

BAB V

SIMPULAN SARAN

5.1 Simpulan

Penelitian ini berhasil mengembangkan sebuah chatbot multimodal berbasis *Vision AI*, *Optical Character Recognition (OCR)*, *Sentence Transformer*, dan *Small Language Model (SLM)* yang mampu memahami gambar edukatif serta menghasilkan jawaban berbasis metadata visual secara akurat. Melalui tahapan pengembangan yang dirancang menggunakan pendekatan *Human-Centered Design (HCD)*, penelitian ini menunjukkan bahwa model *vision open-source* seperti BLIP-Base dan OCR mampu mengekstraksi informasi visual secara efektif dalam bentuk caption, daftar objek, dan teks yang merepresentasikan isi gambar. Hasil pengujian *embedding* menggunakan *Sentence Transformer* juga membuktikan bahwa representasi vektor yang dihasilkan bersifat stabil dan efektif untuk mendukung pencarian berbasis kemiripan dalam mekanisme *Retrieval-Augmented Generation (RAG)*, sehingga sistem mampu menemukan metadata visual yang paling relevan terhadap setiap pertanyaan pengguna.

Ketika metadata hasil retrieval tersebut digunakan sebagai konteks oleh *Small Language Model*, chatbot mampu menghasilkan jawaban edukatif yang relevan, aman, dan sesuai dengan konteks gambar. Pendekatan RAG terbukti berperan penting dalam membatasi penalaran model bahasa agar seluruh respons yang dihasilkan berlandaskan data visual yang tervalidasi, sehingga *risiko hallucination* dapat ditekan secara signifikan. Dengan demikian, penelitian ini berhasil menjawab seluruh rumusan masalah dan mencapai tujuan penelitian, yaitu mengembangkan sistem yang tidak hanya mampu menghasilkan visual edukatif, tetapi juga memahami, menjelaskan, dan mendukung interaksi pembelajaran berbasis gambar secara aman dan kontekstual.

Luaran utama dari penelitian ini adalah sebuah prototipe chatbot multimodal edukatif bernama “Si Atang” yang dirancang sebagai pendamping belajar berbasis

visual bagi guru dan siswa sekolah dasar di Desa Wisata Tigaraksa. Prototipe ini menjembatani kesenjangan interpretasi, interaksi, dan kepercayaan (*interpretation, interaction, and trust gap*) yang sebelumnya muncul dari penggunaan gambar statis hasil penelitian terdahulu. Melalui integrasi *pipeline Vision-to-Language* yang kokoh, sistem ini memanfaatkan metadata visual dengan tingkat kelengkapan mencapai 0,97 serta akurasi *retrieval (Precision@1)* sebesar 0,60 sebagai dasar *reasoning* yang faktual dan terkontrol.

Selain luaran berupa prototipe sistem, penelitian ini juga menghasilkan kontribusi praktis berupa pendekatan *pipeline multimodal* berbasis model open-source yang ringan dan dapat direplikasi pada konteks pembelajaran serupa. Tingginya tingkat penerimaan guru dengan skor rata-rata 4,93 dari 5,00 menunjukkan bahwa sistem yang dikembangkan dinilai bermanfaat, mudah digunakan, dan relevan dengan kebutuhan pembelajaran berbasis komunitas desa. Secara keseluruhan, hasil penelitian ini menegaskan bahwa integrasi Vision AI, RAG, dan SLM merupakan solusi yang efektif untuk menyediakan pendamping digital edukatif yang kontekstual, aman, dan sesuai bagi guru maupun anak-anak dalam lingkungan pembelajaran berbasis komunitas.

5.2 Saran

Berdasarkan hasil penelitian serta keterbatasan yang ditemukan, berikut beberapa saran yang dapat dijadikan acuan untuk pengembangan praktis maupun kontribusi akademis di masa depan.

5.2.1 Saran Praktis

1. Pengelola desa dan guru disarankan mengikuti pelatihan dasar *prompt engineering*, terutama *instruction-based prompt*, agar mampu menghasilkan konten visual edukatif secara mandiri dan berkelanjutan.
2. Gambar hasil generatif sebaiknya dimanfaatkan dalam kegiatan belajar mengajar dan promosi wisata edukatif, misalnya untuk menjelaskan siklus

tanaman, ekosistem lokal, atau alat pertanian desa agar pembelajaran lebih menarik dan kontekstual.

3. Diperlukan kerja sama antara universitas, pemerintah daerah, dan komunitas desa agar implementasi teknologi AI dapat berlanjut dan menjadi bagian dari strategi digitalisasi Desa Wisata Tigaraksa.

5.2.2 Saran Akademis

1. Penelitian lanjutan dapat menambahkan model generatif lain seperti *Stable Diffusion XL* atau DALL·E 4 dan menggunakan metrik tambahan seperti FID atau *KID Score* untuk analisis kualitas visual yang lebih komprehensif.
2. Penilaian di masa depan disarankan melibatkan lebih banyak siswa, guru, dan masyarakat desa agar evaluasi lebih representatif terhadap persepsi manusia dan kualitas visual edukatif.
3. Peneliti berikutnya dapat menghubungkan Supabase dan chatbot multimodal dengan LMS untuk mendukung pembelajaran kontekstual berbasis *Community-Based Tourism* (CBT).
4. Diperlukan penelitian lanjutan untuk menilai dampak penggunaan konten AI terhadap pemahaman konsep dan efektivitas pembelajaran di sekolah.