

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian, pengembangan, dan evaluasi yang telah dilakukan terhadap asisten virtual berbasis *Retrieval-Augmented Generation (RAG)* pada website Siwur, dapat ditarik beberapa kesimpulan, yaitu:

- a. Penelitian ini berhasil mengembangkan asisten virtual menggunakan pendekatan arsitektur *microservices* yang memisahkan layanan *frontend* dan *backend*. Sisi antarmuka pengguna berhasil diintegrasikan ke dalam ekosistem website Siwur yang berbasis TALL Stack (Tailwind CSS, Alpine.js, Laravel, Livewire), sementara layanan kecerdasan buatan (AI Service) dibangun secara terpisah menggunakan Python (FastAPI).
- b. Integrasi arsitektur RAG dengan Google Gemini Flash 2.0 sebagai LLM dan ChromaDB sebagai *vector store* terbukti mampu menyediakan jawaban yang kontekstual berdasarkan basis pengetahuan spesifik Siwur. Berdasarkan hasil evaluasi kinerja menggunakan kerangka kerja RAGAS terhadap 30 pertanyaan uji, sistem menunjukkan performa sebagai berikut:
 - 1) *Faithfulness* (0.7401): Sebesar 74,01% klaim dalam jawaban yang dihasilkan konsisten dan didukung oleh fakta dari dokumen sumber, mengindikasikan kemampuan sistem dalam meminimalkan halusinasi.
 - 2) *Answer Relevancy* (0.8053): Sebesar 70,53% jawaban yang diberikan dinilai relevan dengan pertanyaan pengguna.

- 3) *Context Precision* (0.4279) dan *Context Recall* (0.4624): Sistem mampu mengambil sekitar 42,79% dokumen yang sangat relevan di peringkat teratas dan menemukan sekitar 46,24% dari total konteks penting yang tersedia. Meskipun masih memiliki ruang untuk optimasi pada sisi retrieval, sistem sudah mampu memberikan jawaban yang cukup akurat untuk kebutuhan dasar.

Selain itu, evaluasi penerimaan pengguna *melalui User Acceptance Test (UAT)* menghasilkan skor rata-rata 83% dengan kategori "Sangat Baik". Hal ini menunjukkan bahwa asisten virtual ini dapat diterima dengan baik oleh pengguna akhir dan membantu efisiensi operasional toko.

5.2 Saran

Berdasarkan hasil evaluasi dan keterbatasan yang ditemukan selama penelitian, terdapat beberapa saran untuk pengembangan sistem selanjutnya:

- a. Mengingat nilai *Context Precision* dan *Context Recall* yang masih di bawah 0.5, disarankan untuk menerapkan teknik pencarian hibrida (*hybrid search*) yang menggabungkan pencarian semantik (vektor) dengan pencarian kata kunci (keyword search/BM25). Selain itu, penggunaan teknik Re-ranking (pengurutan ulang) pada dokumen hasil pencarian dapat meningkatkan relevansi konteks yang diberikan kepada LLM.
- b. Perlu dilakukan eksplorasi lebih lanjut pada strategi pemecahan dokumen (*chunking strategy*). Penggunaan semantic chunking atau pemecahan berdasarkan struktur dokumen yang lebih rapi dapat membantu menjaga

keutuhan konteks informasi, sehingga meningkatkan akurasi pengambilan data.

