

Cloud Era: An AI-driven multilingual chatbot for intelligent multi-cloud resource guidance

P. M. Kandamby, V. P. Wattededara, N. N. Thilakarathne

*Department of Information and Communication Technology, Faculty of Technology,
University of Colombo, Sri Lanka*

Managing multi-cloud setups can feel overwhelming, with tangled configurations, scattered docs, and little multilingual support. Standard Retrieval-Augmented Generation (RAG) tools often miss personal touches and connections, making them less helpful. In this regard, this research proposed Cloud Era, an AI-powered chatbot that delivers smart, tailored advice for handling cloud resources. Our goals were to achieve over 80% troubleshooting and setup success compared to basic RAG, boost recall by more than 40% with mixed retrieval, and cut hallucinations by over 70% using agentic flows, all while supporting English and Sinhala. Built on an Agentic RAG framework inspired by LightRAG, it blends semantic vector searches with dynamic knowledge graphs for better precision than simple RAG. Each agent step—clarifying domains, refining and breaking down questions, reevaluating, checking web relevance via LLM, and reranking—runs through LLM API calls. Validation used an LLM-as-judge approach, comparing generated responses with expert answers to measure accuracy. The setup ran locally on a basic GPT model, with prompts tied to source data, providing detailed explanations and resource links. Testing with 10 users showed 87% satisfaction. For multilingual checks, 120 queries across English, Sinhala, and mixed languages confirmed strong retrieval, including 81% recall for Sinhala. Compared to traditional RAG, our graph-RAG hybrid excelled. Local hosting keeps it quick and accessible, though multi-step reasoning can still cause slips, partly eased by intent checks and hybrid pulls. In sum, Cloud Era advances multi-cloud help by weaving agentic RAG, hybrid retrieval, and inclusivity, bringing clear wins for diverse users. Drawbacks include reliance on quality data and limited real-time API ties, while future work aims to add languages and sharpen reasoning.

Keywords: *Agentic RAG, Hybrid retrieval, Knowledge graphs, AI chatbot, Multilingual support*