

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pengembangan layanan informasi akademik yang cepat dan efisien merupakan salah satu fokus utama dalam modernisasi institusi pendidikan tinggi. Sesuai dengan Undang-Undang Nomor 12 Tahun 2012 tentang Pendidikan Tinggi, setiap perguruan tinggi berkewajiban menyelenggarakan pendidikan dan memberikan layanan informasi yang relevan, akuntabel, dan mudah diakses oleh *civitas academica*. Transformasi digital di lingkungan kampus juga sejalan dengan Peraturan Presiden Nomor 95 Tahun 2018 tentang Sistem Pemerintahan Berbasis Elektronik (SPBE) yang mendorong pemanfaatan teknologi informasi terpadu untuk meningkatkan kualitas layanan publik, termasuk sektor pendidikan.

Perkembangan kecerdasan buatan telah melahirkan berbagai inovasi dalam bidang pemrosesan bahasa, salah satunya adalah *Natural Language Processing* (NLP), yaitu teknologi yang memungkinkan komputer untuk memahami dan memproses bahasa manusia secara alami. Dari perkembangan NLP tersebut, lahirlah *Large Language Model* (LLM) sebagai model kecerdasan buatan yang menggunakan pembelajaran mesin untuk menghasilkan teks yang menyerupai bahasa manusia (Lubis et al., 2024). Untuk meningkatkan kualitas jawaban, digunakan *Retrieval-Augmented Generation* (RAG), yaitu teknik yang menggabungkan *pretrained language model* dengan mekanisme *retrieval*, sehingga sistem dapat mengambil informasi yang relevan dari basis data eksternal melalui pencarian *keyword* dan semantik sebelum menghasilkan jawaban yang lebih akurat dan sesuai dengan konteks (Elysia & Herianto, 2024).

Chatbot merupakan aplikasi berbasis kecerdasan buatan yang dibuat untuk memungkinkan interaksi percakapan antara sistem dan manusia secara daring, baik melalui teks, maupun pilihan respons tertentu. Selain meniru percakapan manusia, *chatbot* juga dapat meningkatkan akurasi respons seiring bertambahnya data yang dipelajari (Wintoro et al., 2022). Pada penelitian ini, struktur data yang dikirimkan ke LLM direpresentasikan menggunakan *Token-Oriented Object Notation* (TOON),

yaitu format representasi data terstruktur yang dirancang khusus untuk meminimalkan penggunaan token pada input LLM dibandingkan dengan JSON. Pemanfaatan teknologi *chatbot* di lingkungan perguruan tinggi sebagai media layanan informasi akademik masih sangat terbatas. Sistem informasi akademik yang berjalan secara konvensional mengharuskan mahasiswa datang langsung ke kampus atau menghubungi pihak universitas secara manual untuk mendapatkan informasi, sehingga proses tersebut dinilai tidak efisien (Prasojo et al., 2024). Berbagai kendala dalam layanan informasi konvensional seperti keterbatasan waktu layanan, waktu tunggu yang lama, dan ketidaktepatan informasi sebenarnya dapat diatasi melalui penerapan teknologi *chatbot* (Herman et al., 2023). Kondisi tersebut juga terlihat pada Program Studi Magister Teknologi Informasi Universitas Malikussaleh. Berdasarkan observasi awal yang dilakukan oleh penulis, informasi akademik yang tersedia masih tersebar di berbagai halaman dan *website* yang digunakan masih bersifat statis. Selain itu belum tersedia fitur tanya jawab interaktif yang dapat membantu mahasiswa memperoleh informasi akademik secara langsung sesuai dengan kebutuhan. Akibatnya, mahasiswa masih harus mencari informasi secara manual.

Implementasi *Large Language Model* (LLM) tidak hanya menghadapi tantangan akurasi, tetapi juga efisiensi token. *Prompt* dan konteks yang panjang meningkatkan biaya pemrosesan karena penggunaan model dihitung berdasarkan jumlah token input dan output. OpenAI bahkan menyebutkan bahwa *prompt caching* dapat menurunkan latensi hingga 80% dan biaya input token hingga 90%, yang menunjukkan bahwa pemborosan token merupakan salah satu permasalahan dalam pemanfaatan LLM.

Berdasarkan permasalahan tersebut, pengembangan *chatbot* akademik berbasis *Large Language Model* (LLM) dapat menjadi solusi untuk meningkatkan layanan informasi akademik yang lebih cepat, efisien, dan interaktif. Dalam meningkatkan ketepatan jawaban, sistem diusulkan menggunakan *Retrieval-Augmented Generation* (RAG), sedangkan untuk menangani data akademik yang bersifat heterogen digunakan mekanisme *hybrid retrieval*. Selain itu, untuk mengurangi pemborosan token pada penggunaan LLM, penelitian ini juga

mengusulkan penggunaan *Token-Oriented Object Notation* (TOON) sebagai format representasi data yang lebih ringkas dan efisien.

Pujiono dkk. (2024) mengimplementasikan *Retrieval-Augmented Generation* (RAG) dan *vector database* pada *chatbot* di konteks lembaga layanan publik. Penelitian tersebut menunjukkan bahwa penggunaan LLM yang dipadukan dengan RAG melalui ekstraksi dokumen berbasis regulasi dapat memberikan respons yang baik terhadap pertanyaan yang diajukan, dengan model terbaik yaitu GPT-4 yang memperoleh rata-rata *cosine similarity score* sebesar 0,404.

Penelitian lain yang mengembangkan *chatbot* layanan universitas dilakukan oleh Eirena & Shah (2025) melalui pengembangan *chatbot* berbasis *Hybrid Retrieval-Augmented Generation* (RAG) untuk layanan pelanggan universitas dengan memadukan local LLM dan *semantic retrieval*. Penelitian tersebut menunjukkan bahwa sistem yang dikembangkan mampu memberikan respons yang akurat, cepat, dan menjaga privasi data dengan nilai *BLEU* sebesar 0,75, *ROUGE-1 F-Measure* sebesar 0,75, *cosine similarity* sebesar 0,85, serta rata-rata *response time* sebesar 150 ms.

Berdasarkan latar belakang masalah tersebut, maka penulis melakukan penelitian dengan judul “**Implementasi Metode *Hybrid Retrieval* Dan *Token-Oriented Object Notation* Pada *Chatbot* Akademik Berbasis *Large Language Model***”. Penelitian ini diharapkan dapat menghasilkan sistem layanan informasi akademik yang lebih cerdas, efisien dan responsif, serta mampu menjawab kebutuhan informasi *civitas academica* secara lebih cepat, tepat, dan kontekstual. Selain itu, penerapan TOON juga diharapkan mampu meningkatkan efisiensi penggunaan token dalam proses pertukaran data antara sistem dan *Large Language Model*, sehingga mampu menekan biaya token dan kinerja *chatbot* menjadi lebih optimal.

1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, maka rumusan masalah dalam penelitian ini adalah:

1. Bagaimana cara membangun sistem informasi akademik yang mampu memberikan jawaban secara otomatis, cepat, dan kontekstual untuk mengakses data terstruktur dan tidak terstruktur pada Program Studi Magister Teknologi Informasi Universitas Malikussaleh?
2. Bagaimana kinerja *chatbot* akademik berbasis LLM dalam memberikan jawaban yang cepat, tepat, dan kontekstual pada Program Studi Magister Teknologi Informasi Universitas Malikussaleh?
3. Bagaimana tingkat efisiensi penggunaan token yang dihasilkan dari format JSON ke format TOON pada proses pertukaran data antara sistem dan LLM?

1.3 Tujuan Penelitian

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

1. Dengan mengimplementasikan metode *hybrid retrieval* pada *chatbot* akademik berbasis *Large Language Model* dengan menggabungkan data terstruktur dan tidak terstruktur, sehingga sistem mampu memberikan jawaban yang otomatis, cepat, dan kontekstual di Program Studi Magister Teknologi Informasi Universitas Malikussaleh.
2. Dengan mengimplementasikan arsitektur *Retrieval Augmented Generation*, metode *Hybrid Retrieval*, dan *Token-Oriented Object Notation* untuk menghasilkan jawaban yang cepat, tepat, dan kontekstual pada Program Studi Magister Teknologi Informasi Universitas Malikussaleh.
3. Dengan mengukur tingkat penghematan token dari format JSON ke format TOON pada proses pertukaran data antara sistem dan LLM pada dataset akademik.

1.4 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah sebagai berikut:

1. Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan kajian keilmuan di bidang kecerdasan buatan, khususnya terkait dengan penerapan LLM, *hybrid retrieval*, dan TOON pada *chatbot* akademik.

2. Bagi Program Studi, penelitian ini diharapkan dapat menjadi solusi alternatif dalam meningkatkan layanan informasi akademik yang jauh lebih cepat, efisien dan interaktif.
3. Bagi mahasiswa, penelitian ini diharapkan dapat memudahkan dalam memperoleh informasi akademik secara lebih cepat, tepat, dan sesuai kebutuhan tanpa harus mencari informasi secara manual.
4. Bagi peneliti selanjutnya, penelitian ini diharapkan dapat menjadi bahan acuan dan perbandingan dalam pengembangan *chatbot* akademik berbasis LLM dengan *hybrid retrieval* dan efisiensi token menggunakan TOON

1.5 Ruang Lingkup dan Batasan Penelitian

Adapun batasan masalah pada penelitian ini adalah sebagai berikut:

1. Penelitian ini difokuskan pada pengembangan *chatbot* akademik untuk layanan informasi di Program Studi Magister Teknologi Informasi Universitas Malikussaleh.
2. Sistem yang dikembangkan menggunakan LLM sebagai komponen utama dalam menghasilkan jawaban.
3. *Hybrid retrieval* pada penelitian ini dibatasi pada penggabungan data terstruktur dari basis data relasional dan data tidak terstruktur dari dokumen akademik.
4. Topik percakapan *chatbot* dibatasi seputar informasi layanan akademik Program Studi Magister Teknologi Informasi Universitas Malikussaleh. Apabila pengguna mengajukan pertanyaan di luar topik tersebut atau jawabannya tidak tersedia dalam basis data dan dokumen, sistem akan memberikan respons penolakan secara sopan.
5. Efisiensi token dibatasi pada pengukuran tingkat penghematan token dari format JSON ke format TOON pada proses pertukaran data dengan LLM.
6. Evaluasi sistem dibatasi pada kinerja *chatbot* dalam memberikan jawaban yang tepat dan pada efisiensi penggunaan toke