BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan *Artificial Intelligence* (AI) dalam beberapa tahun terakhir menunjukkan peningkatan yang sangat signifikan, terutama pada cabang *generative artificial intelligence*. Teknologi ini memungkinkan sistem komputer menghasilkan teks, suara, maupun gambar secara otomatis dengan tingkat realisme yang tinggi. AI generatif tidak lagi berada pada ranah eksperimen terbatas, melainkan telah terintegrasi ke dalam berbagai *platform* digital dan media sosial untuk mendukung interaksi pengguna serta produksi konten yang lebih cepat dan adaptif [1]. Perkembangan tersebut mengubah cara pengguna memproduksi, mengonsumsi, dan memaknai konten digital, karena batas antara konten autentik dan konten sintetis menjadi semakin sulit dibedakan [2]. Transformasi ini menandai pergeseran besar dalam ekosistem digital yang sebelumnya bertumpu pada produksi konten oleh manusia secara langsung.

Di balik kemudahan dan efisiensi yang ditawarkan, AI generatif juga memunculkan persoalan etis dan potensi penyalahgunaan. Salah satu fenomena yang banyak dibahas adalah *deepfake*, yakni teknik manipulasi visual atau audio berbasis pembelajaran mesin yang mampu meniru identitas seseorang secara meyakinkan [2]. Penelitian Mai dkk. terhadap 529 narasumber menunjukkan bahwa tingkat keberhasilan manusia dalam mendeteksi *speech deepfake* hanya sekitar 73% [3]. Angka tersebut memperlihatkan bahwa kemampuan deteksi manual masih memiliki keterbatasan. Studi lain juga menemukan bahwa individu cenderung merasa terlalu yakin dalam menilai keaslian konten digital, meskipun performa deteksinya tidak selalu akurat [4]. Kondisi ini membuka ruang bagi penyalahgunaan teknologi untuk menghasilkan konten tidak pantas yang dapat merugikan individu dari sisi reputasi maupun privasi.

Risiko tersebut menjadi semakin kompleks ketika AI generatif terintegrasi langsung dalam suatu *platform* media sosial. Grok AI sebagai chatbot AI berbasis *Large Language Model* (LLM) yang dikembangkan oleh xAI telah terintegrasi dengan *platform* X dan memungkinkan produksi serta modifikasi konten secara otomatis [5]. Integrasi ini memperluas akses pengguna terhadap kemampuan generatif, namun sekaligus

membuka peluang penyalahgunaan terhadap konten pribadi. Dampak nyata dari fenomena ini terlihat pada kasus Grok AI di *platform* X pada Januari 2026, ketika fitur tersebut dimanfaatkan untuk menghasilkan gambar pornografi palsu melalui manipulasi wajah tanpa persetujuan[6]. Kasus tersebut mendorong Kementerian Komunikasi dan Digital (Komdigi) untuk memblokir sementara akses Grok AI pada 10 Januari 2026 karena dinilai melanggar prinsip keamanan dan etika dalam ekosistem digital [6][7]. Peristiwa ini menunjukkan bahwa penyalahgunaan AI generatif telah menimbulkan konsekuensi sosial dan kebijakan yang nyata.

Implikasi dari peristiwa tersebut berkaitan erat dengan rasa aman pengguna dalam mengunggah foto pribadi. Media sosial, termasuk *platform* X, selama ini menjadi ruang berbagi opini dan dokumentasi visual. Namun, aktivitas berbagi tersebut sangat bergantung pada tingkat kepercayaan pengguna terhadap keamanan platform. Penelitian yang dilakukan oleh Niu dkk. (2025) menganalisis 27.800 gambar dari 6.423 pengguna Twitter (Sekarang X) menemukan adanya potensi *face privacy leakage* pada unggahan wajah [8] yang menunjukkan bahwa media sosial bukan sekadar ruang berbagi, melainkan juga ruang yang rentan terhadap eksploitasi data visual. Ancaman manipulasi visual juga memiliki implikasi terhadap keamanan data biometrik, karena citra wajah yang dimodifikasi dapat memicu pelanggaran identitas digital secara berkelanjutan [9]. Selain itu, Pawestri dan Jumino melakukan penelitian terhadap 156 narasumber pengguna Twitter (Sekarang X) di Indonesia menunjukkan bahwa *information privacy concern* berpengaruh terhadap perilaku perlindungan privasi pengguna [10]. Temuan-temuan ini mengindikasikan bahwa meningkatnya ancaman manipulasi visual dapat memengaruhi persepsi keamanan serta keputusan pengguna dalam membagikan foto pribadi.

1.2 Rumusan Masalah

Integrasi Grok AI di platform X memunculkan potensi penyalahgunaan berupa manipulasi wajah tanpa persetujuan pengguna, yang berdampak langsung pada pengalaman dan rasa aman dalam berbagi konten pribadi. Meskipun penelitian tentang *deepfake* dan AI generatif terus berkembang, studi yang secara spesifik mengkaji dampaknya terhadap persepsi keamanan pengguna dalam konteks platform X masih sangat terbatas. Hal ini menimbulkan pertanyaan mendasar mengenai bagaimana

pengguna memaknai rasa aman mereka dalam mengunggah foto pribadi di tengah ancaman manipulasi visual berbasis AI. Oleh karena itu, penelitian ini bertujuan untuk menggali persepsi pengguna terkait rasa aman dalam membagikan konten di platform X pasca munculnya fenomena penyalahgunaan Grok AI.

1.3 Pertanyaan Penelitian

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

1. Bagaimana penyalahgunaan Grok AI memengaruhi persepsi pengguna terkait rasa aman dalam mengunggah foto pribadi di *platform* X?
2. Apa saja bentuk perubahan perilaku pengguna setelah munculnya fenomena penyalahgunaan Integrasi Grok AI di *platform* X?

1.4 Tujuan Penelitian

Berdasarkan pertanyaan penelitian yang telah dirumuskan, maka tujuan penelitian ini adalah:

1. Untuk menganalisis bagaimana penyalahgunaan Grok AI memengaruhi persepsi pengguna terkait rasa aman dalam mengunggah foto pribadi pada *platform* X.
2. Untuk mengidentifikasi perubahan perilaku pengguna setelah munculnya fenomena penyalahgunaan Grok AI di *platform* X.

1.5 Batasan Masalah

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

1. Penelitian ini hanya menganalisis pengaruh penyalahgunaan Grok AI terhadap persepsi rasa aman pengguna dalam mengunggah foto pribadi serta perubahan perilaku pengguna setelah munculnya fenomena tersebut, seperti keputusan untuk mengunggah, membatasi, atau menghentikan unggahan foto pribadi.
2. Pembahasan dibatasi pada konteks *platform* X dan tidak mencakup analisis teknis mengenai algoritma, mekanisme deteksi *deepfake*, kebijakan internal perusahaan, maupun regulasi hukum secara mendalam.

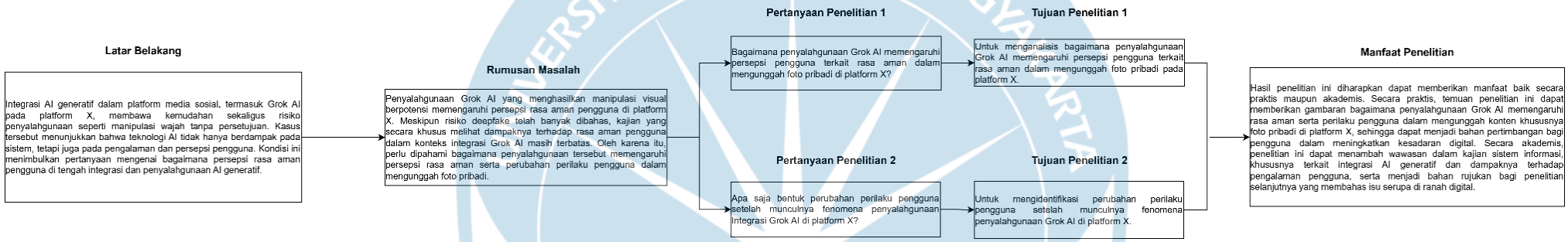
3. Hasil penelitian tidak dapat digeneralisasikan ke *platform* media sosial lain atau populasi pengguna secara statistik karena penelitian menggunakan pendekatan kualitatif yang berfokus pada pengalaman dan pemaknaan subjektif narasumber.
4. Penelitian tidak membedakan pengguna berdasarkan status akun (premium maupun non-premium), sehingga tidak membandingkan persepsi rasa aman atau perubahan perilaku berdasarkan tingkat keanggotaan akun.

1.6 Manfaat Penelitian

Hasil penelitian ini diharapkan dapat memberikan manfaat baik secara praktis maupun akademis. Secara praktis, temuan penelitian ini dapat memberikan gambaran bagaimana penyalahgunaan Grok AI memengaruhi rasa aman serta perilaku pengguna dalam mengunggah konten khususnya foto pribadi di *platform* X, sehingga dapat menjadi bahan pertimbangan bagi pengguna dalam meningkatkan kesadaran digital. Secara akademis, penelitian ini dapat menambah wawasan dalam kajian sistem informasi, khususnya terkait integrasi AI generatif dan dampaknya terhadap pengalaman pengguna, serta menjadi bahan rujukan bagi penelitian selanjutnya yang membahas isu serupa di ranah digital.

1.7 Bagan Keterkaitan

Pada Gambar 1.1 di bawah ini ditampilkan bagan keterkaitan antara latar belakang, rumusan masalah, pertanyaan penelitian, tujuan penelitian, serta metode penelitian:



Gambar 1. 1 Bagan Keterkaitan