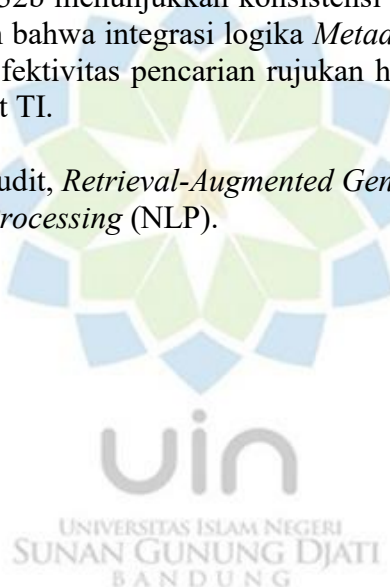


ABSTRAK

Keterbatasan akses terhadap informasi teknis yang akurat dan kontekstual dari dokumen regulasi seperti POJK No. 11/2022 menjadi tantangan operasional signifikan bagi auditor Teknologi Informasi. Proses pencarian leksikal standar (Ctrl+F) terbukti tidak memadai dalam mengatasi kesenjangan leksikal antara bahasa praktisi auditor dengan terminologi kaku regulasi. Penelitian ini bertujuan untuk mengimplementasikan dan mengevaluasi sistem *chatbot* berbasis arsitektur *Retrieval-Augmented Generation* (RAG) yang dioptimasi dengan strategi *Metadata Redirection*. Metodologi penelitian mengikuti kerangka kerja CRISP-DM, di mana dokumen regulasi diolah melalui *advanced chunking* level ayat dan *metadata injection*. Penelitian ini melakukan *ablation study* terhadap 12 skenario eksperimen yang membandingkan performa tiga model LLM (qwen3-32b, llama-3.3-70b, dan gpt-oss-20b), variasi *threshold* jarak semantik, serta perbandingan dataset *Hybrid* dan *Regulasi Only*. Kinerja sistem dievaluasi menggunakan metrik *Information Retrieval* (*Precision@1*, *Recall@3*, *MRR*) dan evaluasi generatif berbasis kerangka RAGAS (*Faithfulness* dan *Answer Relevancy*) dengan llama-3.1-8B Instant sebagai *AI Judge*. Hasil eksperimen menunjukkan bahwa komponen *retriever* mencapai akurasi optimal pada kondisi dataset *Regulasi Only* dengan *threshold* 5.5, menghasilkan *Precision@1* sebesar 58,62%, *Recall@3* 65,52% dan F-1 Score 73,68%. Pada sisi generator, model qwen3-32b menunjukkan konsistensi tertinggi dalam menjaga relevansi jawaban. Temuan ini membuktikan bahwa integrasi logika *Metadata Redirection* dan rekayasa data yang presisi dapat meningkatkan efektivitas pencarian rujukan hukum sekaligus memitigasi risiko halusinasi pada sistem chatbot audit TI.

Kata Kunci: *Chatbot*, Regulasi Audit, *Retrieval-Augmented Generation* (RAG), *Natural Language Processing* (NLP).



ABSTRACT

The limited access to accurate and contextual technical information from regulatory documents, such as POJK No. 11/2022, poses a significant operational challenge for Information Technology (IT) auditors. Standard lexical search processes (Ctrl+F) have proven inadequate in bridging the lexical gap between the practitioners' language and rigid regulatory terminology. This research aims to implement and evaluate a chatbot system based on the Retrieval-Augmented Generation (RAG) architecture, optimized with a Metadata Redirection strategy. The research methodology follows the CRISP-DM framework, where regulatory documents are processed through advanced verse-level chunking and metadata injection. An ablation study was conducted across 12 experimental scenarios, comparing the performance of three Large Language Models (LLM) qwen3-32b, llama-3.3-70b, and gpt-oss-20b alongside variations in semantic distance thresholds and a comparison between Hybrid and Regulation-Only datasets. System performance was evaluated using Information Retrieval metrics (Precision@1, Recall@3, MRR) and generative evaluation based on the RAGAS framework (Faithfulness and Answer Relevancy) with llama-3.1-8B Instant serving as the AI Judge. Experimental results indicate that the retriever component achieves optimal accuracy under the Regulation-Only dataset condition with a 5.5 threshold, yielding a Precision@1 of 58.62%, a Recall@3 of 65.52%, and an F-1 Score of 73.68%. On the generator side, the qwen3-32b model exhibits the highest consistency in maintaining answer relevancy. These findings prove that the integration of Metadata Redirection logic and precise data engineering can enhance the effectiveness of legal reference retrieval while mitigating hallucination risks in IT audit chatbot systems.

Keywords: Chatbot, Audit Regulation, Retrieval-Augmented Generation (RAG), Natural Language Processing (NLP)

