Usulan Ide Tugas Akhir

NIM	2205551154
Nama	I Putu Gede Joni Ananta Udyana
Bidang Keahlian	Sistem Informasi
Topik Sistem Informasi	Frontend Aplikasi Web Video Generator
Judul	Perancangan Antarmuka Aplikasi Web Video Generator Otomatis Berbasis AI Menggunakan LLM, Text-to-Speech, dan Text-to-Image untuk Konten Storytelling
Calon Pembimbing	Prof. Dr. I Made Sukarsa, S.T., M.T. Dr. I Made Suwija Putra, S.T., M.T.
Bahasa Pemrograman	Python dengan *Framework* Flask
Paper Pendukung (SoTA)	Penelitian berjudul "Multimodal Cinematic Video Synthesis Using Text-to-Image and Audio Generation Models" oleh Sridhar S, Nithin A, Vasantha Raj K, dan Shakeel Rifath dari Hindustan Institute of Technology and Science ini membahas pengembangan sistem pembuatan video sinematik berdurasi 60 detik secara otomatis dari input teks. Penelitian ini memanfaatkan Stable Diffusion untuk menghasilkan citra berkualitas tinggi, GPT-2 untuk menyusun narasi menjadi lima adegan (Introduction, Rising Action, Climax, Falling Action, Resolution), serta pipeline audio hibrida yang menggabungkan gTTS untuk voiceover dan musik latar dari YouTube. Sistem dilengkapi teknik interpolasi frame linier, cinematic post-processing, dan sinkronisasi audio-video untuk mencapai hasil setara produksi profesional. Implementasi dilakukan di Google Colab dengan GPU NVIDIA L4 menggunakan Python 3.11, serta menyediakan antarmuka Gradio

	dalam dua mode (Simple dan Advanced) dengan dukungan resolusi hingga 1024x768 dan frame rate 15–30 FPS. Evaluasi menunjukkan kualitas visual tinggi (SSIM 0,85), koherensi narasi yang baik (BLEU 0,72), serta sinkronisasi audio yang efektif (MOS 4,2/5), menjadikannya solusi skalabel untuk aplikasi kreatif, edukasi, dan industri dibandingkan metode sebelumnya. Penelitian yang berjudul "Text2Video: AI-driven Video Synthesis from Text Prompts" oleh Shankar Tejasvi dan Merin Meleet menyajikan sebuah sistem untuk mengubah deskripsi teks menjadi konten video yang koheren dan realistis. Fokus utama dari penelitian ini adalah menggabungkan dua disiplin ilmu, yaitu pemrosesan bahasa alami (NLP) dan visi komputer, untuk menciptakan sebuah alur kerja produksi video yang digerakkan oleh AI. Secara metodologis, sistem ini memanfaatkan model-model yang sudah terlatih (*pre-trained models*) sebagai fondasi untuk menghasilkan video dasar (*generic video*). Namun, yang menjadi nilai lebih dari penelitian ini adalah implementasi teknik *style transfer* menggunakan VideoLORA, yang memungkinkan video yang dihasilkan dapat ditingkatkan dengan gaya visual atau artistik tertentu, sehingga hasilnya menjadi lebih unik dan menarik secara estetika. Arsitektur teknisnya dibangun menggunakan serangkaian pustaka (*libraries*) yang populer di kalangan pengembang AI seperti PyTorch dan PyTorch Lightning untuk proses *deep learning*, serta OpenCV untuk tugas-tugas

	pemrosesan video. Pada intinya, penelitian ini menunjukkan sebuah proses yang efektif untuk tidak hanya mengubah teks menjadi video, tetapi juga memberikan "sentuhan artistik" pada video tersebut, yang sangat relevan untuk aplikasi di industri kreatif, hiburan, hingga media komunikasi. Penelitian berjudul *"A Survey on Generative AI and LLM for Video Generation, Understanding, and Streaming"* merupakan sebuah kajian komprehensif yang mengulas secara mendalam bagaimana teknologi *Generative AI* dan *Large Language Models* (LLM) mentransformasi berbagai aspek dalam teknologi video. Paper ini memetakan lanskap teknologi tersebut ke dalam tiga area utama: *video generation* (penciptaan video), *video understanding* (pemahaman konten video), dan *video streaming* (penyiaran video). Dalam konteks yang paling relevan dengan tugas akhir Anda, bagian *video generation* menyoroti bagaimana teknologi ini dimanfaatkan untuk menghasilkan konten visual yang sangat realistis dari input teks, membahas berbagai pendekatan seperti model difusi (*diffusion models*) yang menjadi dasar untuk sintesis *text-to video*. Selain itu, survei ini juga membahas bagaimana LLM meningkatkan kemampuan system dalam memahami isi video, misalnya untuk membuat takarir (*captioning*) secara otomatis, serta mengoptimalkan pengalaman *streaming* bagi pengguna. Pada akhirnya, penelitian ini tidak hanya memaparkan pencapaian terkini, tetapi juga mengidentifikasi tantangan-tantangan signifikan

	yang masih dihadapi, seperti menjaga konsistensi visual antar adegan, kebutuhan komputasi yang tinggi, dan keterbatasan dataset video berskala besar, yang mana ini semua memberikan konteks dan justifikasi akademis yang kuat untuk proyek yang sedang Anda kerjakan. Penelitian *Advances in AI-Generated Images and Videos* membahas kemajuan teknologi dalam pembuatan konten multimedia sintetis, khususnya gambar dan video, yang didorong oleh perkembangan model generatif seperti *Generative Adversarial Networks* (GANs), *transformers*, dan *diffusion models*. Kajian ini memaparkan teknik-teknik mutakhir untuk menghasilkan citra dan video realistis dari teks, gambar, atau masukan multimodal, serta menjelaskan metode deteksi untuk membedakan konten buatan AI dari yang asli. Berbagai aplikasi potensial diuraikan, mulai dari hiburan, industri kreatif, pendidikan, hingga keamanan dan forensik, diimbangi dengan pembahasan risiko seperti penyebaran *deepfake* dan disinformasi. Penelitian ini juga menyoroti tantangan utama, termasuk kebutuhan sumber daya komputasi besar, konsistensi temporal pada video, generalisasi model, dan aspek etis. Selain menguraikan kumpulan dataset penting yang digunakan di bidang ini, studi tersebut memberikan pandangan ke arah tren masa depan, seperti transparansi dan interpretabilitas model, peningkatan konten multimodal, dan adopsi luas *diffusion models*, sekaligus menawarkan arah

	penelitian untuk mengoptimalkan manfaat teknologi sambil meminimalkan risikonya. Penelitian berjudul *Zero-1-to-A: Zero-Shot One Image to Animatable Head Avatars Using Video Diffusion* membahas metode inovatif untuk menghasilkan avatar kepala yang dapat dianimasikan hanya dari satu gambar masukan, tanpa memerlukan pelatihan ulang khusus (*zero-shot*). Pendekatan ini memanfaatkan model *video diffusion* untuk secara langsung mensintesis rangkaian frame video yang realistis, mempertahankan konsistensi identitas wajah, serta menangkap gerakan kepala dan ekspresi secara alami berdasarkan kondisi pose atau audio yang diberikan. Arsitektur yang diusulkan menggabungkan penyelarasan geometri wajah dengan pembangkitan detail visual berkualitas tinggi, sehingga mampu menghasilkan animasi halus dan ekspresif meski data masukan sangat terbatas. Penelitian ini relevan dengan seminar ide yang mengangkat topik perancangan antarmuka *web video generator* berbasis AI karena teknik ini dapat diintegrasikan sebagai modul pembuatan avatar animasi dari foto pengguna, yang kemudian dapat disinkronkan dengan narasi suara atau teks cerita, sehingga memperkaya variasi konten storytelling yang dihasilkan sistem dan memberikan pengalaman yang lebih interaktif serta personal. Penelitian *STORYAGENT: Customized Storytelling Video Generation via Multi-Agent Collaboration*

	membahas pengembangan sistem pembuatan video storytelling yang dipersonalisasi melalui kolaborasi beberapa agen AI yang masing-masing memiliki peran khusus dalam proses kreatif. Sistem ini memanfaatkan kerangka kerja *multi-agent* di mana setiap agen menangani tahapan tertentu, seperti perencanaan alur cerita, penulisan naskah, pemilihan dan pembuatan aset visual, generasi audio, serta penyuntingan dan perakitan akhir video. Pendekatan ini memungkinkan terciptanya alur kerja terstruktur yang dapat disesuaikan dengan preferensi pengguna, sehingga menghasilkan video yang tidak hanya kohesif secara naratif tetapi juga konsisten secara visual dan auditif. Keterkaitan penelitian ini dengan ide skripsi *Perancangan Antarmuka Aplikasi Web Video Generator Otomatis Berbasis AI* terletak pada penerapan prinsip orkestrasi multi-komponen untuk mengubah *prompt* teks menjadi video storytelling utuh. Konsep multi-agent ini selaras dengan kebutuhan sistem yang mengintegrasikan LLM, Text-to-Speech, dan Text-to-Image dalam alur kerja terhubung, demi menghasilkan konten kreatif yang sinkron dan relevan dengan konteks cerita. Penelitian *Generative AI in Multimodal User Interfaces: Trends, Challenges, and Cross-Platform Adaptability* membahas bagaimana teknologi Generative AI, khususnya multimodal LLM, mengubah desain antarmuka pengguna menjadi lebih adaptif, personal, dan mampu mengintegrasikan berbagai jenis input seperti teks, suara, gambar, dan video secara mulus di berbagai

	perangkat. Topik utamanya mencakup *interface dilemma* atau tantangan merancang UI yang optimal untuk multimodal AI, pengembangan *lightweight framework* agar dapat berjalan efisien di perangkat dengan keterbatasan sumber daya seperti ponsel, serta isu etis seperti privasi, retensi konteks, dan transparansi. Hubungannya dengan proyek *Perancangan Antarmuka Aplikasi Web Video Generator Otomatis Berbasis AI* terletak pada kesamaan kebutuhan untuk merancang antarmuka yang mampu menangani input teks dan mengorkestrasi hasil *text-to-speech* serta *text-to-image* dalam satu alur kerja terpadu, sambil tetap responsif, ramah pengguna, dan adaptif di berbagai platform. Pemahaman tren, tantangan, serta teknik optimisasi dari penelitian tersebut dapat menjadi acuan strategis untuk memastikan antarmuka aplikasi video generator yang dikembangkan tidak hanya fungsional, tetapi juga efisien, konsisten, dan siap berkembang mengikuti kemajuan teknologi AI multimodal. Penelitian *Efficient and Aesthetic UI Design with a Deep Learning-Based Interface Generation Tree Algorithm* membahas metode baru pembuatan antarmuka pengguna menggunakan algoritma *interface generation tree* berbasis Transformer yang dirancang untuk meningkatkan efisiensi dan estetika desain UI. Pendekatan ini memodelkan komponen UI dalam struktur hierarkis berbentuk pohon, mengenkripsi dan mendekripsi strukturnya dengan Transformer, serta memanfaatkan *markup language*

	khusus dan dataset antarmuka web dan mobile dunia nyata untuk pelatihan. Hasil eksperimen menunjukkan peningkatan signifikan dalam akurasi, kemiripan desain, dan skor kepuasan pengguna dibandingkan metode konvensional, termasuk dukungan *reinforcement learning* untuk menyesuaikan desain secara adaptif berdasarkan umpan balik pengguna. Hubungannya dengan proyek *Perancangan Antarmuka Aplikasi Web Video Generator Otomatis Berbasis AI* terletak pada kebutuhan yang sama untuk menghasilkan UI yang responsif, intuitif, dan estetis secara otomatis di mana pendekatan ini dapat diadaptasi untuk membangun antarmuka web yang mengintegrasikan LLM, *text-to-speech*, dan *text-to-image* dalam alur kerja terpadu, sekaligus memastikan kualitas visual yang konsisten dan pengalaman pengguna yang optimal di berbagai perangkat. Penelitian *Perancangan Sistem Informasi Penjualan pada Warung Ibu Neny Berbasis Website Menggunakan Framework Flask* membahas pengembangan sistem informasi berbasis web untuk mengotomatisasi proses penjualan, pemasaran, dan administrasi yang sebelumnya dilakukan secara manual. Sistem dirancang menggunakan metode *waterfall* dengan tahap kebutuhan, analisis, desain menggunakan UML, implementasi menggunakan Python dan Flask, pengujian dengan *black box testing*, serta pemeliharaan untuk memastikan kinerja yang optimal. Hasilnya adalah aplikasi web yang memudahkan admin mengelola produk,

	kategori, dan transaksi, sekaligus mempermudah pelanggan memesan barang secara daring tanpa harus datang langsung. Hubungannya dengan proyek ini terletak pada kesamaan konsep perancangan antarmuka dan sistem berbasis web yang berorientasi pada kemudahan interaksi pengguna, integrasi teknologi terkini, serta optimalisasi proses yang sebelumnya manual menjadi otomatis. Penelitian *Perancangan dan Implementasi Aplikasi Web untuk Pembuatan dan Pengelolaan Video Pembelajaran Interaktif* membahas pengembangan aplikasi berbasis web yang memungkinkan pembuatan, pengeditan, dan pengelolaan video pembelajaran secara interaktif dengan memanfaatkan teknologi multimedia dan integrasi elemen interaktif. Proyek ini menggunakan metodologi pengembangan perangkat lunak yang mencakup analisis kebutuhan, perancangan antarmuka berbasis web yang responsif, serta integrasi fitur seperti *text-to-speech*, kuis interaktif, dan pelacakan kemajuan pengguna. Hasil implementasi menunjukkan bahwa aplikasi dapat meningkatkan efektivitas pembelajaran melalui kombinasi media visual, audio, dan interaksi langsung, sekaligus mempermudah pendidik dalam menyusun materi yang menarik. Relevansinya dengan proyek *Perancangan Antarmuka Aplikasi Web Video Generator Otomatis Berbasis AI* terletak pada kesamaan tujuan untuk memadukan teknologi multimedia dalam sebuah antarmuka yang mudah digunakan, di mana konsep integrasi teks, suara, dan

	visual dapat diadaptasi untuk menghasilkan konten storytelling secara otomatis berbasis AI. Penelitian *Pengembangan Aplikasi Web untuk Pembuatan Video Pembelajaran Interaktif* membahas perancangan platform berbasis web yang menggabungkan media visual, audio, dan elemen interaktif untuk meningkatkan efektivitas pembelajaran. Sistem ini dilengkapi fitur seperti *text-to-speech*, kuis, dan pelacakan kemajuan yang memudahkan pendidik membuat materi menarik secara digital. Relevansinya dengan proyek ini terletak pada integrasi multimodal teks, suara, dan gambar dalam satu antarmuka yang interaktif dan mudah digunakan untuk menghasilkan konten storytelling otomatis. Penelitian *Rancang Bangun Aplikasi Pembuatan Video Animasi Berbasis Web Menggunakan Teknologi AI* membahas pengembangan sistem web yang memungkinkan pembuatan video animasi secara otomatis melalui input teks, dengan memanfaatkan teknologi *text-to-speech* dan *text-to-image*. Aplikasi ini dirancang dengan antarmuka yang intuitif, mendukung kustomisasi karakter, latar, dan narasi, serta dioptimalkan untuk menghasilkan video secara cepat dan efisien. Hubungannya dengan proyek ini terletak pada kesamaan fokus dalam menggabungkan teknologi AI multimodal untuk menghasilkan konten storytelling secara otomatis dalam satu platform web yang mudah diakses.

	Penelitian *Pengembangan Platform AI untuk Pembuatan Video Otomatis Berbasis Web* membahas rancangan sistem yang mengintegrasikan *natural language processing*, *text-to-speech*, dan *image generation* untuk mengubah teks menjadi video secara otomatis. Platform ini dirancang dengan antarmuka yang responsif, fitur pengaturan gaya visual, serta kemampuan sinkronisasi audio dan gambar secara real-time. Keterkaitannya dengan proyek ini terlihat pada kesamaan tujuan dalam menghadirkan solusi pembuatan konten storytelling berbasis AI yang praktis, multimodal, dan terintegrasi dalam satu aplikasi web. Penelitian *Evaluasi Rancangan Antarmuka HCI Modern Berbasis Kecerdasan Buatan* membahas penerapan AI dalam desain antarmuka pengguna untuk meningkatkan personalisasi, efisiensi, dan interaksi yang adaptif melalui teknologi seperti *machine learning* dan *natural language processing*. Hasil penelitian menunjukkan bahwa integrasi AI mampu mempercepat navigasi, memberikan rekomendasi yang relevan, serta menyesuaikan tampilan secara real-time berdasarkan perilaku pengguna, meskipun tetap menghadapi tantangan privasi, keamanan data, dan aksesibilitas bagi pengguna non-teknis. Rekomendasi pengembangan meliputi peningkatan transparansi, penguatan keamanan data, penyederhanaan antarmuka, serta edukasi pengguna agar lebih memahami cara kerja AI. Keterkaitannya dengan proyek *Perancangan Antarmuka Aplikasi Web Video Generator Otomatis*

	Berbasis AI terletak pada pemanfaatan AI multimodal untuk menciptakan pengalaman interaktif yang adaptif dan ramah pengguna dalam menghasilkan konten storytelling secara otomatis. Penelitian *Integrasi Teknologi AI dalam Desain Antarmuka Multimodal* membahas penggabungan *natural language processing*, *text-to-speech*, dan *computer vision* untuk menciptakan antarmuka yang mampu memahami dan merespons input dari berbagai media secara terpadu. Hasilnya menunjukkan peningkatan interaktivitas dan efisiensi penggunaan, meskipun terdapat tantangan dalam optimasi performa dan sinkronisasi antar-modality. Keterkaitannya dengan proyek *Perancangan Antarmuka Aplikasi Web Video Generator Otomatis Berbasis AI* terlihat pada fokus yang sama untuk memadukan berbagai teknologi AI dalam satu platform yang responsif dan intuitif untuk menghasilkan konten storytelling otomatis.

GAMBARAN UMUM SISTEM

1. Pendahuluan

Setiap individu maupun bisnis memerlukan konten kreatif yang menarik untuk menjangkau audiens secara efektif. Pada awalnya, pembuatan video secara manual mungkin menjadi pilihan, namun seiring meningkatnya permintaan akan konten yang bervariasi dan berkualitas tinggi, metode tradisional ini menjadi tidak efisien karena memakan waktu, membutuhkan keahlian teknis, dan sulit dilakukan secara konsisten dalam skala besar. Perkembangan teknologi *Artificial Intelligence* (AI) generatif kemudian menghadirkan solusi untuk mengotomatisasi sebagian besar proses produksi video, sehingga memungkinkan pembuatan konten yang lebih cepat, fleksibel, dan hemat sumber daya.

Sistem *Text-to-Video* modern kini mampu menghasilkan visual, narasi, dan audio langsung dari teks, namun pendekatan yang sepenuhnya mengandalkan model *AI Text-to-Image* generatif masih memiliki keterbatasan. Tantangan seperti menjaga konsistensi visual antar-adegan, memastikan kesesuaian visual dengan konteks cerita, serta kebutuhan sumber daya komputasi yang besar sering kali mengurangi kualitas hasil akhir. Selain itu, alur kerja yang kurang terintegrasi dapat menyebabkan aset visual, audio, dan narasi tidak sinkron, sehingga video yang dihasilkan terasa terputus-putus dan kurang profesional.

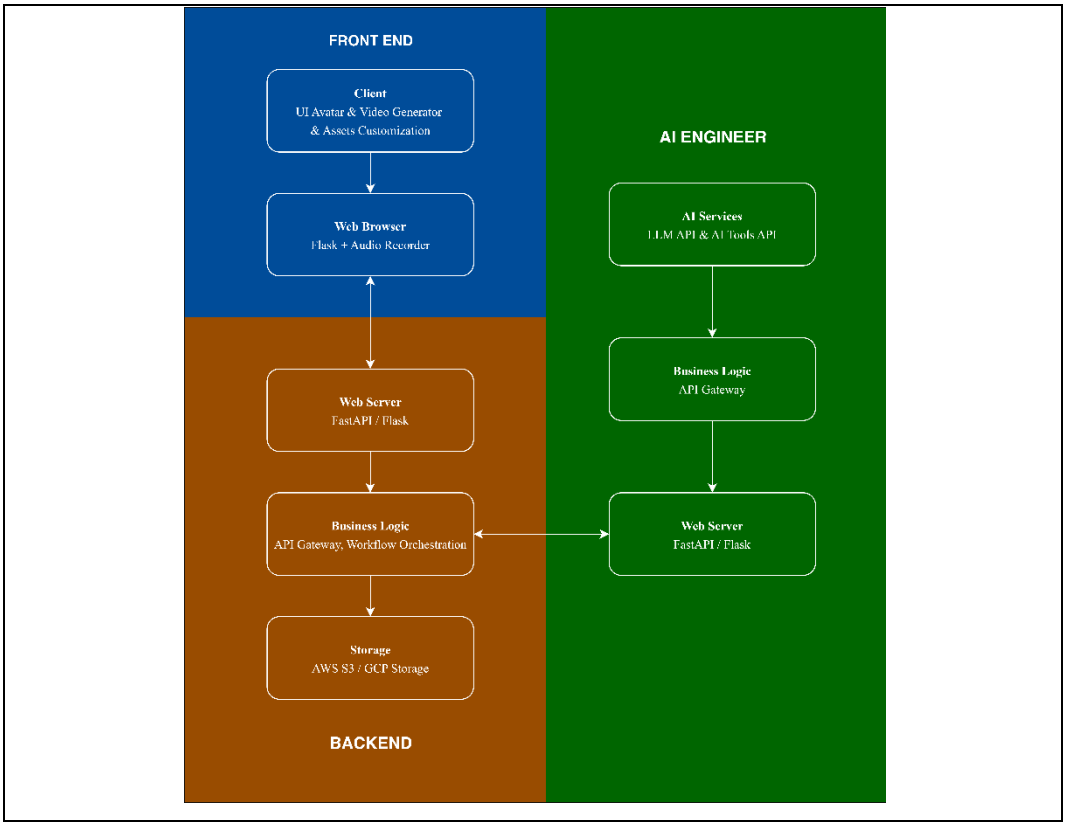
Untuk mengatasi permasalahan tersebut, penelitian ini mengusulkan pengembangan aplikasi web *video generator* berbasis *AI* dengan alur kerja terintegrasi yang mendukung dua jalur utama pembuatan konten. Jalur pertama berfokus pada pembuatan video naratif, dimulai dari teks cerita yang diproses menjadi naskah terstruktur oleh *Large Language Model* (*LLM*), lalu diubah menjadi narasi audio melalui *Text-to-Speech*, serta dilengkapi visual yang dihasilkan dari *Text-to-Image* maupun pencarian *stock video* dari basis data daring seperti Pixabay dan Pexels dengan kata kunci yang diekstraksi otomatis. Jalur kedua adalah pembuatan *avatar* bergaya *podcast*, di mana sistem memanfaatkan gambar atau referensi visual untuk menciptakan karakter animasi yang dapat menyampaikan narasi secara ekspresif dan sinkron dengan audio.

Keunggulan utama sistem ini adalah kemampuannya untuk menyesuaikan hasil *generate* secara dinamis, memastikan posisi suara, transisi visual, dan penempatan aset hasil *generate* selaras dengan alur cerita. Proses *video rendering* kemudian menggabungkan seluruh aset digital menjadi video utuh, lengkap dengan *subtitle* otomatis, yang siap diunduh dalam format *.mp4* atau *.mov*. Pendekatan hibrida ini tidak hanya menjamin koherensi narasi dan sinkronisasi optimal antara teks, visual, dan audio, tetapi juga memberikan fleksibilitas bagi pengguna dalam memilih metode visual yang sesuai dengan kebutuhan. Dengan antarmuka yang intuitif, sistem ini diharapkan menjadi solusi yang praktis dan efektif bagi kreator non-spesialis untuk memproduksi konten kreatif berkualitas tinggi secara efisien.

.

2. Arsitektur

Sistem *video* dan *avatar generator* berbasis kecerdasan buatan dirancang dengan pendekatan modular yang membagi fungsionalitas ke dalam tiga bagian utama: Front End, Backend, dan AI Engineer. *Front End* berperan sebagai antarmuka bagi pengguna untuk menulis atau mengunggah naskah, menyesuaikan aset visual, mengatur sinkronisasi audio, serta meninjau pratinjau hasil. Seluruh permintaan dari pengguna dikirim ke *Backend*, yang bertanggung jawab mengelola alur kerja, menyimpan dan mengatur data pada layanan cloud seperti AWS S3 atau GCP Storage, serta memanggil layanan AI Service melalui *API request* sesuai kebutuhan. AI Engineer bertindak sebagai lapisan komputasi terpisah yang menangani seluruh proses berbasis *machine learning* atau AI, mulai dari pemrosesan teks menggunakan *Large Language Model*, konversi teks ke audio melalui *Text-to-Speech*, pembuatan visual dengan *Text-to-Image*, hingga pencarian dan pengolahan stock video dari basis data daring. Pemisahan ini memastikan setiap lapisan memiliki peran yang jelas namun tetap terintegrasi, sehingga sistem menjadi fleksibel, mudah diskalakan, dan dapat dikembangkan secara berkelanjutan.



Arsitektur sistem dimulai dari Front End, di mana pengguna berinteraksi dengan antarmuka (*user interface*) untuk membuat konten, menyesuaikan aset visual, dan merekam audio langsung melalui browser. Setiap permintaan dari pengguna diterima dan diolah oleh Backend, yang berfungsi sebagai pusat orkestrasi. Lapisan *Business Logic* di Backend bertanggung jawab untuk mengatur alur kerja, mengelola penyimpanan aset pada layanan cloud seperti AWS S3, serta memfasilitasi komunikasi dengan layanan AI Engineer Service. Untuk menjalankan fungsi cerdas, Backend memanggil AI Engineer melalui *API*, yang menyediakan akses ke *Large Language Model* (LLM) dan berbagai alat AI lainnya.

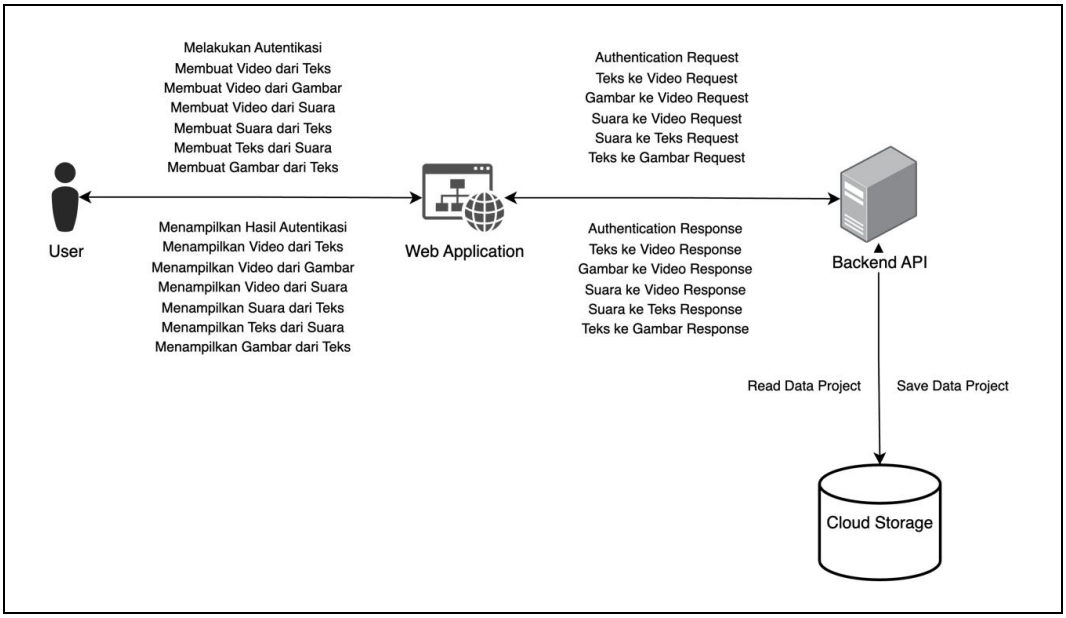
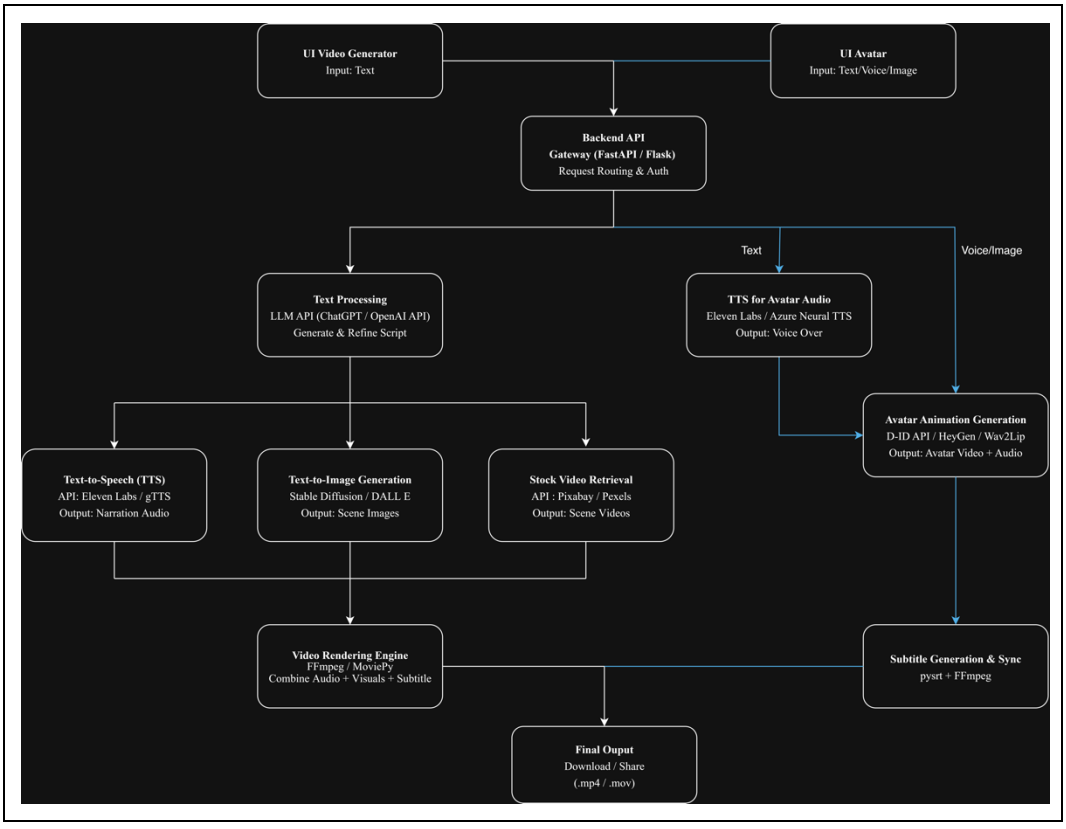


Diagram di atas menggambarkan arsitektur sistem tingkat tinggi (*high-level architecture*) yang umum diterapkan pada aplikasi modern, dengan pemisahan yang jelas antara lapisan antarmuka pengguna (*client-side*) dan logika aplikasi di sisi server (*server-side*). Alur kerja dimulai dari Pengguna yang berinteraksi dengan Aplikasi Web untuk menampilkan data, menangkap input, dan memberikan pengalaman interaksi yang intuitif. Saat pengguna melakukan suatu tindakan, Website mengirimkan permintaan (*request*) ke Backend API, yang berperan sebagai pusat kendali atau "otak" dari sistem.

Di sisi Backend, seluruh logika bisnis (*business logic*) dieksekusi, termasuk validasi data, pengaturan alur proses, dan pemanggilan layanan eksternal. Backend juga berinteraksi dengan lapisan penyimpanan data, yaitu *Cloud Storage*, untuk membaca (*read*) maupun menulis (*write*) data yang diperlukan, seperti menyimpan aset multimedia atau mengambil data hasil pemrosesan AI. Setelah semua proses selesai, Backend mengirimkan respons (*response*) kembali ke Aplikasi Web, yang kemudian menampilkannya kepada pengguna dalam format yang sesuai.

3. Workflow Aplikasi

Workflow aplikasi ini dirancang untuk mengalirkan proses pembuatan video berbasis kecerdasan buatan secara menyeluruh, mulai dari masukan awal pengguna hingga video akhir siap digunakan. Proses dimulai dari penerimaan *input* yang dapat berupa teks, gambar, atau suara. Selanjutnya, sistem mengolah *input* tersebut melalui serangkaian tahap yang saling terhubung mencakup pengolahan naskah, pembangkitan elemen visual, pembuatan narasi audio, hingga penggabungan seluruh aset multimedia. Setiap tahapan dirancang agar berjalan terstruktur, efisien, dan sinkron, sehingga narasi, gambar, audio, dan elemen tambahan lainnya saling mendukung secara harmonis. Melalui alur kerja *end-to-end* ini, pengguna memiliki fleksibilitas untuk memilih jenis aset visual yang digunakan, baik dari sumber stok maupun hasil generasi AI, sambil tetap mendapatkan hasil akhir yang konsisten dan berkualitas. Pendekatan ini memastikan siapa pun, termasuk kreator konten tanpa latar belakang teknis, dapat memproduksi video yang terlihat profesional.



Workflow aplikasi ini dirancang untuk mengakomodasi dua jalur utama pembuatan konten, yaitu pembuatan video naratif dan pembuatan *avatar* bergaya *podcast*. Pada jalur pembuatan video naratif, proses dimulai ketika pengguna memasukkan teks cerita yang kemudian diproses oleh *Large Language Model* (*LLM*) untuk diubah menjadi naskah yang terstruktur. Naskah ini diproses secara paralel oleh tiga layanan inti: *Text-to-Speech* untuk menghasilkan narasi audio, *Text-to-Image* untuk membangkitkan visual berbasis AI, dan layanan pencarian *stock video* untuk memperoleh klip relevan. Sementara itu, pada jalur pembuatan *avatar*, sistem menerima *input* berupa teks, suara, atau gambar referensi yang kemudian dianimasikan dan disinkronkan dengan audio yang dihasilkan secara terpisah. Seluruh aset digital yang dihasilkan, baik dari alur video naratif maupun *avatar*, digabungkan menggunakan *Video Rendering Engine*, dilengkapi dengan *subtitle* otomatis, dan diolah menjadi video utuh dalam format *.mp4* atau *.mov* yang siap diunduh. Tahapan-tahapan ini akan dijabarkan lebih detail pada bagian berikutnya untuk memberikan gambaran menyeluruh tentang bagaimana setiap komponen bekerja secara terintegrasi dalam aplikasi.

4. Daftar Fitur

Fitur	Deskripsi
Text-to-Speech (TTS)	Mengubah naskah teks hasil text processing menjadi narasi suara dengan kualitas tinggi menggunakan API seperti Eleven Labs atau gTTS.
Speech-to-Text (STT)	Mengubah input suara dari pengguna menjadi teks otomatis menggunakan API seperti Whisper atau layanan sejenis, sehingga memudahkan pengguna yang ingin memberi narasi lewat suara.
Text-to-Image	Menghasilkan gambar atau ilustrasi dari deskripsi teks menggunakan model AI seperti Stable Diffusion atau DALL·E, yang akan digunakan sebagai aset visual.
Text-to-Video – AI Image	Membuat video secara penuh menggunakan rangkaian gambar yang dihasilkan oleh AI (image

	generation), disinkronkan dengan narasi suara. Cocok untuk konten kreatif dengan gaya visual unik.
Text-to-Video – Stock Video	Membuat video dengan memanfaatkan klip dari basis data gratis seperti Pixabay atau Pexels, berdasarkan kata kunci yang diekstrak dari naskah, lalu menggabungkannya dengan narasi suara.
Avatar Video – Podcast Style	Menghasilkan avatar digital yang dapat berbicara dan bergerak layaknya pembawa acara podcast. Pengguna cukup memberikan foto referensi, teks, atau rekaman suara, kemudian sistem akan menganimasikan wajah dan gerakan bibir avatar agar sinkron dengan audio yang dihasilkan dari TTS atau suara asli pengguna. Fitur ini cocok untuk membuat konten personal, video edukasi, atau siaran visual tanpa perlu perekaman kamera langsung.
Authentication	Memastikan keamanan akses sistem melalui proses registrasi akun baru dengan verifikasi email, login menggunakan kredensial yang valid, manajemen sesi berbasis JWT, reset kata sandi melalui tautan email, logout untuk mengakhiri sesi aktif, serta dukungan multi-faktor autentikasi menggunakan OTP untuk perlindungan tambahan.
Asset Picker & Editor	Memungkinkan pengguna untuk langsung memilih, mengganti, atau mengedit aset visual dan audio yang digunakan dalam video, baik dari hasil *generate* AI maupun dari pustaka stok yang tersedia. Fitur ini menyediakan pratinjau instan dan opsi penyesuaian seperti pemotongan, pengaturan warna, atau penggantian klip, sehingga pengguna

	memiliki kendali penuh atas kualitas dan kesesuaian aset dengan alur cerita.

5. Batasan Masalah

Dalam pengembangan aplikasi ini, saya, I Putu Gede Joni Ananta (2205551154), berperan fokus pada pengembangan Front-End yang bertugas membangun antarmuka pengguna (*User Interface*) berbasis web. Lingkup pekerjaan saya meliputi perancangan dan implementasi UI/UX untuk fitur utama aplikasi, seperti form *input* teks ide cerita, tampilan *preview* hasil gambar atau video, serta komponen indikator proses berupa *loading state* dan *progress bar*. elain itu, lingkup juga mencakup pengembangan fitur *Asset Picker & Editor* yang memungkinkan pengguna memilih, mengganti, dan melakukan pengeditan sederhana terhadap aset visual maupun audio secara langsung di antarmuka. Integrasi antarmuka dengan *backend* dilakukan melalui pemanggilan API agar data dapat ditampilkan secara real-time sesuai perkembangan proses di *server*. Tanggung jawab saya berhenti pada penyediaan antarmuka yang responsif dan optimal di berbagai perangkat, serta memastikan format data yang dikirimkan sesuai standar yang ditentukan tim *backend*. Seluruh proses pemrosesan teks, pembangkitan gambar, pembuatan audio, hingga perakitan video dilakukan oleh rekan tim lain sesuai pembagian peran.

Rekan saya, I Gusti Ngurah Putu Astrawan (2205551071), berperan sebagai Backend Developer yang bertanggung jawab membangun logika *server* dan mengelola seluruh proses pemrosesan data. Lingkup tugasnya meliputi pembuatan API *endpoint* untuk menerima dan mengirim data antara *frontend* dan *server*, mengatur alur kerja dari input teks hingga tahap rendering video, serta mengatur komunikasi dengan berbagai API pihak ketiga yang digunakan. Selain itu, backend bertanggung jawab mengoptimalkan performa pemrosesan, termasuk penggunaan *caching*, *batching request*, dan manajemen sumber daya, agar proses pembuatan video dapat berjalan efisien. Tanggung jawab backend berhenti pada penyediaan data yang sudah diproses dan siap ditampilkan di antarmuka pengguna.

Rekan lainnya, Sultan Azizul Haromain (2205551155), berperan sebagai AI Engineer/AI Integration yang memfokuskan pekerjaannya pada pemilihan, integrasi, dan pengoptimalan API berbasis AI. Lingkup pekerjaannya mencakup penelitian dan pemilihan API yang sesuai untuk text generation, text-to-speech, dan image generation; perancangan pipeline AI yang mengubah teks menjadi narasi

suara dan visual; serta *fine-tuning prompt* agar hasil yang diperoleh lebih sesuai dengan kebutuhan pengguna. Selain itu, ia bertanggung jawab menguji dan memastikan kualitas hasil keluaran model AI, serta menerapkan strategi penghematan penggunaan kredit API agar biaya operasional tetap terkendali. Tanggung jawabnya berhenti pada penyediaan aset multimedia yang sudah siap diproses lebih lanjut di backend. Pembagian tugas yang jelas ini, setiap anggota tim dapat bekerja secara fokus pada bidang keahliannya masing-masing, sekaligus memastikan seluruh komponen sistem saling terintegrasi dengan baik. Pendekatan ini diharapkan menghasilkan aplikasi yang tidak hanya fungsional dan efisien, tetapi juga memberikan pengalaman pengguna yang optimal dari awal hingga akhir proses pembuatan video.

Dosen Pembimbing 1,	Dosen Pembimbing 2,
Prof. Dr. I Made Sukarsa, S.T., M.T.	Dr. I Made Suwija Putra, S.T., M.T.
NIP. 197510242003121010	NIP. 198808072014041001

DAFTAR PUSTAKA

André Costa, F. S. (2024). Towards an AI-Driven User Interface Design for Web Applications. *Procedia Computer Science*, 179-186.

Ariel Han, Z. C. (2023). Design implications of generative AI systems for visual storytelling for young learners. *Design implications of generative AI systems for visual storytelling for young learners*, 470-474.

Cesa Akbar, E. P. (2024). Perancangan Sistem Informasi Penjualan Pada Warung Ibu Neny Berbasis Website Menggunakan Framework Flask. *Jurnalnya Orang Pintar Komputer*, 523-531.

Hessen Bougueffa, M. K.-A.-L. (2024). Advances in AI-Generated Images and Videos. *Advances in AI-Generated Images and Videos*, 173-208.

Jan Bieniek, M. R. (2024). Generative AI in Multimodal User Interfaces: Trends, Challenges, and Cross-Platform Adaptability. *Generative AI in Multimodal User Interfaces: Trends, Challenges, and Cross-Platform Adaptability*, 1-13.

Jonathan Ho, W. C. (2022). Imagen Video: High Definition Video Generation With Diffusion Models. *Imagen Video: High Definition Video Generation With Diffusion Models*, 1-18.

Lizhen Wang, X. Z. (2023). StyleAvatar: Real-time Photo-realistic Portrait Avatar from a Single Video. *StyleAvatar: Real-time Photo-realistic Portrait Avatar from a Single Video* , 1-10.

Panwen Hu, J. J. (2024). Storyagent: Customized Storytelling Video Generation Via Multi-Agent Collaboration. *Storyagent: Customized Storytelling Video Generation Via Multi-Agent Collaboration*, 1-20.

Pengyuan Zhou, L. W. (2024). A Survey on Generative AI and LLM for Video Generation, Understanding, and Streaming. *A Survey on Generative AI and LLM for Video Generation, Understanding, and Streaming*, 1-16.

Shankar Tejasvi, M. M. (2024). Text2Video: AI-driven Video Synthesis from Text Prompts . *International Research Journal of Engineering and Technology*, 87-93.

Shiyu Duan, R. Z. (2024). Efficient and Aesthetic UI Design with a Deep Learning-Based Interface Generation Tree Algorithm. *Efficient and Aesthetic UI Design with a Deep Learning-Based Interface Generation Tree Algorithm*, 1-5.

Sofyan Nur Salim, A. W. (2024). Evaluasi Rancangan Antarmuka HCI Modern Berbasis Kecerdasan Buatan. *Jurnal Ilmiah KOMPUTASI*, 531-538.

Sridhar S, N. A. (2025). Multimodal Cinematic Video Synthesis Using Text-to-Image and Audio Generation Models. *Multimodal Cinematic Video Synthesis Using Text-to-Image and Audio Generation Models*, 1-11.

Yuzhe Cai, S. M. (2024). Low-code LLM: Graphical User Interface over Large Language Models. *Low-code LLM: Graphical User Interface over Large Language Models*, 1-14.

Zhenglin Zhou, F. M.-S. (2025). Zero-1-to-A: Zero-Shot One Image to Animatable Head Avatars Using Video Diffusion. *Zero-1-to-A: Zero-Shot One Image to Animatable Head Avatars Using Video Diffusion*, 1-17.