

Sistem *question answering* berbasis *Large Language Model* memiliki potensi besar untuk mendukung perolehan informasi bencana yang cepat dan akurat. Penelitian ini bertujuan untuk merancang dan mengevaluasi sebuah AI Agent yang mampu mengorkestrasi beberapa fungsi (*tools*) spesifik dalam konteks kebencanaan, meliputi tanya jawab pengetahuan dengan *Retrieval Augmented Generation* (RAG) pada *knowledge base megathrust*, penyajian data gempa bumi terkini, serta manajemen laporan pengguna (mengirim, memverifikasi, dan melihat laporan bencana). Berdasarkan implementasi sistem tersebut, evaluasi orkestrasi dilakukan menggunakan metrik Akurasi, Presisi, Recall, dan F1-Score untuk mengukur kemampuan agent dalam memilih dan menggunakan *tools* yang tepat. Hasil evaluasi menunjukkan performa agen yang menjanjikan, dengan F1-Score tertinggi dicapai oleh *tool megathrust knowledge* (0.943) dan *latest earthquake* (0.926), menandakan kemampuan yang sangat baik dalam tugas pencarian informasi kebencanaan. Namun, tantangan signifikan ditemukan pada *tool send report* yang memiliki presisi rendah (0.550) akibat tingginya *false positive* dalam deteksi kebutuhan pelaporan. Kelemahan utama teridentifikasi pada kasus *no tools* (tanpa pemanggilan *tool*) yang memiliki recall terendah (0.471), mengindikasikan kecenderungan agent untuk terlalu proaktif dalam memanggil *tools* bahkan ketika tidak diperlukan. Analisis tambahan menunjukkan bahwa optimalisasi parameter RAG (seperti ukuran *chunk*) dan efisiensi penggunaan token merupakan pertimbangan penting untuk meningkatkan performa sistem. Meskipun arsitektur *AI Agent* yang dikembangkan terbukti efektif dalam menjawab pertanyaan kebencanaan, sistem masih memerlukan perbaikan pada mekanisme konfirmasi pengguna, terutama dalam alur pelaporan bencana yang membutuhkan validasi lebih ketat. Penelitian ini memberikan kontribusi dalam pengembangan sistem AI yang dapat mendukung mitigasi bencana melalui penyediaan informasi yang akurat dan responsif, sekaligus mengidentifikasi area-area yang memerlukan pengembangan lebih lanjut untuk implementasi praktis di lapangan.

Kata kunci : *megathrust, question answering system, ai agent, retrieval augmented generation*

ABSTRACT

Question answering systems based on Large Language Models (LLM) with AI Agent as orchestrator have great potential to support rapid and accurate disaster information acquisition. This research aims to design and evaluate an AI Agent capable of orchestrating several specific functions (tools) in disaster contexts, including knowledge question answering with Retrieval Augmented Generation (RAG) on megathrust knowledge base, presenting current earthquake data, and user report management (sending, verifying, and viewing disaster reports). Based on the system implementation, orchestration evaluation was conducted using Accuracy, Precision, Recall, and F1-Score metrics to measure the agent's ability to select and use appropriate tools. The evaluation results show promising agent performance, with the highest F1-Score achieved by megathrust knowledge tool (0.943) and latest earthquake (0.926), indicating excellent capability in disaster information retrieval tasks. However, significant challenges were found in the send report tool which has low precision (0.550) due to high false positives in detecting reporting needs. The main weakness was identified in no tools cases (without tool invocation) which had the lowest recall (0.471), indicating the agent's tendency to be overly proactive in calling tools even when unnecessary. Additional analysis shows that optimizing RAG parameters (such as chunk size) and token usage efficiency are important considerations for improving system performance. Although the developed AI Agent architecture proved effective in answering disaster-related questions, the system still requires improvements in user confirmation mechanisms, particularly in disaster reporting workflows that need stricter validation. This research contributes to the development of AI systems that can support disaster mitigation through accurate and responsive information provision, while identifying areas that require further development for practical field implementation.

Keywords : *megathrust, question answering system, ai agent, retrieval augmented generation*