

## ABSTRAK

Penelitian ini bertujuan mengembangkan sistem pembuatan *flashcard* otomatis dari dokumen PDF menggunakan arsitektur *Retrieval-Augmented Generation* (RAG) untuk meminimalkan risiko *hallucination* pada *Large Language Models* (LLM). Sistem dirancang khusus untuk materi biologi tingkat SD–SMA, dengan mengintegrasikan Google Gemini 2.5 Flash sebagai *generator* dan FAISS sebagai *vector store* untuk *retrieval* berbasis kemiripan semantik. Dokumen PDF diproses melalui ekstraksi, pembersihan, *chunking*, dan *embedding* menggunakan model multilingual. Evaluasi menggunakan framework RAGAS menunjukkan tingkat *faithfulness* sebesar 89%, membuktikan bahwa konten *flashcard* akurat dan sesuai dengan sumber. Meski demikian, skor *answer relevancy* sebesar 66, akan tetapi sistem tetap memberikan informasi yang tetap relevan dan sesuai konteks. Sistem juga menunjukkan stabilitas kecepatan pemrosesan rata-rata 12,20 detik per topik. *Prototype* aplikasi berbasis Streamlit memungkinkan pengguna menghasilkan dan memverifikasi *flashcard* secara transparan. Penelitian ini membuktikan bahwa RAG efektif dalam menghasilkan materi edukasi faktual dan andal.

Kata kunci: *Artificial Intelligence*, *Flashcards*, *Retrieval-Augmented Generation* (RAG), *Large Language Models*, RAGAS.

## **ABSTRACT**

*This study aims to develop an automated flashcard generation system from PDF documents using a Retrieval-Augmented Generation (RAG) architecture to minimize the risk of hallucination in Large Language Models (LLMs). Specifically designed for biology materials at the elementary to high school levels, the system integrates Google Gemini 2.5 Flash as the generator and FAISS as the vector store for semantic similarity-based retrieval. PDF documents are processed through extraction, cleaning, chunking, and embedding using a multilingual model. Evaluation using the RAGAS framework demonstrated a faithfulness score of 89%, proving that the flashcard content is accurate and grounded in the source material. However, the answer relevancy score was 66%, although the system continued to provide information that remained relevant and contextually appropriate. The system also exhibited stable processing speeds, averaging 12.20 seconds per topic. A Streamlit-based prototype allows users to generate and verify flashcards transparently. This research demonstrates that RAG is effective in producing factual and reliable educational materials.*

*Keywords: Artificial Intelligence, Flashcards, Retrieval-Augmented Generation (RAG), Large Language Models, Ragas Evaluation.*