UNIVERSITAS
GADJAH MADA
Perancangan Sistem Tanya Jawab Dokumen Teks Berbasis Retrieval-Augmented Generation Menggunakan Model Gemini dan OpenAI
Nandha Dhesmawan, Ir. Ardhi Wicaksono Santoso., S.Kom., M.Cs.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

INTISARI

PERANCANGAN SISTEM TANYA JAWAB DOKUMEN TEKS BERBASIS *RETRIEVAL-AUGMENTED GENERATION* MENGGUNAKAN MODEL GEMINI DAN OPENAI

Nandha Dhesmawan

22/494595/SV?20878

Sistem pencarian dokumen hukum pada website JDIH DIY saat ini masih mengandalkan pencarian berbasis kata kunci, sehingga menyulitkan pengguna menemukan pasal berdasarkan substansi isinya secara cepat. Penelitian ini merancang sistem manajemen pengetahuan dokumen hukum DIY berbasis RAG dengan infrastruktur *cloud-native* serta membandingkan kinerja Google Gemini-2.5-flash dan OpenAI GPT-4o-mini berdasarkan *response time*, *token usage*, dan akurasi jawaban. Sistem dibangun menggunakan n8n sebagai pengatur alur kerja, PostgreSQL dengan pgvector sebagai basis data vektor, dan Open WebUI sebagai antarmuka. Pengujian dilakukan terhadap 15 pertanyaan dalam tiga kategori. Hasil menunjukkan Gemini unggul dalam *response time* (8.049 ms lebih cepat dari OpenAI 16.260 ms) dan akurasi (skor 1,4 banding 1,1 dari skala 0-2). OpenAI unggul dalam efisiensi *token* (12.726 *token* banding 13.735 *token*). Berdasarkan hasil tersebut, Google Gemini-2.5-flash merupakan model paling optimal untuk implementasi di lingkungan Pemerintah Daerah DIY karena unggul pada kecepatan dan akurasi, dua aspek terpenting dalam domain hukum.

Kata kunci: *Retrieval-Augmented Generation*, *Large Language Model*, Google Gemini, OpenAI GPT, Dokumen Hukum.

UNIVERSITAS
GADJAH MADA
Perancangan Sistem Tanya Jawab Dokumen Teks Berbasis Retrieval-Augmented Generation Menggunakan Model Gemini dan OpenAI
Nandha Dhesmawan, Ir. Ardhi Wicaksono Santoso., S.Kom., M.Cs.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

ABSTRACT

DESIGN AND DEVELOPMENT OF A RETRIEVAL-AUGMENTED GENERATION-BASED DOCUMENT QUESTION-ANSWERING SYSTEM USING GEMINI AND OPENAI MODELS

Nandha Dhesmawan

22/494595/SV?20878

The legal document search system on the JDIH DIY website currently relies on keyword-based search, making it difficult for users to find specific articles based on substantive content quickly. This research designs a knowledge management system for DIY legal documents based on RAG with cloud-native infrastructure and compares the performance of Google Gemini-2.5-flash and OpenAI GPT-4o-mini based on response time, token usage, and answer accuracy. The system was built using n8n for workflow orchestration, PostgreSQL with pgvector as a vector database, and Open WebUI as an interface. Testing was conducted on 15 questions in three categories. Results show Gemini excels in response time (8.049 ms compared to OpenAI's 16.260 ms) and accuracy (score 1.4 vs 1.1 on a 0-2 scale). OpenAI excels in token efficiency (12,726 tokens compared to 13,735 tokens). Based on these results, Google Gemini-2.5-flash is the most optimal model for implementation in the DIY Regional Government environment as it excels in speed and accuracy—the two most important aspects in the legal domain.

***Keywords**: Retrieval-Augmented Generation, Large Language Model, Google Gemini, OpenAI GPT, Legal Documents*