

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil perancangan, implementasi, dan evaluasi sistem *chatbot* informasi akademik FPMIPA UPI, dapat ditarik beberapa kesimpulan, diantaranya.

1. Penelitian ini telah berhasil membangun sebuah arsitektur *Retrieval-Augmented Generation* (RAG) yang lengkap dari awal hingga akhir. Proses ini dimulai dari pengolahan dokumen sumber menjadi basis pengetahuan terstruktur, di mana setiap potongan informasi (*chunk*) diperkaya dengan metadata serta hubungan kontekstual yang jelas, baik secara hierarkis (*Parent-Child*) maupun sekuensial. Di atas fondasi ini, dijalankan sebuah pipeline *retrieval multi-layer* yang secara sistematis melakukan dekomposisi kueri, *query rerouting*, perluasan konteks berdasarkan struktur data, dan *reranking* untuk mendapatkan informasi yang paling baik. Informasi tersebut kemudian diproses oleh LLM Gemini 2.5 Pro, yang kinerjanya dipandu oleh *prompt* terstruktur untuk memastikan jawaban tetap faktual dan aman, dan pada akhirnya diwujudkan dalam sebuah *prototipe chatbot* interaktif yang dilengkapi memori percakapan.
2. Pada evaluasi pertanyaan faktual, *chatbot* menunjukkan kinerja yang cukup baik, ditandai dengan kemampuan tinggi untuk menghasilkan jawaban yang konsisten dengan sumber (*Faithfulness* 0.990) dan relevan dengan pertanyaan (*Answer Relevancy* 0.967). Namun, akurasi faktualnya (*F1 Score* 0.757) masih berada di tingkat sedang. Kelemahan utamanya terletak pada presisi jawaban (*Precision* 0.714) yang lebih rendah dibandingkan kemampuannya menangkap informasi (*Recall* 0.850). Hal ini berarti sistem cenderung menemukan fakta-fakta kunci, tetapi sering

kali menyertakannya dengan detail tambahan yang tidak diminta oleh pengguna, sehingga membuat jawaban kurang ringkas dan tepat sasaran.

3. Untuk evaluasi pada pertanyaan penalaran, *chatbot* menunjukkan kemampuan yang baik dalam menjaga kesesuaian terhadap konteks (*Faithfulness* 1.000), yang berarti sistem tidak berhalusinasi atau mengarang informasi saat melakukan penalaran. Jawaban yang diberikan juga sangat relevan dengan maksud pertanyaan. Akan tetapi, sama seperti pada pertanyaan faktual, tantangan utamanya adalah kelengkapan dan presisi informasi. Dengan *F1 Score* 0.702, sistem masih kesulitan menyajikan jawaban yang sepenuhnya lengkap, terutama ketika informasi yang dibutuhkan berasal dari format kompleks seperti tabel. Meskipun mampu melakukan penalaran logis seperti membandingkan dua data, hasil akhirnya sangat bergantung pada keberhasilan tahap *retrieval* dalam menemukan semua konteks yang diperlukan.
4. Kinerja *chatbot* dalam menangani pertanyaan di luar konteks adalah sangat baik dan merupakan kekuatan terbesarnya. Dengan skor sempurna (1.000) di semua metrik evaluasi, sistem secara konsisten dan akurat menolak untuk menjawab pertanyaan yang informasinya tidak ada dalam basis datanya. Kemampuan ini menunjukkan adanya mekanisme pengaman yang efektif untuk mencegah halusinasi, sehingga menjadikan *chatbot* sebagai sistem yang lebih aman dan dapat dipercaya.
5. Dari perspektif pengguna, *chatbot* dinilai berkinerja baik dengan tingkat akurasi 83,3% dalam menjawab 30 pertanyaan skenario. UAT mengonfirmasi temuan dari evaluasi kuantitatif, di mana sebagian besar kesalahan (16,7%) yang dirasakan oleh pengguna berupa jawaban yang kurang tepat karena informasi yang disajikan tidak lengkap (parsial) atau konteks yang diambil kurang relevan. Hal ini menegaskan bahwa meskipun secara umum jawaban *chatbot* sudah memuaskan, pengalaman pengguna masih terpengaruh oleh masalah kelengkapan informasi.
6. Secara keseluruhan, sistem yang dibangun sudah mampu memberikan jawaban yang relevan dan kontekstual untuk pertanyaan berbasis informasi

akademik, dengan performa yang relatif stabil pada berbagai jenis pertanyaan. Namun, optimasi *strategi retrieval*, penyempurnaan *chunking*, serta peningkatan filter data pengetahuan tetap diperlukan untuk mencapai hasil yang lebih optimal.

5.2 Saran

Penelitian ini masih memiliki sejumlah keterbatasan sehingga belum sepenuhnya sempurna. Oleh karena itu, penulis menyampaikan beberapa saran yang diharapkan dapat menjadi acuan perbaikan dan pengembangan pada penelitian sejenis di masa mendatang. Saran-saran tersebut dirangkum sebagai berikut.

1. Peningkatan Lanjutan pada Proses *Parsing* dan Ekstraksi Tabel: Seperti yang teridentifikasi dalam evaluasi, kegagalan dalam mengekstrak informasi dari tabel merupakan salah satu kendala utama yang menurunkan akurasi faktual *chatbot*. Meskipun telah digunakan *library* modern seperti *docling*, tabel dengan format non-standar atau kompleks masih menjadi tantangan. Oleh karena itu, disarankan agar penelitian selanjutnya fokus pada pengembangan *pipeline* ekstraksi yang lebih baik lagi.
2. Perbaikan Strategi *Chunking*: Strategi *chunking* yang saat ini digunakan masih berpotensi memotong konteks penting atau menggabungkan informasi yang tidak relevan dalam satu *chunk*. Saran perbaikannya adalah beralih dari *hybrid chunking* ke strategi *structural chunking*. Metode ini memecah dokumen berdasarkan struktur logisnya, seperti judul, sub-judul, paragraf, atau daftar. Dengan demikian, setiap *chunk* akan berisi satu unit informasi yang koheren dan lengkap, meminimalkan risiko jawaban yang parsial dan meningkatkan kualitas konteks yang diterima oleh model bahasa.
3. Penambahan Konteks dengan Metadata dan Implementasi Filter Lanjutan: Untuk meningkatkan relevansi hasil pencarian, disarankan untuk memperkaya setiap *chunk* data dengan metadata yang terstruktur. Metadata

ini dapat mencakup kategori seperti tipe_informasi (misalnya, "akademik", "kemahasiswaan", "fasilitas"), peran_target ("mahasiswa", "dosen", "umum"), atau topik_pembahasan ("beasiswa", "magang", "skripsi"). Dengan adanya metadata ini, sistem *retrieval* dapat dimodifikasi untuk melakukan filter awal berdasarkan kategori tersebut sebelum menjalankan pencarian semantik. Mekanisme ini akan memastikan bahwa pencarian hanya dilakukan pada subset data yang paling relevan, secara signifikan mengurangi kemungkinan konteks yang tertukar atau tidak sesuai.

4. Optimalisasi Sistem Pengambilan Informasi (*Retrieval*) dengan *Hybrid Search*: Sistem *retrieval* yang hanya mengandalkan pencarian semantik (vektor) terkadang kesulitan menemukan istilah spesifik seperti nama, kode mata kuliah, atau singkatan. Untuk mengatasi hal ini, disarankan untuk mengimplementasikan arsitektur *Hybrid Search*, yang menggabungkan pencarian semantik dengan pencarian berbasis kata kunci (*keyword search*) seperti BM25.
5. Implementasi Memori Persisten untuk Riwayat Percakapan: Mengganti mekanisme *in-memory store* dengan solusi yang lebih permanen ke basis data lainnya, untuk memastikan riwayat percakapan pengguna tidak hilang dan dapat dianalisis untuk pengembangan lebih lanjut.
6. Integrasi ke Platform Resmi: Mengintegrasikan chatbot ke dalam situs web resmi FPMIPA UPI atau portal mahasiswa agar dapat diakses dan dimanfaatkan secara luas.