fisiologis, analisis temporal video, serta model pembelajaran mendalam. Hasil kajian menunjukkan bahwa belum terdapat satu metode yang mampu mendeteksi seluruh jenis *Deepfake* secara konsisten karena setiap model manipulasi memiliki karakteristik yang berbeda.

2.2.4 Disinformasi dan Hoaks Digital

Disinformasi merupakan informasi yang sengaja dibuat salah dengan tujuan menyesatkan atau memanipulasi opini publik. Media sosial menjadi saluran utama penyebaran hoaks karena keterbukaan platform dalam pertukaran informasi tanpa pengawasan dan penyaringan yang ketat (Sabela, Saputra, & Kuncoro, 2026).

Lin et al. (2024) menjelaskan bahwa sistem deteksi media hasil *AI* dikembangkan melalui berbagai pendekatan, seperti analisis artefak visual, analisis frekuensi citra, deteksi inkonsistensi semantik, analisis metadata, hingga model pembelajaran mendalam (*deep learning*). Akan tetapi, peningkatan kualitas *AI* generatif menyebabkan karakteristik manipulasi semakin sulit dikenali sehingga performa sistem deteksi harus terus diperbarui mengikuti perkembangan teknologi pembangkit konten. Kondisi tersebut menunjukkan bahwa penyelesaian masalah disinformasi tidak hanya bergantung pada teknologi deteksi, tetapi juga memerlukan peningkatan kemampuan masyarakat dalam mengevaluasi informasi digital secara kritis.

2.2.5 Literasi Digital

Literasi digital merupakan kemampuan seseorang dalam memahami, mengevaluasi, dan menggunakan informasi digital secara kritis. Tingkat literasi digital menjadi faktor penentu kemampuan seseorang dalam mengenali dan menghindari dampak negatif dari konten manipulatif, termasuk *Deepfake* (Payne, 2026).

Dhahir dkk. (2024) menunjukkan bahwa terdapat hubungan positif antara tingkat literasi digital dengan kemampuan seseorang dalam mengidentifikasi video *Deepfake*. Individu yang memiliki literasi digital lebih tinggi cenderung lebih kritis dalam mengevaluasi sumber informasi, memperhatikan kejanggalan visual, serta tidak mudah mempercayai konten yang beredar di media sosial. Temuan ini menunjukkan bahwa peningkatan literasi digital dapat menjadi salah satu strategi efektif dalam mengurangi dampak penyebaran disinformasi berbasis *AI.*

2.2.6 Metode Deskriptif Kuantitatif

Metode penelitian kuantitatif merupakan pendekatan ilmiah yang bertujuan untuk mengukur, menjelaskan, dan menguji hubungan antar variabel melalui data numerik yang dianalisis secara statistik, sehingga memungkinkan peneliti menarik kesimpulan yang objektif, teruji, dan dapat di generalisasikan ke populasi yang lebih luas (Hair et al., 2022). Metode deskriptif kuantitatif secara khusus bertujuan untuk menggambarkan dan menganalisis fenomena secara sistematis berdasarkan data numerik yang dikumpulkan dari lapangan.

2.2.7 Kuesioner sebagai Instrumen Penelitian

Kuesioner merupakan instrumen pengumpulan data yang berisi serangkaian pertanyaan atau pernyataan yang disusun secara sistematis untuk memperoleh informasi dari responden. Dalam penelitian ini, kuesioner disusun dengan menyertakan materi stimulus berupa pasangan

konten (asli dan *Deepfake*) yang harus dinilai oleh responden, disertai pertanyaan mengenai pengetahuan dan pengalaman responden terkait teknologi *Deepfake* (Mamuaya et al., 2025).

2.2.8 Pengetahuan tentang Teknologi *Deepfake*

Pengetahuan tentang teknologi *Deepfake* mengacu pada sejauh mana seseorang memahami konsep, cara kerja, serta ciri-ciri konten yang dihasilkan oleh teknologi *Deepfake*. Pemahaman ini mencakup kesadaran akan keberadaan teknologi *Deepfake*, pengetahuan bahwa konten tersebut dihasilkan melalui algoritma *AI* seperti *Generative Adversarial Network* (*GAN*), serta kemampuan mengenali indikator visual maupun audio yang membedakan konten asli dari konten manipulasi (Payne, 2026; Juditha, 2025). Semakin tinggi tingkat pengetahuan seseorang mengenai karakteristik dan mekanisme kerja *Deepfake,* semakin besar pula potensi mereka untuk lebih waspada dan kritis ketika mengonsumsi konten digital. Dalam penelitian ini, pengetahuan tentang *Deepfake* diposisikan sebagai variabel bebas (X1) yang diduga berkontribusi terhadap kemampuan deteksi konten *Deepfake.*

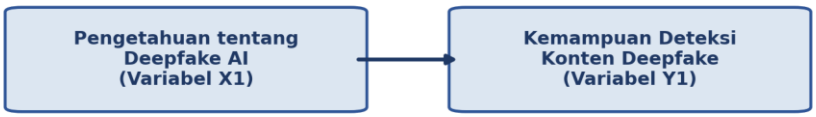
2.2.9 Kemampuan Deteksi Konten Digital

Kemampuan deteksi konten digital merupakan kesanggupan seseorang untuk mengidentifikasi dan membedakan konten asli dari konten hasil manipulasi atau rekayasa digital. Kemampuan ini dipengaruhi oleh berbagai faktor, termasuk tingkat literasi digital,

pengalaman terhadap teknologi, serta pengetahuan mengenai ciri-ciri teknis konten manipulatif (Altuncu, Franqueira, & Li, 2024). Semakin realistis hasil manipulasi yang dihasilkan oleh teknologi *AI*, semakin sulit pula bagi individu untuk mendeteksinya secara visual tanpa bantuan pengetahuan atau alat bantu teknis (Payne, 2026). Dalam penelitian ini, kemampuan deteksi diposisikan sebagai variabel terikat (Y1) yang diukur secara empiris melalui uji stimulus, yaitu dengan meminta responden mengidentifikasi konten asli dari pasangan konten asli dan *Deepfake* yang disajikan dalam kuesioner.

2.3 Kerangka Berpikir

Berdasarkan tinjauan pustaka dan landasan teori yang telah diuraikan, penelitian ini disusun atas dasar asumsi bahwa tingkat pengetahuan masyarakat tentang teknologi *Deepfake AI* (variabel X1) berperan terhadap kemampuan mereka dalam mendeteksi konten *Deepfake AI* (variabel Y1). Kerangka berpikir penelitian ini digambarkan pada Gambar 2.1 berikut.



Gambar 2. 1 Kerangka Berpikir Penelitian

2.4 Hipotesis Penelitian

Berdasarkan kerangka berpikir di atas, maka hipotesis yang diajukan dalam penelitian ini adalah sebagai berikut:

H0: Tidak terdapat hubungan yang signifikan antara tingkat pengetahuan masyarakat tentang *Deepfake AI* dengan kemampuan mereka dalam mendeteksi konten *Deepfake AI.*

Ha: Terdapat hubungan yang signifikan antara tingkat pengetahuan masyarakat tentang *Deepfake AI* dengan kemampuan mereka dalam mendeteksi konten *Deepfake AI.*