BAB II

LANDASAN TEORI

2.1 Penelitian Terdahulu

Tabel 2.1 Penelitian Terdahulu

Penelitian Terdahulu		
1	Nama Jurnal	*International Journal of Educational Technology in Higher Education*
	Judul	*Embracing the future of Artificial Intelligence in the classroom: the relevance of AI literacy, prompt engineering, and critical thinking in modern education*
	Penulis	Ferreira, J.J., Monteiro, M., Esteves, J., Fernandes, R. [14]
	Hasil & Temuan	Penelitian ini mengidentifikasi bahwa AI literacy dan *prompt engineering* adalah hal penting dalam pendidikan modern. *Prompt engineering* ditekankan sebagai keterampilan kunci untuk mendapatkan respons spesifik dari sistem AI yang dapat memperkaya pengalaman edukatif. Kelebihan: memberikan kerangka konseptual untuk integrasi AI dalam pendidikan. Kekurangan: tidak spesifik membahas T2I untuk konteks lokal atau komunitas
2	Nama Jurnal	*Behaviour & Information Technology*
	Judul	*Design guidelines for prompt engineering text-to-image generative models*
	Penulis	Liu, V., & Chilton, L.B. [17]
	Hasil & Temuan	Studi selama 3 bulan mengidentifikasi 6 jenis prompt modifiers yang digunakan praktisi T2I: *subject, style, format, perspective, parameter, dan negative prompts.*

		Penelitian ini memberikan kerangka konseptual sistematis untuk praktik T2I generation dan menjadi dasar kategorisasi metode prompt. Kelebihan: klasifikasi prompt yang tervalidasi. Kekurangan: fokus pada komunitas online profesional, belum diuji efektivitasnya dalam konteks pendidikan formal atau edukasi anak.
3	Nama Jurnal	*Design and Technology Education*
	Judul	*Prompt engineering in higher education: A systematic review to help inform curricula*
	Penulis	D. Lee and E. Palmer [18]
	Hasil & Temuan	Penelitian menguji berbagai model T2I (*DALL-E, Midjourney, Stable Diffusion*) dengan variasi prompt dan settings dalam konteks pendidikan seni dan desain. Temuan menunjukkan T2I dapat mengubah proses pembelajaran kreatif dengan mempercepat ideasi dan eksplorasi visual. Kelebihan: fokus pada proses pembelajaran. Kekurangan: target mahasiswa desain, belum ada evaluasi sistematis tentang efektivitas metode prompt berbeda atau aplikasi untuk anak-anak.
4	Nama Jurnal	*International Journal of Art & Design Education*
	Judul	*Graphic Design Education in the Era of Text-to-Image Generation: Transitioning to Contents Creator*
	Penulis	Hwang, Y. [19]
	Hasil & Temuan	Studi menggunakan Midjourney dan DALL-E dalam pendidikan desain grafis berdasarkan *Technology Pedagogical Content Knowledge (TPACK) framework.* Mahasiswa menggunakan *GenAI* untuk membuat konten visual dengan aturan terstruktur. TPACK terbukti efektif sebagai framework. Kelebihan: framework edukasi yang

		jelas dan aplikatif. Kekurangan: fokus pada mahasiswa desain grafis yang sudah berpengalaman, bukan pendidikan dasar atau konteks non-formal.
5	Nama Jurnal	*New Media & Society*
	Judul	*Enhancing children's understanding of algorithmic biases in and with text-to-image generative AI*
	Penulis	Vartiainen, H., Kahila, J., Tedre, M., López-Pernas, S., & Pope, N. [20]
	Hasil & Temuan	Penelitian *co-design* dan *design-based research* tentang integrasi topik AI (termasuk T2I) dalam pendidikan sekolah untuk anak-anak usia 10-12 tahun. Fokus pada pemahaman algoritmik dan literasi AI melalui *hands-on activities*. Anak-anak mampu mengidentifikasi bias representasi dan memahami keterbatasan AI. Kelebihan: sangat relevan untuk konteks pendidikan anak. Kekurangan: fokus pada *AI literacy* dan *critical thinking,* bukan pada evaluasi kualitas output visual untuk konten edukatif.
6	Nama Jurnal	Cogent Education
	Judul	*Using artificial intelligence in craft education: crafting with text-to-image generative models*
	Penulis	Vartiainen, H., Tedre, M., & Valtonen, T. [21]
	Hasil & Temuan	*Workshop hands-on* dengan 28 guru tentang pembuatan kreatif menggunakan T2I generatif (DALL-E 2). Hasil menunjukkan T2I menginspirasi guru untuk mempertimbangkan AI dalam konteks pembelajaran. Guru mengidentifikasi potensi dan tantangan integrasi T2I. Kelebihan: perspektif guru sebagai pengguna dan fasilitator, metodologi workshop praktis. Kekurangan:

		belum ada panduan praktis tentang metode prompt terbaik atau evaluasi kualitas output untuk berbagai kebutuhan edukatif.
7	Nama Jurnal	*TechTrends*
	Judul	*Generative Artificial Intelligence Images for Instruction: A Content Analysis*
	Penulis	Campbell, O., Lambie, G., Hayes, B., Hartshorne, R. [22]
	Hasil & Temuan	Analisis konten tentang penggunaan gambar GenAI (termasuk T2I) untuk instruksi pembelajaran. Penelitian menemukan bahwa kemudahan membuat gambar dengan prompt sederhana memiliki potensi besar untuk memperkaya materi, namun perlu panduan untuk memastikan kualitas edukatif, akurasi, dan hubungan dengan tujuan pembelajaran. Kelebihan: fokus pada penggunaan pembelajaran Kekurangan: tidak membandingkan model atau metode prompt secara sistematis, belum ada framework evaluasi terstruktur.
8	Nama Jurnal	*International Journal of Human-Computer Interaction*
	Judul	*A taxonomy of prompt modifiers for text-to-image generation*
	Penulis	Oppenlaender, J. [23]
	Hasil & Temuan	Tiga studi konsekutif dengan 270+ partisipan mengeksplorasi apakah partisipan dapat: (1) menilai kualitas prompt, (2) menulis prompt efektif, dan (3) memperbaiki prompt untuk AI art. Temuan menunjukkan prompt engineering adalah keterampilan kreatif yang dapat dipelajari dan ditingkatkan melalui praktik dan feedback. Kelebihan: metodologi sistematis untuk evaluasi prompt. Kekurangan: fokus pada seni, bukan

		aplikasi edukatif atau konteks lokal spesifik.
9	Nama Jurnal	*Computers and Education: Artificial Intelligence*
	Judul	*AI literacy and its implications for prompt engineering strategies*
	Penulis	Knoth, N., Tolzin, A., Janson, A., Marco, J. [24]
	Hasil & Temuan	Penelitian tentang hubungan antara literasi AI dan efektivitas prompt engineering dalam penggunaan Large Language Models. Menunjukkan bahwa pemahaman tentang cara kerja AI, limitasi, dan bias memengaruhi kemampuan membuat prompt efektif. *Framework AI literacy* dikembangkan untuk pendidik. Kelebihan: menghubungkan literasi AI dengan kompetensi *prompt engineering*. Kekurangan: fokus pada LLM teks (ChatGPT, dll), bukan T2I dan belum aplikatif untuk konteks komunitas lokal.
10	Nama Jurnal	Frontiers in Education
	Judul	Prompt engineering as a new 21st century skill
	Penulis	Federiakin,D., Morelov,D., Zlatikin, O., Maur, A. [25]
	Hasil & Temuan	Penelitian ini mengidentifikasi prompt engineering sebagai keterampilan abad 21 yang esensial untuk berbagai profesi. Menunjukkan bahwa teknik prompting yang tepat dapat meningkatkan performa AI secara signifikan (hingga 40-60% *improvement* dalam *task completion*). Framework kompetensi *prompt engineering* untuk kurikulum pendidikan diusulkan. Kelebihan: perspektif pendidikan yang komprehensif. Kekurangan: bersifat *review*, tidak spesifik membahas T2I atau menyediakan panduan praktis evaluasi visual.

11	Nama Jurnal	*Computers & Education*
	Judul	*Evaluating AI-powered text-to-image generators for anatomical illustration: A comparative study*
	Penulis	Geoffroy P. J. C. Noel [26]
	Hasil & Temuan	Penelitian ini membandingkan kualitas gambar edukatif yang dihasilkan oleh Stable Diffusion, Midjourney, dan DALL-E 3 untuk materi IPA sekolah dasar. Hasil menunjukkan bahwa prompt instruction-based menghasilkan akurasi konsep tertinggi, terutama pada materi anatomi dan ekosistem. Stable Diffusion unggul pada detail tekstur, sementara DALL-E lebih presisi.
12	Nama Jurnal	Education and Information Technologies
	Judul	Diffusion Models and Generative Artificial Intelligence: Frameworks, Applications and Challenges
	Penulis	Pranjal Kumar [27]
	Hasil & Temuan	Penelitian ini menemukan bahwa Diffusion Models merupakan pendekatan generatif paling efektif saat ini dengan kinerja unggul dalam berbagai domain seperti sintesis gambar, video, hingga desain molekul.
13	Nama Jurnal	*Proceedings of the IEEE 9th International Conference for Convergence in Technology (I2CT), Pune, India, 2024.*
	Judul	*A Comparative Study of Text-to-Image Generative Models*
	Penulis	S. S. Rajbhandari; N. Jha; R. Pandey [28]
	Hasil & Temuan	Penelitian ini membandingkan kinerja beberapa model text-to-image dengan menilai kualitas visual, kesesuaian gambar terhadap deskripsi, serta stabilitas hasil antarmodel. Temuan utama menunjukkan bahwa model berbasis difusi cenderung menghasilkan detail dan konsistensi visual yang lebih baik dibandingkan metode generatif lain, sementara model yang lebih ringan menawarkan kecepatan tetapi kurang akurat dalam mengikuti prompt. Kelebihannya adalah memberikan perbandingan komprehensif yang membantu pengguna memilih model sesuai kebutuhan, namun kelemahannya terletak pada cakupan eksperimen yang terbatas.
14	Nama Jurnal	*Education and Information Technologies*

	Judul	*Prompt Engineering in Education: A Systematic Review of Approaches and Educational Applications*
	Penulis	Yufeng Qian [29]
	Hasil & Temuan	Penelitian ini menemukan bahwa efektivitas alat *generative AI* dalam pendidikan sangat ditentukan oleh kualitas prompt engineering, yang bukan hanya sekadar merancang input, tetapi mengarahkan AI untuk menghasilkan output yang relevan, akurat, dan mengedukasi. Melalui penelitian ini mengidentifikasi dua pendekatan utama dalam strategi *prompting: technique-based*, yang berfokus pada pencapaian tujuan belajar spesifik, serta *process-based* yang menstimulasi keterlibatan kognitif dan kolaborasi manusia–AI.
15	Nama Jurnal	*Procedia Computer Science*
	Judul	*Comprehensive review of generative artificial intelligence: Mechanisms, models, and applications.*
	Penulis	Pahuja, Kukreja, & Singh [30]
	Hasil & Temuan	Penelitian ini memberikan tinjauan menyeluruh tentang mekanisme dan aplikasi *generative AI*, mencakup *Transformer*s, dan terutama *Diffusion Models*. Studi ini menegaskan bahwa *Diffusion Models* menawarkan stabilitas dan kualitas tertinggi untuk sintesis gambar serta konten multimodal berkat proses *denoising* bertahap. Penulis juga mencatat tren meningkatnya model multimodal besar (LLM + visi) yang memperluas penggunaan AI di bidang kesehatan, pendidikan, dan industri kreatif. Namun, penelitian menyoroti keterbatasan seperti kebutuhan komputasi tinggi, bias data, dan kurangnya standar evaluasi, sehingga diperlukan metrik kualitas seperti CLIPScore untuk menilai output secara lebih objektif.

Berdasarkan penelitian terdahulu yang dirangkum dalam Tabel 2.1, dapat diidentifikasi bahwa *text-to-image generation* (T2I) dan *prompt engineering* telah menjadi topik penting dalam pendidikan, khususnya dalam konteks *creative learning*, literasi AI dan desain berbasis teknologi. Tetapi, pola umum dari literatur yang ada menunjukkan bahwa sebagian besar studi masih berfokus pada

high-resource context yakni lingkungan pendidikan formal di negara maju dengan infrastruktur digital dan sumber daya manusia yang cukup. Fokus tersebut menyebabkan kurangnya penelitian yang menyoroti penerapan teknologi serupa di lingkungan masyarakat dengan keterbatasan sumber daya, seperti desa wisata di Indonesia.

Ferreira et al. dan Knoth et al. menekankan *AI literacy* sebagai kompetensi kunci abad ke-21 serta relevansi *prompt engineering* dalam meningkatkan kemampuan berpikir kritis peserta didik [14]. Keduanya memberi kontribusi penting dalam membangun landasan teoretis tentang hubungan antara *critical thinking*, *AI literacy* dan *prompt-based learning*. Akan tetapi, pendekatan tersebut masih bersifat konseptual karena berbasis kurikulum pendidikan dan fokus pada *Large Language Models* (LLM) seperti ChatGPT sehingga belum memberikan panduan yang dapat diterapkan dalam konteks *text-to-image* maupun pendidikan non-formal.

Liu dan Chilton memberikan kontribusi signifikan melalui penelitian tentang *prompt modifiers* untuk T2I, yang mengklasifikasikan berbagai gaya dan struktur prompt (seperti *subject-type*, *style*, *parameter*, dan *composition*) [17]. Ini penting karena menyediakan kerangka berpikir yang sistematis bagi penelitian selanjutnya. Namun, penelitian tersebut tidak membahas bagaimana variasi prompt tersebut dapat memengaruhi hasil belajar, daya tarik visual bagi anak-anak atau efektivitas edukatif gambar yang dihasilkan yang justru menjadi inti permasalahan di Desa Wisata Tigaraksa.

Penelitian Ringvold dan Hwang membawa pendekatan eksperimental dengan mengintegrasikan T2I seperti *DALL-E, Midjourney* dan *Stable Diffusion* ke dalam pendidikan seni dan desain [19]. Hasil mereka menunjukkan bahwa T2I mampu mempercepat eksplorasi ide visual dan meningkatkan kreativitas mahasiswa. Walaupun demikian, kedua studi tersebut beroperasi pada konteks dengan subjek yang sudah memiliki latar belakang desain, sehingga belum menjawab kebutuhan pendidikan dasar. Selain itu, evaluasi keberhasilan masih

menggunakan ukuran artistik dan estetika umum, bukan indikator seperti pemahaman konsep, relevansi atau keterlibatan anak sebagai audiens utama.

Penelitian Vartiainen et al. menunjukkan pendekatan yang lebih dekat dengan konteks penelitian ini, karena menerapkan metode *co-design* dan *participatory workshop* dalam pendidikan berbasis *Generative AI* [20]. Pendekatan tersebut menekankan kolaborasi antara guru, siswa dan AI sebagai rekan kreatif (*co-creator*). Meski demikian, konteks penelitian mereka masih terbatas pada sekolah dasar di Eropa dengan dukungan literasi digital tinggi, sehingga belum mempertimbangkan kendala di masyarakat pedesaan Indonesia seperti keterbatasan infrastruktur teknologi dan sumber daya manusia non-teknis.

Oppenlaender dan Campbell et al. memperluas cakupan penelitian ke arah *prompt engineering as a creative practice*, menunjukkan bahwa variasi struktur dan intensi prompt dapat memengaruhi kualitas T2I [23]. Namun, penelitian mereka berhenti pada aspek *creative process* tanpa memberikan kerangka evaluatif yang menilai sejauh mana hasil visual tersebut relevan dengan tujuan pembelajaran atau efektif dalam menyampaikan pesan edukatif. Tidak adanya *evaluation framework* yang eksplisit. Inilah yang kemudian membatasi kontribusi T2I sebagai instrumen, bukan sekadar alat eksplorasi visual.

Penelitian Noel memberikan contoh evaluasi komparatif kualitas visual dengan membandingkan *Stable Diffusion, Midjourney dan DALL-E 3* dalam pembuatan ilustrasi [26]. Temuan menunjukkan bahwa prompt *Instruction Based* menghasilkan akurasi konsep tertinggi untuk materi biologi seperti anatomi dan ekosistem. Studi ini memperlihatkan bahwa model T2I memiliki potensi besar untuk menghasilkan visual edukatif, tetapi akurasi masih menjadi tantangan.

Dari perspektif teknis, Kumar dan Pahuja et al. menegaskan bahwa *diffusion models* memberikan stabilitas dan kualitas gambar yang lebih tinggi dibandingkan model generatif lainnya [27]. Namun, keduanya mencatat bahwa model-model ini membutuhkan komputasi tinggi dan evaluasi kualitas yang

objektif, seperti CLIPScore untuk mengatasi bias dan variasi visual yang tidak konsisten.

Rajbhandari et al. memperkuat temuan tersebut melalui studi komparatif yang menunjukkan bahwa model difusi lebih unggul dalam detail visual dan konsistensi semantik, walaupun model yang lebih ringan lebih efisien secara komputasi [28]. Sementara itu, Qian menegaskan bahwa kualitas output AI dalam pendidikan sangat ditentukan oleh strategi *prompting*, baik teknik berbasis tujuan maupun pendekatan berbasis kolaborasi [31].

Penelitian-penelitian terdahulu cenderung memusatkan perhatian pada tiga dimensi utama: (1) eksplorasi teknis T2I dan parameter prompt, (2) pengembangan literasi AI dalam pendidikan formal, dan (3) integrasi T2I sebagai alat bantu dalam hal kreativitas. Namun, dari sisi metodologis dan aplikatif, masih terdapat kekosongan pada empat area kritis. Pertama, belum ada studi yang secara eksplisit menghubungkan variasi metode prompt (deskriptif, instruktif, dan komposisional) dengan kualitas hasil visual dalam konteks edukasi anak-anak, terutama dalam pembelajaran berbasis konten lokal. Kedua, belum ditemukan perbandingan lintas model T2I (misalnya *Gemini vs Stable Diffusion*) yang mempertimbangkan faktor aksesibilitas, dukungan bahasa, biaya operasional dan kelayakan adopsi oleh pengguna non-teknis seperti guru desa atau pengelola wisata. Ketiga, mayoritas penelitian menilai kualitas T2I dengan metrik teknis (seperti CLIPScore atau FID), tapi jarang melalui penilaian manusia yang mempertimbangkan konteks sosial dan edukatif lokal. Keempat, belum ada model penelitian yang mengintegrasikan T2I dengan pendekatan partisipatif berbasis masyarakat, padahal konteks seperti Desa Wisata Tigaraksa justru menuntut pendekatan kolaboratif (pendidik, pengelola desa, dan anak-anak).

Jika dibandingkan dengan 15 penelitian terdahulu, penelitian ini memiliki fokus berbeda dalam lima aspek yang menjawab kesenjangan metodologis dan aplikatif yang telah diidentifikasi yaitu pertama, dari sisi metodologi prompt engineering, penelitian ini merupakan yang pertama melakukan evaluasi sistematis terhadap tiga metode *prompt engineering (descriptive,*

instruction-based, compositional) dalam konteks pendidikan komunitas pedesaan Indonesia. Berbeda dengan Liu & Chilton [17] yang mengklasifikasikan *prompt modifiers* tanpa menguji efektivitas edukatif atau Oppenlaender [23] yang fokus pada *creative practice* tanpa framework evaluasi, penelitian ini merancang dan menguji ketiga metode prompt berdasarkan kerangka teoretis yang jelas dan kemudian mengevaluasi pengaruhnya terhadap kualitas visual edukatif melalui metrik kuantitatif (CLIPScore) dan penilaian kualitatif (human evaluation).

Kedua, penelitian ini melakukan perbandingan komprehensif antara dua model T2I dengan karakteristik berbeda yaitu Stable Diffusion (*open-source*, kontrol teknis tinggi) dan Gemini Nano Banana (*multimodal transformer*, aksesibilitas dan memahami bahasa natural) yang tidak hanya mengevaluasi performa teknis, tetapi juga mempertimbangkan kelayakan adopsi oleh pengguna non-teknis. Berbeda dengan Noel [26] yang membandingkan *Stable Diffusion, Midjourney, dan DALL-E 3* tanpa mempertimbangkan faktor aksesibilitas atau dukungan bahasa lokal, penelitian ini secara eksplisit menjawab pertanyaan: "Model mana yang paling bisa untuk diadopsi guru desa dengan keterbatasan literasi teknis dan infrastruktur?"

Ketiga, framework evaluasi penelitian ini mengintegrasikan pendekatan objektif (*CLIPScore*) dengan penilaian subjektif dari pengguna lokal (guru, pengelola desa, anak-anak) membuat validasi yang tidak dijumpai pada studi terdahulu. Penelitian [20] dan [22] menggunakan metrik teknis seperti akurasi klasifikasi tanpa melibatkan evaluator dari komunitas sasaran, sementara [21] melibatkan guru dalam *workshop* namun tanpa metrik kuantitatif sebagai pembanding objektif. Kombinasi evaluasi dalam penelitian ini memastikan bahwa hasil tidak hanya valid secara teknis, tetapi juga relevan secara edukatif.

Keempat, konteks aplikasi penelitian ini yaitu Desa Wisata Tigaraksa sebagai lingkungan pembelajaran berbasis komunitas dengan fokus edukasi pertanian untuk anak-anak SD belum dieksplorasi dalam literatur T2I. Mayoritas penelitian terdahulu berlokasi di lingkungan pendidikan formal negara maju fokus pada konteks creative design untuk mahasiswa [18]. Penelitian ini memilih

konteks pedesaan Indonesia dengan keterbatasan sumber daya untuk mendemonstrasikan bahwa teknologi Generative AI dapat diadaptasi mendukung pembelajaran visual di komunitas yang selama ini belum dalam literatur teknologi pendidikan.

Kelima, penelitian ini tidak berhenti pada evaluasi eksperimental, tetapi juga menghasilkan prototype sistem berbasis web yang memungkinkan guru dan pengelola desa menghasilkan visual edukatif secara mandiri hanya dengan menuliskan prompt dalam bahasa Indonesia. Pendekatan ini menunjukkan komitmen terhadap keberlanjutan dan adopsi praktis, yang jarang dijumpai dalam penelitian T2I terdahulu yang cenderung berhenti pada tahap evaluasi model tanpa mempertimbangkan implementasi sistem yang siap digunakan oleh komunitas non-teknis.

Dengan demikian, penelitian ini tidak hanya mengisi gap metodologis dan teoretis yang teridentifikasi dalam literatur, tetapi juga memberikan kontribusi berupa panduan berbasis bukti dan sistem yang dapat langsung diterapkan oleh Desa Wisata Tigaraksa dan komunitas serupa di Indonesia untuk memproduksi konten edukatif yang menarik, relevan secara budaya dan berkelanjutan.

2.2 Teori yang berkaitan

2.2.1 Generative Artificial Intelligence dan Text-to-Image (T2I) Generation

Generative Artificial Intelligence (GenAI) merupakan cabang dari kecerdasan buatan yang berfokus pada kemampuan sistem untuk menghasilkan konten baru berdasarkan pola dan data yang telah dipelajari sebelumnya. Teknologi ini bekerja dengan memanfaatkan deep neural networks dan model probabilistik yang belajar dari distribusi data untuk menciptakan representasi baru yang menyerupai data asli.

Salah satu implementasi paling menonjol dari *GenAI* adalah *Text-to-Image (T2I) Generation,* yaitu proses menghasilkan gambar dari deskripsi teks (prompt). *Text-to-Image Generation* merupakan bentuk penerapan *Generative Artificial*

Intelligence yang menghasilkan gambar berdasarkan instruksi teks. Proses ini diawali dengan pengubahan teks menjadi representasi numerik, kemudian dilanjutkan dengan tahapan difusi yang bertujuan membentuk gambar sesuai makna deskriptif dari teks tersebut.

Model seperti Stable Diffusion menggunakan pendekatan latent diffusion, sedangkan model Gemini Nano Banana (Flash) berbasis multimodal transformer yang memahami konteks bahasa alami. T2I telah menjadi salah satu inovasi yang paling relevan dalam pendidikan visual karena memungkinkan siapa pun menciptakan ilustrasi edukatif hanya dengan teks sederhana tanpa kemampuan desain profesional [32]

2.2.1.1 Stable Diffusion

Gambar 2.1 Logo Stable Diffusion

Stable Diffusion merupakan model *open-source* yang dikembangkan oleh Stability AI dan menggunakan mekanisme latent diffusion, di mana gambar dihasilkan melalui proses bertahap dari noise acak menuju representasi visual yang bermakna. Model ini beroperasi dalam *latent space* yang menghubungkan hubungan semantik antara bahasa dan visual [33]. Proses ini memungkinkan sistem untuk "memahami" makna deskriptif dari prompt dan mengubahnya menjadi bentuk visual yang realistis dan detail. Kelebihan utama *Stable Diffusion* terletak pada tingkat fleksibilitas dan kontrol teknis yang tinggi, karena pengguna dapat mengatur berbagai parameter seperti *guidance scale*, *inference steps*, dan *seed* untuk menyesuaikan gaya, warna serta komposisi gambar [34]. Selain itu,

karena bersifat *open* dan dapat dijalankan secara lokal, model ini sangat sesuai untuk penelitian akademik yang memerlukan transparansi proses.

2.2.1.2 Gemini Nano Banana (Flash)

Gambar 2.3 Logo Gemini Nano Banana (Flash)

Gemini Nano Banana (Flash), dikembangkan oleh Google DeepMind sebagai model *multimodal transformer* yang menggabungkan kemampuan bahasa alami dan visual [35]. Berbeda dari *Stable Diffusion* yang memerlukan struktur prompt teknis, Gemini mampu memahami perintah dalam bentuk kalimat alami (*instruction-following*) seperti "buat gambar anak-anak sedang belajar menanam pohon di kebun" . Kemampuan ini menjadikan Gemini lebih mudah digunakan oleh pengguna non-teknis seperti guru, siswa atau pengelola desa yang tidak terbiasa dengan terminologi teknis AI. Selain itu, model ini dirancang dengan fokus pada efisiensi dan pemahaman konteks budaya, termasuk dukungan terhadap bahasa Indonesia dan konteks lokal yang membuatnya sangat relevan untuk penelitian berbasis komunitas.

2.2.2 Prompt Engineering

Prompt engineering merupakan praktik menyusun instruksi teks agar sistem kecerdasan buatan (AI) dapat menghasilkan keluaran yang sesuai dengan tujuan pengguna. Menurut Liu & Chilton, *prompt* berfungsi sebagai *semantic bridge* antara niat manusia dan mesin, sementara Oppenlaender menekankan pentingnya struktur linguistik yang eksplisit agar model mampu memahami konteks secara semantik dan visual. Kualitas keluaran model Text-to-Image (T2I) sangat bergantung pada kejelasan, urutan logis serta kedalaman konteks emosional dari *prompt* yang diberikan [36].

Dalam penelitian ini, *prompt engineering* dikategorikan menjadi empat metode utama. Pertama, *descriptive prompt* yang berfokus pada penjabaran detail visual secara eksplisit seperti warna, bentuk, latar hingga gaya artistik. Kedua, *instruction-based prompt*, yang menekankan pada perintah atau tujuan tertentu, misalnya menghasilkan infografis edukatif dengan gaya visual anak-anak. Ketiga, *compositional prompt*, yang berorientasi pada pengaturan tata letak, skala dan hubungan antar elemen visual agar hasil gambar tampak harmonis dan mudah dipahami. Keempat, agentic prompt yang memberi peran atau identitas sosial tertentu kepada AI selama proses generasi, sehingga hasilnya memiliki gaya, tone, dan konteks komunikasi yang lebih spesifik.

Prompt yang efektif harus mencerminkan keseimbangan antara *cognitive load* dan *semantic structure*, di mana model diarahkan secara presisi tanpa membatasi kreativitas visual yang dihasilkan [37]. Dalam konteks pendidikan, keberhasilan *prompt engineering* diukur dari sejauh mana gambar yang dihasilkan mampu menyampaikan pesan pembelajaran secara jelas, relevan dengan konteks dan menarik secara pedagogis.

Secara teoretis, penelitian ini memperluas pemahaman tentang *prompt engineering* dengan memasukkan dimensi konteks lokal sebagai variabel mediator yang berpengaruh terhadap efektivitas hasil visual. Dengan demikian, kualitas gambar tidak hanya dinilai berdasarkan ketepatan teknis hasil generasi, tetapi juga

sejauh mana gambar tersebut mencerminkan budaya, nilai serta kebutuhan masyarakat setempat [38].

Prompt pada dasarnya adalah seni menjelaskan konteks dengan cara yang paling konkret dan terstruktur agar mesin dapat "memahami" dan menerjemahkannya menjadi keluaran yang sesuai dengan yang diinginkan, Berikut metode-metode prompt :

2.2.2.1 Descriptive Prompt

Pendekatan descriptive prompt berfokus pada penyusunan teks yang menggambarkan detail visual secara eksplisit mencakup warna, bentuk, tekstur, gaya artistik, pencahayaan, latar belakang hingga ekspresi emosional dari subjek yang digambarkan. Tujuan utamanya adalah membantu model AI memahami konteks visual secara spesifik sehingga hasil gambar yang dihasilkan tidak ambigu. Sebagai contoh, prompt seperti "*seorang petani sedang menanam padi di sawah Tigaraksa pada pagi hari, dengan sinar matahari lembut, burung-burung beterbangan di langit, dan gunung di kejauhan*". Berikut contoh dari prompt deksriptif :

Gambar 2.4 Contoh Descriptive Prompt

2.2.2.2 Instruction-Based Prompt

Berbeda dari pendekatan deskriptif yang berorientasi pada “apa” yang harus digambar, *instruction-based prompt* berfokus pada “bagaimana” hasil itu seharusnya dibuat. Pendekatan ini menekankan aspek *command structure* yakni instruksi yang mengarahkan gaya, format, tone atau tujuan tertentu dari output. Dengan kata lain, prompt tidak hanya menjelaskan objek, tetapi juga fungsi dan konteks penggunaannya.

Sebagai contoh, instruksi seperti *“Buat infografis edukatif bertema Cara Menanam dan Memanen Kakao di Desa Tigaraksa dengan gaya ilustrasi anak-anak, warna cerah, dan layout vertikal langkah demi langkah. Tampilkan*

lima tahap utama: (1) persiapan lahan, (2) penanaman bibit, (3) perawatan tanaman, (4) panen buah, dan (5) pengolahan awal. Sertakan ikon kecil seperti cangkul, air, daun, buah kakao, dan matahari. Latar belakang adalah pemandangan desa tropis yang hijau dan cerah, dengan anak-anak tersenyum belajar bersama petani lokal. Tulisan edukatif mudah dibaca, suasana positif dan inspiratif". Berikut contoh dari *prompt instruction-based:*

Gambar 2.5 Contoh Instruction Prompt

2.2.2.3 Compositional Prompt

Pendekatan *compositional prompt* digunakan untuk mengatur tata letak (*layout*) dan hubungan antar elemen visual agar hasil gambar memiliki keseimbangan estetika dan fungsionalitas. Komposisi visual memiliki peran penting dalam

menarik perhatian dan memandu fokus pengamat terhadap pesan utama dalam gambar. Karena itu, metode ini tidak hanya mendeskripsikan objek, tetapi juga menentukan struktur seperti posisi, skala, perspektif, pencahayaan, dan interaksi antar elemen.

Sebagai contoh, prompt seperti *"Letakkan dua anak di tengah sedang menyiram tanaman, di belakangnya ada kebun kopi, dan di sisi kiri terdapat papan bertuliskan 'Belajar Bertani di Desa Tigaraksa"*. Berikut contoh dari *prompt compositional:*

Gambar 2.6 Contoh Compositional Prompt

2.2.2.4 Agentic Prompt

Pendekatan *agentic prompt* melibatkan *role assignment* kepada model AI selama proses generasi. Alih-alih hanya memberi tahu "apa" yang harus digambar,

metode ini menempatkan AI seolah-olah memiliki identitas, tujuan atau niat kreatif tertentu. Dengan memberikan "peran" ini, AI dapat memahami konteks sosial dan emosional dari prompt, sehingga menghasilkan keluaran yang lebih konsisten dengan *tone* dan tujuan komunikasi yang diinginkan.

Sebagai contoh, prompt seperti *"Bertindaklah sebagai ilustrator buku anak-anak dan buat gambar tentang kegiatan belajar mengenal tanaman hias di taman desa, dengan gaya lembut dan karakter yang ceria"*.Pendekatan agentic prompt tidak berdiri sendiri, tetapi justru merupakan kombinasi dari tiga metode utama sebelumnya *descriptive*, *instruction-based* dan *compositional*. Berikut contoh dari *prompt agentic:*

Gambar 2.7 Contoh Agentic Prompt

2.2.3 Evaluasi Model Text-to-Image (T2I)

Evaluasi hasil visual dari model *Text-to-Image* (T2I) merupakan proses penting untuk menentukan sejauh mana model mampu menghasilkan gambar yang akurat, menarik dan relevan dengan deskripsi teks yang diberikan. Karena model generatif bersifat multimodal menggabungkan pemahaman bahasa dan visual. proses evaluasinya pun bersifat multidimensi, melibatkan aspek teknis, estetika, dan persepsi manusia.

Secara umum, terdapat dua kategori utama dalam evaluasi yaitu *objective evaluation* dan *subjective evaluation*. Evaluasi objektif dilakukan dengan menggunakan metrik matematis yang mampu mengukur performa model secara kuantitatif, sedangkan evaluasi subjektif menilai hasil berdasarkan persepsi dan interpretasi manusia terhadap makna dan kualitas gambar.

Dengan demikian, proses evaluasi ini dapat digunakan untuk membandingkan performa antar model T2I (Gemini dan Stable Diffusion) dan menilai efektivitas berbagai metode penyusunan prompt seperti deskriptif, instruktif atau komposisional dalam menghasilkan visual yang paling relevan dan berkualitas, berikut evaluasi :

2.2.3.1 CLIPScore (Contrastive Language–Image Pretraining Score)

CLIPScore merupakan metrik yang menilai sejauh mana hasil gambar yang dihasilkan model *Text-to-Image (T2I)* merepresentasikan makna dari *prompt* yang diberikan. Metrik ini didasarkan pada arsitektur CLIP (*Contrastive Language–Image Pretraining*) yaitu model yang dilatih untuk memahami hubungan antara teks dan gambar dan bukan hanya kesamaan visual permukaan seperti warna atau bentuk, melainkan kesamaan makna yang dipahami oleh model.

Secara teknis, perhitungan CLIPScore melibatkan dua komponen utama: *Text Encoder dan Image Encoder*. Pertama, *Text Encoder* mengubah teks prompt

menjadi representasi vektor yang merepresentasikan makna dari kalimat tersebut. Kedua, *Image Encoder* mengubah gambar hasil generasi menjadi vektor yang merepresentasikan ciri-ciri visual dalam ruang semantik yang sama. Kedua vektor tersebut kemudian dinormalisasi dan dibandingkan menggunakan *cosine similarity*, yaitu ukuran kesamaan arah antara dua vektor [39].

Nilai CLIPScore berkisar antara -1 hingga 1, di mana skor yang lebih tinggi menunjukkan kesesuaian yang lebih kuat antara makna teks dan gambar. Skor mendekati 1 menunjukkan bahwa gambar sangat sesuai dengan makna deskripsi teks. Skor mendekati 0 menunjukkan bahwa hubungan antara teks dan gambar lemah atau ambigu. Skor negatif menunjukkan bahwa gambar bertentangan dengan makna teks atau model gagal memahami konteks prompt.

Dan juga terdapat penelitian yang menunjukkan bahwa CLIPScore memiliki bias karena model dasarnya dilatih terutama menggunakan caption berbahasa Inggris, sehingga kurang optimal ketika digunakan untuk menilai visual berbahasa Indonesia atau domain edukatif spesifik.

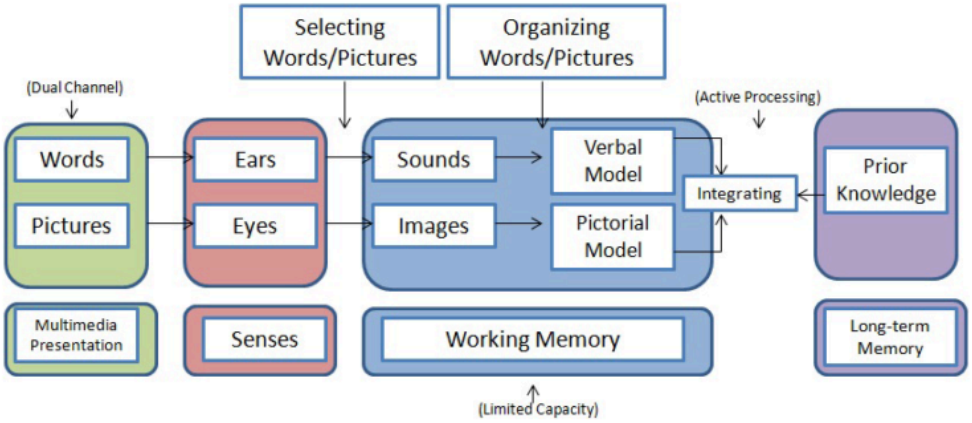
2.2.3.2 Human Evaluation

Selain metrik matematis, evaluasi berbasis manusia (human-based evaluation) tetap menjadi pendekatan yang sangat penting dalam menilai kualitas hasil visual dari model Text-to-Image (T2I), terutama dari sisi persepsi, pemahaman makna, dan relevansi konteks.

Pendekatan ini memungkinkan penilaian yang lebih baik karena manusia dapat mengenali detail semantik, nuansa visual, serta ketepatan representasi objek yang sering kali tidak dapat ditangkap oleh algoritma kuantitatif. Penilaian manusia juga mampu mengidentifikasi aspek-aspek seperti kejelasan pesan visual, daya tarik gambar bagi audiens tertentu (misalnya anak-anak), kesesuaian budaya dan konteks lokal, serta potensi misinformasi visual seperti hal-hal yang berada di luar jangkauan metrik otomatis seperti CLIPScore [40].

Selain itu, evaluasi manusia dapat menangkap elemen subjektif seperti kemudahan dipahami yang sangat penting dalam konteks pembelajaran visual. Dengan demikian, human-based evaluation bukan hanya berfungsi sebagai pelengkap metrik teknis, tetapi menjadi komponen kunci untuk memastikan bahwa gambar yang dihasilkan T2I benar-benar bermanfaat bagi pengguna dalam konteks dunia nyata.

2.2.4 Teori Pembelajaran Visual dan Multimedia Learning



Gambar 2.8 Cognitive Theory of Multimedia Learning

Teori pembelajaran visual menyatakan bahwa manusia memproses informasi lebih efektif ketika teks dan gambar disajikan secara bersamaan. Otak manusia memiliki dua saluran utama untuk memproses informasi *verbal* dan *visual*, sebagaimana dijelaskan dalam Gambar.

Sementara itu, *Cognitive Theory of Multimedia Learning* berargumen bahwa pembelajaran akan lebih efektif jika materi disusun dengan mempertimbangkan beban kognitif dan hubungan antara teks dan visual. Prinsip seperti *coherence* (menghindari elemen yang tidak relevan), *modality* (menggabungkan teks dan audio) dan *contiguity* (menyajikan teks dekat dengan gambar) menjadi panduan penting dalam desain pembelajaran modern [41]

Secara teoretis, pembelajaran visual membantu anak anak memahami konsep abstrak dengan lebih mudah karena visual mampu menggantikan narasi panjang

dengan representasi intuitif. Teori ini juga mendukung pengembangan media interaktif, infografis dan ilustrasi edukatif yang kini banyak diterapkan dalam pendidikan berbasis teknologi.

2.3 Framework/Algoritma yang digunakan

2.3.1 Comparative Evaluation Framework for Text-to-Image Generation using Prompt Engineering

Comparative Evaluation Framework for Text-to-Image Generation using Prompt Engineering adalah Kerangka yang dibangun untuk mengintegrasikan proses penyusunan prompt, generasi gambar dan evaluasi kualitas hasil visual ke dalam satu sistem eksperimental yang terukur.

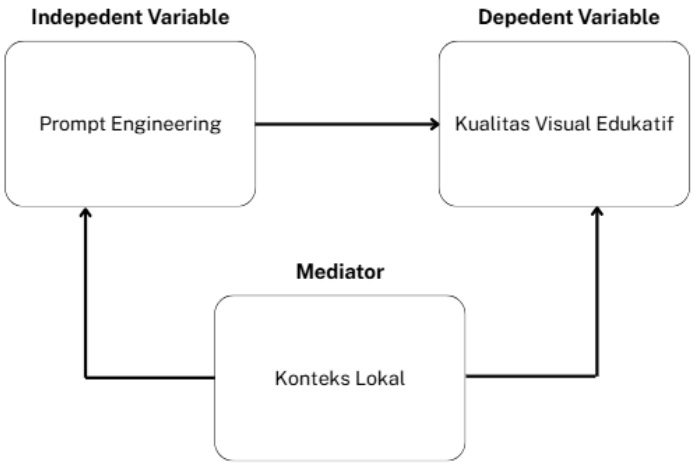
Pada tahap awal, framework ini memanfaatkan pendekatan *prompt engineering* sebagai mekanisme utama untuk menguji bagaimana variasi prompt (descriptive, instruction, dan compositional) memengaruhi kualitas gambar yang dihasilkan. Tahap kedua menggunakan model *Text-to-Image (T2I)* melalui dua model, yaitu *Stable Diffusion* (berbasis *Latent Diffusion Model*) dan *Gemini Nano Banana* (berbasis *multimodal transformer*). Keduanya merepresentasikan dua paradigma berbeda yaitu model *Open Source* yang fleksibel dan efisien secara komputasi, serta Gemini Nano yang merupakan model tertutup dengan kemampuan konteks bahasa yang tinggi. Tahap ketiga adalah *evaluation framework* yang menilai performa kedua model melalui dua pendekatan dengan evaluasi kuantitatif (menggunakan CLIPScore untuk mengukur kesamaan distribusi dan kesesuaian semantik) serta evaluasi kualitatif (menggunakan penilaian manusia untuk menilai relevansi konteks dan daya edukatif gambar).

Ketiga komponen tersebut saling terhubung dalam satu siklus proses yang bersifat iterasi, di mana hasil evaluasi digunakan untuk merefleksikan efektivitas metode prompt yang diuji. Dengan demikian, *framework* ini tidak hanya berfungsi sebagai mekanisme teknis untuk membandingkan dua model T2I, tetapi juga sebagai pendekatan konseptual untuk memahami hubungan antara bahasa, visualisasi dan konteks edukatif di lingkungan masyarakat. [42].

Dengan pendekatan ini, penelitian tidak hanya menghasilkan perbandingan performa model AI secara teknis tetapi juga membangun fondasi untuk penerapan *prompt engineering* sebagai sarana literasi AI dan pembelajaran visual yang sesuai untuk Desa Wisata Tigaraksa.

2.3.2 Kerangka Konseptual Penelitian

Kerangka konseptual penelitian ini dibangun untuk menjelaskan hubungan antara *Prompt Engineering* sebagai variabel independen (IV), *Kualitas Visual Edukatif* sebagai variabel dependen (DV), serta *Konteks Lokal* sebagai variabel mediator. Hubungan ketiga variabel tersebut mencerminkan upaya penelitian ini dalam memahami bagaimana metode penyusunan *prompt* dapat memengaruhi hasil visual edukatif yang dihasilkan oleh model *Text-to-Image (T2I*) dengan mempertimbangkan aspek budaya dan lingkungan sosial Desa Wisata Tigaraksa sebagai faktor penengah yang berperan penting.



Gambar 2.9 Kerangka Konseptual Penelitian

Secara teoretis, *Prompt Engineering* berperan sebagai komponen utama yang menentukan arah dan struktur komunikasi antara manusia dan sistem AI generatif. Variasi metode seperti *descriptive*, *instruction-based*, dan *compositional* dirancang untuk menguji sejauh mana bentuk instruksi memengaruhi hasil visual yang dihasilkan oleh model. Keluaran dari model kemudian dievaluasi

berdasarkan *Kualitas Visual Edukatif*, yang meliputi relevansi, kejelasan konsep dan daya tarik visual yang mendukung proses pembelajaran anak-anak di lingkungan edukasi berbasis komunitas.

Sementara itu, Konteks Lokal berperan sebagai mediator yang menjembatani hubungan antara metode *prompt* dan hasil visual yang dihasilkan. Dalam penelitian ini, konteks lokal mencakup nilai budaya, lingkungan sosial, serta karakteristik visual yang mewakili identitas Desa Wisata Tigaraksa. Integrasi konteks lokal diyakini dapat memperkuat makna edukatif dari visual yang dihasilkan, sehingga gambar tidak hanya relevan secara teknis, tetapi juga kontekstual, komunikatif, dan mudah dipahami oleh audiens sasaran.

Dengan demikian, hubungan antarvariabel dalam penelitian ini dapat dijelaskan sebagai berikut: semakin tepat metode *Prompt Engineering* yang digunakan, semakin tinggi Kualitas Visual Edukatif yang dihasilkan oleh model T2I terutama apabila proses generasi gambar mempertimbangkan Konteks Lokal sebagai faktor mediasi.

2.4 Tools/software yang digunakan

2.4.1 Stable Diffusion

Stable Diffusion merupakan model *Text-to-Image (T2I)* open-source yang dikembangkan oleh Stability AI. Model ini menggunakan pendekatan *Latent Diffusion Model (LDM)* yang di mana proses pembangkitan gambar dilakukan dalam *latent space* berukuran lebih kecil dibandingkan ruang piksel asli. Hal ini menjadikan Stable Diffusion lebih efisien secara komputasi tanpa mengorbankan kualitas gambar yang dihasilkan.

2.4.2 Gemini Nano Banana

Gemini Nano Banana merupakan model *multimodal generative AI* dari Google DeepMind yang mampu memproses teks dan menghasilkan keluaran visual berdasarkan pemahaman semantik lintas-modalitas. Model ini menggunakan arsitektur *transformer-based*, memungkinkan sistem untuk memahami hubungan antara teks dan gambar secara kontekstual. Tidak seperti Stable Diffusion yang lebih fokus pada fleksibilitas teknis, Gemini mengedepankan integrasi bahasa alami dan kemudahan penggunaan, sehingga sangat sesuai untuk pengguna non-teknis dalam konteks pendidikan.

2.4.3 Hugging Face

Gambar 2.10 Logo Hugging Face

Hugging Face merupakan platform open-source yang menyediakan berbagai model *AI pre-trained models* serta ekosistem untuk pengembangan aplikasi berbasis machine learning dan Generative AI. Platform ini memiliki *library* bernama *Transformers*, yang menjadi fondasi dalam pemanggilan dan integrasi model bahasa maupun *multimodal* seperti *Stable Diffusion* dan *CLIP.* Dalam

konteks penelitian ini, Hugging Face digunakan sebagai repositori model dan API integrasi untuk mengakses model *Text-to-Image* (T2I) *Stable Diffusion* .

2.4.4 Jupyter Notebook

Gambar 2.11 Logo Jupyter Notebook

Jupyter Notebook merupakan salah satu alat paling populer dalam dunia penelitian berbasis data dan *machine learning*. Platform ini memungkinkan peneliti menulis dan menjalankan kode pemrograman secara interaktif, sekaligus menampilkan grafik, tabel, maupun output visual lainnya dalam satu alur kerja yang rapi. Setiap sel dapat berisi kode, catatan penjelasan, atau hasil visualisasi, sehingga proses eksperimen menjadi lebih terstruktur dan mudah dilacak.