

BAB I

PENDAHULUAN

A. Latar Belakang

Perkembangan *Artificial Intelligence* (AI) telah membawa transformasi signifikan dalam berbagai sektor, termasuk dalam bidang pendidikan digital. Pemanfaatan teknologi AI memungkinkan proses analisis data dan penyediaan informasi akademik dilakukan secara lebih cepat, efisien, dan otomatis. Integrasi AI dalam sistem pembelajaran berbasis teknologi informasi juga terbukti mampu meningkatkan efektivitas layanan pendidikan serta mempermudah akses terhadap sumber pengetahuan bagi mahasiswa dan akademisi (Firdaus dkk, 2024). Oleh karena itu, AI menjadi salah satu teknologi yang semakin penting dalam mendukung proses pembelajaran modern.

Salah satu perkembangan penting dalam teknologi AI adalah munculnya *Large Language Model* (LLM) yang mampu memahami, memproses, dan menghasilkan teks dengan kompleksitas linguistik yang tinggi. Kemampuan ini menjadikan LLM banyak dimanfaatkan dalam berbagai aktivitas akademik, seperti pencarian informasi, penyusunan materi pembelajaran, serta pemberian respons terhadap pertanyaan akademik. Dengan kemampuan tersebut, LLM berperan sebagai alat bantu yang dapat meningkatkan efisiensi dalam memperoleh informasi akademik. Hal ini menunjukkan bahwa LLM memiliki potensi besar dalam mendukung kegiatan pembelajaran di era digital (Firdaus dkk, 2024; Zhang dkk, 2026).

Di antara berbagai platform LLM yang tersedia, ChatGPT dan Gemini AI merupakan dua sistem yang memiliki tingkat adopsi tinggi di kalangan pengguna global. Tingginya penggunaan kedua platform ini menunjukkan bahwa AI telah menjadi salah satu sumber informasi digital yang penting bagi mahasiswa dan akademisi. Namun demikian, performa model AI sangat dipengaruhi oleh karakteristik tugas dan konteks pertanyaan yang diberikan. Akibatnya, kualitas respons yang dihasilkan dapat bervariasi tergantung pada tingkat kompleksitas pertanyaan. Oleh karena itu, diperlukan evaluasi empiris

untuk memastikan ketepatan informasi dan relevansi respons yang diberikan (Jonathan dkk, 2025).

Pemanfaatan ChatGPT dan Gemini AI dalam konteks akademik terus meningkat seiring dengan kebutuhan mahasiswa terhadap akses informasi yang cepat dan responsif. Kedua model ini mampu membantu menyederhanakan konsep yang kompleks serta menyajikan informasi secara lebih terstruktur. Meskipun demikian, kualitas respons yang dihasilkan tidak selalu konsisten karena dipengaruhi oleh formulasi pertanyaan dan karakteristik materi. Kondisi ini menimbulkan kebutuhan untuk melakukan evaluasi terhadap tingkat akurasi dan kesesuaian informasi yang dihasilkan dalam konteks akademik. Dengan demikian, pengukuran performa model AI menjadi hal yang penting untuk dilakukan (Hidayatullah, 2025; Jonathan dkk, 2025).

Penelitian ini secara khusus menggunakan GPT-5.3 dan Gemini 3 sebagai objek penelitian. Pemilihan kedua model ini didasarkan pada pertimbangan bahwa keduanya merupakan versi terbaru yang tersedia pada saat penelitian dilakukan, serta *Large Language Model* terus mengalami perkembangan dalam kemampuan pemrosesan bahasa alami dan kualitas respons yang dihasilkan (Firdaus dkk, 2024; Zhang dkk, 2026). Selain itu, performa model AI dipengaruhi oleh perbedaan karakteristik dan pendekatan pengembangan yang digunakan, sehingga diperlukan analisis komparatif untuk mengukur kualitas respons yang dihasilkan (Jonathan dkk, 2025). Dengan demikian, penggunaan kedua model ini diharapkan dapat memberikan gambaran yang lebih representatif mengenai performa *Large Language Model* dalam konteks akademik.

Untuk melakukan evaluasi secara objektif, diperlukan metode yang mampu mengukur tingkat kesesuaian antara respons model AI dengan referensi akademik. Salah satu pendekatan yang dapat digunakan adalah *Sentence-BERT* (SBERT) yang dikombinasikan dengan *Cosine Similarity* untuk mengukur kemiripan semantik antar teks. Pendekatan ini memungkinkan representasi makna kalimat dalam bentuk vektor sehingga

dapat dibandingkan secara matematis. Dengan demikian, tingkat kesamaan makna antara dua teks dapat diukur secara lebih akurat dan objektif (Gusdevi dkk, 2025; Reimers & Gurevych, 2019). Oleh karena itu, metode ini relevan digunakan dalam evaluasi jawaban akademik berbasis AI.

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk mengevaluasi dan membandingkan performa GPT-5.3 dan Gemini 3 dalam menghasilkan respons akademik. Evaluasi dilakukan menggunakan pendekatan *semantic similarity* untuk mengukur tingkat kesesuaian antara respons model AI dan jawaban rujukan akademik. Penelitian ini diharapkan dapat memberikan gambaran mengenai tingkat ketepatan dan konsistensi kedua model dalam konteks akademik. Selain itu, hasil penelitian juga diharapkan dapat menjadi referensi dalam pemanfaatan AI sebagai alat bantu pembelajaran yang lebih terukur dan objektif.

B. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana tingkat kedekatan semantik jawaban ChatGPT dan Gemini AI terhadap kunci jawaban akademik pada mata kuliah *Data Mining* menggunakan metode *Cosine Similarity*?
2. Bagaimana perbandingan performa ChatGPT dan Gemini AI dalam menjawab pertanyaan akademik mata kuliah *Data Mining* berdasarkan hasil evaluasi kemiripan semantik menggunakan metode *Cosine Similarity*?

C. Batasan Masalah

Agar pembahasan dalam penelitian ini lebih terfokus dan tidak meluas dari permasalahan yang telah dirumuskan, maka batasan masalahnya yaitu:

1. Penelitian ini hanya berfokus pada dua model *Large Language Model* (LLM), yaitu GPT-5.3 dan Gemini 3.
2. Jenis pertanyaan yang digunakan dalam penelitian ini terbatas pada istilah akademik mata kuliah *Data Mining*.
3. Evaluasi jawaban hanya dilakukan pada aspek kedekatan semantik

antara jawaban LLM dan jawaban kunci, yang diukur menggunakan metode *Cosine Similarity*.

4. Evaluasi aspek lain seperti *reasoning*, *factual consistency*, bias, atau performa komputasi tidak menjadi fokus penelitian.
5. Format pertanyaan dan jawaban yang digunakan adalah teks (*text-based*). Penelitian tidak mencakup input atau output dalam bentuk gambar, tabel, grafik, atau kode program yang tidak bisa dievaluasi secara semantik.
6. Setiap pertanyaan dijalankan sebanyak tiga kali pada masing-masing model LLM untuk menilai konsistensi dan stabilitas keluaran dalam kondisi terkontrol.

D. Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diuraikan sebelumnya, maka tujuan penelitian ini adalah:

1. Mengukur dan menganalisis tingkat kedekatan semantik antara jawaban yang dihasilkan oleh ChatGPT dan Gemini AI dengan kunci jawaban akademik pada mata kuliah *Data Mining* menggunakan metode *Cosine Similarity*.
2. Menganalisis dan membandingkan performa ChatGPT dan Gemini AI dalam menjawab pertanyaan akademik mata kuliah *Data Mining* berdasarkan nilai kemiripan semantik yang dihasilkan.

E. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat baik secara teoretis maupun praktis, sebagai berikut:

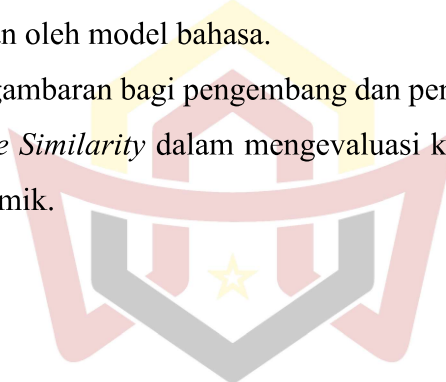
a. Manfaat Teoritis

1. Memberikan kontribusi terhadap pengembangan ilmu pengetahuan pada bidang *Large Language Model* (LLM) dalam evaluasi kualitas jawaban akademik yang dihasilkan oleh sistem kecerdasan buatan.
2. Menambah kajian akademik terkait evaluasi performa *Large Language Model* (LLM) seperti ChatGPT dan Gemini AI menggunakan pendekatan kemiripan semantik.
3. Menjadi dasar teoretis dan referensi bagi penelitian selanjutnya yang

membahas penerapan metode *Cosine Similarity* dalam penilaian kualitas jawaban AI pada konteks pendidikan tinggi.

b. Manfaat Praktis

1. Memberikan informasi bagi pendidik dan mahasiswa mengenai tingkat keandalan ChatGPT dan Gemini AI dalam menjawab pertanyaan akademik pada mata kuliah *Data Mining*.
2. Menjadi bahan pertimbangan bagi institusi pendidikan dalam merumuskan kebijakan pemanfaatan *Large Language Model* (LLM) sebagai alat bantu pembelajaran secara tepat, terukur, dan bertanggung jawab.
3. Menjadi dasar pertimbangan etika akademik dalam penggunaan AI di lingkungan perkuliahan, khususnya terkait batasan dan validitas jawaban yang dihasilkan oleh model bahasa.
4. Memberikan gambaran bagi pengembang dan peneliti mengenai performa metode *Cosine Similarity* dalam mengevaluasi kualitas jawaban AI pada konteks akademik.



UIN IMAM BONJOL
PADANG