

ABSTRAK

Kebutuhan yang meningkat akan manajemen layanan TI yang efisien di perguruan tinggi mengungkap berbagai keterbatasan pada model helpdesk tradisional yang masih bergantung pada proses manual dan operator manusia. Sistem lama kerap menghadapi waktu respons yang lambat, akumulasi permintaan layanan, serta kelelahan staf. Untuk menjawab tantangan ini, penelitian ini memperkenalkan chatbot cerdas berbasis Large Language Model (LLM) yang mengintegrasikan API DeepSeek-v3 dengan strategi Retrieval-Augmented Generation (RAG), disesuaikan dengan konteks *IT service desk* di salah satu universitas di Indonesia. Solusi ini dirancang menggunakan arsitektur modular, yang secara otomatis mengelola pemasukan dokumen melalui pra-pemrosesan multi-format, *chunking* dinamis, dan pembuatan *embedding*. Data tersebut disimpan di Supabase untuk mendukung pengambilan informasi top-k secara efisien. Ketika pengguna mengajukan pertanyaan, segmen pengetahuan yang relevan diambil secara *real-time* dan diolah menjadi respons yang koheren oleh model bahasa DeepSeek-v3. Sistem dievaluasi melalui pengujian fungsional, benchmarking performa, serta kuesioner *User Experience Questionnaire* (UEQ). Hasil menunjukkan sistem berjalan stabil selama 24/7 dengan akurasi pengambilan informasi lebih dari 98% dan waktu respons rata-rata $2,1 \pm 0,2$ detik, atau 35% lebih cepat dibanding sistem *helpdesk* sebelumnya. Survei terhadap 26 pengguna menunjukkan skor UEQ tinggi ($4,5 \pm 0,3$ dari 5), mencerminkan kepuasan dalam aspek efisiensi, keandalan, dan inovasi. Selain itu, sembilan strategi prompting dibandingkan secara statistik untuk meningkatkan kualitas respons. Hasil penelitian ini menunjukkan potensi dan skalabilitas integrasi LLM dalam layanan TI akademik, serta memberikan panduan praktis untuk implementasi chatbot RAG di lingkungan pendidikan yang terbatas sumber daya.

Kata Kunci: *IT helpdesk, large language models, retrieval-augmented generation, chatbot*