

Integrasi Generative AI dan Large Language Model pada Aplikasi Web untuk Transformasi Sketsa Menjadi Gambar Realistis dan Evaluasi Otomatis

Muhammad Yulasfi Pahrizal*✉, Pramudya Prima Insan✉, Nasruddin Bin Idris✉
Program Studi Sistem Informasi, Universitas Mulia, Indonesia

Abstrak: Penelitian ini bertujuan untuk merancang dan membangun sebuah aplikasi generative artificial intelligence (GenAI) berbasis web yang mampu mentransformasi sketsa dua dimensi menjadi gambar realistis secara otomatis, serta dilengkapi fitur AI review sebagai media evaluasi dan pembelajaran menggambar sketsa. Penelitian ini mengintegrasikan layanan GenAI berbasis diffusion model untuk proses generasi gambar dengan large language model (LLM) yang menganalisis sketsa, prompt, dan hasil gambar, tanpa melakukan pelatihan atau pengembangan arsitektur model AI baru. Metode penelitian yang digunakan adalah penelitian terapan (applied research) dengan pendekatan rekayasa perangkat lunak model waterfall. Aplikasi dibangun menggunakan Next.js sebagai frontend, FastAPI sebagai backend, PostgreSQL sebagai basis data, MinIO sebagai object storage, serta Docker dan Nginx untuk mendukung deployment berbasis container. Studi ini menggunakan pengujian fungsional dengan metode black box testing teknik equivalence partitioning yang menunjukkan seluruh 44 skenario uji pada sembilan modul sistem berjalan sesuai kebutuhan fungsional dengan tingkat keberhasilan 100%. Evaluasi penerimaan pengguna (User Acceptance Test) terhadap 15 responden menghasilkan skor rata-rata 4,38 dari skala 5 (87,59%) dengan kategori sangat baik, dimana fitur AI review memperoleh nilai tertinggi. Kesesuaian struktur sketsa dan prompt dinilai pengguna 81,33%; kualitas visual belum divalidasi evaluator independen. Secara praktis, aplikasi ini dapat langsung dimanfaatkan oleh desainer, ilustrator, dan kreator digital untuk mempercepat eksplorasi konsep visual tanpa memerlukan keahlian ilustrasi tingkat lanjut. Secara sosial, aplikasi ini berkontribusi pada demokratisasi proses kreatif dengan meningkatkan aksesibilitas produksi visual bagi masyarakat yang memiliki keterbatasan kemampuan menggambar. Nilai originalitas penelitian ini terletak pada penambahan fitur AI review berbasis LLM sebagai mekanisme evaluasi dan umpan balik otomatis yang tidak ditemukan pada aplikasi maupun penelitian transformasi sketsa pembandingan yang ditinjau.

Kata Kunci: AI Review; FastAPI; Generative Artificial Intelligence; Large Language Model; Next.js

Abstract: This study aims to design and develop a web-based generative artificial intelligence (GenAI) application that automatically transforms two-dimensional sketches into realistic images while providing an AI review feature for sketch evaluation and drawing guidance. This study integrates a diffusion-based GenAI service for image generation with a large language model (LLM) that analyzes sketches, prompts, and generated images, without developing or training a new AI model architecture. This study adopted an applied research approach following

the waterfall software development model. The application was built with Next.js for the frontend, FastAPI for the backend, PostgreSQL for the database, MinIO as object storage, and Docker with Nginx for containerized deployment. For functional testing this study uses the black-box method with equivalence partitioning that showed all 44 test scenarios across nine system modules met the functional requirements, achieving a 100% success rate. User Acceptance Testing (UAT) involving 15 respondents produced an average score of 4.38 out of 5 (87.59%), categorized as very good, with the AI review feature receiving the highest rating. Users rated sketch-structure and prompt fidelity at 81.33%; visual quality has not been validated by independent evaluators. Practically, the application can be used directly by designers, illustrators, and digital creators to accelerate visual concept exploration without requiring advanced illustration skills. Socially, it contributes to the democratization of the creative process by improving access to visual production for people with limited drawing ability. The originality of this research lies in the addition of an LLM-based AI review feature as an automated evaluation and feedback mechanism that was not found in the sketch-to-image applications and prior studies reviewed.

Keywords: AI Review; FastAPI; Generative Artificial Intelligence; Large Language Model; Next.js

1. PENDAHULUAN

Perkembangan teknologi kecerdasan buatan (*artificial intelligence*) mengalami kemajuan yang sangat pesat dan telah memberikan dampak signifikan di berbagai sektor, termasuk industri kreatif. Salah satu cabang AI yang berkembang pesat adalah *generative artificial intelligence* (GenAI), yaitu teknologi yang mampu menghasilkan konten visual secara otomatis berdasarkan masukan tertentu. Kehadiran GenAI membuka peluang baru dalam proses kreatif, khususnya pada tahap perancangan visual dalam produksi desain dua dimensi (2D).

Dalam industri desain, proses penciptaan konsep visual seperti desain karakter dan lingkungan umumnya dimulai dari sketsa manual. Tahapan ini membutuhkan waktu yang relatif lama serta keterampilan ilustrasi yang memadai untuk menghasilkan visual yang dapat digunakan sebagai *concept art*. Bagi desainer pemula maupun desainer dengan keterbatasan kemampuan ilustrasi, proses ini sering menjadi hambatan dalam mengembangkan ide secara cepat dan efisien, sementara tuntutan industri menghendaki kecepatan produksi dan fleksibilitas desain tanpa mengurangi kualitas hasil.

Pemanfaatan GenAI dalam mentransformasi sketsa menjadi gambar realistis mulai banyak diterapkan sebagai solusi atas permasalahan tersebut. Perkembangan model *diffusion* belakangan ini telah mendorong kemajuan pesat pada berbagai subbidang pembangkitan gambar, mulai dari *style transfer*, *image restoration*, hingga *image editing* [1]. Platform seperti Playground AI mampu mengubah sketsa hitam-putih menjadi visual yang lebih detail dan estetik dalam waktu singkat, dan terbukti dapat mempercepat proses pencarian konsep visual dalam produksi desain serta membantu desainer mengeksplorasi variasi visual secara efisien [2]. Penelitian lain mengembangkan aplikasi berbasis GenAI bernama PixAi yang berfokus pada efisiensi desain karakter bagi desainer dengan keterbatasan kemampuan ilustrasi, mampu mentransformasi sketsa kasar menjadi visual yang layak digunakan sebagai *concept art* atau aset produksi animasi 2D [3]. Pendekatan serupa diterapkan melalui *sketch-guided latent diffusion model* yang menjadikan sketsa sebagai panduan dalam sintesis gambar wajah beresolusi tinggi, sehingga gambar yang dihasilkan tetap konsisten dengan struktur komposisi sketsa masukan [4]. Selain pendekatan berbasis *diffusion model*, penerapan *generative adversarial networks* (GANs)

turut dimanfaatkan untuk mengukur tingkat kemiripan antara sketsa dan foto asli menggunakan *structural similarity index* (SSIM), sehingga membantu pengguna pemula memahami kualitas dan realisme sketsa yang dihasilkan [5].

Berdasarkan uraian tersebut, penerapan GenAI dalam transformasi sketsa ke gambar realistis memiliki potensi besar dalam meningkatkan efisiensi dan kualitas proses desain 2D. Namun demikian, terdapat celah penelitian (*research gap*) yang perlu diperhatikan: hampir seluruh sistem yang dikembangkan pada studi-studi tersebut [2], [3], [4], [5] hanya berfokus pada proses generasi gambar semata, tanpa memberikan umpan balik atau evaluasi terhadap sketsa yang dibuat maupun kesesuaian antara prompt yang diberikan dengan desain yang dihasilkan. Kondisi ini menyebabkan pengguna, terutama pemula, tidak dapat mengetahui aspek apa yang perlu diperbaiki pada sketsa mereka agar dapat menghasilkan output yang lebih baik.

Oleh karena itu, penelitian ini mengusulkan pengembangan sebuah aplikasi berbasis GenAI bernama Mulia Draw2Pic yang tidak hanya mampu mentransformasi sketsa menjadi gambar realistis, tetapi juga dilengkapi dengan fitur *AI review*. Fitur ini dirancang untuk menganalisis sketsa yang diunggah pengguna, mengevaluasi kesesuaian antara prompt yang diberikan dengan hasil desain yang dihasilkan, serta memberikan umpan balik dan saran perbaikan yang dapat digunakan sebagai panduan belajar menggambar sketsa. Dibandingkan dengan penelitian terdahulu, kontribusi utama penelitian ini terletak pada tiga aspek. Pertama, dari sisi metode, penelitian ini mengintegrasikan LLM sebagai evaluator visual-tekstual [6] pada domain transformasi sketsa, yang belum diterapkan pada penelitian sejenis [2], [3], [4], [5], [7]. Kedua, dari sisi implementasi, sistem dilengkapi dengan mekanisme *versioning* dan manajemen riwayat generasi yang memungkinkan proses eksplorasi visual dilakukan secara iteratif, bukan sekadar transformasi satu arah seperti pada [2] dan [3]. Ketiga, dari sisi hasil, evaluasi yang dihasilkan bersifat kualitatif dan *actionable* dalam bentuk bahasa alami, menggantikan pendekatan metrik numerik seperti SSIM yang digunakan pada [5]. Ringkasan posisi penelitian ini terhadap beberapa penelitian terdahulu ditunjukkan pada Tabel 1.

Berdasarkan uraian latar belakang tersebut, rumusan masalah penelitian ini adalah:

- Bagaimana merancang dan membangun aplikasi berbasis GenAI yang mampu mentransformasi sketsa dua dimensi menjadi gambar realistis untuk mendukung proses desain dalam industri kreatif.
- Bagaimana merancang dan mengimplementasikan fitur *AI review* yang mampu menganalisis sketsa, mengevaluasi kesesuaian prompt dengan hasil gambar, serta memberikan umpan balik dan saran perbaikan sebagai media pembelajaran menggambar sketsa.

Sejalan dengan rumusan masalah ini, penelitian ini bertujuan merancang dan membangun aplikasi GenAI yang mengintegrasikan model generatif *pre-trained* melalui Replicate API, menyediakan input sketsa dan prompt teks, mengimplementasikan fitur *AI review* berbasis *large language model*, serta menguji efektivitas dan penerimaan sistem melalui pengujian fungsional dan evaluasi pengguna, sehingga dapat mendukung proses desain dalam industri kreatif, khususnya bidang animasi dan ilustrasi digital.

Tabel 1. Ringkasan Posisi Penelitian terhadap Penelitian Terdahulu

No	Penelitian	Fokus Penelitian	Perbandingan dengan Penelitian Ini
1	Playground AI [2]	Mengubah sketsa hitam-putih menjadi visual yang lebih detail dan estetis.	Penelitian ini membangun aplikasi web mandiri yang terintegrasi, bukan bergantung pada antarmuka platform pihak ketiga.
2	PixAi [3]	Mentransformasi sketsa kasar menjadi <i>concept art</i>	Penelitian ini melengkapi kekurangan PixAi dengan menambahkan fitur <i>AI review</i> sebagai umpan balik otomatis.

		untuk efisiensi desain karakter.	
3	GANs & SSIM [5]	Mengukur kemiripan sketsa dan foto asli menggunakan metrik SSIM.	Penelitian ini menggantikan skor numerik dengan evaluasi berbasis LLM yang menghasilkan umpan balik bahasa alami.
4	SketchyGAN [7]	Mengembangkan arsitektur model generatif untuk sketsa berbagai kategori objek.	Penelitian ini tidak mengembangkan model baru, melainkan fokus pada implementasi aplikasi praktis berbasis model <i>pre-trained</i> .
5	LLM sebagai Evaluator [6]	Meneliti kemampuan LLM sebagai evaluator otomatis pada domain teks rekomendasi.	Penelitian ini mengadaptasi konsep tersebut ke domain visual, yaitu evaluasi sketsa dan kesesuaian hasil gambar.
6	Sketch-Guided Latent Diffusion Model [4]	Memandu sintesis gambar wajah beresolusi tinggi menggunakan sketsa sebagai kondisi pada model <i>latent diffusion</i> .	Penelitian ini mengadopsi prinsip panduan sketsa berbasis <i>diffusion model</i> yang serupa, namun diterapkan pada domain desain karakter dan lingkungan secara umum, serta dilengkapi fitur AI review yang tidak tersedia pada penelitian tersebut.

2. METODE PENELITIAN

Penelitian ini menggunakan metode penelitian terapan (*applied research*) dengan pendekatan rekayasa perangkat lunak (*software engineering*), karena penelitian berfokus pada perancangan, pengembangan, implementasi, dan pengujian aplikasi berbasis web yang memanfaatkan teknologi GenAI dan LLM. Proses penelitian dilakukan melalui beberapa tahapan, meliputi identifikasi kebutuhan sistem, studi literatur terhadap teknologi dan penelitian terkait, perancangan arsitektur sistem, implementasi aplikasi, integrasi layanan AI berbasis API, hingga pengujian dan evaluasi sistem.

2.1. Metode Pengumpulan Data

- Observasi langsung yang dilakukan terhadap proses pembuatan sketsa digital dan alur transformasi gambar untuk memahami kebutuhan pengguna dalam eksplorasi visual.
- Studi literatur yang dilakukan terhadap jurnal ilmiah, buku, dan dokumentasi teknis terkait GenAI, *large language model*, *image-to-image generation*, dan pengembangan aplikasi web modern.
- Studi dokumentasi yang dilakukan terhadap dokumentasi resmi teknologi yang digunakan, yaitu Next.js, FastAPI, Docker, PostgreSQL, MinIO, dan Nginx.
- Kuesioner yang digunakan sebagai instrumen *User Acceptance Test* (UAT) menggunakan skala Likert 1-5 yang dibagikan secara daring melalui Google Form kepada calon pengguna aplikasi, meliputi mahasiswa, desainer/ilustrator, dan pengguna umum, setelah masing-masing responden mencoba langsung semua fitur pada aplikasi.

2.2. Teknik Analisis Data

- Evaluasi persepsi kualitas hasil generatif, yaitu penilaian pengguna terhadap lima indikator pada kuesioner UAT: (1) keberhasilan transformasi sketsa menjadi gambar realistis (realisme dan kualitas visual), (2) kesesuaian komposisi hasil dengan struktur sketsa, (3) kesesuaian hasil dengan *prompt*, (4) manfaat iterasi *generate* ulang, dan (5) kewajaran waktu proses. Evaluasi ini bersifat subjektif dan dilakukan oleh pengguna, bukan oleh panel ahli independen maupun metrik objektif, sehingga dilaporkan terpisah dari pengujian fungsional.

- b. Pengujian fungsional yang dilakukan menggunakan metode *black box testing* dengan teknik *equivalence partitioning* [8] terhadap seluruh modul sistem.
- c. Evaluasi fitur AI review yang dilakukan secara kualitatif untuk menilai konsistensi struktur output, relevansi saran perbaikan prompt, dan kemampuan model dalam mempertahankan format yang ditentukan. Hal ini sejalan dengan temuan [9] bahwa konsistensi LLM sebagai evaluator masih bervariasi antar-model maupun antar-kriteria. Evaluasi tersebut dioperasionalkan melalui (i) tiga skenario *black box* pada modul AI review untuk memverifikasi konsistensi format keluaran (skor sketsa, skor prompt, saran perbaikan, dan penyimpanan versi), serta (ii) lima pernyataan UAT tentang kemudahan, relevansi, kebermanfaatan saran, pemahaman kekurangan, dan manfaat sebagai media belajar, yang dianalisis menggunakan rata-rata, simpangan baku, dan proporsi responden yang setuju (skor 4–5) untuk menilai konsistensi penilaian antar-responden. Uji pengulangan (*repeat-run*) terhadap keluaran LLM tidak dilakukan pada penelitian ini.
- d. Evaluasi penerimaan pengguna yang dilakukan melalui UAT dengan menghitung persentase kelayakan, sebagaimana lazim diterapkan pada penelitian pengujian penerimaan pengguna sistem informasi berbasis web [10], menggunakan rumus berikut.

$$\text{Persentase Kelayakan} = (\text{Skor Aktual} \div \text{Skor Ideal}) \times 100\% \quad (1)$$

dengan Skor Aktual adalah jumlah skor jawaban seluruh responden pada suatu pernyataan dan Skor Ideal adalah jumlah responden dikalikan skor maksimum (5), sehingga untuk 15 responden Skor Ideal per pernyataan adalah 75. Persentase suatu aspek dihitung dari rata-rata skor seluruh pernyataan pada aspek tersebut dibagi 5, dan sebagai pelengkap dilaporkan simpangan baku serta proporsi responden yang setuju (skor 4–5).

- e. Hasil persentase kemudian diinterpretasikan ke dalam lima kategori kelayakan, yaitu: 81%-100% (Sangat Baik); 61%-80% (Baik); 41%-60% (Cukup); 21%-40% (Kurang); dan 0%-20% (Sangat Kurang).

2.3. Instrumen, Responden, dan Prosedur UAT

Instrumen UAT terdiri atas 23 pernyataan berskala Likert 1–5 (1 = sangat tidak setuju, 5 = sangat setuju) yang dikelompokkan ke dalam enam aspek: kanvas dan alat gambar (4 pernyataan), fitur *generate* gambar (5), fitur AI review (5), riwayat dan *versioning* (2), tampilan dan UI/UX umum (2), serta penerimaan pengguna (5). Pernyataan disusun sesuai dengan fitur yang dibangun. Selain itu, instrumen memuat lima pertanyaan karakteristik responden, yaitu usia, jenis kelamin, latar belakang, frekuensi menggambar sketsa, dan pengalaman menggunakan aplikasi AI generatif.

Responden dipilih dengan teknik non-probabilitas (*purposive*), yaitu calon pengguna aplikasi yang bersedia mencoba seluruh fitur sebelum mengisi kuesioner. Jumlah responden ($n = 15$) tidak dimaksudkan untuk generalisasi statistik, melainkan untuk evaluasi penerimaan awal. Setiap responden mencoba langsung fitur kanvas, transformasi gambar, AI review, dan riwayat versi, kemudian mengisi kuesioner secara daring melalui Google Form.

Validitas isi instrumen terbatas pada kesesuaian pernyataan dengan fitur aplikasi. Reliabilitas konsistensi internal dihitung dengan *Cronbach's alpha* (ambang 0,70) dan validitas butir dengan korelasi item-total terkoreksi terhadap r -tabel ($n = 15$, $\alpha = 0,05$, $r = 0,514$); hasilnya dilaporkan pada Subbab 3.4.

2.4. Metode Pengembangan Perangkat Lunak

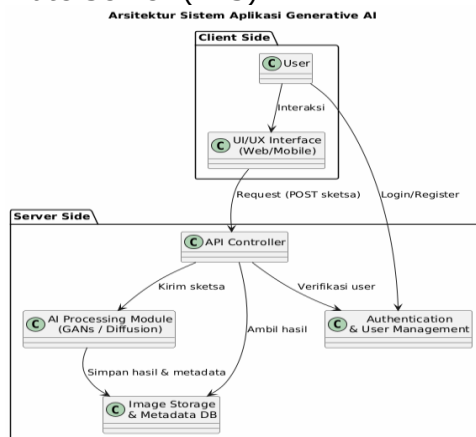
Pengembangan perangkat lunak dilakukan menggunakan metode *waterfall* [11], yang penerapannya juga banyak ditemukan pada penelitian pengembangan sistem informasi berbasis web sejenis [12], dan terdiri atas enam tahapan berurutan, yaitu:

- Requirement gathering and analysis*, dilakukan untuk mengidentifikasi kebutuhan fungsional seperti fitur upload sketsa, generasi gambar, AI review, dan penyimpanan riwayat, serta kebutuhan non-fungsional seperti performa dan kemudahan penggunaan.
- Design*, dilakukan untuk merancang *activity diagram*, *entity relationship diagram*, arsitektur frontend-backend, dan struktur database.
- Implementation*, menerjemahkan seluruh rancangan ke dalam kode program menggunakan Next.js pada sisi frontend dan FastAPI pada sisi backend.
- Integration and testing*, menggabungkan seluruh komponen sistem dan menguji fungsionalitasnya secara menyeluruh.
- Verification*, melibatkan pengujian langsung oleh calon pengguna untuk memastikan sistem memenuhi kebutuhan yang diharapkan.
- Operation and maintenance*, mencakup pengoperasian sistem serta pemeliharaan berkala berupa perbaikan bug, pembaruan dependensi, dan pengembangan fitur tambahan.

2.5. Arsitektur Sistem

Arsitektur sistem dirancang menggunakan pendekatan *client-server* dengan komponen yang didistribusikan sebagai container Docker, sebagaimana ditunjukkan pada Gambar 1. Pengguna berinteraksi dengan antarmuka web yang mengirimkan permintaan berupa sketsa dan prompt ke *API Controller*. *API Controller* meneruskan data ke *AI Processing Module* untuk proses generasi gambar, sekaligus berkoordinasi dengan modul *Authentication & User Management* untuk proses login/registrasi. Hasil generasi beserta metadata disimpan pada komponen *Image Storage & Metadata DB*.

Adapun Nginx berperan sebagai *reverse proxy* yang mendistribusikan lalu lintas HTTP ke layanan Next.js sebagai frontend sekaligus *backend for frontend* yang terhubung ke PostgreSQL melalui Prisma ORM, dan ke layanan FastAPI yang menangani proses generate gambar dan *AI review* melalui Replicate API. Gambar hasil transformasi disimpan secara permanen pada MinIO sebagai *object storage*, sehingga gambar tetap dapat diakses meskipun URL sementara dari Replicate API telah kedaluwarsa. Proses *deployment* dilakukan secara otomatis melalui *pipeline* CI/CD yang terdiri atas GitHub Actions untuk *build* dan *push image* Docker ke GitHub Container Registry, serta Jenkins untuk *deployment* pada *Virtual Private Server* (VPS).



Gambar 1. Arsitektur Sistem Aplikasi Mulia Draw2Pic

3. HASIL DAN PEMBAHASAN

Penelitian ini menghasilkan sebuah aplikasi berbasis GenAI bernama Mulia Draw2Pic yang mampu mentransformasi sketsa dua dimensi menjadi gambar realistis secara otomatis, sekaligus dilengkapi dengan fitur *AI review* sebagai komponen evaluatif berbasis *large language model*. Aplikasi dikembangkan dengan arsitektur *client-server* yang terpisah secara fungsional antara frontend berbasis React dengan framework Next.js dan backend berbasis Python dengan framework FastAPI yang menyediakan RESTful API.

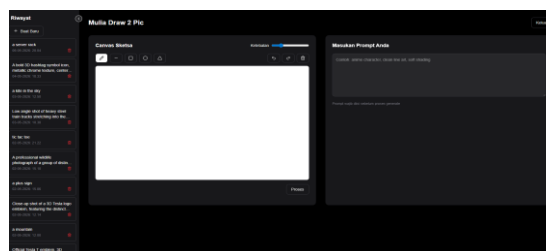
3.1. Implementasi Sistem

Proses transformasi sketsa ke gambar realistis memanfaatkan model generatif berbasis *diffusion*, yaitu google/nano-banana-2, yang diakses melalui platform Replicate API dan mampu menganalisis sketsa tangan sebagai panduan komposisi untuk menghasilkan gambar realistis sesuai deskripsi prompt. Fitur *AI review* diimplementasikan menggunakan model google/gemini-3-flash yang diakses melalui mekanisme *streaming* pada Replicate API, mampu menganalisis konten gambar sketsa dan hasil transformasi secara bersamaan untuk menghasilkan evaluasi terstruktur berupa penilaian kualitas sketsa, evaluasi prompt, perbandingan hasil gambar dengan prompt, diagnosis penyebab ketidaksempurnaan, dan rekomendasi perbaikan prompt.

Penyimpanan gambar dikelola menggunakan MinIO sebagai *object storage self-hosted* yang berjalan pada infrastruktur VPS yang sama dengan layanan backend dan frontend, sehingga gambar hasil transformasi tetap dapat diakses meskipun URL sementara dari Replicate API telah kedaluwarsa. Manajemen pengguna mengimplementasikan autentikasi berbasis email dan password menggunakan Prisma ORM dengan database PostgreSQL, dilengkapi mekanisme sesi berbasis HTTP Cookie httpOnly serta proteksi route menggunakan Next.js Middleware. Setiap sesi generasi disimpan sebagai entitas *Generation*, dan setiap iterasi *generate* atau *review* dalam satu sesi disimpan sebagai entitas *GenerationVersion* dengan nomor versi *auto-increment*, sehingga pengguna dapat memuat ulang sesi sebelumnya dan berpindah antar-versi melalui komponen *timeline*. Seluruh komponen sistem di-*deploy* menggunakan Docker Container dengan Nginx sebagai *reverse proxy*, serta *pipeline* CI/CD yang terdiri atas GitHub Actions untuk *build* dan *push image* ke GitHub Container Registry dan Jenkins untuk proses *deployment* ke VPS.

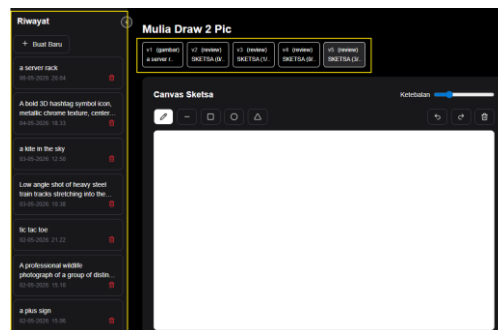
3.2. Antarmuka dan Alur Penggunaan Aplikasi

Halaman utama aplikasi, sebagaimana ditunjukkan pada Gambar 2, terdiri atas empat area utama, yaitu sidebar riwayat generasi yang dapat dilipat di sisi kiri, header berisi judul aplikasi dan tombol keluar, *timeline versioning horizontal* pada bagian atas konten utama, serta dua panel konten yang terdiri atas kanvas sketsa dengan fitur *AI review* pada panel kiri dan input prompt beserta tampilan hasil gambar pada panel kanan. Kanvas digital diimplementasikan menggunakan HTML5 Canvas yang menyediakan lima alat gambar, yaitu *Pen*, *Line*, *Rect*, *Circle*, dan *Triangle*, dengan mekanisme *snapshot-restore* untuk pratinjau bentuk secara *realtime*, pengaturan ketebalan garis, serta operasi *undo*, *redo*, dan *clear*.



Gambar 2. Halaman Utama Aplikasi Mulia Draw2Pic

Proses *generate* gambar dimulai ketika pengguna menekan tombol Proses setelah mengisi prompt dan menggambar sketsa. Sistem melakukan validasi terhadap prompt dan isi kanvas, mengonversi kanvas menjadi data base64 PNG, kemudian mengirimkannya ke *endpoint* /api/generate. Backend FastAPI memanggil model google/nano-banana-2 melalui Replicate API, mengunggah hasil gambar ke MinIO, dan mencatat metadata sebagai GenerationVersion baru sebelum menampilkan hasil kepada pengguna. Proses *AI review* mengikuti alur serupa: setelah gambar hasil generate tersedia, sistem mengonversi sketsa dan gambar hasil menjadi base64, mengirimkannya ke *endpoint* /api/review, kemudian menampilkan evaluasi dari model google/gemini-3-flash secara progresif melalui *streaming* sebelum hasil review disimpan sebagai versi baru pada *timeline*, sebagaimana ditunjukkan pada Gambar 3. Fitur ini memungkinkan pengguna melakukan iterasi secara terstruktur: setelah menerima evaluasi, pengguna dapat memperbaiki sketsa atau prompt dan melakukan *generate* ulang, dengan seluruh riwayat versi tetap tersimpan dan dapat dipulihkan kembali.



Gambar 3. Timeline Versioning dan Riwayat Generasi

3.3. Pengujian Fungsional Sistem (Black Box Testing)

Pengujian sistem dilakukan menggunakan metode *black box testing* dengan teknik *equivalence partitioning* [8], yaitu teknik pengujian yang skenario uji beserta kriteria keberhasilannya dirancang secara sistematis di muka berdasarkan partisi kelas input pada spesifikasi fungsional sistem, sehingga hasil setiap skenario bersifat objektif dan dapat diverifikasi ulang, tidak bergantung pada penilaian subjektif penguji. Dengan karakteristik ini, validitas pengujian lebih ditentukan oleh kelengkapan cakupan skenario terhadap seluruh kelas input dan fungsi sistem. Eksekusi pengujian dilakukan oleh satu penguji dengan latar belakang profesional sebagai animator digital yang memahami konteks alur kerja sketsa-ke-gambar, sehingga memvalidasi kesesuaian hasil sistem terhadap 44 skenario uji secara kompeten; penilaian persepsi pengguna yang bersifat subjektif ditempatkan terpisah pada tahap UAT yang melibatkan 15 responden. Pengujian mencakup sembilan modul fungsional dengan ringkasan hasil pengujian yang ditunjukkan pada Tabel 2.

Tabel 2. Ringkasan Pengujian Black Box Testing Aplikasi Mulia Draw2Pic

N o	Modul Fungsional	Jumlah Skenario Uji	Contoh Hasil yang Diharapkan	Keterangan
1	Autentikasi (Registrasi, Login, Logout)	9	Pengguna terdaftar/masuk sesuai validasi data; email duplikat, umur <18, dan password tidak sesuai kriteria ditolak sistem.	Berhasil
2	Kanvas Digital &	9	Lima alat gambar (Pen, Line, Rect, Circle, Triangle) beserta undo, redo, clear, dan pengaturan ketebalan garis berfungsi baik.	Berhasil

Alat Gambar				
3 Validasi Input	5	Sistem menampilkan alert dan menghentikan proses generate/review/registrasi jika field wajib (prompt, kanvas, data registrasi) kosong.	Berhasil	
4 Fitur Generate Gambar	5	Sketsa berhasil ditransformasi menjadi gambar realistis, tersimpan ke MinIO dan metadata GenerationVersion tercatat di database.	Berhasil	
5 Fitur AI Review	3	Evaluasi streaming tampil bertahap dengan skor sketsa, skor prompt, dan saran perbaikan, lalu tersimpan sebagai versi baru.	Berhasil	
6 Manajemen Riwayat & Versioning	6	Sesi dapat dimuat ulang, berpindah versi, dihapus, dan dibuat baru; pemulihan versi bertipe review menampilkan gambar yang benar.	Berhasil	
7 Fitur Download	1	Gambar hasil transformasi berhasil diunduh ke perangkat lokal pengguna dalam format PNG.	Berhasil	
8 Pipeline CI/CD	3	Push tag Git memicu build otomatis oleh GitHub Actions ke GHCR, dan Jenkins berhasil melakukan deployment tanpa downtime.	Berhasil	
9 Penanganan Kesalahan Umum	3	Akses tanpa sesi valid dialihkan ke login, kegagalan API Replicate mengembalikan error terkontrol, dan akses data pengguna lain ditolak (403).	Berhasil	

Berdasarkan hasil pengujian, seluruh 44 skenario uji pada sembilan modul fungsional menghasilkan *output* yang sesuai dengan hasil yang diharapkan, sehingga persentase keberhasilan pengujian dihitung menggunakan rumus

$$\text{Persentase} = (\text{Skor Aktual} \div \text{Skor Ideal}) \times 100\% = (44/44) \times 100\% = 100\%(2)$$

Hasil ini menunjukkan bahwa seluruh fitur, mulai dari autentikasi, kanvas digital, generate gambar, *AI review*, manajemen riwayat, download, *pipeline* CI/CD, hingga penanganan kesalahan, telah berjalan sesuai kebutuhan fungsional sistem. Namun demikian, tingkat keberhasilan 100% tersebut hanya menyatakan bahwa 44 skenario fungsi berjalan sesuai spesifikasi dan tidak dapat diartikan sebagai bukti bahwa kualitas gambar yang dihasilkan sudah optimal; kualitas hasil generatif dibahas secara terpisah pada Subbab 3.5.

3.4. Evaluasi Penerimaan Pengguna (User Acceptance Test)

Selain pengujian fungsional, evaluasi sistem juga dilakukan melalui *User Acceptance Test* (UAT) untuk mengukur penerimaan dan kepuasan pengguna terhadap aplikasi Mulia Draw2Pic [10]. Pengujian melibatkan 15 responden yang merupakan calon pengguna aplikasi, terdiri atas 10 responden laki-laki dan 5 responden perempuan dengan rentang usia 19-36 tahun (rata-rata 23,3 tahun), yang berasal dari kalangan mahasiswa, desainer/ilustrator, dan karyawan/pekerja. Setiap responden diminta mencoba langsung fitur kanvas, transformasi gambar, dan *AI review* sebelum mengisi kuesioner berskala Likert 1-5 melalui Google Form. Karakteristik responden dirinci pada Tabel 3.

Tabel 3. Karakteristik Responden UAT (n = 15)

Karakteristik	Kategori	Jumlah	Persentase
Jenis kelamin	Pria	10	66,7%
	Wanita	5	33,3%
Latar belakang	Mahasiswa	10	66,7%

Karakteristik	Kategori	Jumlah	Persentase
Usia	Desainer/ilustrator	2	13,3%
	Karyawan/pekerja	3	20,0%
	19–36 tahun (rata-rata 23,3; median 22; SD 4,0)	15	100%
Frekuensi menggambar sketsa	Sering (hampir setiap minggu)	2	13,3%
	Cukup sering (beberapa kali sebulan)	5	33,3%
	Jarang (beberapa kali setahun)	6	40,0%
Pengalaman aplikasi AI generatif	Tidak pernah	2	13,3%
	Sering	4	26,7%
	Jarang	9	60,0%
	Belum pernah	2	13,3%

Responden didominasi mahasiswa (66,7%) dan laki-laki (66,7%) dengan usia relatif muda (median 22 tahun). Sebanyak 8 dari 15 responden (53,3%) jarang atau tidak pernah menggambar sketsa dan 11 responden (73,3%) jarang atau belum pernah menggunakan aplikasi AI generatif, sehingga sampel merepresentasikan pengguna pemula yang menjadi target aplikasi. Di sisi lain, hanya dua responden berlatar belakang desainer/ilustrator sehingga penilaian praktisi profesional masih terbatas. Rekapitulasi skor rata-rata untuk setiap aspek evaluasi ditunjukkan pada Tabel 4.

Tabel 4. Rekapitulasi Skor User Acceptance Test per Aspek Evaluasi

Aspek Evaluasi	Rata-rata Skor	Persentase	Kategori
Kanvas & Alat Gambar	4,25	85,00%	Sangat Baik
Fitur Generate Gambar	4,17	83,47%	Sangat Baik
Fitur AI Review	4,52	90,40%	Sangat Baik
Riwayat & Versioning	4,37	87,33%	Sangat Baik
Tampilan & UI/UX Umum	4,50	90,00%	Sangat Baik
Penerimaan Pengguna (Acceptance)	4,51	90,13%	Sangat Baik
Rata-rata Keseluruhan	4,38	87,59%	Sangat Baik

Seluruh aspek evaluasi memperoleh persentase kelayakan di atas 80%, sehingga berada pada kategori Sangat Baik. Aspek Fitur *AI Review* memperoleh persentase tertinggi (90,40%), yang mengindikasikan bahwa fitur *AI review* sebagai kontribusi utama penelitian ini dinilai sangat relevan, mudah dipahami, dan bermanfaat sebagai media evaluasi maupun pembelajaran sketsa. Aspek fitur *generate* gambar memperoleh persentase terendah (83,47%), terutama dipengaruhi oleh persepsi sebagian responden terhadap waktu proses generate gambar yang dirasa belum sepenuhnya optimal. Secara keseluruhan, hasil UAT memperoleh rata-rata skor 4,38 dari skala 5 dengan persentase kelayakan 87,59%, yang menunjukkan bahwa aplikasi Mulia Draw2Pic diterima dengan sangat baik oleh pengguna. Distribusi jawaban menunjukkan dispersi yang rendah: simpangan baku butir rata-rata per aspek berkisar 0,56–0,81, 89,3% dari 345 jawaban bernilai 4–5, dan hanya dua jawaban (0,6%) bernilai di bawah 3, keduanya pada pernyataan kewajaran waktu proses *generate*. Reliabilitas konsistensi internal seluruh instrumen tergolong baik (*Cronbach's alpha* = 0,854 untuk 23 butir). Namun, korelasi item-total terkoreksi berkisar 0,00–0,71 dan hanya 9 dari 23 butir melampaui r-tabel 0,514; validitas butir dengan demikian dinyatakan terbatas, yang kemungkinan dipengaruhi efek langit-langit (89,3% jawaban bernilai 4–5) dan ukuran sampel yang kecil.

3.5. Evaluasi Kualitas Hasil Generatif

Agar kualitas keluaran tidak disamakan dengan kebenaran fungsional, kualitas hasil generatif dievaluasi terpisah melalui lima indikator persepsi pengguna yang diambil dari butir UAT, yaitu keberhasilan transformasi (realisme dan kualitas visual), kesesuaian komposisi dengan struktur sketsa, kesesuaian dengan *prompt*, manfaat iterasi, dan kewajaran waktu proses. Hasilnya ditunjukkan pada Tabel 5.

Tabel 5. Indikator Persepsi Pengguna terhadap Kualitas Hasil Generatif (n = 15)

Indikator	Rata-rata	SD	Persentase	Setuju (skor 4–5)	Skor ≤ 3
Keberhasilan transformasi (realisme dan kualitas visual)	4,33	0,72	86,67%	13 (86,7%)	2
Kesesuaian komposisi dengan struktur sketsa	4,07	0,80	81,33%	11 (73,3%)	4
Kesesuaian hasil dengan <i>prompt</i>	4,07	0,70	81,33%	12 (80,0%)	3
Manfaat iterasi <i>generate</i> ulang	4,47	0,64	89,33%	14 (93,3%)	1
Kewajaran waktu proses <i>generate</i>	3,93	1,16	78,67%	9 (60,0%)	6

Keberhasilan transformasi dinilai paling tinggi (86,67%), sedangkan kesesuaian terhadap struktur sketsa dan terhadap *prompt* masing-masing 81,33%. Pada kedua indikator tersebut tidak ada responden yang memberi skor di bawah 3, tetapi 4 responden (26,7%) bersikap netral terhadap kesesuaian struktur sketsa dan 3 responden (20,0%) terhadap kesesuaian *prompt*, sehingga persepsi positif dominan namun belum seragam. Kewajaran waktu proses memiliki dispersi terbesar di seluruh instrumen (SD 1,16) dengan hanya 9 dari 15 responden yang setuju.

Indikator di atas merupakan persepsi pengguna yang sebagian besar bukan praktisi profesional (Tabel 3), bukan pengukuran objektif. Penelitian ini belum menggunakan panel evaluator independen dengan rubrik terstruktur maupun metrik otomatis seperti *structural similarity index* (SSIM), *learned perceptual image patch similarity* (LPIPS), atau skor kesesuaian teks–gambar berbasis CLIP. Oleh karena itu, pernyataan bahwa sistem menghasilkan gambar “realistis” dan “konsisten dengan struktur sketsa” dibatasi sebagai temuan persepsi pengguna dan bukan bukti kuantitatif atas kualitas citra. Evaluasi lanjutan direkomendasikan memakai rubrik 1–5 pada empat dimensi (kesesuaian struktur sketsa, kesesuaian *prompt*, kualitas visual, dan tingkat realisme) oleh sekurang-kurangnya tiga evaluator independen disertai ukuran kesepakatan antar-penilai.

3.6. Evaluasi Kualitas dan Konsistensi AI Review

Kualitas dan konsistensi fitur *AI review* dievaluasi pada dua tingkat. Pertama, konsistensi format keluaran diverifikasi melalui tiga skenario *black box* pada modul *AI review* (Tabel 2) yang seluruhnya sesuai harapan, yaitu evaluasi tampil bertahap melalui *streaming* dengan skor sketsa, skor *prompt*, dan saran perbaikan, lalu tersimpan sebagai versi baru. Kedua, persepsi pengguna terhadap isi evaluasi diukur melalui lima pernyataan UAT yang hasilnya ditunjukkan pada Tabel 6.

Tabel 6. Evaluasi Persepsi Pengguna terhadap Fitur AI Review (n = 15)

Pernyataan	Rata-rata	SD	Persentase	Setuju (skor 4–5)	Skor ≤ 3
<i>AI review</i> mudah dipahami dan digunakan	4,40	0,74	88,00%	13 (86,7%)	2

Pernyataan	Rata-rata	SD	Persentase	Setuju (skor 4–5)	Skor ≤ 3
Evaluasi relevan dengan sketsa dan <i>prompt</i>	4,47	0,74	89,33%	13 (86,7%)	2
Saran perbaikan <i>prompt</i> membantu dan dapat langsung diterapkan	4,40	0,63	88,00%	14 (93,3%)	1
Membantu memahami kekurangan sketsa atau <i>prompt</i>	4,53	0,64	90,67%	14 (93,3%)	1
Bermanfaat sebagai media belajar menggambar	4,80	0,56	96,00%	14 (93,3%)	1

Seluruh pernyataan memperoleh persentase di atas 88%, simpangan baku berkisar 0,56–0,74, dan tidak ada jawaban di bawah 3, yang menunjukkan penilaian antar-responden yang cukup seragam. Manfaat sebagai media belajar menjadi butir tertinggi di seluruh instrumen (96,00%; 13 dari 15 responden memberi skor 5), sedangkan kemudahan penggunaan dan kebermanfaatan saran *prompt* relatif lebih rendah (88,00%). Dengan 2 responden bersikap netral pada kemudahan dan relevansi, ruang perbaikan terletak pada kejelasan penyajian evaluasi.

Temuan ini terbatas pada persepsi dan format keluaran. Penelitian ini tidak membandingkan isi *AI review* dengan penilaian ahli (*ground truth*) dan tidak menguji keluaran yang sama secara berulang, padahal keluaran LLM bersifat probabilistik dan konsistensinya dilaporkan bervariasi antar-model maupun antar-kriteria [9]. Dengan demikian, kata “konsisten” pada fitur ini hanya mengacu pada konsistensi format dan keseragaman persepsi pengguna. Evaluasi lanjutan perlu mengulang review pada sampel sketsa yang sama beberapa kali untuk mengukur variansi skor dan menghitung kesepakatannya dengan penilaian ahli.

3.7. Pembahasan

Sistem yang dibangun memiliki sejumlah kelebihan. Pertama, arsitektur yang bersih dan terstruktur dengan pemisahan yang jelas antara *frontend* Next.js, *backend* layanan AI FastAPI, dan *backend-for-frontend* menghasilkan sistem yang mudah dipahami dan dikembangkan lebih lanjut. Kedua, sistem *versioning* yang komprehensif memberikan nilai tambah signifikan karena setiap iterasi generasi dan review tersimpan serta dapat dimuat ulang, menjadikan aplikasi ini bukan sekadar alat sekali pakai melainkan lingkungan kerja yang mendukung proses iteratif. Ketiga, penggunaan MinIO sebagai *object storage self-hosted* memastikan persistensi gambar tanpa bergantung pada URL sementara dari Replicate API yang dapat kedaluwarsa. Keempat, *pipeline* CI/CD berbasis GitHub Actions dan Jenkins meminimalkan risiko kesalahan manusia dalam proses *deployment*. Kelima, fitur *AI review* mampu memberikan evaluasi yang komprehensif dan kontekstual karena model dapat menganalisis gambar dan teks secara bersamaan, bukan hanya berdasarkan teks *prompt* semata. Keenam, kemudahan akses terhadap proses transformasi visual yang ditawarkan aplikasi ini sejalan dengan wacana demokratisasi dalam praktik pembuatan gambar berbasis GenAI, yang menyoroti pergeseran akses, keterampilan, dan keberagaman dalam produksi visual kontemporer [13], sehingga pengguna dengan keterbatasan kemampuan ilustrasi tetap dapat berpartisipasi dalam proses kreatif secara setara.

Di samping kelebihan tersebut, sistem juga memiliki sejumlah kelemahan yang perlu diperhatikan. Pertama, waktu respons proses *generate* gambar sangat bergantung pada ketersediaan dan performa Replicate API, sehingga pengguna dapat mengalami waktu tunggu tanpa kepastian ketika terjadi antrean tinggi. Kedua, layanan Replicate API juga dikenakan biaya berdasarkan konsumsi sumber daya; berdasarkan data penggunaan

selama pengembangan, biaya model google/gemini-3-flash mencapai sekitar USD 0,03 dan model google/nano-banana-2 mencapai sekitar USD 1,88 untuk menghasilkan 28 gambar, dengan total biaya pengujian sebesar USD 1,91. Ketiga, kanvas digital saat ini juga hanya mendukung input melalui *mouse* dan belum mendukung *touch events* untuk layar sentuh. Keempat, gambar sketsa masih disimpan sebagai data base64 pada database PostgreSQL sehingga berpotensi membengkakkan ukuran basis data. Kelima, belum terdapat mekanisme *rate limiting* pada *endpoint* FastAPI yang dapat melindungi sistem dari penyalahgunaan.

Dibandingkan dengan penelitian-penelitian sebelumnya, sistem yang dikembangkan memiliki perbedaan fundamental dalam hal orientasi dan tujuan. Penelitian [7] dan [14] berfokus pada pengembangan arsitektur model generatif baru dari sudut pandang akademis dan teknis, tanpa membangun aplikasi yang dapat langsung digunakan oleh pengguna umum. Penelitian PixAi [3] memiliki orientasi yang paling dekat dengan penelitian ini, yaitu membangun aplikasi praktis bagi desainer, namun belum menyertakan mekanisme evaluasi atau umpan balik otomatis. Penelitian [5] menggunakan SSIM sebagai metrik evaluasi kemiripan, tetapi bersifat numerik dan kurang *actionable* bagi pengguna awam. Penelitian ini mengisi celah tersebut dengan mengintegrasikan fitur *AI review* berbasis *large language model* yang memberikan evaluasi kualitatif dan saran perbaikan yang dapat langsung diterapkan, sejalan dengan temuan [6] mengenai efektivitas LLM sebagai evaluator konten. Kemampuan fitur ini dalam menilai kesesuaian prompt turut relevan dengan temuan bahwa pemilihan dan penyusunan kata kunci dalam prompt secara signifikan memengaruhi kualitas visual yang dihasilkan model generatif [15], sehingga evaluasi otomatis terhadap struktur prompt oleh fitur *AI review* memiliki dasar yang kuat dalam literatur. Sistem ini sekaligus dilengkapi dengan *versioning*, manajemen riwayat, dan *pipeline* CI/CD, sehingga menghasilkan sebuah sistem yang dapat dioperasikan dan dikembangkan lebih lanjut, bukan sekadar prototipe penelitian.

Skor tertinggi fitur *AI review* (90,40%) dapat dijelaskan oleh beberapa faktor. Pertama, umpan balik diberikan atas sketsa dan *prompt* milik responden sendiri bersama hasil gambarnya, sehingga terasa relevan (89,33% responden setuju). Kedua, mayoritas responden jarang atau tidak pernah menggambar (53,3%), sehingga nilai edukatif umpan balik terasa besar, terlihat dari butir media belajar yang mencapai 96,00%. Ketiga, keluaran yang terstruktur dan ditampilkan secara progresif melalui *streaming* memberi respons yang cepat dipahami. Penjelasan ini bersifat interpretatif karena penelitian tidak membandingkan kondisi dengan dan tanpa *AI review*, dan skor yang tinggi dapat dipengaruhi oleh dominasi pengguna pemula serta kecenderungan menilai positif (*acquiescence bias*).

Fitur *generate* gambar memperoleh skor terendah (83,47%) terutama karena butir kewajaran waktu proses (78,67%, hanya 9 dari 15 responden setuju, dengan dispersi terbesar) serta butir kesesuaian struktur sketsa dan *prompt* (masing-masing 81,33%). Waktu proses bergantung pada antrean dan latensi layanan Replicate API sehingga tidak sepenuhnya dapat dikendalikan aplikasi. Kesesuaian struktur dan *prompt* yang belum seragam dapat dikaitkan dengan sifat probabilistik model generatif pra-latih dan dengan tidak adanya kondisi struktural eksplisit pada alur ini, berbeda dengan model sketch-guided yang dilatih pada pasangan sketsa-gambar untuk mempertahankan geometri [4]. Hipotesis ini perlu diuji dengan evaluasi objektif. Pada aspek kanvas, butir kelengkapan alat gambar (76,00%) merupakan yang terendah di seluruh instrumen dan selaras dengan keterbatasan input yang hanya mendukung *mouse* dengan lima alat dasar.

Perbandingan dengan penelitian terdahulu pada studi ini bersifat kualitatif. Penelitian [4] melaporkan evaluasi kualitas sintesis pada model yang dilatih khusus, dan penelitian [5] memakai SSIM, sedangkan penelitian ini tidak mengukur kualitas citra secara objektif sehingga hasilnya tidak dapat dibandingkan secara numerik dengan keduanya. Kontribusi

penelitian ini bersifat terapan-integratif: komponen *diffusion* dan LLM adalah layanan pra-latih, dan kebaruannya terletak pada rancangan alur yang memasangkan sketsa, *prompt*, dan hasil gambar sebagai masukan evaluasi multimodal beserta *versioning* iteratif dalam sistem yang dapat dioperasikan, bukan pada algoritma baru. Perbedaan dengan [3] dan [5] pada Tabel 1 tetap berupa perbedaan cakupan fitur.

Keterbatasan penelitian berdampak langsung pada validitas, skalabilitas, dan keamanan sistem, sebagaimana dirangkum pada Tabel 7. Sebagai gambaran biaya, 28 gambar menghabiskan sekitar USD 1,88 atau sekitar USD 0,067 per gambar, sehingga secara ekstrapolasi linear 1.000 generasi memerlukan sekitar USD 67 untuk model generasi saja.

Tabel 7. Keterbatasan Penelitian, Dampaknya, dan Arah Pengembangan

Keterbatasan	Dampak	Arah pengembangan
Ketergantungan pada Replicate API dan biaya per generasi	Skalabilitas dan ketersediaan: latensi di luar kendali aplikasi (kewajaran waktu 78,67%) dan biaya tumbuh linear terhadap jumlah pengguna	Antrean tugas, <i>caching</i> , kuota per pengguna, serta penyedia atau model alternatif
Belum ada <i>rate limiting</i> pada <i>endpoint</i> FastAPI	Keamanan: <i>endpoint</i> dapat disalahgunakan untuk menghabiskan kredit API dan membebani server	<i>Rate limiting</i> per pengguna/IP, kuota harian, dan <i>session expiration</i>
Sketsa disimpan sebagai base64 di PostgreSQL	Skalabilitas: ukuran basis data membengkak (base64 sekitar sepertiga lebih besar daripada biner) dan memperlambat <i>backup</i> serta kueri	Menyimpan sketsa di MinIO dan hanya menyimpan referensinya di basis data
Input hanya <i>mouse</i> dengan lima alat dasar	Validitas penilaian kanvas (kelengkapan alat 76,00%) dan aksesibilitas pada layar sentuh atau <i>stylus</i>	Dukungan <i>pointer/touch events</i> , <i>stylus</i> , dan penambahan alat
$n = 15$, sampel non-probabilitas, dua responden desainer/ilustrator, validitas butir terbatas	Validitas eksternal dan validitas butir terbatas sehingga hasil UAT tidak dapat digeneralisasi	Sampel lebih besar dan berstrata, <i>System Usability Scale</i> , dan revisi butir instrumen
Kualitas citra belum diukur objektif dan <i>AI review</i> belum diuji pengulangan	Validitas konstruk klaim realisme, kesesuaian struktur, dan konsistensi <i>AI review</i>	Panel evaluator independen, metrik objektif, uji <i>repeat-run</i> , dan perbandingan dengan penilaian ahli

Secara praktis, hasil ini menunjukkan bahwa pengguna pemula dapat menerima alur sketsa-*generate-review*-iterasi sebagai sarana eksplorasi visual dan belajar menggambar. Secara ilmiah, penelitian ini memperluas gagasan LLM sebagai evaluator [6] ke domain visual, tetapi bukti yang disajikan masih pada tingkat persepsi pengguna sehingga memerlukan studi terkontrol lanjutan.

4. KESIMPULAN

Penelitian ini berhasil merancang dan membangun aplikasi berbasis web bernama Mulia Draw2Pic yang memanfaatkan teknologi *generative artificial intelligence* untuk mentransformasi sketsa dua dimensi menjadi gambar realistis secara otomatis, menggunakan arsitektur *client-server* dengan Next.js sebagai *frontend*, FastAPI sebagai *backend* layanan AI, PostgreSQL sebagai *database*, MinIO sebagai *object storage*, serta Docker dan Nginx sebagai infrastruktur *deployment*. Integrasi layanan GenAI melalui Replicate API berhasil diimplementasikan dan menghasilkan transformasi sketsa menjadi

gambar yang berjalan sesuai spesifikasi fungsional (Subbab 3.3) dan dinilai positif oleh pengguna pada aspek fitur *generate* gambar dalam evaluasi UAT (Subbab 3.4), sementara fitur *AI review* berhasil diimplementasikan sebagai komponen evaluatif yang mampu menganalisis kualitas sketsa, mengevaluasi kesesuaian prompt dan hasil gambar, serta memberikan saran perbaikan secara otomatis sehingga memiliki nilai edukatif bagi pengguna. Namun, kualitas hasil generatif (realisme dan kesesuaian struktur sketsa) baru didukung oleh persepsi pengguna (81,33%–86,67%; Subbab 3.5) dan belum oleh evaluator independen atau metrik objektif, dan konsistensi fitur *AI review* baru diverifikasi pada tingkat format keluaran serta persepsi pengguna (Subbab 3.6), bukan melalui uji pengulangan.

Sistem manajemen riwayat dan *versioning* yang diterapkan memungkinkan pengguna menyimpan, memuat ulang, dan membandingkan setiap iterasi hasil *generate* maupun *review* secara terstruktur, sehingga mendukung proses eksplorasi visual secara bertahap. Pengujian fungsional menggunakan metode *black box testing* dengan teknik *equivalence partitioning* [8] menunjukkan bahwa seluruh modul aplikasi berjalan sesuai kebutuhan fungsional sistem, sementara infrastruktur *deployment* berbasis Docker Container, GitHub Actions, Jenkins, dan Nginx berhasil mendukung proses deployment yang terintegrasi dan konsisten pada lingkungan VPS. Hasil evaluasi penerimaan pengguna melalui UAT menunjukkan tingkat penerimaan yang sangat baik, dengan fitur *AI review* memperoleh penilaian tertinggi yang mengonfirmasi relevansi dan kebermanfaatannya sebagai media evaluasi maupun pembelajaran sketsa, sedangkan aspek fitur *generate* gambar memperoleh penilaian relatif terendah terutama terkait waktu proses *generate* gambar. Secara keseluruhan, penelitian ini menghasilkan aplikasi GenAI berbasis web yang tidak hanya berfungsi sebagai alat transformasi visual dari sketsa menjadi gambar realistis, tetapi juga sebagai media evaluasi dan pembelajaran interaktif berbasis AI bagi pengguna, khususnya desainer pemula dan kreator digital. Penelitian selanjutnya disarankan untuk menambahkan dukungan input berbasis layar sentuh dan stylus, memperluas pilihan model GenAI untuk berbagai gaya visual, memindahkan penyimpanan sketsa ke *object storage*, menambahkan mekanisme keamanan seperti *rate limiting* dan *session expiration*, serta memperluas evaluasi pengguna dengan jumlah responden yang lebih besar dan metode pengukuran *usability* terstandar seperti *System Usability Scale*. Penelitian lanjutan juga disarankan mengevaluasi kualitas hasil generatif dengan panel evaluator independen berbasis rubrik (kesesuaian struktur sketsa, kesesuaian *prompt*, kualitas visual, dan tingkat realisme) disertai metrik objektif dan ukuran kesepakatan antar-penilai, menguji konsistensi *AI review* melalui uji pengulangan dan perbandingan dengan penilaian ahli, serta memperbaiki validitas butir instrumen pada sampel yang lebih besar.

PERNYATAAN KONTRIBUSI PENULIS (CREDIT)

Muhammad Yulasfi Pahrizal: Konseptualisasi, Metodologi, Perangkat lunak, Validasi, Penulisan – draf awal, Penulisan – peninjauan dan penyuntingan. **Pramudya Prima Insan:** Konseptualisasi, Kurasi data, Analisis formal, Penulisan – draf awal. **Nasruddin Bin Idris:** Investigasi, Perangkat lunak, Supervisi, Penulisan – peninjauan dan penyuntingan.

PERNYATAAN KONFLIK KEPENTINGAN

Para penulis menyatakan bahwa tidak terdapat kepentingan finansial atau hubungan pribadi yang diketahui yang dapat memengaruhi hasil penelitian yang dilaporkan dalam artikel ini.

PERNYATAAN PENDANAAN

Penulis tidak menerima dukungan finansial khusus untuk penelitian, kepenulisan, dan/atau publikasi artikel ini.

PERNYATAAN PENGGUNAAN AI GENERATIF DAN TEKNOLOGI BERBANTUAN AI DALAM PROSES PENULISAN

Para penulis menggunakan ChatGPT-5.6 untuk membantu penyusunan dan penyuntingan bahasa pada naskah ini. Seluruh konten yang dibantu oleh AI telah ditinjau dan disunting oleh penulis, dan para penulis bertanggung jawab sepenuhnya atas isi akhir yang dipublikasikan.

REFERENCES

- [1] H. Chen *et al.*, "Comprehensive exploration of diffusion models in image generation: a survey," *Artif. Intell. Rev.*, vol. 58, no. 4, Apr. 2025, doi: 10.1007/s10462-025-11110-3.
- [2] R. A. Warman, "Mengubah Gambar Sketsa Menjadi Visual yang Menakjubkan : Membangun Dunia Animasi dengan Kecerdasan Buatan," *Journal of Animation and Games Studies*, vol. 10, no. 1, Apr. 2024.
- [3] Tommy Hari Prihartanto, Januar Ivan Halimawan, and Akkapurlaura, "Pixai: Aplikasi berbasis Generative AI Sebagai Alat Untuk Desain Karakter Yang Efisien," *Jurnal Seni dan Reka Rancang: Jurnal Ilmiah Magister Desain*, vol. 7, no. 3, pp. 463–472, Nov. 2024, doi: 10.25105/jsrr.v7i3.21673.
- [4] Y. Peng, C. Zhao, H. Xie, T. Fukusato, and K. Miyata, "Sketch-Guided Latent Diffusion Model for High-Fidelity Face Image Synthesis," *IEEE Access*, vol. 12, pp. 5770–5780, 2024, doi: 10.1109/ACCESS.2023.3346408.
- [5] A. Satriawan, B. Imran, and S. Erniwati, "Identifikasi Kemiripan Foto Asli Dan Sketsa Menggunakan Model Generative Adversarial Networks (GANs)," *Jurnal Kecerdasan Buatan dan Teknologi Informasi*, vol. 2, no. 3, pp. 122–127, Sep. 2023.
- [6] X. Zhang *et al.*, "Large Language Models as Evaluators for Recommendation Explanations," in *RecSys 2024 - Proceedings of the 18th ACM Conference on Recommender Systems*, Beijing: Association for Computing Machinery, Inc, Oct. 2024, pp. 33–42. doi: 10.1145/3640457.3688075.
- [7] W. Chen and J. Hays, "SketchyGAN: Towards Diverse and Realistic Sketch to Image Synthesis," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Salt Lake City, UT, USA, 2018, pp. 9416–9425, doi: 10.1109/CVPR.2018.00981.
- [8] A. S. Wulandari, A. Saepudin, M. P. Kinanti, Z. Sudesi, A. Saifudin, and Y. Yulianti, "Pengujian Aplikasi Sistem Informasi Akademik Berbasis Web Menggunakan Metode Black Box Testing Equivalence Partitioning," *Jurnal Teknologi Sistem Informasi dan Aplikasi*, vol. 5, no. 2, p. 102, May 2022, doi: 10.32493/jtsi.v5i2.17561.
- [9] H. Seo *et al.*, "Large Language Models as Evaluators in Education: Verification of Feedback Consistency and Accuracy," *Applied Sciences (Switzerland)*, vol. 15, no. 2, Jan. 2025, doi: 10.3390/app15020671.
- [10] H. Yakub, B. Daniawan, A. Wijaya, and L. Damayanti, "Sistem Informasi E-Commerce Berbasis Website Dengan Metode Pengujian User Acceptance Testing," *JSITIK: Jurnal Sistem Informasi dan Teknologi Informasi Komputer*, vol. 2, no. 2, pp. 113–127, Apr. 2024, doi: 10.53624/jsitik.v2i2.362.
- [11] R. S. Pressman and B. R. Maxim, *Software Engineering: A Practitioner's Approach*, 9th ed. New York, NY, USA: McGraw-Hill Education, 2020.

- [12] B. Fachri, C. Rizal, and Supiyandi, "Penerapan Metode Waterfall Dalam Perancangan Sistem Informasi Merdeka Belajar Kampus Merdeka Berbasis Web," *Jurnal Komputer Teknologi Informasi dan Sistem Informasi (JUKTISI)*, vol. 2, no. 3, pp. 591–597, 2024, doi: 10.62712/juktisi.v2i3.147.
- [13] M. Bak Herrie, N. R. Maleve, L. Philipsen, and A. B. Staunæs, "Democratization and generative AI image creation: aesthetics, citizenship, and practices," *AI Soc.*, vol. 40, no. 5, pp. 3495–3507, Jun. 2025, doi: 10.1007/s00146-024-02102-y.
- [14] L. Wang, V. Sindagi, and V. Patel, "High-Quality Facial Photo-Sketch Synthesis Using Multi-Adversarial Networks," in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG 2018)*, Xi'an, China, 2018, pp. 83–90, doi: 10.1109/FG.2018.00022.
- [15] J. Oppenlaender, "A taxonomy of prompt modifiers for text-to-image generation," *Behaviour and Information Technology*, vol. 43, no. 15, pp. 3763–3776, 2024, doi: 10.1080/0144929X.2023.2286532.