BAB I

PENDAHULUAN

1.1. Latar Belakang

Perkembangan kecerdasan buatan telah melahirkan teknologi *Deepfake*, yaitu teknik memanipulasi video dan audio yang memungkinkan wajah dan suara seseorang diganti dengan orang lain secara sangat realistis. Teknologi ini tidak hanya mampu menghasilkan konten yang terlihat meyakinkan, tetapi juga semakin mudah digunakan oleh siapa saja karena banyak aplikasinya yang tersedia secara gratis, sehingga penyalahgunaannya pun semakin sulit dibendung [1][2]. Salah satu contoh nyata dari dampak teknologi ini dijelaskan oleh Zahro dkk. [3], di mana sempat viral sebuah video yang memperlihatkan Presiden Joko Widodo seolah-olah fasih berbicara dalam bahasa Mandarin sebuah konten yang sepenuhnya merupakan hasil rekayasa *Deepfake* namun berhasil menipu banyak kalangan masyarakat. Fenomena semacam ini tidak terlepas dari peran media sosial yang kini menjadi medium utama penyebaran *Deepfake* secara masif dan cepat.

Di antara semua platform media sosial yang ada, Facebook menjadi platform yang paling banyak digunakan sekaligus paling rawan. Menurut data Surfshark [4], menjelaskan dari kumpukan data yang mereka lakukan dari tahun 2017 sampai 2025 tercatat bahwa Facebook memimpin dalam penipuan terkait *Deepfake* pada tahun 2025. Disi lain Data Reportal menjelaskan [5], Facebook tercatat sebagai platform media sosial dengan posisi pertama terbanyak di dunia dan menempati posisi keempat dunia dalam jumlah pengguna Facebook per april 2026, sekaligus peringkat kedua dunia dalam jangkauan iklan di platform tersebut [6]. Data ini juga didukung juga oleh dataIndonesia.id [7], di mana Indonesia tercatat sebagai negara dengan pengguna Facebook terbanyak ketiga di dunia per Juli 2025. Dari sisi Gen Z nya sendiri menduduki peringkat ke dua yang sedangkan Milenial menduduki peringkat ke satu dengan penggunaan

Facebook di indonesia menurut penejelasan dari Kompas.com [8]. Besarnya basis pengguna ini menjadikan Facebook sebagai ruang digital yang sangat strategis dalam kehidupan informasi masyarakat Indonesia. Chesney dkk. [9], menjelaskan bahwa sistem berbagi konten yang terbuka di media sosial menciptakan kondisi yang sangat ideal bagi penyebaran *Deepfake*, karena konten yang menipu bisa menjangkau jutaan orang jauh sebelum sempat dicek kebenarannya.

Generasi Z mereka yang lahir antara tahun 1997 hingga 2012 merupakan kelompok yang paling sering terpapar konten *Deepfake*, namun justru memiliki kemampuan identifikasi yang paling rendah. Temuan ini juga dibahas oleh kolaborasi internasional antara Departemen Teknik Mesin dan Industri di Universitas Gadjah Mada (UGM) dan *Deakin University* [10], mereka mengkhawatirkan mengingat Gen Z sering dipuja sebagai generasi yang melek digital, dari temuan mereka mengatakan bahwa dari 647 Gen Z yang berpartisipasi, 83% dari Gen Z cenderung menganggap informasi yang diberikan sebagai berita yang benar meskipun sebenarnya hoaks. Sementara, 71 orang atau sekitar 11%, cenderung menganggap informasi yang diberikan sebagai hoaks meski sebenarnya berita benar. Mirisnya hanya 42 orang atau sekitar 6% yang dapat membedakan berita benar dan berita palsu dengan tepat dan akurat.

Temuan mereka ini sejalan dengan sebuah penelitian dari Stanford *University* di AS yang menemukan bahwa meski Gen Z merupakan *digital natives*, 82% partisipan Gen Z bahkan gagal membedakan berita dan iklan. Di Indonesia, masih lemahnya kemampuan Gen Z dalam mendeteksi hoaks bisa jadi disebabkan oleh masih masih kurangnya kemampuan literasi digital yang mereka miliki konsep yang akan dibahas lebih lanjut pada bagian dasar teori literasi digital. Berbagai akademisi mengamati banyaknya celah dalam edukasi literasi digital di berbagai jenjang pendidikan Indonesia.

Universitas Cambridge [11], melalui *Misinformation Susceptibility Test*, yang melibatkan lebih dari 8.000 partisipan, menemukan bahwa hanya 11% dari kelompok usia 18 hingga 29 tahun yang mampu mengidentifikasi berita palsu dengan benar, jauh lebih rendah dibandingkan 36% pada kelompok usia 65 tahun ke atas. Sultana dkk. [12], yang memperjelas temuan ini dengan menunjukkan

bahwa kemampuan deteksi *Deepfake* oleh manusia secara umum masih sangat rendah, bahkan ketika dibantu oleh teknologi sekalipun. Hancock dkk. [13], menjelaskan mengapa hal ini terjadi konten audiovisual cenderung dinilai lebih kredibel dibandingkan teks karena kedekatannya dengan persepsi inderawi manusia, sehingga menjadikannya lebih efektif dalam membentuk keyakinan yang keliru pada penontonnya. Temuan Shahid dkk. [14], memperkuat kondisi ini dalam penelitiannya terhadap 36 pengguna media sosial di India, sebagian besar pengguna tidak menyadari bahwa video yang mereka tonton telah dimanipulasi, bahkan sebagian dari mereka secara sadar tetap menyebarkan konten yang diketahui palsu apabila isinya sesuai dengan keyakinan yang mereka pegang sebelumnya.

Westerlund [1] menjelaskan juga meskipun pengguna awam umumnya mengandalkan tanda-tanda visual seperti gerakan mulut yang tidak wajar, arah mata yang tidak natural, atau distorsi pada tepi wajah untuk mendeteksi *Deepfake*, dalam praktiknya kemampuan ini masih jauh dari memadai. Meskipun indikator-indikator ini secara teknis dapat dikenali, Mutmainnah dkk. [15], menegaskan bahwa teknologi *Deepfake* kini telah berkembang sedemikian rupa sehingga konten yang dihasilkan hampir tidak dapat dibedakan dari konten asli oleh pengguna yang tidak terlatih menjadikan ketergantungan pada pengamatan visual semata sebagai strategi yang semakin tidak andal.

Ketidakmampuan Generasi Z dalam mengidentifikasi *Deepfake* telah menimbulkan dampak nyata yang serius. Kristiyenda dkk. [16], mencatat bahwa pada awal 2025, pelaku penipuan menggunakan *Deepfake* wajah Presiden Prabowo Subianto dan mantan Menteri Keuangan Sri Mulyani untuk mengelabui korban melalui Facebook dengan dalih program bantuan pemerintah sebuah kasus yang membuktikan bahwa ancaman ini bukan sekadar teoretis. Hancock dkk. [13], menjelaskan bahwa paparan terhadap konten *Deepfake* tidak hanya menimbulkan kerugian material, tetapi juga membawa konsekuensi sosial yang lebih dalam meningkatnya ketidakpastian masyarakat terhadap media secara umum, yang pada akhirnya mendorong penurunan kepercayaan terhadap informasi bahkan terhadap konten yang sebenarnya asli. Kondisi ini menciptakan

apa yang disebut sebagai *liar's dividend* atau disebut dengan menyangkal kebenaran bukti, rekaman, atau informasi yang sah dengan menuduhnya sebagai berita bohong (*fake news)*, di mana masyarakat menjadi apatis dan skeptis terhadap semua konten digital, sehingga disinformasi justru semakin sulit diatasi.

Dari penelitian di atas belum mampu menjawab pertanyaan mendasar yang justru paling relevan untuk konteks Indonesia yaitu bagaimana sesungguhnya Generasi Z Indonesia secara subjektif mengalami, merasakan, dan menjalani proses menghadapi konten *Deepfake* di Facebook mulai dari pengalaman pertama kali terpapar, cara mereka berupaya mengidentifikasi konten tersebut berdasarkan tanda-tanda visual yang mereka amati, hingga bagaimana pengalaman itu mengubah kepercayaan mereka terhadap informasi.

1.2. Rumusan Masalah

Generasi Z merupakan kelompok dengan tingkat paparan tertinggi terhadap konten Deepfake di media sosial dari pada generasi lainnya. Namun justru menunjukkan kemampuan deteksi yang masih rendah terhadap konten yang kini semakin realistis dan sulit dibedakan dari konten asli, khususnya di platform Facebook. Ketidakmampuan ini membuka ruang bagi manipulasi opini publik, penipuan finansial berkedok tokoh publik, hingga menurunnya kepercayaan masyarakat terhadap informasi di media sosial.

Apabila persoalan ini tidak segera diteliti, kesenjangan literasi digital pada Generasi Z berisiko membuat mereka justru ikut menyebarkan misinformasi berbasis kecerdasan buatan tanpa disadari. Padahal generasi ini justru diharapkan menjadi agen literasi digital bagi kelompok yang lebih rentan di sekitar mereka. Penelitian-penelitian sebelumnya pun belum menjawab persoalan ini secara utuh untuk konteks Indonesia, karena belum ada yang menggali pengalaman subjektif Generasi Z secara menyeluruh mulai dari paparan awal, proses identifikasi, hingga dampaknya terhadap kepercayaan informasi pada satu platform yang sama, yaitu Facebook.

1.3. Pertanyaan Penelitian

Berdasarkan pernyataan pada rumusan masalah di atas, maka pertanyaan penelitian yang dikemukakan yaitu :

1. Bagaimana pengalaman umum Generasi Z pada saat terpapar video *Deepfake* di media sosial Facebook?
2. Bagaimana proses identifikasi yang dilakukan Generasi Z ketika mencurigai suatu konten sebagai *Deepfake* berbasis kecerdasan buatan di media sosial Facebook?
3. Bagaimana Dampak video *Deepfake* mempengaruhi kepercayaan Generasi Z terhadap informasi yang mereka dapatkan?

1.4. Tujuan Penelitian

Tujuan yang ingin dicapai dari penelitian ini adalah :

1. Mendeskripsikan pengalaman umum Generasi Z pada saat pertama kali terpapar konten *Deepfake* di media sosial Facebook.
2. Mengeksplorasi proses identifikasi yang dilakukan Generasi Z ketika mencurigai suatu konten sebagai *Deepfake* berbasis kecerdasan buatan di media sosial Facebook, termasuk strategi dan langkah-langkah yang mereka tempuh dalam memverifikasi keaslian konten tersebut.
3. Mengeksplorasi bagaimana paparan konten *Deepfake* mempengaruhi kepercayaan Generasi Z terhadap informasi yang mereka peroleh dari Facebook.

1.5. Batasan Masalah

Agar penelitian ini lebih terfokus dan tidak meluas, maka batasan masalah yang diterapkan adalah sebagai berikut :

1. Penelitian ini hanya membahas pengalaman Generasi Z dalam mengidentifikasi konten *Deepfake* dari tiga dimensi sesuai pertanyaan penelitian, pertama pengalaman umum terpapar, kedua proses identifikasi *Deepfake*, dan ketiga dampak terhadap kepercayaan informasi. Penelitian tidak mencakup dimensi psikologis klinis.
2. Penelitian difokuskan pada platform media sosial Facebook, fokus tunggal pada platform Facebook merupakan keterbatasan penelitian ini, mengingat karakteristik penyebaran dan jenis konten Deepfake dapat berbeda pada platform lain seperti TikTok, Instagram, atau YouTube. Dengan demikian, temuan penelitian ini tidak dapat digeneralisasikan pada pengalaman pengguna di platform media sosial lain.
3. Penelitian ini tidak bertujuan untuk menggeneralisasi temuan ke seluruh populasi Generasi Z, melainkan untuk menggali pengalaman subjektif secara mendalam sesuai dengan pendekatan kualitatif yang digunakan.
4. Penelitian ini berfokus pada konten *Deepfake* berbasis video dan audio yang ditemukan di Facebook, dan tidak mencakup konten manipulatif dalam bentuk foto diam maupun teks.
5. Penelitian ini hanya berfokus pada pengalaman Generasi Z sebagai pengguna yang terpapar dalam mengidentifikasi konten *Deepfake* yang lahir di tahun 1997 - 2012 Generasi Z yang berstatus sebagai mahasiswa aktif dan berdomisili di Indonesia, dan tidak mencakup pengalaman dari sisi pembuatan *Deepfake*.

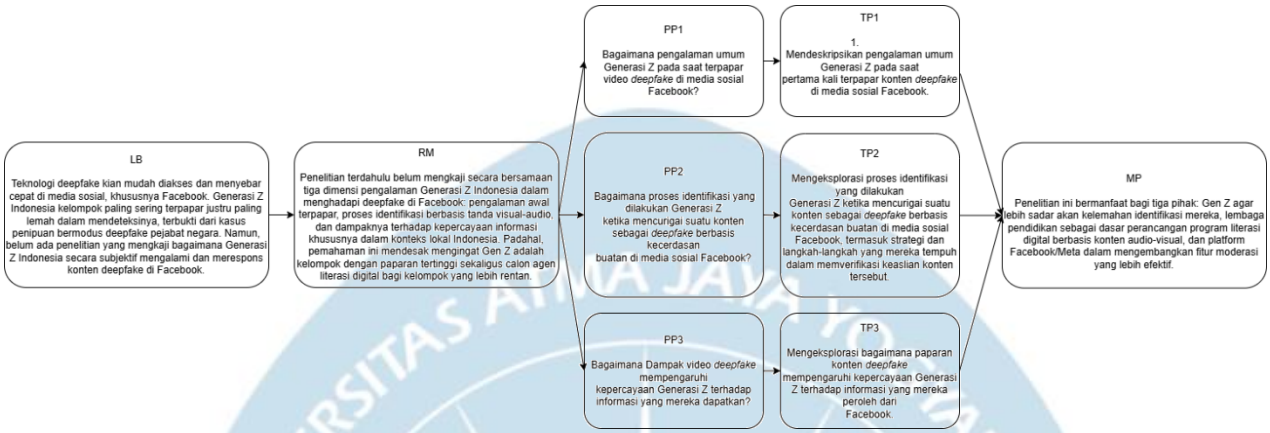
1.6. Manfaat Penelitian

Hasil penelitian ini diharapkan dapat membantu Generasi Z memahami pola dan kelemahan mereka sendiri dalam mengidentifikasi konten *Deepfake* di Facebook, sehingga mereka dapat lebih waspada dan tidak mudah tertipu. Selain itu, temuan penelitian ini dapat menjadi cermin bagi Gen Z untuk menyadari bahwa kemampuan identifikasi yang selama ini mereka andalkan masih bersifat intuitif dan perlu diperkuat dengan strategi verifikasi yang lebih sistematis.

Hasil penelitian ini diharapkan dapat memberikan manfaat bagi lembaga pendidikan sebagai dasar pertimbangan dalam merancang program literasi digital dan berpikir kritis. Selain itu, temuan penelitian ini dapat dijadikan rujukan dalam mengembangkan materi pembelajaran yang bersifat praktis dan teknis dalam mengenali serta memverifikasi keaslian konten audio-visual yang relevan dan kontekstual, khususnya terkait kemampuan identifikasi konten manipulatif berbasis AI di kalangan mahasiswa.

Hasil penelitian ini juga diharapkan dapat menjadi referensi bagi platform Facebook/Meta dalam memahami perilaku pengguna Generasi Z ketika berhadapan dengan konten *Deepfake* guna mendukung pengembangan fitur moderasi yang lebih efektif.

1.7. Bagan Keterkaitan



Gambar 1. 1. Bagan Keterikatan