

BAB 1

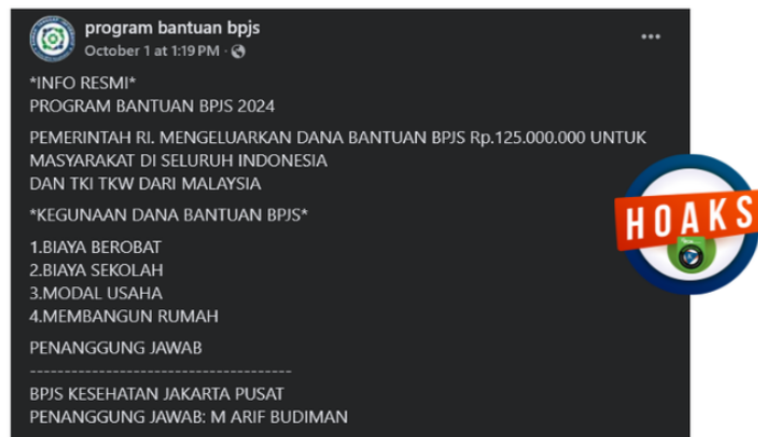
PENDAHULUAN

1.1 Latar Belakang Masalah

Perkembangan teknologi kecerdasan buatan (AI) telah membawa dampak signifikan di berbagai bidang, terutama pada bidang keamanan siber (*cybersecurity*) dalam penyebaran berita deephoaks. Di tengah kemajuan ini, muncul pula tantangan baru yang mengancam integritas informasi dan keamanan data digital. Salah satu cabang utama dalam AI yang sangat relevan dengan perkembangan ini adalah Natural Language Processing (NLP). NLP merupakan subbidang dari AI yang memanfaatkan machine learning (ML) untuk memungkinkan komputer berkomunikasi menggunakan bahasa manusia [1].

Salah satu inovasi penting dalam NLP adalah Generative AI (GenAI), yakni teknologi AI yang mampu menghasilkan data baru berdasarkan data yang telah ada sebelumnya [2]. GenAI dapat menciptakan berbagai jenis konten, termasuk teks, gambar, audio, dan video. Seiring dengan perkembangan teknologi, produk berbasis GenAI semakin banyak bermunculan dan dapat dengan mudah diakses melalui internet. Beberapa contoh dari teknologi ini antara lain ChatGPT untuk pembuatan teks, DALL-E untuk generasi gambar, serta AIVA untuk pembuatan musik [3, 4]. Pada awalnya, GenAI dikembangkan sebagai alat bantu bagi pengguna dalam mengembangkan ide-ide kreatif. Namun, kemampuannya juga berpotensi untuk disalahgunakan, terutama dalam konteks serangan siber dan penyebaran misinformasi. Salah satu bentuk penyalahgunaan tersebut adalah penciptaan dan penyebaran konten hoaks yang disintesis oleh AI.

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A



Gambar 1.1. Contoh Berita Hoaks

Hoaks adalah informasi palsu yang sengaja dibuat dan disebar untuk menyesatkan, memanipulasi, atau bahkan merugikan individu atau kelompok tertentu [5], Gambar 1.1 menunjukkan contoh berita hoaks yang dapat ditemukan di dunia maya. Dalam konteks GenAI, muncul fenomena teks hoaks yang dihasilkan oleh AI. Istilah "deepfake text" seringkali digunakan untuk merujuk pada teks yang diproduksi secara sintetis dengan tingkat realisme tinggi oleh model GenAI seperti GPT, Llama, dan Gemini [4, 6]. Model-model ini mampu menghasilkan teks secara otomatis dengan kualitas yang sangat mirip dengan tulisan manusia, sehingga semakin sulit dibedakan. Akibatnya, pembuatan dan penyebaran teks hoaks di internet menjadi jauh lebih mudah dilakukan. Fenomena ini menimbulkan ancaman serius bagi keamanan siber karena dapat digunakan untuk tujuan jahat seperti phishing, rekayasa sosial, atau penyebaran propaganda.

Menurut Nevada Today (2023), teknologi deepfake dapat dimanfaatkan untuk menyebarkan informasi palsu yang berujung pada hoaks, menciptakan kebingungan dalam isu-isu penting, serta digunakan sebagai alat untuk melecehkan, mengintimidasi, atau merendahkan seseorang secara daring [7]. Hal ini menunjukkan bagaimana deepfake menjadi ancaman yang serius pada bidang keamanan siber karena kemampuannya dalam memanipulasi informasi. Sejalan dengan hal tersebut, laporan dari Antara News (2024) mengungkapkan bahwa kasus penipuan di Indonesia akibat penggunaan 'deepfake' meningkat hingga 1550 persen pada periode 2022–2023, yang semakin memperkuat urgensi deteksi terhadap konten manipulatif ini[8].

Seiring dengan perkembangan AI, perbedaan antara teks yang dihasilkan

oleh AI dan teks yang ditulis oleh manusia semakin sulit untuk dideteksi. Hal ini disebabkan oleh kemampuan penulisan AI yang dapat terus ditingkatkan menggunakan data terbaru yang tersedia di internet (Kompas.com, 2025) [9]. Tantangan ini semakin diperburuk dengan luasnya jalur penyebaran berita hoaks melalui media sosial seperti WhatsApp, Instagram, dan Facebook, yang memungkinkan konten palsu menyebar dengan cepat dan menjangkau audiens yang lebih luas. Salah satu topik utama yang sering dimanfaatkan dalam penyebaran hoaks adalah isu sensitif, terutama yang berkaitan dengan agama dan politik. Selain itu, rendahnya tingkat literasi digital masyarakat juga menjadi faktor pendukung penyebaran hoaks, di mana banyak individu masih memiliki keterbatasan dalam memverifikasi informasi secara kritis [10]. Kondisi ini menggarisbawahi kebutuhan mendesak akan mekanisme deteksi yang efektif untuk menjaga keamanan informasi dan siber.

Untuk mengatasi masalah penyebaran teks hoaks ini, diperlukan model yang mampu melakukan deteksi terhadap konten yang dihasilkan oleh GenAI. Dalam penelitian ini, model yang akan digunakan adalah model berbasis Bidirectional Encoder Representations from Transformers (BERT). BERT merupakan model bahasa berbasis deep learning yang dapat memahami dan menganalisis teks beserta konteks dalam suatu kalimat secara dua arah (bidirectional). Model BERT juga dapat dilatih untuk melakukan klasifikasi teks, analisis sentimen, atau menjawab pertanyaan [11].

Melihat dampak signifikan dari penyalahgunaan generative AI dalam berbagai bidang, terutama dalam keamanan siber dan integritas informasi, pengembangan model deteksi teks hoaks berbasis IndoBERT untuk Bahasa Indonesia menjadi topik penelitian yang sangat relevan. Penelitian ini diharapkan dapat memberikan kontribusi dalam mengurangi penyebaran dan penyalahgunaan teks hoaks, serta meningkatkan keamanan dalam penggunaan teknologi AI dan literasi digital masyarakat.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah di atas, rumusan masalah dalam penelitian ini adalah:

1. Bagaimana cara membangun model IndoBERT untuk mendeteksi berita hoaks berbahasa Indonesia?

2. Teknik apa yang digunakan dalam proses mengumpulkan dataset berita?
3. Bagaimana hasil evaluasi performa model deteksi berita hoaks berbasis IndoBERT yang dikembangkan, khususnya dalam konteks upaya keamanan siber?

1.3 Batasan Permasalahan

Pada bagian ini dijabarkan batasan yang diterapkan dalam penelitian ini agar pelaksanaan penelitian menjadi lebih terfokus kepada aspek-aspek berikut:

1. Penelitian ini hanya berfokus pada deteksi berita hoaks dan berita fakta berbasis teks.
2. Model yang dikembangkan terbatas dalam berita berbahasa Indonesia saja.
3. Dataset yang digunakan terbatas pada dataset kaggle (berita fakta) dan dataset hasil scraping turnbackhoax (berita hoaks)

1.4 Tujuan Penelitian

1. Mengimplementasikan metode *fine-tuning* dalam membangun model IndoBERT untuk mendeteksi berita hoaks berbahasa Indonesia.
2. Teknik dalam mengumpulkan dataset terbagi menjadi web-scraping dan menggunakan dataset yang sudah ada.
3. Mengevaluasi matriks performa model yang dibuat dalam membedakan berita hoaks dengan berita fakta.

1.5 Manfaat Penelitian

Memberikan solusi teknologi yang dapat diintegrasikan dalam platform media untuk mendeteksi dan mengurangi dampak disinformasi.

1. Menghasilkan teknik kategorisasi berita deephoaks.
2. Menghasilkan dataset berita deephoaks.
3. Menghasilkan akurasi model berita deephoaks.

1.6 Sistematika Penulisan

Berisikan uraian singkat mengenai struktur isi penulisan laporan penelitian, dimulai dari Pendahuluan hingga Simpulan dan Saran.

Sistematika penulisan laporan adalah sebagai berikut:

- **Bab 1 PENDAHULUAN**
Menjelaskan tentang masalah dan potensi ancaman yang ditimbulkan oleh perkembangan GenAI. Dampak yang dibawa beserta ancaman dalam bidang keamanan siber seperti penyebaran hoaks.
- **Bab 2 LANDASAN TEORI**
Menjelaskan teori keamanan siber, ancaman dan serangan keamanan siber, peran ai dalam keamanan siber, konsep dasar deephoaks, model BERT, dan penelitian terkait sebelumnya.
- **Bab 3 METODOLOGI PENELITIAN**
Menguraikan teknik yang digunakan dalam membangun model, proses pengumpulan data (dataset online dari kaggle dan teknik webscraping), proses pembersihan data, proses pelatihan model dengan teknik fine-tuning, dan hasil evaluasi dari model dengan confusion matrix.
- **Bab 4 HASIL DAN DISKUSI**
Pembahasan dan analisa dari model yang menunjukkan hasil akurasi model sebesar 95%, F1 sebesar 95%, precision sebesar 95%, recall sebesar 95%, dan ROC-AUC sebesar 99%.
- **Bab 5 KESIMPULAN DAN SARAN**
Mendapatkan kesimpulan bahwa model memiliki potensi besar sebagai alat pertahanan dalam upaya keamanan siber dengan tingkat akurasi sebesar 95%. Hasil dari penelitian juga menghasilkan saran untuk menggunakan sumber dataset yang lebih bervariasi untuk berita hoaks agar dapat menghindarkan terjadinya bias serta pengembangan multimodel yang mencakup gambar dalam upaya pendeteksian berita hoaks dan keamanan siber.