

PERANCANGAN CHATBOT BERBASIS ARTIFICIAL INTELLIGENCE UNTUK LAYANAN INFORMASI PPDB DI SMK MUTIARA 17 AGUSTUS KOTA BEKASI

Hanifan Urfan Fiidin, Ahmad Hanafi Musyaroat, Abdul Samee, Husni Thamrin

Teknik Informatika, Universitas Muhammadiyah Surakarta

Jl. A. Yani, Mendungan, Kec. Kartasura Indonesia

L208240010@student.ums.ac.id

ABSTRAK

Layanan informasi Penerimaan Peserta Didik Baru (PPDB) di SMK Mutiara 17 Agustus Kota Bekasi masih dilakukan secara manual, yang mengakibatkan banyak pertanyaan berulang dari orang tua dan calon peserta didik, keterlambatan respons, dan ketidakkonsistenan informasi. Kondisi ini menunjukkan bahwa sistem layanan informasi yang lebih cepat, akurat, dan interaktif sangat diperlukan. Dengan menggunakan metode Hybrid Retrieval-Augmented Generation (RAG) sebagai solusi layanan informasi PPDB yang efektif, penelitian ini bertujuan untuk mengembangkan dan menerapkan chatbot berbasis AI. Metode hybrid retrieval yang menggabungkan Sentence-BERT (SBERT) dan BM25 untuk pencarian informasi di database pengetahuan digunakan untuk membangun sistem. Selanjutnya, melalui platform Ollama dan penyimpanan vektor Qdrant, model bahasa besar (LLM) diintegrasikan dengan sistem. Dengan menggunakan Gradio, antarmuka chatbot dapat diakses melalui browser. Hasil pengujian menunjukkan bahwa model Llama3:8b memiliki waktu respons rata-rata sekitar 4 detik, dan DeepSeek-R1:8b memiliki waktu respons rata-rata sekitar 27 detik. Dibandingkan dengan metode pencarian tunggal, metode hybrid RAG menunjukkan kualitas retrieval dan relevansi jawaban yang lebih baik. Dengan menggunakan data PPDB yang tersedia, sistem dapat menjawab pertanyaan pengguna secara kontekstual dan real-time. Chatbot berbasis Hybrid RAG ini menunjukkan kemampuan untuk menjadi alternatif untuk solusi digitalisasi layanan administrasi sekolah yang interaktif dan mudah digunakan. Mereka juga menunjukkan kemampuan untuk meningkatkan efisiensi dan efektivitas layanan informasi PPDB.

Kata kunci : PPDB, Chatbot, Retrieval-Augmented Generation, Ollama, Qdrant, Large Language Model.

1. PENDAHULUAN

Pada era digital, kemajuan teknologi informasi telah berdampak besar pada banyak hal, termasuk pendidikan. Teknologi di sekolah meningkatkan kinerja administrasi dan mendukung pembelajaran. Proses Penerimaan Peserta Didik Baru (PPDB) adalah salah satu layanan administrasi yang memiliki peran strategis setiap tahun.

Kegiatan rutin tahunan PPDB membutuhkan informasi yang cepat, akurat, dan responsif kepada masyarakat dan calon siswa. Informasi yang paling sering ditanyakan termasuk persyaratan pendaftaran, biaya pendidikan, pilihan jurusan, jadwal pendaftaran, fasilitas sekolah, dan prosedur registrasi ulang. Jumlah pertanyaan yang masuk secara bersamaan dapat meningkatkan beban kerja panitia PPDB.

Studi kasus di SMK Mutiara 17 Agustus Kota Bekasi menunjukkan bahwa guru dan karyawan sekolah masih melakukan layanan informasi PPDB secara manual. Guru yang tergabung dalam panitia PPDB harus memperhatikan tanggung jawab mereka sebagai tenaga pendidik dan tanggung jawab untuk memberikan informasi kepada orang tua dan calon peserta didik. Beberapa masalah nyata disebabkan oleh keadaan ini, termasuk keterlambatan respons terhadap pertanyaan yang masuk, ketidakkonsistenan informasi yang disampaikan, penumpukan pertanyaan serupa yang berulang, dan penurunan kualitas pelayanan selama jam sibuk pendaftaran. Permasalahan ini berdampak langsung pada kepuasan

orang tua dan calon siswa tentang kemampuan mereka untuk mendapatkan informasi yang mereka butuhkan.

Penelitian ini bertujuan untuk mengidentifikasi masalah dan menciptakan sistem chatbot berbasis kecerdasan buatan untuk layanan informasi PPDB di SMK Mutiara 17 Agustus di Kota Bekasi. Penelitian ini akan menggunakan metode Hybrid Retrieval-Augmented Generation (RAG) dengan kombinasi SBERT dan BM25, dan akan mengevaluasi kinerja sistem chatbot berdasarkan relevansi jawaban dan waktu respons.

Penelitian ini menyarankan pengembangan chatbot berbasis Hybrid Retrieval-Augmented Generation (RAG) yang menggabungkan dua teknik pencarian informasi: Sentence-BERT (SBERT) untuk pencarian berbasis kemiripan semantik dan BM25 untuk pencarian berbasis kata kunci. Untuk menyelesaikan masalah ini, chatbot ini akan dibuat. Tujuan dari kombinasi kedua pendekatan ini adalah untuk menghasilkan pencarian yang lebih luas dan relevan daripada pencarian yang dilakukan hanya dengan satu pendekatan. Platform Ollama digunakan sebagai engine LLM lokal, Qdrant sebagai penyimpanan vektor (vector store), dan Gradio sebagai antarmuka browser yang dapat diakses. Informasi pengetahuan sistem didasarkan pada dokumen resmi dan data PPDB dari SMK Mutiara 17 Agustus Kota Bekasi. Pendekatan ini diharapkan akan memungkinkan chatbot untuk menjawab pertanyaan pengguna secara otomatis, konsisten, dan secara real-

time tanpa bergantung pada ketersediaan panitia PPDB secara langsung.

2. TINJAUAN PUSTAKA

2.1. Chatbot

Chatbot adalah program komputer yang memiliki kemampuan untuk berkomunikasi dengan orang lain melalui suara atau GUI, sehingga mereka dapat memahami pertanyaan pengguna dan secara otomatis memberikan respons. Dimungkinkan untuk menggunakan chatbot berbasis kecerdasan buatan (AI) sebagai sistem layanan dan informasi online untuk sekolah; ini dapat meningkatkan efisiensi pengelolaan sekolah dan mengurangi jumlah pekerjaan yang harus dilakukan oleh guru [1].

Chatbot berbasis aturan dan berbasis kecerdasan buatan (AI) menggunakan teknik pembelajaran mesin dan pemrosesan bahasa alami untuk memahami dan memberikan respons yang lebih dinamis dan kontekstual. Dengan menggunakan teknologi pemrosesan bahasa natural (NLP), chatbot dapat digunakan dalam sistem informasi penerimaan siswa baru yang berbasis web untuk memberikan informasi yang lebih efektif dan efisien kepada siswa yang akan datang [2].

2.2. Artificial Intelligence (AI)

Kecerdasan buatan (AI) adalah bidang ilmu komputer yang berfokus pada pembuatan sistem yang dapat meniru kemampuan kecerdasan manusia seperti memahami bahasa, memecahkan masalah, belajar, dan bernalar. AI telah mengubah banyak aspek kehidupan, seperti pendidikan, kesehatan, bisnis, dan layanan publik. Dalam bidang pendidikan, AI digunakan untuk mengotomatiskan proses administrasi, meningkatkan pengalaman belajar, dan memberi orang tua dan siswa layanan yang lebih responsif.

Sesuai dengan kemajuan teknologi, AI semakin berkembang di bidang pendidikan. Chatbot berbasis kecerdasan buatan telah terbukti mampu meningkatkan kualitas interaksi antara siswa dan sistem informasi akademik. Selain itu, lembaga pendidikan juga dapat dengan mudah menangani banyak pertanyaan masyarakat [3].

2.3. Natural Language Processing (NLP)

Salah satu percabangan kecerdasan buatan adalah pemrosesan bahasa alami (NLP). Proses ini menggabungkan linguistik komputasional dengan pembelajaran mesin dan pembelajaran mendalam (deep learning) untuk memproses bahasa alami, baik teks maupun suara. Proses pemrosesan bahasa alami (NLP) sebagai subbidang ilmu kecerdasan buatan yang mempelajari cara manusia dan komputer berkomunikasi melalui bahasa alami. NLP merupakan dasar pengembangan aplikasi chatbot [4].

NLP digunakan dalam sistem chatbot untuk memproses teks pengguna agar dapat dipahami oleh mesin. Tokenisasi, penghapusan stop word, stemming, ekstraksi fitur, dan klasifikasi intent adalah beberapa

langkah dalam proses pemrosesan bahasa natural. Hidayat et al. membuat chatbot konsultasi akademik dengan NLP. Mereka menunjukkan bahwa teknik ini dapat meningkatkan respons sistem terhadap pertanyaan pengguna. Selain itu, teknik representasi semantik berbasis vektor, yang merupakan bagian dari proses pemrosesan bahasa modern, memungkinkan mesin memahami kata dan kalimat secara kontekstual [5].

2.4. Large Language Model (LLM)

Untuk berbagai tujuan, seperti menerjemahkan teks, membuat konten, membuat tanya jawab otomatis, dan mengembangkan chatbot, Large Language Model (LLM) adalah model kecerdasan buatan yang dilatih menggunakan data teks berskala besar dengan miliaran parameter. Model GPT-3 menunjukkan sebagai salah satu LLM terbaik, yang menunjukkan kemampuan belajar satu kali dalam berbagai tugas pemrosesan bahasa [6].

Sistem chatbot LLM yang dikembangkan pada penelitian ini dijalankan secara lokal menggunakan platform Ollama dengan model Llama 3 dan DeepSeek-R1. Penggunaan LLM lokal menawarkan keunggulan seperti tidak perlu membayar API, kecepatan inferensi yang lebih tinggi, dan perlindungan privasi data yang lebih baik. Touvron et al. memperkenalkan model Llama 2 sebagai LLM open-source yang memiliki kemampuan untuk bersaing dengan model komersial dalam berbagai standar NLP. Ini membentuk fondasi untuk pengembangan Llama 3, yang digunakan dalam penelitian ini [7].

2.5. Retrieval-Augmented Generation (RAG)

Model generatif bahasa (generasi) dan mekanisme pencarian informasi (retrieval) adalah dua komponen utama sistem kecerdasan buatan yang digabungkan dalam arsitektur Retrieval-Augmented Generation (RAG). Arsitektur ini memulai proses pencarian dokumen atau konteks yang relevan dari basis data. Kemudian, sistem menggunakannya sebagai landasan bagi model bahasa untuk menghasilkan jawaban. RAG dapat membantu dalam tugas-tugas pengolahan bahasa yang membutuhkan pengetahuan mendalam karena dapat mengurangi halusinasi model dan meningkatkan akurasi jawaban berdasarkan data nyata [8].

RAG telah terbukti meningkatkan respons sistem chatbot. Albert dan Voutama membuat chatbot berbasis PDF dengan menggunakan RAG dan Ollama lokal. Mereka berhasil membuat sistem yang mampu menjawab pertanyaan pengguna dengan benar berdasarkan konten dokumen. Dalam penelitian yang sama [9], RAG digunakan untuk chatbot kesehatan dan memperoleh hasil yang memuaskan. Untuk meningkatkan relevansi dokumen yang diambil, RAG digunakan dalam mekanisme hybrid retrieval yang menggabungkan pencarian berbasis vektor semantik dan algoritma BM25 [10].

2.6. Penerimaan Peserta Didik Baru (PPDB)

Lembaga pendidikan mengadakan Penerimaan Peserta Didik Baru (PPDB) setiap awal tahun ajaran untuk menyambut siswa baru. Proses PPDB mencakup berbagai tahapan, mulai dari pengumuman pendaftaran, pengumpulan berkas persyaratan, seleksi, pengumuman hasil seleksi, hingga daftar ulang. Informasi yang sering ditanyakan oleh masyarakat dalam proses PPDB meliputi syarat pendaftaran, biaya pendidikan, pilihan jurusan, jadwal pendaftaran, fasilitas sekolah, dan prosedur registrasi.

Pelaksanaan PPDB di sekolah menengah kejuruan (SMK) memiliki tantangan tersendiri karena luasnya informasi yang harus disampaikan kepada calon peserta didik, termasuk penjelasan mengenai berbagai program keahlian yang tersedia. Dalam kondisi sumber daya manusia yang terbatas, pemberian layanan informasi PPDB secara manual oleh guru yang juga merangkap sebagai panitia PPDB seringkali menimbulkan permasalahan berupa ketidakfokusan dalam pelaksanaan kegiatan belajar mengajar (KBM). Penerapan teknologi chatbot berbasis AI diharapkan dapat menjadi solusi alternatif yang mampu memberikan layanan informasi PPDB secara otomatis, cepat, dan akurat kepada masyarakat.

2.7. Penelitian Terdahulu

Beberapa penelitian terdahulu yang relevan dengan penelitian ini telah dilakukan dan menjadi landasan pengembangan sistem chatbot berbasis AI untuk layanan informasi PPDB.

Dalam penelitian berjudul "Penerapan Artificial Intelligence pada Aplikasi Chatbot sebagai Sistem Pelayanan dan Informasi Online pada Sekolah" berhasil mengembangkan chatbot berbasis AI untuk layanan informasi sekolah [1].

Hasil penelitian menunjukkan bahwa chatbot mampu memberikan respons yang cepat dan akurat terhadap pertanyaan-pertanyaan umum dari siswa dan orang tua, sehingga meningkatkan efisiensi layanan administrasi sekolah.

Penelitian berjudul "Designing a Chatbot with NLP Technology in a Website-Based New Student Admission Information System". Penelitian ini mengembangkan chatbot khusus untuk sistem informasi penerimaan siswa baru berbasis website menggunakan teknologi NLP. Hasilnya menunjukkan bahwa chatbot berhasil mempercepat proses penyampaian informasi kepada calon siswa dan mengurangi antrean pertanyaan langsung kepada panitia penerimaan [2].

Dalam penelitian berjudul "Pengembangan Chatbot Berbasis PDF Menggunakan Local Retrieval-Augmented Generation (RAG) dan Ollama" mengembangkan chatbot yang memanfaatkan arsitektur RAG lokal untuk menjawab pertanyaan berdasarkan konten dokumen PDF. Penelitian ini membuktikan bahwa kombinasi RAG dan LLM lokal melalui Ollama menghasilkan sistem chatbot yang

akurat dan efisien tanpa bergantung pada layanan cloud berbayar [9].

Penelitian berjudul "Rancangan dan Implementasi Chatbot Layanan Informasi Pendaftaran Pascasarjana di Perguruan Tinggi" yang mengimplementasikan chatbot untuk melayani informasi pendaftaran mahasiswa baru di perguruan tinggi. Hasil pengujian menunjukkan bahwa chatbot mampu meningkatkan kepuasan pengguna dalam mendapatkan informasi pendaftaran secara mandiri tanpa harus menghubungi petugas secara langsung [10].

Dalam penelitiannya berjudul "Pengembangan Chatbot Berbasis AI untuk Mendukung Pelayanan Perpustakaan" mengembangkan chatbot AI untuk layanan perpustakaan sekolah. Penelitian ini menunjukkan bahwa chatbot berbasis AI dapat diadaptasi untuk berbagai jenis layanan informasi di lingkungan pendidikan, tidak hanya terbatas pada layanan akademik semata [11].

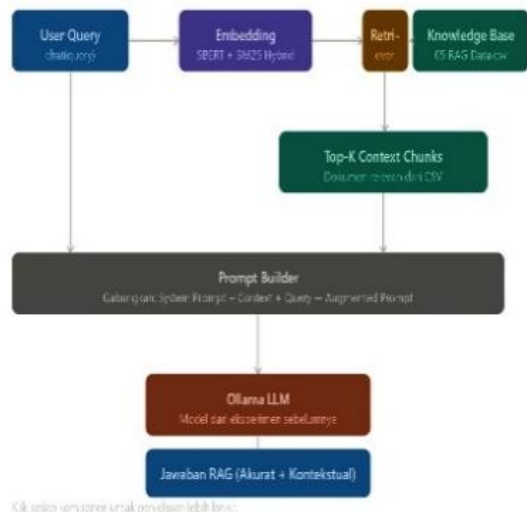
Berdasarkan kajian terhadap penelitian-penelitian terdahulu tersebut, terdapat beberapa perbedaan yang menjadi kebaruan (*novelty*) dalam penelitian ini. Penelitian ini secara khusus berfokus pada layanan informasi PPDB di jenjang SMK dengan mengintegrasikan arsitektur RAG hybrid yang menggabungkan dense retrieval berbasis vektor semantik (BAAI/bge-m3) dengan sparse retrieval berbasis BM25, serta mekanisme reranking menggunakan model CrossEncoder (BAAI/bge-reranker-v2-m3). Pendekatan ini diharapkan menghasilkan sistem chatbot dengan tingkat akurasi dan relevansi respons yang lebih tinggi dibandingkan penelitian-penelitian sebelumnya.

3. METODE PENELITIAN

Penelitian terapan (*applied research*) ini berfokus pada pembuatan dan penggunaan sistem chatbot berbasis AI. Dengan metode pengembangan sistem (*system development*), pendekatan kuantitatif digunakan. Sistem dirancang, dibangun, diimplementasikan, dan dievaluasi berdasarkan parameter performa yang ditetapkan. Analisis kebutuhan, perancangan sistem, implementasi, dan pengujian adalah bagian dari siklus pengembangan perangkat lunak [12].

Secara keseluruhan, sistem bekerja melalui beberapa tahapan utama yang saling terhubung, mulai dari menerima pertanyaan pengguna hingga menghasilkan jawaban yang relevan dan kontekstual. Sistem chatbot dimulai dengan input User Query, yang kemudian diproses melalui modul Embedding menggunakan model BAAI/bge-m3 untuk menghasilkan representasi vektor semantik, seperti yang ditunjukkan pada Gambar X. Secara bersamaan, reranking dilakukan pada dokumen Knowledge Base, yang terdiri dari dataset FAQ RAG. Hasil pencarian kemudian disaring menjadi Top-K Context Chunks yang paling relevan dengan menggunakan model DeepSeek/LLM. Prompt Builder kemudian

menggabungkan potongan konteks yang dipilih menjadi Prompt yang ditingkatkan, yang menggabungkan sistem prompt, konteks, dan pertanyaan pengguna. Prompt yang dibangun ini kemudian dikirim ke Ollama LLM untuk diproses oleh model yang telah dipilih sebelumnya, yang menghasilkan Jawaban RAG yang kontekstual dan akurat[13]. Dan berikut di bawah ini gambar dari kerja chatbot.



Gambar 1. arsitektur kinerja chatbot

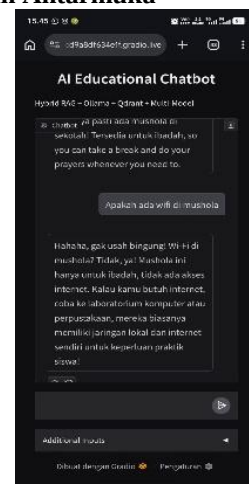
Metode Hybrid Retrieval-Augmented Generation (RAG), yang terdiri dari berbagai tahapan terintegrasi, digunakan untuk mengembangkan sistem chatbot. Sistem menangani pertanyaan yang diajukan oleh pengguna melalui antarmuka berbasis Gradio menggunakan metode pencarian hybrid yang menggabungkan SBERT untuk pencarian berbasis kemiripan semantik dan BM25 untuk pencarian berbasis kata kunci. Selanjutnya, hasil pencarian digunakan untuk memeriksa sumber pengetahuan yang ada di Qdrant. Dokumen resmi PPDB SMK Mutiara 17 Agustus Kota Bekasi berisi informasi tentang persyaratan pendaftaran, biaya pendidikan, jurusan, jadwal, dan prosedur registrasi. Dengan menggunakan mekanisme Top-K Context Chunks, dokumen yang paling relevan dipilih dan digabungkan dengan pertanyaan pengguna pada tahap Prompt Builder untuk membuat pesan tambahan yang dikirimkan ke model LLM melalui platform Ollama.

Dua model LLM untuk perbandingan digunakan dalam penelitian ini: Llama3:8b dan DeepSeek-R1:8b. Gambar 1 menunjukkan struktur sistem secara keseluruhan. Pengujian sistem dilakukan dalam dua aspek: pertama, menguji waktu respons untuk mengetahui seberapa cepat sistem menghasilkan jawaban; kedua, menguji relevansi jawaban dengan membandingkan output sistem dengan jawaban referensi yang telah ditetapkan. Hasil pengujian digunakan sebagai dasar analisis untuk menentukan model LLM terbaik untuk sistem chatbot layanan informasi PPDB.

4. HASIL DAN PEMBAHASAN

Hasil implementasi sistem menunjukkan bahwa Dengan penggabungan Ollama, Qdrant, dan antarmuka berbasis Gradio, chatbot berbasis Hybrid Retrieval-Augmented Generation (RAG) berhasil digunakan. Sistem dapat secara interaktif menerima pertanyaan pengguna dan memberikan jawaban kontekstual berdasarkan data yang tersimpan pada knowledge base. Tampilan antarmuka chatbot ditunjukkan pada Gambar 4. Pada proses pengujian, pengguna mengajukan pertanyaan terkait fasilitas sekolah, yaitu mengenai ketersediaan Wi-Fi di mushola. Sistem berhasil memahami konteks pertanyaan dan memberikan jawaban secara natural serta komunikatif. Jawaban yang dihasilkan tidak hanya memberikan informasi utama, tetapi juga memberikan saran tambahan terkait lokasi alternatif yang menyediakan akses internet, seperti laboratorium komputer dan perpustakaan sekolah. Ini menunjukkan bahwa integrasi retrieval dan Large Language Model (LLM) dapat menghasilkan respons yang lebih interaktif yang mirip dengan percakapan manusia. Selain itu, metode hybrid retrieval membantu sistem menemukan konteks informasi yang relevan sebelum Ollama memulai proses generasi jawaban. Metode ini memungkinkan chatbot membuat tanggapan yang lebih sesuai dengan kebutuhan pengguna. dibandingkan model generatif biasa tanpa retrieval. Implementasi antarmuka menggunakan Gradio juga mempermudah pengguna dalam berinteraksi dengan sistem melalui perangkat desktop maupun mobile. Meskipun hasil pengujian menunjukkan performa yang cukup baik, masih terdapat beberapa keterbatasan pada sistem, terutama ketika pengguna memberikan pertanyaan yang terlalu umum, ambigu, atau tidak tersedia pada knowledge base. Dalam beberapa kondisi, chatbot masih dapat menghasilkan jawaban yang kurang spesifik atau terlalu panjang sehingga diperlukan pengembangan lebih lanjut pada proses retrieval maupun optimasi prompt agar kualitas jawaban menjadi lebih akurat dan konsisten.

4.1. Tampilan Antarmuka

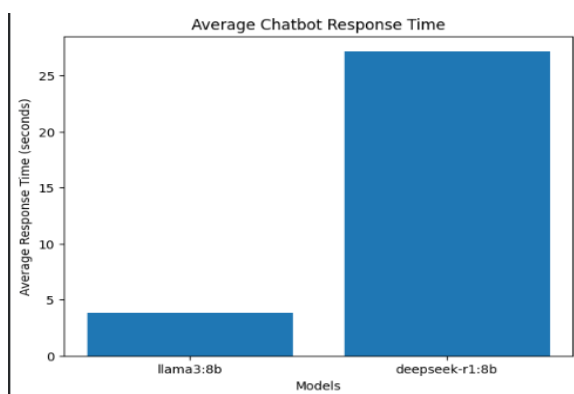


Gambar 2. Tampilan sistem

Antarmuka chatbot dikembangkan Untuk menyediakan pengalaman pengguna yang interaktif dan mudah digunakan, Gradio Tampilan utama sistem terdiri dari area percakapan, atau chat room, di mana pengguna dapat mengajukan pertanyaan kepada chatbot berbasis AI secara langsung. Untuk memberikan jawaban yang lebih kontekstual dan relevan, sistem menggunakan pendekatan Hybrid Retrieval-Augmented Generation (RAG), yang menggabungkan Ollama dan Qdrant. Sebelum model LLM menghasilkan jawaban, pertanyaan pengguna akan diproses melalui mekanisme retrieval untuk mendapatkan informasi yang sesuai dari database pengetahuan. Hasil pengujian menunjukkan bahwa chatbot mampu memberikan jawaban yang informatif dan komunikatif secara real-time. Selain itu, Gradio membuat akses sistem melalui browser pada perangkat desktop dan mobile lebih mudah.

4.2. Analisa Kinerja Chatbot

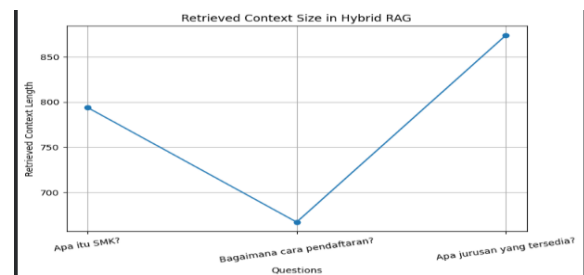
Analisis kinerja chatbot dilakukan dengan membandingkan rata-rata waktu respons antara model Llama3:8b dan DeepSeek-R1:8b. Berdasarkan hasil pengujian yang ditunjukkan pada Gambar 5, model Llama3:8b memiliki rata-rata waktu respons sekitar 4 detik, sedangkan model DeepSeek-R1:8b membutuhkan waktu respons yang lebih lama, yaitu sekitar 27 detik. Hasil tersebut menunjukkan bahwa model Llama3:8b mampu memberikan respons yang lebih cepat dibandingkan DeepSeek-R1:8b pada implementasi chatbot berbasis RAG yang digunakan. Waktu respons yang lebih rendah memberikan pengalaman interaksi yang lebih baik bagi pengguna karena jawaban dapat diterima secara lebih real-time. Sementara itu, model DeepSeek-R1:8b memerlukan waktu pemrosesan yang lebih tinggi karena proses reasoning yang lebih kompleks dalam menghasilkan jawaban. Oleh karena itu, pemilihan model LLM perlu disesuaikan dengan kebutuhan sistem, baik dari sisi kecepatan respons maupun kualitas jawaban yang dihasilkan.



Gambar 3. avarage chatbot response

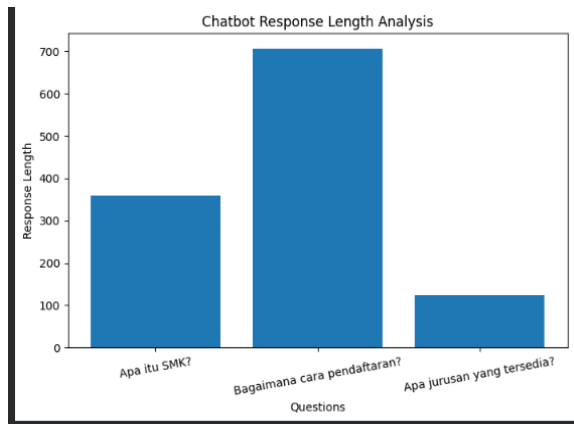
Analisis retrieval context dilakukan untuk melihat jumlah konteks informasi yang berhasil

diambil oleh sistem Hybrid Retrieval-Augmented Generation (RAG) berdasarkan beberapa pertanyaan pengguna. Berdasarkan hasil pengujian pada Gambar 6, setiap pertanyaan menghasilkan ukuran retrieved context yang berbeda-beda sesuai dengan kompleksitas dan ketersediaan informasi pada knowledge base. Pertanyaan “Apa itu SMK?” menghasilkan retrieved context sekitar 795 karakter, sedangkan pertanyaan “Bagaimana cara pendaftaran?” menghasilkan context yang lebih kecil, yaitu sekitar 667 karakter. Sementara itu, pertanyaan “Apa jurusan yang tersedia?” menghasilkan retrieved context terbesar, yaitu sekitar 873 karakter. Perbedaan ukuran retrieved context menunjukkan bahwa sistem Hybrid RAG mampu menyesuaikan jumlah konteks yang diambil berdasarkan relevansi informasi terhadap pertanyaan pengguna. Semakin banyak informasi relevan yang ditemukan pada knowledge base, maka semakin besar pula context yang digunakan dalam proses generasi jawaban oleh model LLM. Dengan mekanisme tersebut, chatbot dapat menghasilkan jawaban yang lebih kontekstual dan sesuai dengan kebutuhan pengguna.



Gambar 4. retrieved context

Analisis panjang respons chatbot dilakukan untuk mengetahui jumlah karakter jawaban yang dihasilkan sistem terhadap beberapa pertanyaan pengguna. Berdasarkan hasil pengujian pada Gambar 7, panjang respons yang dihasilkan berbeda-beda tergantung pada jenis pertanyaan dan banyaknya informasi relevan yang ditemukan pada knowledge base. Pertanyaan “Bagaimana cara pendaftaran?” menghasilkan respons terpanjang, yaitu sekitar 705 karakter. Hal ini menunjukkan bahwa sistem memberikan penjelasan yang lebih detail terkait prosedur pendaftaran. Sementara itu, pertanyaan “Apa itu SMK?” menghasilkan panjang respons sekitar 360 karakter, sedangkan pertanyaan “Apa jurusan yang tersedia?” menghasilkan respons yang lebih singkat, yaitu sekitar 125 karakter. Perbedaan panjang respons menunjukkan bahwa chatbot mampu menyesuaikan jawaban berdasarkan konteks dan kebutuhan informasi dari pengguna. Semakin kompleks pertanyaan yang diberikan, maka sistem cenderung menghasilkan jawaban yang lebih panjang dan informatif. Dengan pendekatan Hybrid RAG, chatbot dapat menghasilkan respons yang lebih relevan dan kontekstual sesuai dengan data yang tersedia pada knowledge base.



Gambar 5. chatbot response length

Evaluasi kinerja chatbot pendidikan yang diusulkan dilakukan dengan mengajukan tiga pertanyaan representatif yang umum ditanyakan oleh calon siswa. Hasil pengujian disajikan pada Tabel 1, yang mencakup tiga parameter utama, yaitu panjang respons (response length), panjang konteks (context length), dan waktu respons (response time). Berdasarkan hasil pengujian, pertanyaan "Apa itu SMK?" menghasilkan respons terpanjang dengan 289 karakter dan waktu respons sebesar 3,76 detik, dengan panjang konteks yang diambil sebesar 794 token. Pertanyaan "Bagaimana cara pendaftaran?" menghasilkan respons sepanjang 225 karakter dengan waktu respons 2,81 detik dan konteks sebesar 667 token. Sementara itu, pertanyaan "Apa jurusan yang tersedia?" memiliki respons terpendek dengan 138 karakter dan waktu respons tercepat sebesar 2,15 detik, meskipun memiliki panjang konteks terbesar yaitu 874 token. Secara keseluruhan, rata-rata waktu respons chatbot berada di kisaran 2,15 hingga 3,76 detik, yang menunjukkan bahwa sistem mampu memberikan jawaban secara real-time dengan latensi yang tergolong rendah dan dapat diterima untuk kebutuhan interaksi pengguna. Perbedaan waktu respons yang terjadi dipengaruhi oleh kompleksitas pertanyaan serta panjang konteks yang diambil dari basis pengetahuan. Hasil ini mengindikasikan bahwa chatbot yang diusulkan memiliki kinerja yang baik dalam memproses dan merespons pertanyaan seputar informasi sekolah secara efisien.

Question	Response Length	Context Length	Response Time (s)
Apa itu SMK?	289	794	3.76
Bagaimana cara pendaftaran?	225	667	2.81
Apa jurusan yang tersedia?	138	874	2.15

Gambar 6. table performance evaluation

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, chatbot berbasis Hybrid Retrieval-Augmented Generation (RAG) menggunakan Ollama dan Qdrant berhasil

dikembangkan untuk memberikan layanan informasi pendidikan secara interaktif dan kontekstual. Sistem mampu melakukan retrieval informasi dari knowledge base sebelum menghasilkan jawaban sehingga respons yang diberikan menjadi lebih relevan dan akurat. Hasil pengujian menunjukkan bahwa model Llama3:8b memiliki rata-rata waktu respons sekitar 4 detik, lebih cepat dibandingkan DeepSeek-R1:8b yang mencapai sekitar 27 detik. Selain itu, retrieved context yang diperoleh sistem berada pada rentang 667 hingga 873 karakter, sedangkan panjang respons chatbot berkisar antara 125 hingga 705 karakter tergantung kompleksitas pertanyaan pengguna. Dengan demikian, metode hybrid retrieval terbukti mampu meningkatkan kualitas konteks dan relevansi jawaban chatbot. Meskipun demikian, sistem masih memiliki keterbatasan pada pertanyaan yang ambigu atau di luar knowledge base sehingga diperlukan pengembangan lebih lanjut untuk meningkatkan akurasi dan kualitas respons chatbot.

DAFTAR PUSTAKA

- [1] E. Erlina et al., "Penerapan Artificial Intelligence pada Aplikasi Chatbot sebagai Sistem Pelayanan dan Informasi Online pada Sekolah," *J. Inf. Syst. Technol.*, vol. 4, pp. 221–230, Nov. 2023, doi: 10.37253/joint.v4i3.6296.
- [2] M. Fauzan, R. Imanda, and M. Hasbi, "Designing an Chatbot with NLP Technology in a Website-Based New Student Admission Information System," *J. Appl. Informatics Comput.*, vol. 8, pp. 358–366, Nov. 2024, doi: 10.30871/jaic.v8i2.8489.
- [3] Suprihartini, Y., Rahayu, D. A., Utama, I. W. K., Suwarma, D. M., Marlina, M., & Irawan, I. (2025). Analisis penggunaan chatbot berbasis AI sebagai asisten virtual dalam pembelajaran daring. *Jurnal Review Pendidikan dan Pengajaran*, 8(2), 4778–4784. <https://doi.org/10.31004/jrpp.v8i2.44966>
- [4] Puspitasari, A., Paradhita, A. N., Tineka, Y. W., Sulistyowati, V., Noriska, N. K. S., & Haryanto. (2024). Natural language processing (NLP) technology for chatbot website. *Jurnal Penelitian Pendidikan IPA*, 10(Special Issue). <https://doi.org/10.29303/jppipa.v10iSpecialIssue.8241>
- [5] Hidayat, S., Silvanie, A., & Subekti, R. (2022). Chatbot untuk konsultasi akademik menggunakan natural language processing (NLP) di IBI-K 57. *Journal of Information System, Applied, Management, Accounting and Research*, 6(2), 396–410.
- [6] Brown, T., et al. (2020). "Language models are few-shot learners." *Advances in Neural Information Processing Systems*, 33, 1877–1901.
- [7] Touvron, H., et al. (2023). "Llama 2: Open foundation and fine-tuned chat models." *arXiv preprint arXiv:2307.09288*.

-
- [8] P. Lewis et al., "Retrieval-augmented generation for knowledge-intensive NLP tasks," *Advances in Neural Information Processing Systems*, vol. 33, 2020.
- [9] Albert, G. D., & Voutama, A. (2025). Pengembangan chatbot berbasis PDF menggunakan local retrieval-augmented generation (RAG) dan Ollama. *Jurnal Informatika dan Teknik Elektro Terapan*, 13(2). <https://doi.org/10.23960/jitet.v13i2.6361>
- [10] Samudra, G., Turmudi Zy, A., & Ermanto. (2025). Implementation of retrieval augmented generation (RAG) in the design of digestive health chatbot. *JSAT: Journal Scientific and Applied Informatics*, 8(1), 181–188.
- [11] M. Qaulan, Wahyuni, and P. Adytia, "Penerapan Chatbot Berbasis AI untuk Pelayanan Perpustakaan di STMIK Widya Cipta Dharma," *TEMATIK*, vol. 12, pp. 23–30, Jun. 2025, doi: 10.38204/tematik.v12i1.2283.
- [12] G. D. Albert and A. Voutama, "Pengembangan chatbot berbasis PDF menggunakan local retrieval-augmented generation (RAG) dan Ollama," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 2, 2025.
- [13] M. Mustaqim, A. Gunawan, Y. B. Pratama, and I. Zaliman, "Pengembangan chatbot layanan publik menggunakan machine learning dan natural language processing," *Journal of Information Technology and Society*, vol. 1, no. 1, pp. 1–4, 2023.