

ABSTRACT

With the rapid advancement of information and communication technology in the digital era, the global business landscape has undergone significant changes, including a major shift toward digital transformation through e-commerce platforms. However, as the number of products and the demand for accurate information from customers increase, new challenges arise in providing responsive and relevant customer service. Systems are required to quickly and accurately respond to customer inquiries. Therefore, AI-based chatbot technology is increasingly being adopted to address these challenges. This research proposes the development of an e-commerce chatbot using a Retrieval-Augmented Generation (RAG) approach, which combines the capabilities of a Large Language Model (LLM) with a vector-based document retrieval system. The LLM model used is Llama-3.3-70B-Instruct, which is enriched with relevant information through semantic search on a knowledge base stored in a vector storage system in the form of FAISS. With this approach, the chatbot can provide answers based on up-to-date factual data without requiring fine-tuning, while also minimizing the risk of producing assumptive or speculative responses. The results of this research show that integrating LLM and RAG significantly improves customer service efficiency in e-commerce platforms. This is proven through evaluations using performance metrics with the results of a fairly high metrics score, indicating that the chatbot is capable of delivering accurate and relevant answers according to user needs.

Keywords: e-commerce, chatbot, Large Language Model, Retrieval-Augmented Generation, LLaMA, FAISS, dense retrieval, performance metrics