

Implementasi Chatbot Informasi Akademik Berbasis Retrieval-Augmented Generation (RAG) Menggunakan LLaMA 3.1

Oleh :

Muhammad Saddam Heykal Bustomy (221080200131)

Dosen Pembimbing :

Dr. Mochamad Alfian Rosid, S.Kom., M.Kom.

Program Studi Informatika

Universitas Muhammadiyah Sidoarjo

2025



Latar Belakang

- Di era digital saat ini, perkembangan teknologi semakin pesat terutama Kecerdasan Buatan (AI).
- Proses pencarian informasi ini menjadi kurang efisien dan kurang efektif dikarenakan informasi tersebar di berbagai platform (Website prodi, fakultas, bahkan bagian administrasi).
- Membangun Chatbot berbasis AI sebagai pusat informasi yang memberikan jawaban secara cepat dan akurat dari sumber data resmi
- Menyediakan akses informasi akademik yang mudah langsung dari web prodi dan meningkatkan efisiensi mahasiswa dalam pencarian informasi.

Rumusan Masalah

Rumusan masalah yang diangkat dalam penelitian ini adalah:
“Bagaimana merancang dan membangun chatbot berbasis AI yang mampu memberikan informasi akademik secara cepat dan kontekstual kepada mahasiswa Program Studi Informatika?”

Tujuan

Mengembangkan sebuah sistem chatbot berbasis AI yang dapat meningkatkan efisiensi dalam memperoleh informasi akademik dari Program Studi Informatika Universitas Muhammadiyah Sidoarjo.

Batasan Masalah

- Chatbot hanya akan melayani pertanyaan seputar informasi umum akademik yang berkaitan dengan Program Studi Informatika.
- Sumber data yang digunakan terbatas pada hasil web scraping dari situs web resmi program studi dan dokumen PDF.
- Sistem tidak mencakup kemampuan untuk melakukan tindakan administratif seperti pengisian formulir atau validasi data pribadi mahasiswa.

Penelitian Terdahulu

Hidayat et al., 2021

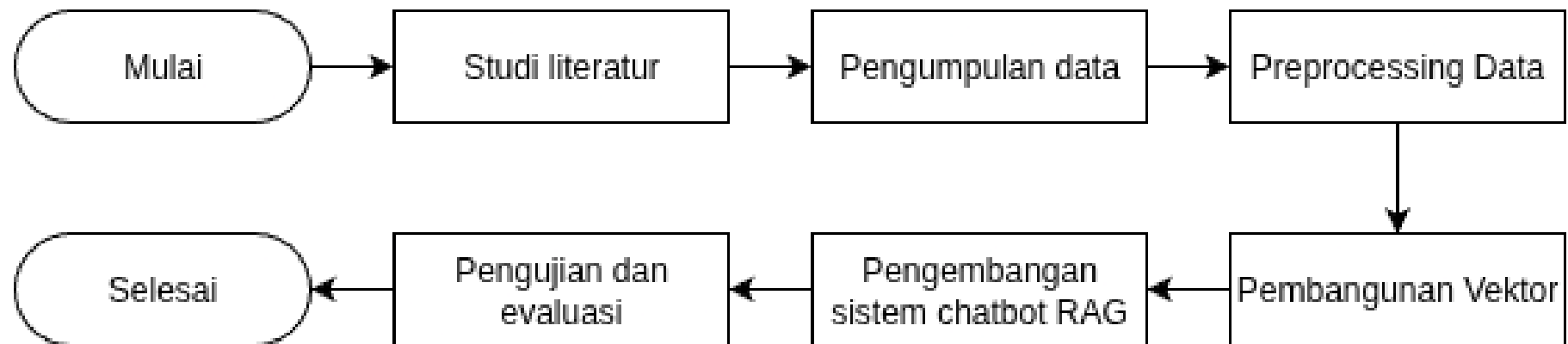
Judul : Optimalisasi Layanan Sistem Informasi Mahasiswa dengan Integrasi Telegram : Chatbot Retrieval-Augmented-Generation berbasis Large Language Model

Milasanti D, 2024

Judul : Sistem Chatbot Berbasis Large Language Model (LLM) Dan Retrieval Augmented Generation (RAG) Pada Artikel Ilmiah Garuda Kemdikbud

Kesimpulan : Penelitian sebelumnya dengan model LLM lokal memiliki kelemahan pada aspek latensi dengan waktu respon yang lambat, serta akurasi yang kurang optimal.

Tahapan Penelitian



Rancangan Penelitian

➤ Pengumpulan Data

Sumber Data :

Website resmi Prodi	Dokumen PDF
➤ Data : Semua informasi yang ada pada web resmi program studi Informatika. ➤ Metode : Web Scraping → Ekstraksi teks dari halaman web	➤ Data : Buku panduan akademik, SOP, panduan skripsi, dll. ➤ Metode : Parsing dokumen PDF → Konversi isi dokumen ke teks yang terstruktur

Rancangan Penelitian

➤ Preprocessing Data

Sebelum	Sesudah
“Menghasilkan lulusan yang profesional, unggul, inovatif, dan kompetitif dalam rekayasa perangkat lunak dan sistem cerdas yang adaptif terhadap perkembangan IPTEKS berdasarkan nilai-nilai Islam untuk kesejahteraan masyarakat tingkat ASEAN pada tahun 2038”	Menghasilkan lulusan yang profesional, unggul, inovatif, dan kompetitif dalam rekayasa perangkat lunak dan sistem cerdas yang adaptif terhadap perkembangan IPTEKS berdasarkan nilai-nilai Islam untuk kesejahteraan masyarakat tingkat ASEAN pada tahun 2038

Hasil preprocessing :

- Teks dibersihkan dari tag HTML
- Teks dipecah menjadi potongan (chunk) dengan ukuran maksimal 4000 karakter
- Menggunakan overlap 200 karakter agar konteks tetap terhubung

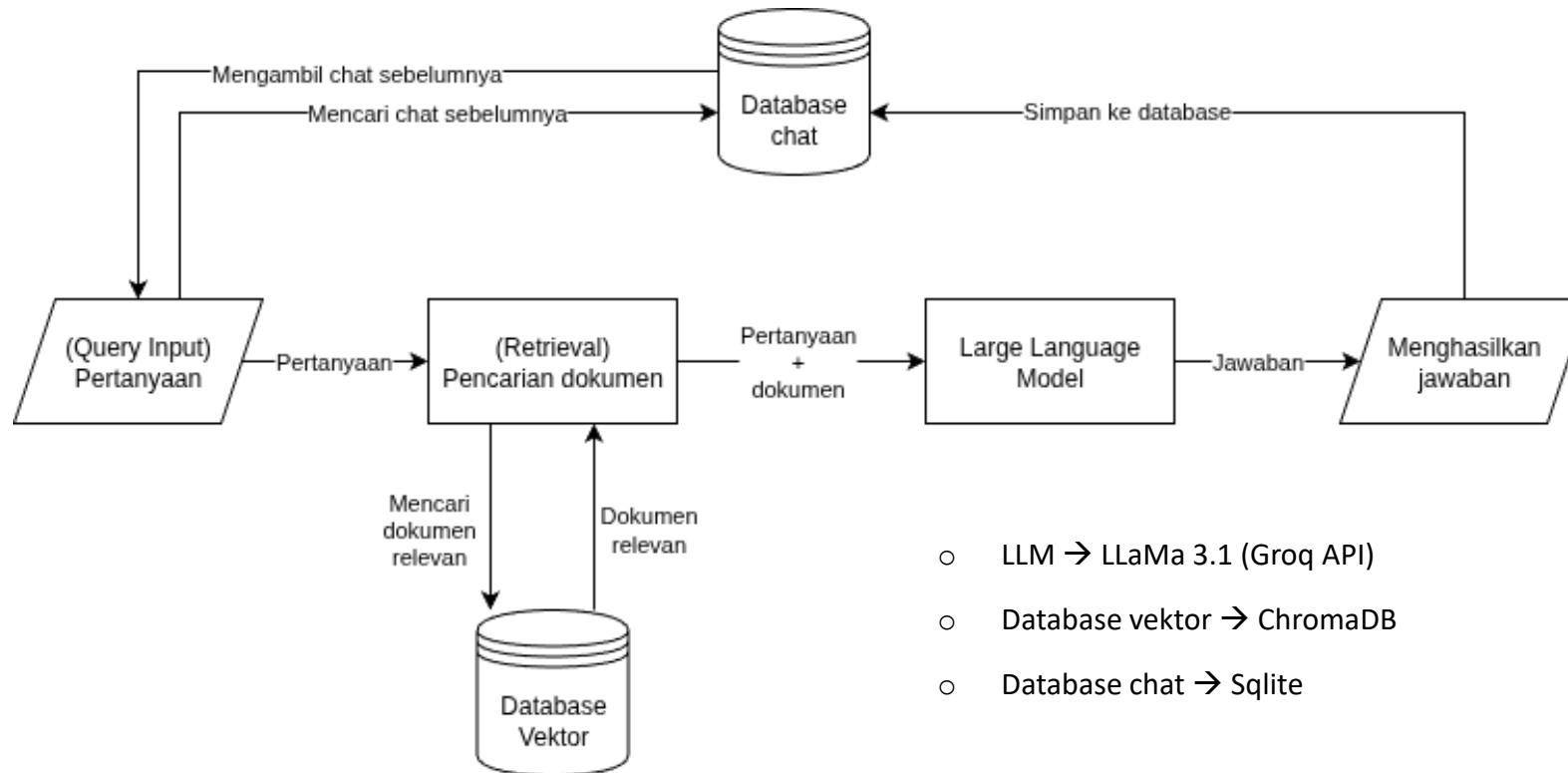
Rancangan Penelitian

➤ Pembangunan Vektor

Sebelum	Sesudah
Menghasilkan lulusan yang profesional, unggul, inovatif, dan kompetitif dalam rekayasa perangkat lunak dan sistem cerdas yang adaptif terhadap perkembangan IPTEKS berdasarkan nilai-nilai Islam untuk kesejahteraan masyarakat tingkat ASEAN pada tahun 2038	[-0.02805283, 0.0016733863, -0.13027878, -0.09528997, 0.020507494, -0.0108989095, 0.063497745, -0.010579597, -0.035166822, -0.007186834, -0.007837213, -0.0024443567, 0.08700824, -0.006801269, -0.011802809, -0.031749826, -0.026138144, -0.04108551, -0.087224066, 0.013060821, 0.0004179242, -0.006775316, -0.039658986, -0.032068517, 0.07313148, 0.0340987, -0.027446775, 0.00012398965, -0.031950794, -0.04718184, 0.0034216747, -0.028119756, -0.027272059, -0.037347674, -0.047374934, -0.06768788, 0.072472334, -0.055526186, -0.02520834, -0.07217245, 0.01854235, -0.018051617, 0.01461787, -0.015421444, 0.009143037, -0.011913811, 0.036699366, 0.0022164185, 0.064935096, 0.0106698135, -0.05006927, 0.011274555, 0.03964463, -0.064298384, 0.04749779, 0.052654978, -0.004527706, 0.005267031, 0.026171189, -0.058371577, 0.11699647, 0.09592061, -0.07412396]

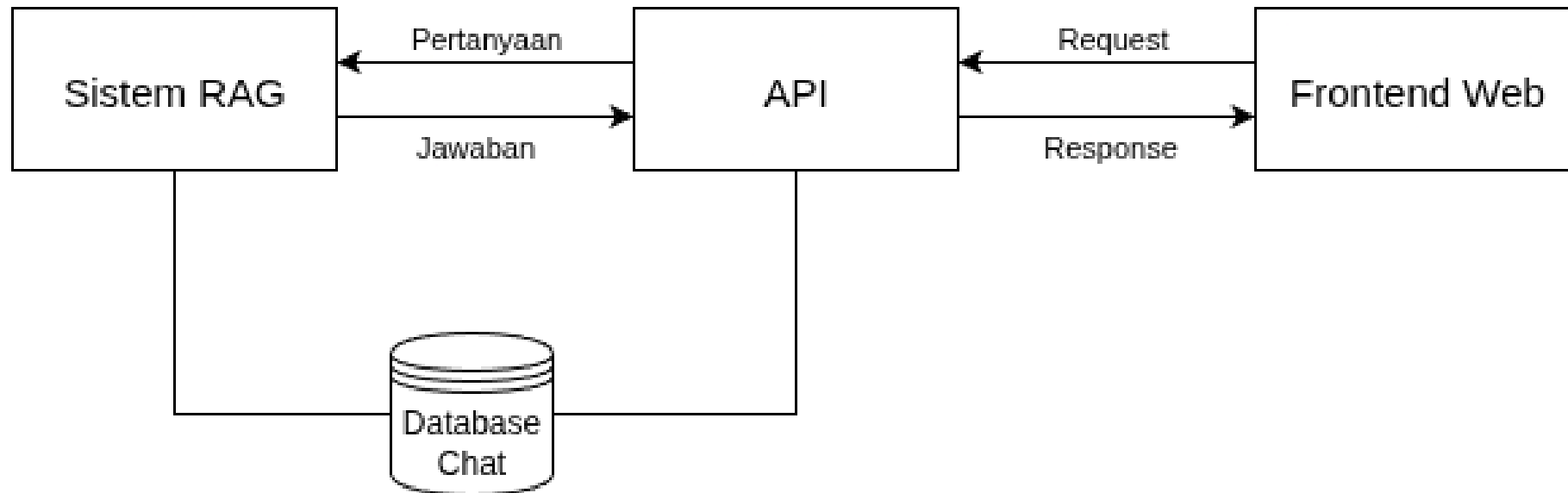
Arsitektur RAG

Retrieval Augmented Generation (RAG)



- LLM → LLaMa 3.1 (Groq API)
- Database vektor → ChromaDB
- Database chat → Sqlite

Arsitektur Sistem



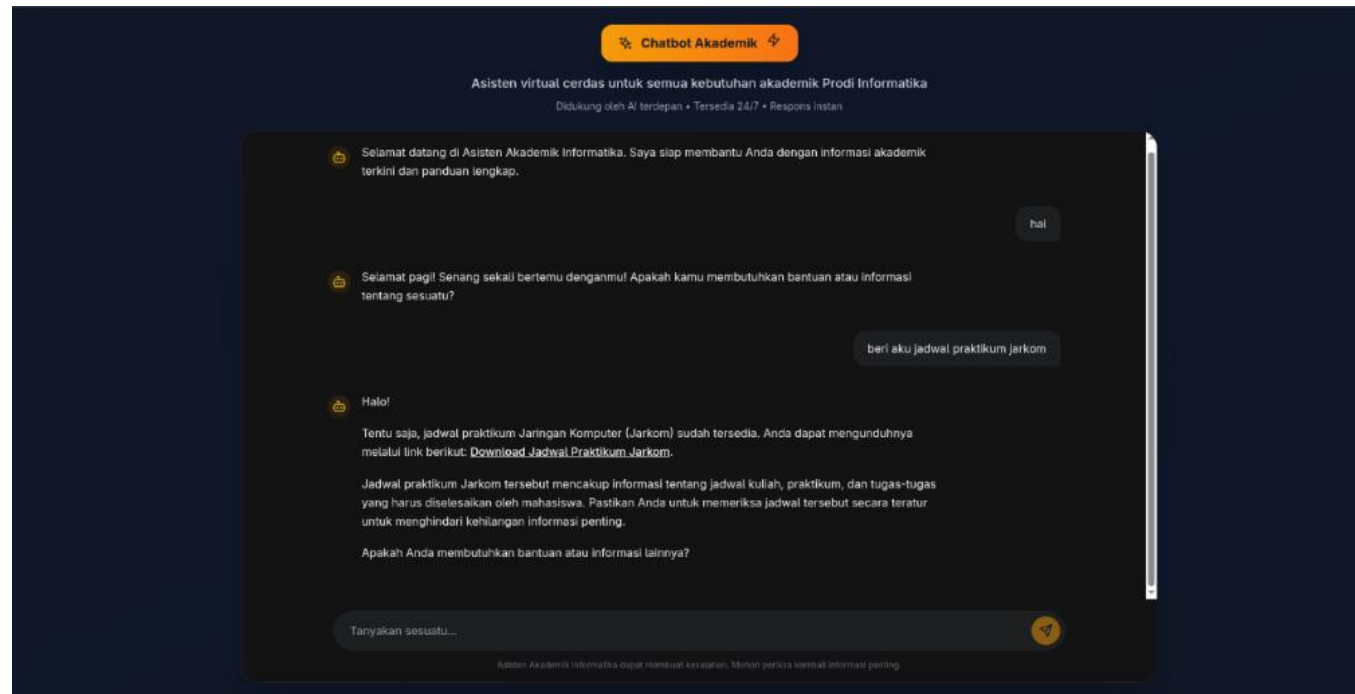
Tahap Pengujian

Pengujian difokuskan pada dua aspek:

1. Kualitas jawaban → dinilai berdasarkan akurasi, relevansi, dan kejelasan dengan pertanyaan seputar informasi program studi
2. Kinerja sistem → Menguji stabilitas : menjalankan kode sederhana yang mengirim pertanyaan secara bersamaan

Hasil & Pembahasan

➤ Hasil Implementasi Antarmuka Chatbot



Gambar 3.1 Implementasi Antarmuka Chatbot pada Website

Hasil & Pembahasan

➤ Hasil Evaluasi Kualitas Jawaban

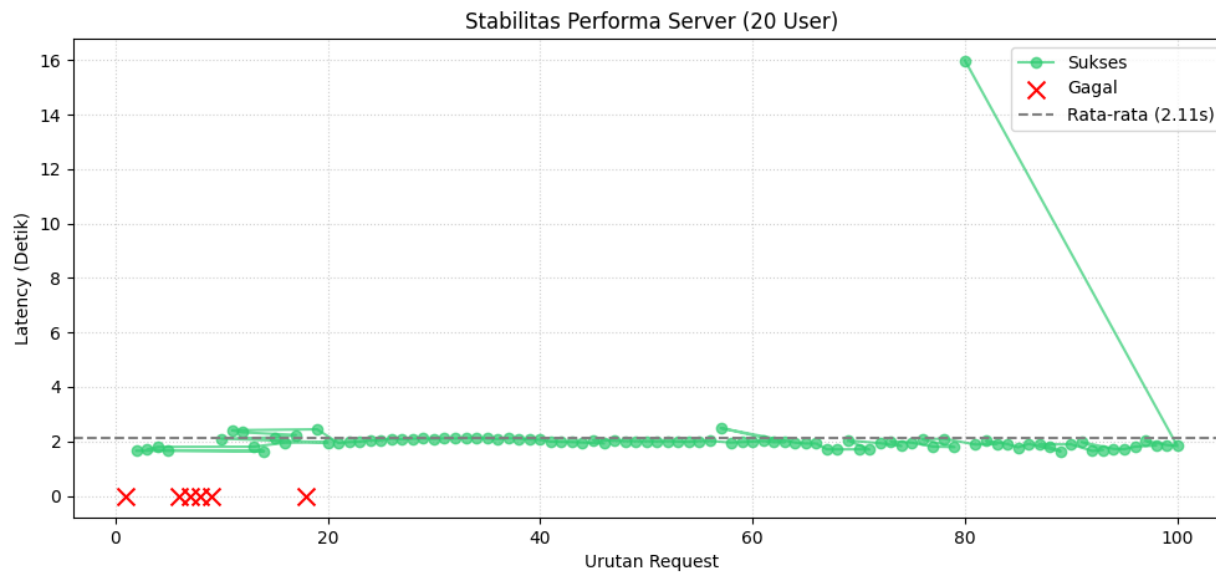
1. Evaluasi jawaban menggunakan kerangka kerja RAGAS
2. Evaluasi mengindikasikan bahwa retriever system sangat efektif dalam mengambil jawaban yang relevan.

Metrik Evaluasi	Nilai
Faithfulness	0.77
Context Precision	0.86
Context Recall	0.75

- Faithfulness = ketepatan jawaban
- Context Precision = relevansi dokumen
- Context Recall = pencarian informasi

Hasil & Pembahasan

➤ Hasil Evaluasi Kinerja Sistem



Parameter	Hasil Pengujian
Total Request	100
Sukses Rate	94% (94 Sukses, 6 Gagal)
Throughput	4.14 request/detik
Rata – Rata Latensi	2.11 detik
P95 (95% User)	2.27 detik
Max Latensi	15.99 detik

Kesimpulan

Berdasarkan penelitian yang telah dilakukan, disimpulkan bahwa chatbot akademik berbasis Retrieval-Augmented Generation (RAG) menggunakan model LLaMA 3.1 melalui Groq API berhasil meningkatkan efisiensi akses informasi mahasiswa dibandingkan situs web statis. Berdasarkan evaluasi, sistem terbukti mampu menyajikan jawaban yang akurat secara kontekstual dengan skor Context Precision mencapai 0.86 dan Faithfulness 0.77. Selain itu, rata-rata latensi respon sebesar 2.11 detik dinilai sangat memadai untuk kebutuhan pencarian informasi umum, memastikan pengguna mendapatkan jawaban tanpa menunggu waktu yang lama. Integrasi arsitektur LangGraph dan basis data vektor ChromaDB pada sistem ini menghasilkan respon yang solutif melalui penyertaan tautan sumber data yang relevan.

Referensi

- 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>
- Djoelianto, A. D., Kautsar, I. A., & Rosid, M. A. (2022). Development of Web Service and Telegram Bot for Location-Based Health Service Information System. Procedia of Engineering and Life Science, 2(2). <https://doi.org/10.21070/pels.v2i2.1280>
- Elysia, S., & Herianto. (2024). Chatbot Berbasis Retrieval Augmented Generation (RAG) untuk Peningkatan Layanan Informasi Sekolah. Journal TIFDA (Technology Information and Data Analytic), 1(2), 52–58. <https://doi.org/10.70491/tifda.v1i2.52>
- Fanani, I. (2025). Implementasi Retrieval Augmented Generation Untuk Evaluasi Proposal Tugas Akhir Mahasiswa. Jurnal Teknologi Komputer Dan Informatika, 3(2). <https://doi.org/10.59820/tekomin.v3i2.336>
- Hidayat, L. R., Pasek, G., Wijaya, S., & Dwiyanaputra, R. (2021). Optimalisasi Layanan Sistem Informasi Mahasiswa dengan Integrasi Telegram : Chatbot Retrieval-Augmented-Generation berbasis Large Language Model (Optimization of Student Information System Services with Telegram Integration: Chatbot Retrieval-Augmented Generation based on Large Language Model). <http://jtika.if.unram.ac.id/index.php/JTIKA/>
- Ichsanudin, I., Pratama, R., & Sisephaputra, B. (2024). Pengembangan Sistem Helpdesk Menggunakan Chatbot Dengan Metode Retrieval-Augmented Generation (RAG). Journal of Informatics and Computer Science, 06.
- Kim, C. Y., Lee, C. P., & Mutlu, B. (2024). Understanding Large-Language Model (LLM)-powered Human-Robot Interaction. ACM/IEEE International Conference on Human-Robot Interaction, Llm, 371–380. <https://doi.org/10.1145/3610977.3634966>
- Milasanti D. (2024). SISTEM CHATBOT BERBASIS LARGE LANGUAGE MODEL (LLM) DAN RETRIEVAL AUGMENTED GENERATION (RAG) PADA ARTIKEL ILMIAH GARUDA KEMDIKBUD.
- Naikar, S., & Indraj, C. (2025). AI-Driven Innovation in Information Retrieval: Transforming Library Operations and User Engagement. International Conference on Global Perspectives on Open Science: Bridging Research, Access and Innovations, May, 1–7. http://papers.ssrn.com/abstract_id=5261229
- Neumann, A. T., Yin, Y., Sowe, S., Decker, S., & Jarke, M. (2025). An LLM-Driven Chatbot in Higher Education for Databases and Information Systems. IEEE Transactions on Education, 68(1), 103–116. <https://doi.org/10.1109/TE.2024.3467912>

Referensi

- Neupane, S., Hossain, E., Keith, J., Tripathi, H., Ghiasi, F., Golilarz, N. A., Amirlatifi, A., Mittal, S., & Rahimi, S. (2024). From Questions to Insightful Answers: Building an Informed Chatbot for University Resources. <http://arxiv.org/abs/2405.08120>
- Nu'aeni H. (2025). PENGEMBANGAN CHATBOT UNTUK SISTEM INFORMASIPERPUSTAKAAN DIGITAL (DIGILIB) MENGGUNAKANRETRIEVAL-AUGMENTED GENERATION (RAG).
- Swacha, J., & Gracel, M. (2025). Retrieval-Augmented Generation (RAG) Chatbots for Education: A Survey of Applications. In Applied Sciences (Switzerland) (Vol. 15, Issue 8). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/app15084234>
- Thakkar, P., & Yadav, A. (2025). Towards LLM Based Multi Agent Systems For Recommendations. July. <https://doi.org/10.13140/RG.2.2.16103.76967>
- Vidivelli, S., Ramachandran, M., & Dharunbalaji, A. (2024). Efficiency-Driven Custom Chatbot Development: Unleashing LangChain, RAG, and Performance-Optimized LLM Fusion. Computers, Materials and Continua, 80(2), 2423–2442. <https://doi.org/10.32604/cmc.2024.054360>
- Wang, J., & Duan, Z. (2024). Agent AI with LangGraph: A Modular Framework for Enhancing Machine Translation Using Large Language Models. <http://arxiv.org/abs/2412.03801>
- Zubaidi, A., Mardiansyah, A. Z., Wedashwara, W., & Jatmika, A. H. (2021). Integrasi Sistem Informasi Akademik dan BOT Telegram Sebagai Media Pengaksesan Informasi di Universitas Mataram. Jurnal Teknologi Informasi, Komputer, Dan Aplikasinya (JTika), 3(2), 251–260. <http://jtika.if.unram.ac.id/index.php/JTIKA/>
- Zulfa, A. A., Ibrahim, T., & Arifudin, O. (2025). Peran Sistem Informasi Akademik Berbasis Web Dalam Upaya Meningkatkan Efektivitas Dan Efisiensi Pengelolaan Akademik Di Perguruan Tinggi. Jurnal Tahsinia, 6(1), 115–134.

Sekian, Terimakasih

