

BAB IV

SIMPULAN DAN SARAN

4.1 Simpulan

Selama mengerjakan proyek laporan Project Independen MBKM dengan judul "Pengembangan Chatbot Berbasis Large Language Model dengan Metode Retrieval-Augmented Generation untuk Layanan Informasi di Universitas Multimedia Nusantara", penulis mengalami pertumbuhan baik secara teknis maupun personal. Bagian kesimpulan dan saran memegang peranan penting dalam sebuah laporan atau karya tulis. Tujuan utamanya adalah untuk merangkum inti dari temuan penelitian dan memberikan rekomendasi yang konstruktif untuk pengembangan di masa depan.

Berdasarkan tahapan yang telah dilakukan, chatbot ini dibangun dengan menggunakan pendekatan RAG di mana informasi tidak hanya diambil dari pengetahuan model, tetapi juga dari dokumen aktual seperti *handbook*, situs kampus, dan regulasi akademik lainnya. Pada laporan yang telah dikerjakan, proyek ini menjawab rumusan masalah sebagai berikut:

1. Bagaimana cara mengembangkan model LLM yang dapat memahami informasi spesifik terkait UMN menggunakan metode RAG?

Pengembangan model LLM dilakukan dengan mengintegrasikan metode RAG agar dapat menangani pertanyaan yang berkaitan langsung dalam konteks Universitas Multimedia Nusantara (UMN). Model ini dibangun dengan mengumpulkan data dari sumber internal seperti *handbook MyUMN*, situs halaman UMN, dan dokumen akademik lainnya yang kemudian diproses ke dalam bentuk vektor menggunakan embedding dan disimpan dalam basis data vektor (ChromaDB). Melalui proses retrieval yang terstruktur, sistem mampu mengekstraksi informasi kontekstual dan menghasilkan jawaban menggunakan model LLM (Llama 8B) sehingga menjawab rumusan masalah pertama.

2. Bagaimana cara meningkatkan pemahaman model terhadap berbagai layanan, kebijakan, dan informasi akademik yang tersedia di UMN?

Dalam menjawab kebutuhan pada rumusan masalah kedua, sistem dirancang agar mampu mengenali berbagai variasi bahasa alami dari mahasiswa serta memahami berbagai kategori layanan akademik dan administratif. Hal ini dicapai melalui desain dataset yang mencakup topik-topik seperti KRS, registrasi, beasiswa, dan prosedur administrasi. Evaluasi sistem menunjukkan bahwa penggunaan metode RAG berhasil meningkatkan relevansi dan kedalaman jawaban karena model tidak hanya menghasilkan respons dari parameter internal, tetapi juga dari dokumen kontekstual yang terkait.

3. Bagaimana chatbot dapat diintegrasikan dengan Web UI agar dapat digunakan oleh mahasiswa dan staf secara praktis?

Dalam menjawab rumusan masalah ketiga, sistem chatbot dikembangkan dengan antarmuka berbasis Streamlit sebagai solusi Web UI yang dapat diakses oleh mahasiswa dan staf secara praktis. Antarmuka ini dirancang dengan tampilan minimalis dan fungsi yang responsif, sehingga pengguna dapat langsung mengajukan pertanyaan tanpa perlu instalasi tambahan. Proses integrasi dilakukan tanpa menggunakan WhatsApp API, agar akses tetap terbuka dan mudah untuk diuji lintas platform.

Secara keseluruhan, proyek ini berhasil menunjukkan bahwa penerapan LLM dengan metode RAG dapat memberikan solusi informasi yang relevan dan adaptif terhadap kebutuhan pengguna kampus. Selain menghasilkan sistem prototipe yang fungsional, proyek ini juga memberikan pembelajaran teknis mendalam bagi penulis dalam bidang pemrosesan bahasa alami, pengelolaan vektor, dan desain sistem berbasis AI. Proyek ini diharapkan dapat menjadi landasan awal bagi pengembangan chatbot universitas yang lebih canggih dan terintegrasi di masa mendatang.

4.1.1 Penguatan Kompetensi Teknis

Penulis mendalami implementasi Large Language Model (LLM), khususnya model Llama-8B untuk menciptakan chatbot yang memahami informasi kampus

secara dinamis. Melalui eksperimen dengan metode Retrieval-Augmented Generation (RAG), penulis menguasai tahapan seperti:

a. Implementasi Large Language Model (LLM), khususnya model Llama-8B

Dalam proyek ini, penulis memanfaatkan model LLaMA 3.1 8B melalui layanan Ollama yang berjalan di lokal. Setelah itu, dilakukan fine-tuning seperti QLoRA, pendekatan yang digunakan adalah inference-only dengan memasukkan konteks dari dokumen handbook mahasiswa secara dinamis. Penulis belajar bagaimana mengatur prompt, membatasi token agar sesuai kapasitas model (4096 token), dan merancang antarmuka interaksi berbasis teks.

b. Penguasaan Metode Retrieval-Augmented Generation (RAG)

Penulis mengimplementasikan metode RAG, yaitu pendekatan yang menggabungkan kemampuan pencarian dokumen (retrieval) dengan kemampuan generatif dari LLM. Sistem memanfaatkan vektor embedding dengan model SentenceTransformers, mengindeks potongan teks dari PDF handbook ke dalam database ChromaDB, dan secara otomatis memilih bagian paling relevan saat menjawab pertanyaan. Penulis memahami secara mendalam bagaimana retrieval relevan membentuk konteks yang meningkatkan akurasi jawaban LLM.

c. Pengumpulan dan Pembersihan Data Institusi

Selama proyek, penulis mengembangkan pipeline ekstraksi teks dari file PDF menggunakan pustaka PyMuPDF (fitz), dilanjutkan dengan pembersihan teks menggunakan import re dan pembagian teks ke dalam potongan kecil agar sesuai untuk embedding dan tokenisasi. Proses ini menunjukkan pentingnya data preparation sebelum data dapat dimanfaatkan secara efektif dalam sistem berbasis AI.

d. Optimasi Embedding untuk Relevansi Konteks

Penulis juga melakukan eksplorasi pada model embedding ringan yang cocok untuk sistem lokal, dan mempelajari pentingnya mengontrol jumlah token dalam konteks melalui tokenizer BERT multilingual. Sistem dibatasi hingga

maksimal 4096 token dan didesain agar mampu menyisipkan konteks terpenting tanpa membebani LLM.

e. Fine-Tuning Model dengan Dataset Dokumen UMN

Fine-tuning adalah proses melatih model LLM pada dataset spesifik untuk meningkatkan kinerjanya pada tugas tertentu. Penulis membuat dataset khusus yang berisi pertanyaan dan jawaban seputar informasi kampus, kemudian menggunakan dataset ini untuk fine-tuning model Llama-8B.

f. Pengembangan Antarmuka Interaktif menggunakan Streamlit

Streamlit digunakan sebagai platform utama untuk membangun antarmuka pengguna chatbot. Penulis merancang layout percakapan seperti tampilan chat modern, menyesuaikan tampilan melalui CSS untuk mendukung mode gelap dan terang, serta memastikan tampilan responsif di perangkat mobile. Selain tampilan, integrasi UI dengan pipeline RAG juga berhasil dilakukan dengan alur yang efisien dan mudah digunakan oleh mahasiswa.

4.1.2 Peningkatan Keterampilan Manajerial

Kompetensi penulis di bidang Kecerdasan Buatan (AI) berkembang pesat melalui pendalaman pipeline Large Language Model (LLM), keahlian manipulasi data dengan Python, dan kemahiran integrasi sistem. Proyek ini tidak hanya memperkaya kemampuan teknis, tetapi juga menempa keterampilan manajerial melalui:

a. Penyusunan Timeline Waktu untuk Menghindari Burnout

Dengan cakupan proyek yang cukup kompleks dan melibatkan eksplorasi teknis, penulis mengatur jadwal pengerjaan secara fleksibel, membagi pekerjaan dalam mingguan, dan memastikan adanya waktu istirahat untuk mencegah kelelahan. Hal ini menjadi pelajaran penting dalam pengelolaan waktu dan beban kerja jangka panjang.

b. Kolaborasi Intensif Dengan Anggota Lintas Keahlian

Selama proyek, penulis berinteraksi dengan dosen pembimbing serta mahasiswa sebagai pengguna potensial chatbot. Dalam proses ini, penulis belajar menyampaikan progres dan kendala secara terstruktur, memahami sudut pandang dari bidang lain, dan menyesuaikan fitur chatbot sesuai kebutuhan pengguna akhir.

c. Adaptasi Terhadap Masukan dan Perubahan Kebutuhan

Seiring berjalannya proyek, terdapat berbagai masukan dari pihak pembimbing mengenai teknis model, antarmuka, maupun usability chatbot. Penulis belajar menyikapi masukan secara terbuka, melakukan iterasi sistem berdasarkan feedback tersebut, serta beradaptasi ketika terjadi perubahan arah pengembangan. Interaksi berkelanjutan dengan dosen pembimbing dan praktisi AI juga menumbuhkan pola pikir analitis, etos kerja yang profesional, serta kemampuan mandiri dalam menuntaskan tantangan teknis.

4.1.3 Pembentukan Pola Pikir dan Etos Kerja Profesional

1. Kemandirian dalam Penyelesaian Masalah Teknis

Dalam proses pengembangan, penulis dihadapkan pada berbagai kendala seperti error saat menjalankan model lokal, error indexing database, hingga isu kompatibilitas model dan embedding. Dengan mengandalkan dokumentasi resmi, forum komunitas, dan debugging mandiri, penulis berhasil menyelesaikan berbagai kendala tersebut secara independen.

2. Etos Kerja dan Tanggung Jawab Proyek

Sebagai proyek individual, penulis bertanggung jawab penuh terhadap hasil akhir sistem. Hal ini melatih konsistensi, disiplin, dan kemampuan menjaga kualitas kerja sepanjang durasi program. Penulis juga terus menjaga dokumentasi teknis dan non-teknis untuk mendukung transparansi dan keberlanjutan proyek.

3. Penguatan Pola Pikir Kritis dan Problem Solving

Melalui interaksi rutin dengan dosen pembimbing dan pengujian sistem secara berkala, penulis terbiasa berpikir sistematis dalam menghadapi tantangan. Penulis

juga melatih kemampuan merancang solusi teknis berdasarkan logika dan data yang ada, serta mempertimbangkan efisiensi dan skalabilitas sistem ke depannya.

4.2 Saran

1. Penyelenggara Lomba

Secara keseluruhan, pelaksanaan lomba karya tulis ilmiah nasional yang diselenggarakan oleh BEM FT UMB pada tahun 2025 berlangsung dengan sangat baik, baik dari segi konsep maupun pelaksanaannya. Lomba ini memiliki tujuan yang jelas dan didukung oleh berbagai kegiatan yang dirancang untuk mengasah kemampuan peserta di bidang teknologi. Sebagai masukan, akan lebih baik jika lomba diselenggarakan secara tatap muka agar suasana kompetisi antar peserta menjadi lebih hidup dan interaktif.

2. Mitra/Perusahaan (Universitas Multimedia Nusantara)

Penulis menyarankan agar pengembangan chatbot berbasis LLM dengan metode RAG ini dapat dikembangkan lebih lanjut dan diimplementasikan secara nyata dalam sistem pelayanan informasi kampus. Selain itu, UMN dapat mempertimbangkan penyediaan dataset institusional yang lebih terstruktur untuk mendukung riset dan inovasi serupa di masa depan.

a. Universitas

UMN dapat memperkuat kurikulum dengan materi-materi seperti Natural Language Processing (NLP), Large Language Model, dan pemrosesan data unstructured agar mahasiswa lebih siap dalam mengembangkan teknologi berbasis AI. Selain itu, dukungan dan pendampingan terhadap proyek-proyek MBKM juga dapat diperluas melalui workshop atau pelatihan intensif sebelum program dimulai. Dalam menjamin keberlanjutan dan peningkatan kualitas program proyek independen, pihak universitas juga perlu memperhatikan beberapa aspek penting, salah satunya adalah penyampaian informasi yang lengkap dan mendetail mengenai program sebaiknya dilakukan sejak awal semester. Meskipun sosialisasi telah dilaksanakan, informasi yang diberikan saat ini masih terlalu umum dan terkesan