

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

AI atau artificial Intelligence adalah bidang multidisiplin yang dapat membuat keputusan tanpa dipengaruhi oleh nilai-nilai pribadi manusia [1]. Saat ini pun, AI sudah memiliki kemampuan untuk generate sebuah text. Munculnya AI ini tentu saja memiliki dampak baik dan dampak buruknya. Dengan adanya AI ini, tentunya pekerjaan seperti merangkum dan pembuatan text tentu akan lebih cepat dilakukan. Tetapi dengan adanya AI yang dapat men generate text, akan muncul beberapa permasalahan seperti potensi perannya sebagai sarana disinformasi [2] . Dari penelitian yang sudah dilakukan, AI yang bisa generate text memiliki beberapa bahaya seperti, potensial bias pada pembuatan konten, dapat membawa resiko terhadap privasi pengguna, dan juga konten yang dibuat dapat merupakan informasi yang salah sehingga dapat merugikan [3].

Saat ini pun sudah ada beberapa penelitian mengenai klasifikasi teks buatan AI dan buatan manusia. Iyab Katib dkk [4] melakukan sebuah studi menggunakan berbagai algoritma, seperti Decision Tree, SVM, XGBoost, CNN, ELM, dan TSA-LSTMRNN. Hasil terbaik diperoleh menggunakan TSA-LSTMRNN, dengan akurasi 93,17% untuk dataset manusia dan 93,83% untuk dataset ChatGPT.

Penelitian kedua dilakukan oleh Kumar dan tim [5] dengan judul “Deep Dive into Language Traits of AI-Generated Abstracts.” Dalam studi ini, digunakan lima metode, yaitu Analisis Diskriminan Linear (LDA), Regresi Logistik (LR), Klasifikasi Support Vector (SVC), Peningkatan Gradien Ekstrem (XGBoost), dan Pengklasifikasi Pohon Ekstra (ETC). Metode Peningkatan Gradien Ekstrem (XGBoost) menghasilkan akurasi tertinggi sebesar 93,4% dengan memanfaatkan fitur semantik (SF), fitur linguistik (LF), dan fitur pragmatik (HBH).

Dengan pemaparan diatas, peneliti ingin melakukan penelitian mengenai deteksi teks buatan Chat-GPT menggunakan model BERT. BERT (Bidirectional Encoder Representation from Transformers) adalah salah satu dari metode NLP (Natural Language Processing) yang dilatih dengan langkah pre-training dan fine tuning. Arsitektur dari BERT model sendiri

terdiri dari multi-layer bidirectional Transformer. Transformer didasarkan konsep self-attention yang dapat membuat model untuk memahami konteks, bukan sekedar urutan kata secara keseluruhan [7]. BERT sendiri menggunakan encoder bidirectional yang berarti BERT membaca seluruh inputan bukan hanya dari satu arah, tetapi dua arah secara bersamaan. Dengan kemampuan membaca secara dua arah, BERT dapat memahami teks dengan lebih akurat [8].

Penelitian ini akan diawali dengan pengumpulan data, selanjutnya data yang telah dikumpulkan akan melalui tahap preprocessing, meliputi pembersihan data (data cleaning) dan tokenisasi menggunakan BERT. Data dibagi menjadi data latih dan data uji, lalu model dilatih menggunakan data latih dan dievaluasi kinerjanya dengan data uji. Dengan penelitian ini, diharapkan dapat dikembangkan model deteksi teks buatan ChatGPT yang lebih akurat dan efisien dengan memanfaatkan BERT sebagai model. Kemampuan BERT dalam memahami konteks serta hubungan antar kata diharapkan mampu memberikan kinerja yang lebih unggul dibandingkan metode NLP lainnya.

### 1.2. Rumusan masalah

1. Seberapa tinggi tingkat *accuracy* dari mendeteksi Teks Buatan ChatGPT yang dibuat dengan model BERT (Bidirectional Encoder Representation from Transformers)?
2. Bagaimana pengaruh jumlah kata terhadap akurasi dalam mendeteksi Teks Buatan ChatGPT?
3. Bagaimana akurasi dari model *pre-trained* BERT dalam mendeteksi teks buatan ChatGPT ketika diterapkan dengan optimisasi menggunakan *Stochastic Gradient Descent* (SGD) dan *AdamW*?

### 1.3. Tujuan

Penelitian ini bertujuan untuk merinci sasaran yang dicapai secara jelas sebagai berikut.

1. Mengetahui seberapa tinggi tingkat *accuracy* dalam mendeteksi Teks Buatan ChatGPT yang dibuat dengan model BERT (Bidirectional Encoder Representation from Transformers)?
2. Mengetahui pengaruh jumlah kata terhadap akurasi dalam mendeteksi Teks Buatan ChatGPT

3. Mengetahui akurasi dari model *pre-trained* BERT dalam mendeteksi teks buatan ChatGPT ketika diterapkan dengan optimisasi menggunakan *Stochastic Gradient Descent* (SGD) dan *AdamW*

#### 1.4. Batasan Masalah

1. Bahasa yang digunakan dalam penelitian ini adalah python
2. Model yang digunakan adalah BERT (Bidirectional Encoder Representation from Transformers)
3. Dataset yang digunakan adalah gabungan dari dataset “ChatGPT Classification Dataset” dan “GPT vs. Human: A Corpus of Research Abstracts”
4. Dataset terdiri dari dua :
  - a. Dataset dengan kata kurang dari 200
  - b. Dataset dengan kata lebih dari 200
5. Masing-masing dataset memiliki berjumlah dari 452 yang terdiri dari :
  - a. Data buatan chat-gpt berjumlah 226
  - b. Data buatan manusia berjumlah 226