

ABSTRAK

Penelitian ini bertujuan mengembangkan *chatbot* AI berbasis *Large Language Model* (LLM) dengan integrasi *LangChain* dan *Retrieval-Augmented Generation* (RAG) untuk meningkatkan layanan pelanggan di PT Telemedia Prima Nusantara. Metode pengembangan menggunakan CRISP-DM, meliputi tahapan pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan deployment. Metode pengumpulan data dilakukan melalui observasi proses layanan, wawancara dengan tim internal, analisis dokumen FAQ, dokumentasi alur pengiriman invoice dan status pembayaran dari sistem Mixradius, riwayat percakapan pelanggan di WhatsApp, serta dokumen teknis internal lainnya. Data tambahan dikumpulkan melalui feedback pengguna internal saat evaluasi sistem dan eksperimen chatbot berbasis interaksi langsung, menggunakan metrik BLEU, ROUGE, dan waktu respons untuk pengukuran performa. Metode pemecahan masalah menggunakan pendekatan *Retrieval-Augmented Generation* (RAG) yang diorkestrasi dengan *framework LangChain*. Informasi dikumpulkan dari basis data vektor menggunakan FAISS, kemudian diolah oleh GPT-3.5 Turbo untuk menghasilkan jawaban. Komponen Skill Router diterapkan untuk klasifikasi intensi dan pengaturan jalur pemrosesan. Hasil evaluasi menunjukkan chatbot mampu memberikan jawaban dengan akurasi 93,33%, BLEU-2 sebesar 0,518, ROUGE-L sebesar 0,683, dan waktu respons rata-rata 1,55 detik. Hasil penelitian ini menunjukkan sistem terbukti efektif dalam menjawab pertanyaan pelanggan secara otomatis, efisien, dan kontekstual. Rekomendasi pengembangan selanjutnya mencakup perluasan basis pengetahuan, deteksi intensi otomatis, eskalasi ke agen manusia, integrasi monitoring, serta optimalisasi *embedding* dan *vector store* alternatif.

Kata Kunci: *Chatbot AI, Large Language Model (LLM), LangChain, Retrieval-Augmented Generation (RAG), Customer Service, CRISP-DM.*

ABSTRACT

This research aims to develop an AI chatbot based on Large Language Model (LLM) with LangChain and Retrieval-Augmented Generation (RAG) integration to improve customer service at PT Telemedia Prima Nusantara. The development method uses CRISP-DM, which includes the stages of business understanding, data understanding, data preparation, modeling, evaluation, and deployment. Data collection methods were carried out through observation of service processes, interviews with internal teams, analysis of FAQ documents, documentation of invoice delivery and payment status from the Mixradius system, customer conversation history on WhatsApp, and other internal technical documents. Additional data was collected through internal user feedback during system evaluation and experiments with a chatbot based on direct interaction, using BLEU, ROUGE, and response time metrics for performance measurement. The problem-solving method employed the Retrieval-Augmented Generation (RAG) approach, orchestrated using the LangChain framework. Information was collected from the vector database using FAISS, then processed by GPT-3.5 Turbo to generate answers. The Skill Router component was applied for intent classification and processing path configuration. Evaluation results show that the chatbot can provide answers with 93.33% accuracy, a BLEU-2 score of 0.518, a ROUGE-L score of 0.683, and an average response time of 1.55 seconds. These findings demonstrate that the system is effective in automatically, efficiently, and contextually answering customer questions. Further development recommendations include expanding the knowledge base, automatic intent detection, escalation to human agents, monitoring integration, and optimization of alternative embedding and vector stores.

Keywords: *AI Chatbot, Large Language Model (LLM), LangChain, Retrieval-Augmented Generation (RAG), Customer Service, CRISP-DM.*