BAB III

METODE PENELITIAN

3.1 Metode Penelitian

Penelitian ini memilih *Human-Centered Design* (HCD) sebagai metode utama karena pengembangan chatbot multimodal berbasis *Vision AI* dan *Retrieval-Augmented Generation* (RAG) tidak hanya ditentukan oleh ketepatan teknologinya, tetapi juga oleh sejauh mana sistem tersebut dapat dipahami, digunakan, dan memberikan manfaat nyata bagi guru serta siswa di Desa Wisata Tigaraksa. Pendekatan HCD menempatkan manusia sebagai pusat proses perancangan, sehingga seluruh keputusan desain mulai dari pemilihan fitur *Vision AI*, pemanfaatan hasil *Optical Character Recognition* (OCR), strategi *retrieval* data, hingga bentuk interaksi dengan chatbot dibangun berdasarkan pemahaman yang mendalam terhadap kebutuhan dan karakteristik pengguna akhir. Melalui HCD, perancangan sistem mempertimbangkan konteks lokal, tingkat literasi digital, serta kebiasaan belajar guru dan siswa, sehingga teknologi yang dikembangkan tidak bersifat abstrak atau terlalu kompleks. Pendekatan ini memastikan bahwa sistem yang dihasilkan aman digunakan oleh anak-anak, memiliki alur interaksi yang intuitif, serta selaras dengan nilai dan kebutuhan pembelajaran di lingkungan desa. Selain itu, HCD memungkinkan proses iteratif yang melibatkan umpan balik pengguna, sehingga sistem dapat disesuaikan secara berkelanjutan berdasarkan pengalaman nyata di lapangan. Dengan demikian, HCD menjadi kerangka metodologis yang paling tepat untuk menjembatani kompleksitas teknologi AI dengan kebutuhan pengguna di tingkat akar rumput. Pendekatan ini memastikan bahwa chatbot multimodal yang dikembangkan tidak hanya canggih secara teknis, tetapi juga relevan, mudah diakses, dan mampu memberikan dampak signifikan bagi penguatan pembelajaran berbasis visual di Desa Wisata Tigaraksa.

Tabel 3.1 Perbandingan Metode Penelitian

Aspek	HCD (Human-Centered Design)	UCD (User-Centered Design)	Design Thinking
Fokus Utama	Kebutuhan, konteks dan perilaku manusia secara menyeluruh (holistik)	Kebutuhan dan kenyamanan pengguna saat menggunakan produk	Inovasi, kreativitas, dan eksplorasi solusi
Orientasi	Pengalaman dan kemampuan manusia dalam menggunakan sistem	Usability dan efektivitas penggunaan	Pemecahan masalah secara kreatif
Keterlibatan Pengguna	Sangat tinggi dan berkelanjutan dari awal sampai akhir	Tinggi pada tahap desain dan evaluasi	Sedang–tinggi, terutama pada tahap Empathy dan Prototype
Sifat Proses	Iteratif berdasarkan feedback nyata pengguna	Iteratif dengan fokus pada peningkatan usability	Iteratif, eksploratif, mendorong ide kreatif
Konteks Penggunaan	Dipakai saat sistem harus benar-benar sesuai kebutuhan pengguna akhir, termasuk konteks sosial	Cocok untuk produk digital yang menekankan kemudahan penggunaan	Cocok untuk tahap ideasi di proyek inovatif dan produk baru
Kelebihan	Menghasilkan solusi yang sangat relevan, aman, dan mudah dipahami	Menjamin produk mudah digunakan dan tidak membingungkan	Menghasilkan ide-ide kreatif dan keluar dari pola lama
Kelemahan	Membutuhkan interaksi intens dan waktu lebih	Kadang terlalu fokus pada usability	Bisa terlalu abstrak dan kurang teknis untuk implementasi AI
Kesesuaian untuk Chatbot Pendidikan	Sangat cocok karena mempertimbangkan kemampuan anak-anak, konteks lokal, dan alur belajar	Cocok, tapi kurang memperhatikan aspek sosial dan konteks desa	Cocok untuk eksplorasi konsep awal chatbot, kurang kuat untuk implementasi jangka panjang

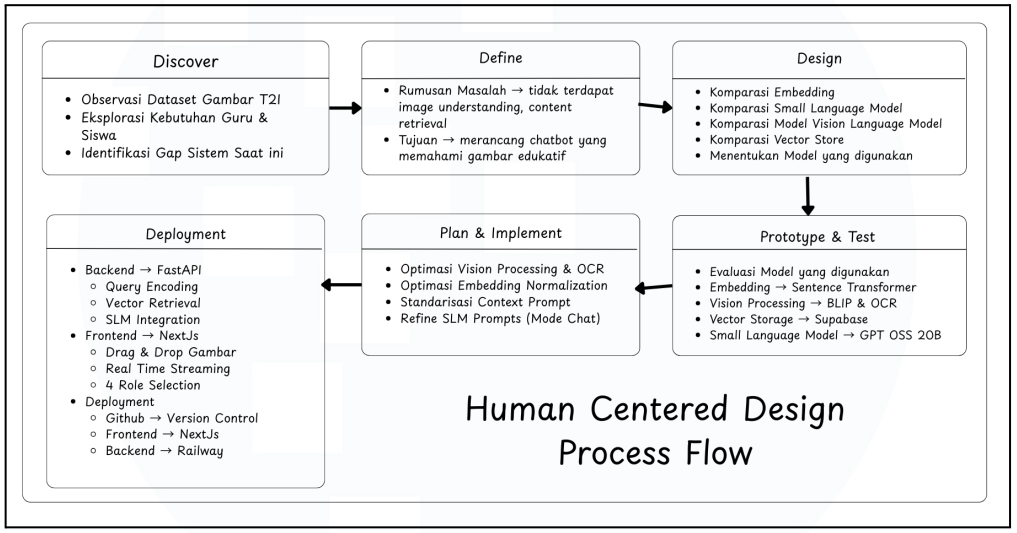
Proses HCD pada penelitian ini mengikuti lima tahap utama yaitu Discover, Define, Design, Prototype & Test, terakhir Plan & Implement. Tahap *Discover* digunakan untuk menggali kebutuhan pengguna di lingkungan belajar desa. Selanjutnya, tahap *Define* merumuskan inti permasalahan yang perlu diselesaikan melalui sistem. Tahap *Design* menghasilkan ide serta rancangan awal chatbot yang sesuai dengan kebutuhan edukatif dan konteks lokal. Pada tahap *Prototype & Test*, sistem awal dibangun dan diuji langsung kepada pengguna untuk menilai kejelasan, kemudahan penggunaan dan efektivitas fitur. Terakhir, tahap *Plan & Implement* memastikan hasil pengembangan dapat diimplementasikan secara lebih matang dan siap digunakan dalam skenario pembelajaran.

Pemilihan Human-Centered Design (HCD) dalam penelitian ini didasarkan pada kebutuhan untuk mengembangkan sistem AI yang tidak hanya berfungsi secara teknis, tetapi juga sesuai dengan konteks sosial, pedagogis dan kognitif pengguna. Dibandingkan pendekatan lain, HCD memberikan kerangka yang lebih holistik karena mempertimbangkan kebutuhan pengguna, pola interaksi serta dampak sistem terhadap proses belajar. Dalam konteks pembelajaran berbasis desa dengan tingkat literasi digital yang beragam, HCD memungkinkan sistem dirancang agar aman, mudah digunakan dan relevan bagi guru maupun siswa. Secara analogi, penerapan HCD dapat diibaratkan seperti membuat sepatu khusus untuk seorang pelari yang tidak cukup hanya kuat secara material, tetapi harus disesuaikan dengan bentuk kaki dan medan yang dilalui. Dengan cara ini, sistem yang dikembangkan tidak hanya canggih secara teknis tetapi juga efektif dan nyaman digunakan dalam praktik pembelajaran.

Melalui pendekatan HCD ini, penelitian tidak hanya menghasilkan solusi berbasis teknologi, tetapi juga memastikan bahwa chatbot yang dikembangkan benar-benar usable, aman digunakan anak-anak dan memberikan manfaat nyata bagi proses belajar di Desa Wisata Tigaraksa.

3.2 Tahapan Penelitian

Tahapan penelitian ini disusun mengikuti alur Human-Centered Design agar sistem yang dibangun benar-benar sesuai dengan kebutuhan guru dan siswa. Alur penelitian digambarkan sebagai berikut:



Gambar 3.1 Diagram Alur Penelitian

3.2.1 Discover

Tahap Discover merupakan tahap awal untuk memahami secara mendalam kebutuhan dan permasalahan nyata dalam proses pembelajaran berbasis gambar di Desa Wisata Tigaraksa. Pada tahap ini, peneliti melakukan observasi terhadap pengurus desa dan kumpulan gambar edukatif yang sebelumnya dihasilkan melalui penelitian *Text-to-Image (T2I)* dan dianalisis kembali melalui eksplorasi awal di *Google Collab*. Melalui analisis awal tersebut, peneliti menemukan bahwa meskipun gambar edukatif sudah tersedia, tetapi guru tetap kesulitan memberikan penjelasan yang terstruktur, sementara siswa membutuhkan penjelasan visual yang sederhana dan mudah dipahami. Temuan ini memberikan gambaran menyeluruh mengenai tantangan pedagogis, keterbatasan metadata visual pada gambar T2I, serta kebutuhan teknologi yang harus dijawab oleh sistem chatbot multimodal.

3.2.2 Define

Tahap Define merupakan tahap perumusan masalah inti berdasarkan temuan pada tahap Discover. Dari hasil analisis sebelumnya, dirumuskan bahwa permasalahan utama adalah belum adanya sistem yang mampu memahami isi gambar secara menyeluruh, mulai dari objek, teks hingga konteks visual dan mengubahnya menjadi penjelasan edukatif yang dapat digunakan oleh guru dan siswa. Guru membutuhkan pendamping digital yang mampu memberikan uraian yang relevan dan terstruktur, sementara siswa memerlukan jawaban yang aman, ringkas dan sesuai dengan tingkat pemahaman mereka. Berdasarkan kebutuhan tersebut, peneliti menetapkan tujuan penelitian, batasan fitur dan fokus pengembangan sistem, yaitu membangun chatbot multimodal berbasis *Vision AI, RAG* dan *Small Language Model (SLM)* yang mampu menjawab pertanyaan pengguna berdasarkan metadata visual secara akurat, terarah dan tidak halusinasi.

3.2.3 Design

Tahap Design difokuskan pada perancangan solusi teknis dan alur interaksi sistem berdasarkan masalah yang telah didefinisikan. Peneliti membandingkan dan merancang arsitektur sistem yang mengintegrasikan BLIP untuk captioning, OCR untuk ekstraksi teks visual dan *Sentence Transformer* untuk *embedding* serta pencarian vektor. Mekanisme *Retrieval-Augmented Generation* dirancang untuk memastikan setiap jawaban chatbot merujuk pada metadata visual yang benar. Selain itu, peneliti merancang *user flow* chatbot yang mencakup banyak mode, rancangan UI, alur percakapan dan struktur prompt yang digunakan untuk mengontrol perilaku SLM. Tahap Design menghasilkan gambaran menyeluruh mengenai bagaimana sistem akan bekerja, terlihat, dan digunakan.

3.2.4 Prototype & Test

Tahap Prototype & Test merupakan tahap realisasi konsep menjadi prototipe fungsional yang kemudian diuji secara internal untuk memastikan seluruh komponen sistem bekerja sesuai rancangan. Pada tahap Prototype, peneliti membangun pipeline *Vision Processing* di *Google Collab* untuk menghasilkan caption, teks OCR dan embedding yang disimpan ke Supabase. Selanjutnya, peneliti mengembangkan backend menggunakan *FastAPI* untuk menangani

proses embedding query, pencarian vektor melalui *RAG* serta reasoning menggunakan Small Language Model. UI pengguna dirancang dengan Next.js untuk memungkinkan interaksi percakapan antara pengguna dan chatbot. Setelah prototipe selesai, tahap Test dilakukan melalui pengujian internal untuk mengevaluasi Vision Processing dan akurasi retrieval RAG. Hasil pengujian ini digunakan untuk memperbaiki pipeline, meningkatkan akurasi pencarian dan memastikan bahwa jawaban chatbot tetap terarah serta tidak halusinasi.

3.2.5 Plan & Implement

Tahap *Plan & Implement* merupakan tahap penyempurnaan dan penerapan sistem berdasarkan hasil evaluasi internal pada tahap sebelumnya. Pada tahap ini, peneliti merencanakan perbaikan sistem dengan mengoptimalkan proses RAG agar mampu mengambil metadata visual yang lebih relevan, menyempurnakan struktur prompt untuk mengurangi potensi halusinasi SLM dan meningkatkan konsistensi alur *Vision Processing* yang dibangun di *Google Colab.* Tahap implementasi mencakup integrasi seluruh komponen seperti *Vision AI, embedding, Supabase Vector, backend FastAPI* dan antarmuka next.js ke dalam prototipe yang stabil dan dapat digunakan untuk pengujian teknis lebih lanjut. Selain itu, peneliti menyusun dokumentasi teknis pipeline serta laporan penelitian untuk memastikan seluruh proses dapat ditinjau, direplikasi dan dikembangkan pada penelitian selanjutnya. Implementasi sistem ini menjadi fondasi awal bagi pengembangan chatbot multimodal yang dapat diperluas penggunaannya untuk mendukung pembelajaran berbasis gambar di masa mendatang.

3.3 Teknik Pengumpulan Data

Teknik pengumpulan data dalam penelitian ini berfokus sepenuhnya pada data visual yang telah dihasilkan pada penelitian sebelumnya, yaitu kumpulan gambar edukatif yang dibuat menggunakan model *Text-to-Image* (T2I). Seluruh gambar tersebut merupakan hasil eksperimen terdahulu yang dirancang khusus untuk merepresentasikan berbagai konteks pembelajaran di Desa Wisata Tigaraksa, mulai dari aktivitas pertanian, objek lingkungan hingga visual edukatif yang

sesuai dengan kebutuhan guru dan siswa sekolah dasar. Gambar-gambar ini dipilih karena telah melalui proses evaluasi sebelumnya, sehingga kualitas visual, relevansi topik dan kesesuaiannya dengan kebutuhan pembelajaran sudah terjamin. Dengan demikian, penelitian ini tidak memerlukan pengambilan gambar baru atau dokumentasi lapangan tetapi memaksimalkan pemanfaatan dataset visual yang telah tersedia.

Setelah seluruh gambar dikumpulkan dari penelitian terdahulu, data visual ini kemudian diproses menggunakan pipeline *Vision AI* untuk menghasilkan metadata yang diperlukan dalam sistem *Retrieval-Augmented Generation (RAG)*. Proses pengolahan terdiri dari tiga tahapan utama. Pertama, ekstraksi *caption* dan deskripsi visual menggunakan model BLIP yang mampu mengenali objek, aktivitas dan konteks semantik dalam gambar. Kedua, ekstraksi teks menggunakan OCR untuk membaca informasi tulisan yang mungkin terdapat pada gambar misalnya label, tanda atau teks edukatif yang merupakan bagian dari visual T2I. Ketiga, seluruh informasi tersebut dikonversi menjadi representasi vektor menggunakan Sentence Transformer untuk menghasilkan embedding yang dapat digunakan pada pencarian berbasis *vector similarity search*. Hasil pengolahan ini berupa metadata terstruktur yang mencakup caption, objek terdeteksi, teks OCR, deskripsi visual dan embedding vektor. Seluruh metadata kemudian disimpan dalam database Supabase sebagai *knowledge store* yang menjadi fondasi proses reasoning dalam chatbot multimodal.

Dengan demikian, teknik pengumpulan data pada penelitian ini tidak hanya mengandalkan pengumpulan gambar dari penelitian sebelumnya, tetapi juga mencakup proses pengubahan gambar menjadi informasi semantik yang dapat dipahami oleh sistem AI. Tahapan ini memastikan bahwa chatbot memiliki landasan visual yang kuat, relevan dan terstruktur, sehingga mampu menjawab pertanyaan berdasarkan fakta visual secara akurat dan bebas dari halusinasi.

3.3.1 Sumber Dataset

Dataset yang digunakan dalam penelitian ini seluruhnya berasal dari penelitian sebelumnya yang berfokus pada pengembangan gambar edukatif menggunakan model *Text-to-Image* (T2I), khususnya Gemini. Pada penelitian tersebut, metode prompt instruction terbukti menghasilkan visual yang paling relevan dan stabil, sehingga digunakan untuk menghasilkan tambahan 100 gambar edukatif baru. Gambar-gambar tersebut merupakan hasil eksperimen terkontrol yang dirancang untuk merepresentasikan berbagai konteks pembelajaran di Desa Wisata Tigaraksa, seperti aktivitas pertanian, flora dan fauna lokal, alat desa, lingkungan sekitar dan visual edukatif yang sesuai dengan kebutuhan siswa sekolah dasar. Setiap gambar pada dataset ini telah melalui proses evaluasi sebelumnya sehingga kualitas visual, relevansi topik dan kesesuaiannya dengan konteks pembelajaran. Karena dataset ini telah tersedia dan tervalidasi, penelitian ini tidak melakukan pengambilan gambar baru maupun dokumentasi lapangan, melainkan memanfaatkan dataset yang sudah ada sebagai fondasi dalam pengembangan sistem chatbot multimodal.

3.3.2 Karateristik Dataset Visual

Dataset visual yang digunakan dalam penelitian ini terdiri dari kumpulan gambar edukatif yang dihasilkan melalui tiga variasi metode *prompt engineering*, yaitu *descriptive, instruction-based,* dan *compositional*. Setiap kategori prompt menghasilkan gaya visual dan tingkat kedetailan yang berbeda sehingga memberikan variasi metadata yang berguna dalam proses analisis *Vision AI*. Gambar dalam dataset memiliki format JPG atau PNG dengan resolusi bervariasi, namun seluruhnya memiliki kualitas visual yang memadai untuk proses ekstraksi caption, objek dan teks menggunakan BLIP serta OCR. Konten visual dalam dataset mencakup berbagai objek dan situasi yang umum ditemui di lingkungan Desa Wisata Tigaraksa seperti tanaman pangan, hewan ternak, peralatan kerja , proses pertanian hingga visual edukatif berbasis desa. Variasi konten ini memungkinkan sistem menghasilkan metadata yang kaya dan mendukung proses vector similarity search yang lebih representatif. Dengan karakteristik tersebut,

dataset ini menjadi fondasi yang kuat untuk pembangunan sistem *Vision AI* dan RAG dalam penelitian ini.

3.4 Teknik Analisis Data

Teknik analisis data pada penelitian ini dilakukan melalui serangkaian proses komputasi yang dijalankan di *Google Colab*, yang berfungsi sebagai lingkungan pemrosesan utama untuk mengekstraksi dan mengolah metadata visual dari kumpulan gambar penelitian sebelumnya. Seluruh tahapan analisis dimulai dengan pemanggilan model *Vision AI* di Colab, yaitu BLIP untuk menghasilkan caption dan deskripsi semantik gambar, OCR untuk mengekstraksi teks yang muncul pada visual, serta *Sentence Transformer* untuk menghasilkan *embedding* yang merepresentasikan isi gambar dalam bentuk vektor numerik. Hasil keluaran model-model tersebut kemudian dianalisis untuk memastikan akurasi semantik caption BLIP, ketepatan hasil OCR, dan konsistensi embedding melalui uji *vector similarity*. Analisis ini dilakukan dengan membandingkan metadata yang dihasilkan dengan isi asli gambar, sehingga peneliti dapat memverifikasi apakah informasi visual berhasil ditangkap dengan benar oleh model-model tersebut.

Selain itu, Google Colab digunakan untuk menguji kelayakan *embedding* melalui proses *similarity checking*, di mana embedding dari gambar yang mirip diuji untuk memastikan apakah nilai kedekatannya konsisten secara matematis (*cosine similarity*). Hasil embedding yang baik akan mengelompokkan gambar-gambar bertema serupa ke dalam ruang vektor yang berdekatan, sehingga memungkinkan proses *retrieval* berbasis RAG berjalan akurat. Seluruh analisis teknis ini dilakukan secara terprogram dalam bentuk notebook yang berisi skrip Python, visualisasi, dan log output model. Tahapan ini memastikan bahwa metadata visual yang dihasilkan tidak hanya lengkap secara teknis, tetapi juga valid secara semantik sehingga dapat digunakan sebagai dasar reasoning oleh model bahasa dalam chatbot multimodal. Dengan demikian, teknik analisis data berbasis *Google Colab* berfungsi sebagai proses verifikasi awal untuk menjamin bahwa sistem

memiliki fondasi data yang kuat, akurat, dan bebas dari kesalahan sebelum digunakan dalam pipeline RAG dan SLM.

3.4.1 Analisis Vision Processing (BLIP, OCR, dan Embedding)

Tahap ini berfokus pada evaluasi kualitas metadata visual yang dihasilkan oleh model Vision AI. Analisis caption dilakukan dengan memeriksa sejauh mana keluaran BLIP mampu menggambarkan objek dan konteks gambar secara akurat, lengkap, dan sesuai fakta visual. Hasil OCR dievaluasi berdasarkan tingkat keberhasilan model dalam mendeteksi teks yang muncul pada gambar T2I. Seluruh metadata ini kemudian dibandingkan dengan konten asli gambar untuk mengidentifikasi kesalahan semantik, kehilangan informasi, atau ketidaktepatan konteks. Selanjutnya, *embedding* dari *Sentence Transformer* dianalisis menggunakan *cosine similarity* untuk mengukur konsistensi representasi vektor di antara gambar yang memiliki tema serupa. Evaluasi ini memastikan bahwa metadata visual yang dihasilkan cukup stabil dan dapat menjadi dasar pengetahuan yang andal dalam sistem RAG.

3.4.2 Analisis Retrieval Berbasis RAG (Vector Similarity Search)

Pada tahap ini, sistem RAG dievaluasi untuk menilai kemampuan model dalam mengambil metadata visual yang paling relevan terhadap pertanyaan yang diajukan pengguna. Analisis dilakukan dengan menguji embedding pertanyaan dan mengukurnya terhadap embedding gambar dalam Supabase Vector menggunakan *vector similarity* (cosine similarity). Kinerja retrieval dinilai menggunakan metrik seperti *Precision@1* untuk memastikan gambar paling relevan muncul sebagai hasil utama. Selain itu, analisis juga mencakup pemeriksaan kecocokan antara pertanyaan pengguna, metadata yang diretrieval, dan jawaban yang diberikan oleh model bahasa. Tahap ini menjadi penentu apakah sistem mampu menghubungkan pertanyaan dengan bukti visual yang tepat, sehingga mencegah terjadinya halusinasi pada tahap reasoning.

3.4.3 Analisis Jawaban Chatbot (SLM Reasoning Evaluation)

Analisis ini dilakukan untuk mengevaluasi kualitas jawaban yang dihasilkan oleh Small Language Model (SLM) ketika diberikan konteks metadata hasil RAG. Fokus analisis mencakup relevansi jawaban terhadap gambar, keakuratan informasi, kejelasan bahasa, serta kesesuaian jawaban dengan kebutuhan edukatif guru dan siswa. Respons chatbot dikaji untuk memastikan bahwa model mengacu pada metadata visual, bukan menghasilkan informasi di luar konteks gambar. Evaluasi dilakukan menggunakan indikator seperti tingkat kesesuaian konteks, ketepatan semantik, dan *hallucination check* terhadap setiap jawaban. Analisis ini membuktikan apakah SLM mampu menyusun penjelasan edukatif yang ramah pengguna, konsisten, dan aman digunakan dalam pembelajaran berbasis visual.