

ABSTRAK

Bimbingan Belajar Griya Sinau Syahir menghadapi kendala promosi digital karena interaksi media sosial rendah dan website *company profile* masih bersifat statis. Di sisi lain, penggunaan *Large Language Model* (LLM) konvensional secara mandiri berisiko memunculkan halusinasi data. Penelitian ini bertujuan mengimplementasikan chatbot berbasis *Retrieval Augmented Generation* (RAG) pada website tersebut. Chatbot dibatasi pada penyampaian informasi seputar profil lembaga, program belajar, biaya, dan prosedur pendaftaran berdasarkan basis pengetahuan berkas JSON. Metode yang digunakan adalah studi kasus dengan pengumpulan data via wawancara dan studi pustaka. Sistem dibangun menggunakan stack Next.js, TypeScript, LangChain.js, model embedding gemini-embedding-001, dan LLM gemini-2.5-flash. Basis pengetahuan dari berkas JSON diurai menjadi 15 chunks menggunakan RecursiveCharacterTextSplitter, dikonversi menjadi vektor 3072 dimensi, dan disimpan dalam Memory VectorStore. Pencarian dokumen relevan menerapkan *Similarity Search* berbasis rumus Cosine Similarity. Kemudian di-evaluasi menggunakan *framework* Promptfoo. Hasil penelitian menunjukkan sistem mampu memilah dokumen relevan berdasarkan skor kemiripan tertinggi yaitu 0,7460 dan 0,6930 menjadi konteks AI. Integrasi Gemini 2.5 Flash menghasilkan jawaban natural yang sesuai data resmi lembaga. Evaluasi otomatis menggunakan Promptfoo mencatatkan pass rate 100%, 3 dari 3 *cases*. Kesimpulannya, chatbot berbasis RAG ini berhasil diimplementasikan dengan andal untuk menyajikan informasi akademik secara cepat, tepat, dan terhindar dari risiko halusinasi data.

Kata kunci: Chatbot, *Retrieval Augmented Generation* (RAG), Website *Company profile*, Google Gemini API

ABSTRACT

Griya Sinau Syahir Tutoring Center faces digital promotion challenges due to low social media engagement and a company profile website that remains static. On the other hand, the independent use of conventional Large Language Models (LLM) carries the risk of data hallucination. This study aims to implement a chatbot based on Retrieval Augmented Generation (RAG) on the website as a solution. The chatbot is limited to delivering information regarding the institution's profile, learning programs, fees, and registration procedures based on a knowledge base derived from JSON files. The method used is a case study with data collection through interviews and literature review. The system was built using the Next.js, TypeScript, and LangChain.js stack, the gemini-embedding-001 embedding model, and the gemini-2.5-flash LLM. The knowledge base from JSON files was parsed into 15 chunks using RecursiveCharacterTextSplitter, converted into 3072-dimensional vectors, and stored in a Memory VectorStore. The relevant document search applies Similarity Search based on the Cosine Similarity formula. Prompt testing was evaluated using the Promptfoo framework. The results show that the system was able to select relevant documents based on the highest similarity scores of 0.7460 and 0.6930 to form the AI context. The integration of Gemini 2.5 Flash produced natural answers consistent with the institution's official data. Automated testing and evaluation using Promptfoo recorded a 100% pass rate, with 3 out of 3 cases passed. In conclusion, this RAG-based chatbot was successfully and reliably implemented to deliver academic information quickly, accurately, and free from the risk of data hallucination.

Keywords: Chatbot, Retrieval Augmented Generation (RAG), Company profile Website, Google Gemini API