

BAB 5 SIMPULAN DAN SARAN

5.1 Simpulan

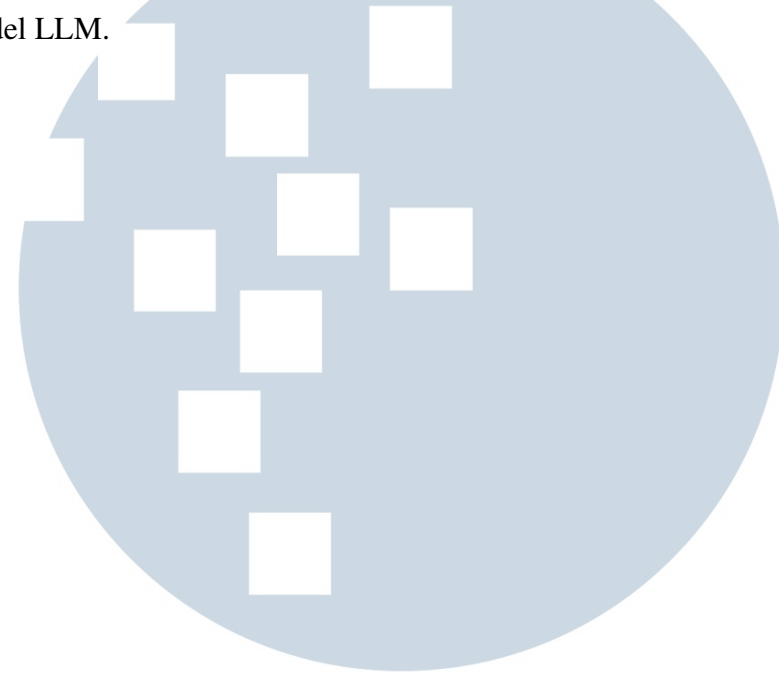
1. Dari penelitian yang dilakukan, ditunjukkan bahwa *Large Language Model* berbasis *Vector Database* yang diimplementasikan menggunakan teknik *Retrieval Augmented Generation* dapat digunakan untuk memberikan jawaban yang relevan dan menutupi kekurangan model GPT 3.5 yang memiliki batasan basis pengetahuan dimana model GPT 3.5 yang hanya memiliki pengetahuan sampai Januari 2022, setelah diimplementasikan teknik *Retrieval Augmented Generation*, LLM dapat menjawab berdasarkan dataset yang sudah disediakan dalam *vector database*. Hal tersebut membuktikan bahwa teknik *Retrieval Augmented Generation* dapat mengurangi sifat halusinasi dari *Large Language Model*.
2. Berdasarkan evaluasi yang dilakukan, model mendapatkan nilai akurasi sebesar 86.667% yang membuktikan bahwa model berhasil dikembangkan dengan baik dan memiliki akurasi yang cukup tinggi.

5.2 Saran

Selama penelitian ini, terdapat beberapa catatan-catatan penting yang dapat dikembangkan pada penelitian selanjutnya.

1. Arsitektur sistem model yang dikembangkan belum bisa menyimpan hasil jawaban-jawaban maupun pertanyaan-pertanyaan yang diajukan sebelumnya, sehingga komunikasi tidak berjalan secara mengalir dengan mengikuti konteks pertanyaan sebelumnya. Dalam penelitian selanjutnya dapat dikembangkan arsitektur sistem yang memungkinkan terjadinya kelanjutan percakapan sesuai dengan konteks percakapan sebelumnya.
2. Analisa perbandingan lanjutan untuk parameter LLM Model dapat dibandingkan lebih lanjut dengan LLM Model lainnya seperti GPT-4 dengan konfigurasi arsitektur aplikasi yang sama karena dalam penelitian ini tidak dilakukan pengetesan terhadap model LLM yang berbeda. Penggunaan GPT-4 sebagai LLM Model dalam aplikasi RAG diharapkan dapat meningkatkan

akurasi karena GPT-4 memiliki tingkat *reasoning* yang lebih bagus dari GPT 3.5 Turbo yang digunakan dalam penelitian ini yang dapat dilihat dari perbandingan jurnal-jurnal yang telah dipaparkan sebelumnya yang mendapatkan akurasi yang cukup tinggi karena menggunakan GPT-4 sebagai Model LLM.



UMMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA