

## ANALISA STRATEGIS DALAM PROSES GENERATE IMAGE-TO-VIDEO PADA PLATFORM AI GENERATIF UNTUK OPTIMALISASI KUALITAS VIDEO

Imam Ainudin Pirmansah<sup>1)</sup>, Dhimas Adi Satria<sup>2)</sup>, Rifai Ahmad Musthofa<sup>3)</sup>

<sup>1,2,3)</sup> Teknologi Informasi Universitas Amikom Yogyakarta

email : imam.af@amikom.ac.id<sup>1)</sup>, dhimas@amikom.ac.id<sup>2)</sup>, ramusthofa07@amikom.ac.id<sup>3)</sup>

### Abstraksi

Perkembangan pesat kecerdasan buatan generatif (AI) telah mentransformasi proses pembuatan konten digital, khususnya dalam menghasilkan video dari gambar statis. Meskipun telah hadir berbagai platform AI image-to-video seperti Kling, Runway, Pixverse, dan Hailuo, hingga kini belum ada benchmarking komprehensif dan sistematis terkait kinerja serta kualitas output dari platform-platform tersebut. Penelitian ini menyajikan analisis komparatif terhadap keempat platform guna mengoptimalkan kualitas video untuk kebutuhan kreatif maupun profesional. Pendekatan kuantitatif dan kualitatif digunakan melalui pengukuran waktu proses, resolusi, frame rate, Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), serta penilaian berbasis skala Likert oleh panelis ahli. Hasil eksperimen menunjukkan bahwa Runway secara konsisten menghasilkan kualitas visual dan sinematografi terbaik, Kling unggul dalam stabilitas karakter dan efisiensi biaya, Hailuo menonjol pada detail tekstur, sementara Pixverse menawarkan workflow tercepat. Temuan ini memberikan panduan praktis bagi pengguna, pengembang, dan pelaku industri dalam memilih dan mengoptimalkan platform AI image-to-video.

### Kata Kunci :

AI generatif, image-to-video, kualitas video, benchmarking, analisis komparatif, PSNR, SSIM

### Abstract

*The rapid advancement of generative artificial intelligence (AI) has transformed digital content creation, particularly in generating videos from static images. Despite the emergence of multiple image-to-video AI platforms such as Kling, Runway, Pixverse, and Hailuo there remains a lack of comprehensive, systematic benchmarking on their performance and output quality. This study presents a comparative analysis of these four platforms to optimize video quality for creative and professional applications. Both quantitative and qualitative approaches are employed, involving metrics such as processing time, resolution, frame rate, Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), and Likert-scale evaluations by expert panelists. Experimental results show that Runway consistently delivers the best visual fidelity and cinematic quality, Kling excels in character stability and cost-efficiency, Hailuo demonstrates superior texture detail, and Pixverse offers the fastest workflow. These findings provide practical guidance for users, developers, and industry practitioners to select and optimize AI-powered image-to-video generation platforms.*

### Keywords :

*generative AI, image-to-video, video quality, benchmarking, comparative analysis, PSNR, SSIM*

### Pendahuluan

Perkembangan pesat kecerdasan buatan generatif (AI) telah merevolusi bidang multimedia digital, khususnya dengan memungkinkan penciptaan konten. Perkembangan pesat kecerdasan buatan generatif (AI) telah merevolusi bidang multimedia digital, khususnya dengan memungkinkan penciptaan konten visual yang realistis dan dinamis secara otomatis [1]. AI kini berperan sebagai alat digital untuk mendukung kreativitas manusia, memperluas peluang inovasi di bidang animasi, hiburan, pendidikan, dan pemasaran. Integrasi AI ke dalam proses kreatif mendorong batas-batas imajinasi kreator dan memungkinkan penciptaan karya yang dulunya tidak terbayangkan. AI tidak akan menggantikan kreator, tetapi justru meningkatkan imajinasi dan memperluas kreativitas para animator [2]. Selain itu, AI telah membawa

perubahan besar dalam industri animasi, mengotomasi tugas-tugas teknis sehingga tim lebih fokus pada aspek kreatif dan percepatan produksi [3].

Tidak seperti gambar statis atau teks, konten video melibatkan dinamika spasial dan temporal yang kompleks dan membutuhkan evaluasi multi-dimensi, mulai dari kualitas presentasi, penyampaian informasi semantik, hingga konsistensi dengan realitas fisik dan niat creator [4]. Evaluasi kualitas video AI pun menghadapi tantangan unik yang hingga kini belum sepenuhnya terjawab, padahal model-model video generatif berkembang sangat pesat dan semakin luas digunakan di berbagai ranah profesional maupun personal [4]. Hal ini menegaskan adanya gap penelitian yang signifikan di bidang evaluasi video AI, terutama terkait

kebutuhan framework penilaian baru yang komprehensif.

Aspek penting lain adalah pengalaman pengguna (Quality of Experience/QoE) yang kini semakin diakui sebagai penentu utama penerimaan dan keberhasilan sistem video berbasis AI [5]. Mengukur dan memahami QoE berarti memahami sejauh mana kepuasan pengguna akhir tercapai saat berinteraksi dengan konten multimedia hasil AI. Sayangnya, kebanyakan model penilaian kualitas gambar dan video konvensional seperti IQA/VQA (misal: PSNR, SSIM) hanya mengukur kesesuaian "rekonstruksi" terhadap konten referensi asli dan kurang mampu menangkap artefak generatif khas AI [6]. Maka, dibutuhkan kombinasi antara penilaian teknis dan persepsi manusia melalui pendekatan subjektif untuk menghasilkan evaluasi yang benar-benar relevan dengan tantangan dan peluang teknologi GenAI saat ini [5].

Dalam beberapa tahun terakhir, sejumlah platform AI image-to-video telah bermunculan seperti Kling, Runway, Pixverse, dan Hailuo masing-masing menawarkan keunggulan dan kemampuan teknis yang unik. Namun, terdapat kesenjangan dalam literatur terkait perbandingan objektif dan komprehensif antar platform berdasarkan metrik kuantitatif maupun kualitatif. Karena kualitas video yang dihasilkan AI kini semakin krusial baik dari sisi realisme visual maupun kesesuaian dengan prompt kreatif evaluasi sistematis menjadi penting untuk seleksi platform dan optimalisasi workflow.

Penilaian kualitas video sendiri merupakan bidang yang terus berkembang dan membutuhkan evaluasi multi-dimensi, mencakup aspek spasial, temporal, gaya, dan semantik. Metrik kuantitatif umum seperti Peak Signal-to-Noise Ratio (PSNR) dan Structural Similarity Index Measure (SSIM) banyak digunakan untuk menilai kualitas gambar dan video secara objektif, sedangkan penilaian kualitatif berbasis skala Likert menangkap persepsi dan kepuasan pengguna. Walaupun metrik-metrik ini penting, studi sebelumnya jarang melakukan benchmarking pada platform AI image-to-video komersial secara publik menggunakan skenario dan prompt standar.

Penelitian ini bertujuan mengisi kesenjangan tersebut dengan melakukan analisis komparatif pada empat platform AI image-to-video terkemuka Kling, Runway, Pixverse, dan Hailuo. Evaluasi mencakup waktu proses, resolusi, frame rate, serta kualitas visual (PSNR dan SSIM), juga penilaian subjektif oleh panelis ahli menggunakan skala Likert. Semua eksperimen menggunakan prompt dan gambar input yang identik untuk menjamin keadilan dan replikasi. Tujuannya adalah menyediakan referensi praktis dan sistematis bagi pengguna, pengembang, dan peneliti yang ingin memilih atau mengoptimalkan workflow image-to-video berbasis AI.

Kontribusi utama makalah ini adalah sebagai berikut:

1. Framework benchmarking terstruktur untuk evaluasi komparatif platform AI image-to-video komersial.
2. Analisis gabungan kuantitatif dan kualitatif performa platform dengan metrik teknis dan human-centered.
3. Rekomendasi dan insight praktis untuk pemilihan platform yang paling sesuai dengan kebutuhan dan prioritas proyek.

### **Tinjauan Pustaka**

Perkembangan teknologi image-to-video telah menarik banyak perhatian dalam beberapa tahun terakhir. [7] mengusulkan pendekatan GAN multi-tahap untuk translasi image-to-video, menggunakan prediksi struktur objek dan pemodelan temporal dengan jaringan konvolusi 3D. Karya mereka menyoroti manfaat pemisahan antara prediksi struktur dan penyempurnaan temporal, menghasilkan video yang lebih realistis dan konsisten secara waktu. Namun, metode mereka sangat bergantung pada data struktur objek (seperti landmark wajah atau skeleton) dan terbatas pada latar belakang statis, sehingga membatasi generalisasi pada skenario video yang lebih dinamis atau tidak terstruktur. Selain itu, studi ini tidak mengevaluasi platform AI publik maupun variasi prompt, sehingga perlu dilakukan benchmarking yang lebih praktis.

[8] Penelitian Fan et al. Memperkenalkan AIGCBench, benchmark komprehensif untuk evaluasi algoritma AI image-to-video (I2V). AIGCBench menggunakan dataset terbuka yang beragam dan sebelas metrik evaluasi di empat dimensi utama: kontrol video, efek gerak, konsistensi temporal, dan kualitas video. Benchmark ini menggabungkan metrik referensi dan non-referensi, divalidasi oleh preferensi manusia. Namun, aplikasi praktisnya dibatasi oleh waktu inferensi yang lambat dan fitur terbatas pada sistem komersial, serta jumlah kasus uji yang relatif sedikit. Riset ini berbeda karena fokus pada output nyata dari platform komersial populer dengan kriteria evaluasi yang lebih praktis dan terstandarisasi.

Syahputra et al. Menilai efektivitas AI generatif, seperti ChatGPT, dalam mendukung guru di Indonesia. Studi mereka menemukan bahwa AI generatif dapat meningkatkan efisiensi pembelajaran dan literasi digital jika didukung pelatihan yang memadai. Namun, tantangan seperti infrastruktur, literasi teknologi yang rendah, dan potensi ketergantungan berlebihan pada AI masih menjadi hambatan. Walaupun penelitian ini berfokus pada pendidikan, hal ini menegaskan pentingnya evaluasi AI generatif di setiap domain secara spesifik [9].

Kemunculan AI generatif telah mendefinisikan ulang pipeline animasi tradisional. Tools AIGC kini aktif terlibat dalam eksekusi teknis seperti inbetween frame, efek kompleks, dan konsistensi gaya, sehingga tugas repetitif makin minim. Platform AI seperti Pictory.ai terbukti efektif meningkatkan hasil belajar, dengan skor N-Gain 0,74 pada mahasiswa ekonomi [10].

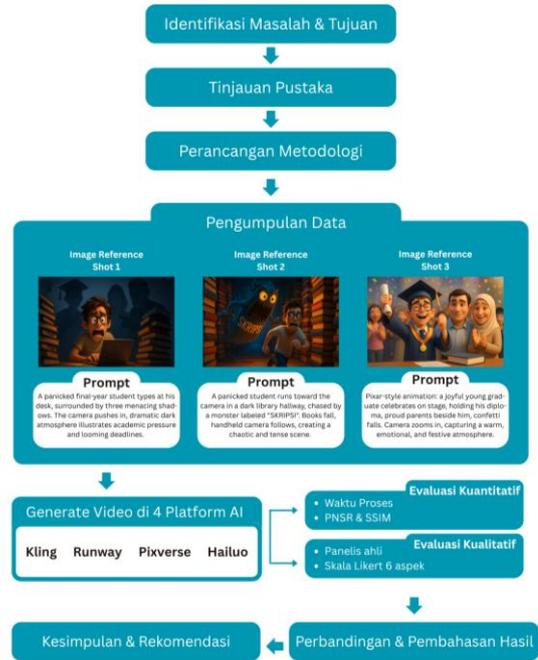
Meski ada kemajuan pada pengembangan model dan framework benchmarking, hingga kini masih jarang dilakukan perbandingan sistematis head-to-head antar platform image-to-video AI komersial, khususnya menggunakan gabungan metrik teknis (PSNR, SSIM) dan penilaian manusia (skala Likert). Penelitian ini berupaya mengisi gap tersebut melalui benchmarking empat platform utama Kling, Runway, Pixverse, dan Hailuo menggunakan prompt identik dan kombinasi evaluasi kuantitatif serta kualitatif, sehingga dapat memberikan panduan praktis bagi pemilihan platform pada aplikasi kreatif dan industri.

### Metode Penelitian

Penelitian ini menggunakan pendekatan eksperimental komparatif untuk mengevaluasi performa dan kualitas output empat platform AI image-to-video komersial: Kling, Runway, Pixverse, dan Hailuo. Metodologi dirancang untuk memberikan penilaian objektif maupun subjektif dengan skenario standar dan input data yang konsisten.

PSNR dan SSIM dipilih sebagai metrik utama karena lebih konsisten dan akurat dalam merepresentasikan persepsi manusia [11]. Algoritma SSIM memiliki korelasi yang kuat dengan Human Vision System [12].

Ada beberapa tahapan yang dilakukan dalam penelitian ini yang tercantum dalam alur penelitian yang ditampilkan pada gambar 1. Alur Penelitian ini menggambarkan proses penelitian yang sistematis, mulai dari identifikasi masalah hingga rekomendasi dan saran penelitian berikutnya. Setiap tahap saling terhubung untuk menjamin hasil penelitian yang kredibel, terukur, dan dapat dijadikan rujukan bagi perkembangan evaluasi AI generatif di bidang image-to-video. Proses ini juga menekankan pentingnya kombinasi antara analisis teknis (angka, metrik) dan penilaian berbasis persepsi manusia (user experience), sehingga hasil akhirnya relevan bagi kebutuhan dunia nyata.



Gambar 1 Alur Penelitian

### Pengumpulan dan Persiapan Data

Tiga gambar statis dipilih sebagai input dasar masing-masing dengan prompt identik seperti yang ditampilkan pada gambar 2. Setiap platform diminta menghasilkan video berdurasi 5 detik untuk setiap pasangan gambar dan prompt, guna menjamin keseragaman evaluasi. Video hasil generate diunduh sesuai format dan resolusi asli dari tiap platform.



Gambar 2 Referensi gambar dan prompt

### Evaluasi Kuantitatif

Analisis kuantitatif berfokus pada tiga metrik utama:

1. Waktu Proses: Waktu total yang dibutuhkan platform untuk menghasilkan video dari satu gambar dan prompt.
2. Peak Signal-to-Noise Ratio (PSNR): PSNR dihitung antara frame representatif dari setiap video hasil generate dan gambar referensinya. Nilai PSNR lebih tinggi berarti distorsi lebih kecil dan kualitas visual lebih baik.
3. Structural Similarity Index Measure (SSIM): SSIM dihitung pada frame yang sama untuk menilai kesamaan struktur, luminansi, dan

kontras. Nilai SSIM antara 0 (tidak mirip) hingga 1 (sangat mirip).

### Evaluasi Kualitatif

Penilaian kualitatif dilakukan oleh 5 panelis ahli (dosen multimedia, praktisi industri, dan mahasiswa tingkat akhir yang berpengalaman di bidang video editing/animasi). Setiap panelis memberi rating pada video hasil generate menggunakan skala Likert 1-5 (1 = Sangat Buruk, 5 = Sangat Baik) pada enam aspek:

1. Stabilitas karakter
2. Stabilitas kamera
3. Konsistensi gaya visual
4. Ketajaman dan detail tekstur
5. Pencahayaan dan bayangan
6. Kesesuaian dengan prompt

Tabel 1 Skala Likert

Nilai	Kriteria
1	Sangat Buruk
2	Buruk
3	Cukup
4	Baik
5	Sangat Baik

Setelah melakukan pengujian kuantitatif, kemudian akan dilakukan pengujian secara kualitatif dengan melakukan perhitungan terhadap hasil survey dari setiap panelis. Untuk perhitungannya menggunakan rumus sebagai berikut:

$$\text{Rumus Index} = (\text{Total skor})/Y \times 100\%$$

Keterangan :

Total = jumlah seluruh skor penilaian panelis untuk suatu aspek/platform.

Y = skor maksimum yang dapat diperoleh (jumlah panelis × nilai maksimum Likert × jumlah aspek/aspek tunggal yang dinilai).

Nilai indeks dinyatakan dalam persentase (%) dan dapat dikategorikan untuk interpretasi mutu, misalnya:

80%–100% = Sangat Baik

60%–79% = Baik

40%–59% = Cukup

20%–39% = Buruk

0%–19% = Sangat Buruk

### Prosedur Eksperimen

1. Setiap platform menghasilkan tiga video (satu per gambar/prompt).
2. Untuk setiap video, metrik kuantitatif (PSNR, SSIM, waktu proses, resolusi, frame rate) dihitung menggunakan perangkat lunak standar (Python/scikit-image, FFmpeg).

3. Semua video dianonimkan dan diacak sebelum dinilai panelis.
4. Hasil dirangkum, dan rerata performa tiap platform dihitung serta dibandingkan.

### Hasil dan Pembahasan

Bagian ini adalah hasil analisis dan pembahasan data pengujian dari penelitian yang membandingkan kualitas empat platform AI (Kling, Runway, Pixverse, Hailuo) dalam menghasilkan video dari gambar (image-to-video). Analisis ini mencakup aspek kuantitatif (SSIM dan PSNR) dan kualitatif (penilaian visual berdasarkan skala likert), serta kelebihan dan kelemahan masing-masing sistem.

Tabel 2 Durasi Generate

Platform	Rata-rata Waktu Generate (Detik)
Kling	1,19
Runway	0,63
Pixverse	0,15
Hailuo	2,30

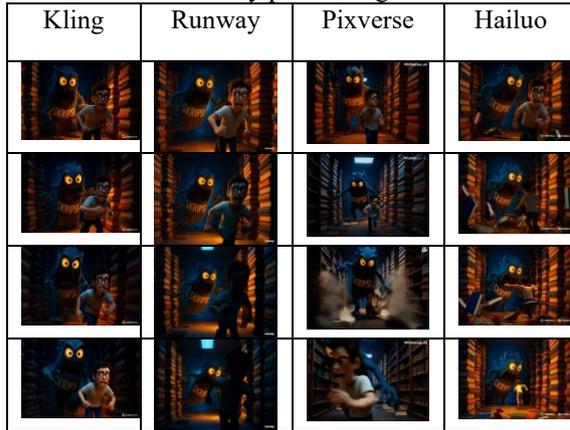
Pixverse paling cepat dalam proses generate, namun resolusi paling rendah. Kling dan Runway menawarkan resolusi tinggi dengan waktu generate cukup efisien. Hailuo memiliki waktu proses terlama.

Tabel 3 Gallery perbandingan Shot 1



Tabel 3 membandingkan kualitas visual dari masing-masing platform pada Shot 1, yang umumnya menjadi pengenalan atau adegan awal. Dalam shot ini, aspek seperti ketepatan ekspresi awal karakter, posisi kamera, dan suasana sangat menentukan kesan pertama. Runway tampil sangat unggul karena mampu menjaga konsistensi visual dengan pencahayaan sinematik dan tekstur yang tajam. Kling juga menunjukkan hasil baik, meskipun terdapat sedikit ketidakstabilan pada transisi antar frame. Hailuo mempertahankan karakter dengan cukup konsisten, meski sedikit tertinggal dalam hal pencahayaan. Sebaliknya, Pixverse tampak belum optimal dalam menjaga tekstur dan gaya visual secara merata, dengan hasil cenderung kasar atau tidak presisi terhadap prompt.

Tabel 4 Gallery perbandingan Shot 2



Tabel 4 menunjukkan bahwa tantangan utama biasanya terletak pada dinamika gerakan kamera dan interaksi karakter dengan lingkungan. Runway kembali menunjukkan performa tinggi, dengan gerakan kamera yang halus dan atmosfer yang konsisten. Kling berada di posisi menengah, menghasilkan pergerakan yang cukup stabil namun kadang disertai fluktuasi warna. Hailuo unggul dalam detail tekstur tetapi kadang gagal mempertahankan sudut kamera secara konsisten. Pixverse dalam shot ini masih menghadapi masalah dalam transisi antar frame dan ekspresi karakter, menghasilkan pengalaman visual yang kurang stabil. Secara keseluruhan, Runway menjadi pilihan terbaik untuk shot dinamis seperti ini.

Tabel 5 Galeri perbandingan shot



Tabel 5 umumnya merupakan puncak emosi atau klimaks dari narasi video. Oleh karena itu, aspek ekspresi wajah, pencahayaan dramatis, dan kesesuaian dengan prompt menjadi sangat penting. Dalam tabel ini, Runway berhasil menjaga intensitas dan emosi dengan sangat baik, didukung oleh pencahayaan yang mendalam dan sudut kamera sinematik. Hailuo mengejar ketertinggalan dengan visual yang cukup kuat, meskipun ekspresi karakter terkadang tampak datar. Kling berada di tengah, namun masih kalah dari Runway dalam hal kedalaman emosional dan atmosfer. Pixverse, lagi-lagi, tidak mampu mempertahankan stabilitas gaya

dan ekspresi, membuat hasil akhirnya terasa kurang kuat secara naratif.

### Pengujian Kuantitatif

Pengujian kuantitatif meliputi penilaian kualitas visual menggunakan metode SSIM (Structural Similarity Index Measure) dan PSNR (Peak Signal-to-Noise Ratio). Hasil pengujian kuantitatif ditampilkan pada tabel 6 dan tabel 7.

Tabel 6 Pengujian metode PSNR

Platform	PSNR Video 1	PSNR Video 2	PSNR Video 3	Rata-rata
<b>Kling</b>	28,10 db	28,47 db	28,26 db	28,28
<b>Runway</b>	28,41 db	28,65 db	28,35 db	28,47
<b>Pixverse</b>	18,24 db	14,5 db	19,43 db	17,39
<b>Hailuo</b>	28,91 db	28,57 db	28,12 db	28,53

Pada Tabel 6 menunjukkan bahwa Hailuo mencetak nilai rata-rata tertinggi (28.53 dB), menunjukkan ketajaman dan kejernihan visual terbaik secara teknis. Runway menyusul dengan (28.47 dB), sangat kompetitif dalam kejernihan gambar. Kling berada di posisi ketiga (28.28 dB), masih dalam kategori tinggi dan stabil. Pixverse jauh tertinggal (17.39 dB), menunjukkan kualitas visual yang kurang baik secara teknis.

Tabel 7 Pengujian metode SSIM

Platform	SSIM Video 1	SSIM Video 2	SSIM Video 3	Rata-rata
<b>Kling</b>	0,35	0,34	0,37	0,35
<b>Runway</b>	0,44	0,35	0,41	0,40
<b>Pixverse</b>	0,37	0,30	0,31	0,33
<b>Hailuo</b>	0,46	0,30	0,40	0,39

Tabel 7 menunjukkan bahwa Hailuo dan Runway konsisten unggul pada SSIM, menandakan struktur visual yang lebih akurat dan mirip referensi. Kling masih cukup baik, sedangkan Pixverse sangat fluktuatif (bervariasi drastis antar video).

### Pengujian Kualitatif

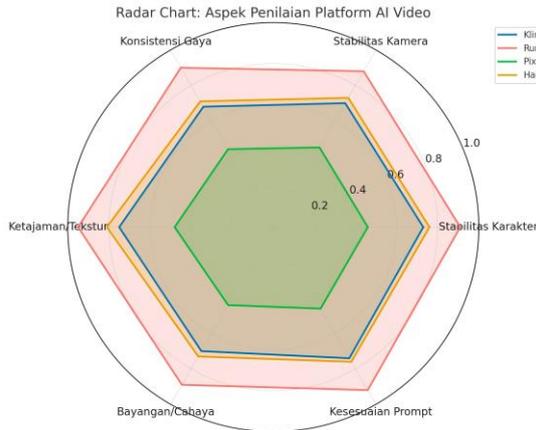
Dalam tahap ini, peneliti melakukan pengujian kuisioner yang diberikan terhap 5 panelis (Dosen, Praktisi industri, dan mahasiswa tingkat akhir). Adapun tampilan hasil pengujian dapat dilihat pada tabel 8 dan gambar 2.

Tabel 8 Total Perhitungan Skala Likert

Platform	Shot 1	Shot 2	Shot 3
<b>Kling</b>	74%	65%	79%
<b>Runway</b>	89%	83%	96%
<b>Pixverse</b>	46%	40%	50%
<b>Hailuo</b>	74%	71%	82%

Hasil perhitungan rata-rata yang ditunjukkan pada tabel 8 bahwa Runway konsisten unggul di seluruh shot (di atas 80% bahkan mendekati sempurna di shot 3). Kling dan Hailuo selalu di atas 65%, dengan

performa terbaik di shot 3. Pixverse selalu paling rendah, tidak pernah menyentuh 60%.



Gambar 3 Aspek penilaian Platform

Gambar 3 di atas menampilkan perbandingan performa empat platform AI video berdasarkan enam aspek utama. Dalam hal stabilitas karakter, Runway unggul dengan tampilan karakter yang konsisten dari awal hingga akhir, disusul oleh Kling dan Hailuo yang cukup stabil, sementara Pixverse mengalami banyak perubahan ekspresi dan bentuk. Pada stabilitas kamera, Runway kembali unggul dengan pergerakan sinematik dan halus, sedangkan Kling dan Hailuo cukup stabil dan Pixverse kurang konsisten. Dari segi konsistensi gaya visual, Runway menonjol dengan tone dan warna yang seragam, diikuti oleh Hailuo, sedangkan Pixverse sering mengalami perubahan mencolok antar frame. Dalam aspek ketajaman dan tekstur, Runway menghasilkan visual paling detail dengan noise minimal, Kling dan Hailuo berada di level sedang, sementara Pixverse cenderung buram. Untuk pencahayaan dan bayangan, Runway menunjukkan pencahayaan yang sinematik dan realistis, Kling dan Hailuo cukup baik, namun Pixverse sering tampak tidak natural. Terakhir, dalam hal kesesuaian dengan prompt, Runway paling akurat mewujudkan isi deskripsi, Hailuo dan Kling cukup relevan namun tidak selalu rinci, dan Pixverse sering menyimpang dari maksud prompt.

Hasil studi komparatif ini menunjukkan keunggulan dan kelemahan unik pada tiap platform AI image-to-video yang diuji. Runway secara konsisten menunjukkan performa unggul baik pada metrik objektif maupun subjektif. Skor PSNR dan SSIM yang tinggi menandakan pemrosesan teknis yang canggih, sehingga video yang dihasilkan tajam dan struktur visualnya sangat mirip referensi. Penilaian Likert panelis menunjukkan Runway unggul pada stabilitas karakter, stabilitas kamera, konsistensi gaya visual, dan kualitas sinematik. Ini menjadikan Runway sangat cocok untuk aplikasi yang membutuhkan visual fidelity tinggi seperti animasi profesional, storytelling, dan konten pemasaran.

Hailuo menyusul dengan performa baik pada metrik kuantitatif dan menonjol pada detail tekstur visual,

meskipun kadang kurang stabil pada sudut kamera dan pencahayaan. Hailuo direkomendasikan untuk kebutuhan detail visual tinggi seperti visualisasi produk atau video edukasi.

Kling menawarkan solusi stabil dan andal, dengan skor kuat di penilaian objektif maupun kualitatif, meski tidak menjadi yang terbaik di aspek tertentu. Keseimbangan stabilitas karakter, efisiensi biaya, dan kemudahan penggunaan membuat Kling cocok untuk proyek kreatif umum dengan anggaran terbatas. Namun, inkonsistensi warna dan transisi frame kadang membatasi untuk proyek sinematik.

Pixverse, meski unggul pada kecepatan workflow, secara konsisten berada di bawah platform lain pada kualitas teknis dan penilaian panelis. Pixverse lebih cocok untuk prototyping cepat, eksplorasi ide, atau visualisasi konsep awal, daripada produksi final. Kemudahan penggunaan dan proses cepat cocok untuk pengguna dengan waktu terbatas, tapi kualitas output masih perlu ditingkatkan untuk adopsi luas.

AI telah mentransformasi pipeline produksi animasi, mempercepat tahapan produksi dan memungkinkan kreator lebih fokus pada keputusan kreatif ketimbang tugas-tugas manual [3].

Temuan ini konsisten dengan literatur sebelumnya yang menyoroti pentingnya evaluasi multi-dimensi pada kualitas video AI [8]. Berbeda dari studi sebelumnya yang lebih fokus pada algoritma atau dataset sintesis, penelitian ini memberikan insight praktis berbasis pengalaman nyata, dengan kombinasi metrik teknis dan penilaian manusia. Penilaian panelis ahli memberi validasi tambahan sesuai ekspektasi pengguna dan standar industri. [7]

## Kesimpulan dan Saran

Penelitian ini menyajikan benchmarking komprehensif terhadap empat platform AI image-to-video komersial terkemuka Kling, Runway, Pixverse, dan Hailuo dengan pendekatan kombinasi metrik kuantitatif (PSNR, SSIM, waktu proses) dan kualitatif (skor Likert panelis ahli). Hasil eksperimen menunjukkan:

1. Runway secara konsisten memberikan kualitas visual dan performa sinematik terbaik, sehingga sangat direkomendasikan untuk aplikasi profesional dan kreatif yang menuntut fidelity visual tinggi dan konsistensi.
2. Hailuo menonjol dalam aspek detail tekstur dan kekayaan visual, cocok untuk kebutuhan visualisasi produk, dokumentasi fiktif, atau konten edukasi berbasis visual detail.
3. Kling menawarkan solusi stabil dengan performa baik dan efisiensi biaya, sangat cocok untuk workflow kreatif umum yang memerlukan keseimbangan antara kualitas dan affordability, meskipun tidak selalu unggul dalam setiap aspek spesifik.
4. Pixverse unggul pada aspek kecepatan workflow dan kemudahan penggunaan, sehingga dapat menjadi pilihan utama untuk rapid prototyping,

eksplorasi ide, atau kebutuhan produksi cepat, meski kualitas visualnya belum menyamai tiga platform lainnya.

Benchmarking ini membuktikan bahwa evaluasi multi-dimensi menggabungkan aspek teknis dan persepsi manusia sangat penting dalam menentukan platform terbaik sesuai kebutuhan spesifik. Framework evaluasi yang dikembangkan dalam penelitian ini juga dapat diadaptasi pada benchmarking platform AI generatif lain di masa mendatang.

#### Saran untuk penelitian selanjutnya:

Terdapat beberapa batasan: eksperimen hanya pada video pendek 5 detik dengan prompt/gambar terbatas; hasil bisa berbeda pada video lebih panjang atau konten lebih kompleks. Penilaian kualitatif hanya oleh panelis terbatas, sehingga subjektivitas mungkin terjadi. Selain itu, perubahan cepat pada platform AI komersial bisa mempengaruhi replikasi hasil di masa mendatang.

Penelitian selanjutnya perlu memperluas sampel data, menggunakan prompt dan tipe konten yang lebih beragam, serta melibatkan lebih banyak platform seiring perkembangan teknologi. Pengembangan framework benchmarking open-source akan sangat bermanfaat untuk evaluasi yang lebih transparan dan ketat.

#### Daftar Pustaka

- [1] M. Kristianto, D. Yuono, F. Teknik, U. Tarumanagara, F. Teknik, and U. Tarumanagara, "Interaksi manusia dan ai sebagai pendekatan desain ruang kreatif," vol. 5, no. 2, pp. 1699–1710, 2023, doi: 10.24912/stupa.v5i2.24310.
- [2] O. Article, "Future of animation with artificial intelligence," vol. 4, no. December, pp. 180–187, 2023, doi: 10.29121/shodhkosh.v4.i2SE.2023.5.
- [3] Y. Chen, Y. Wang, T. Yu, and Y. Pan, "The Effect of AI on Animation Production Efficiency : An Empirical Investigation Through the Network Data Envelopment Analysis," 2024.
- [4] X. Liu, X. Xiang, Z. Li, and Y. Wang, "A Survey of AI-Generated Video Evaluation," pp. 1–59.
- [5] A. Ghildyal, S. Member, Y. Chen, S. Zadtootaghaj, and N. B. Senior, "Quality Prediction of AI Generated Images and Videos : Emerging Trends and Opportunities," pp. 1–16.
- [6] A. Susanto, D. Sinaga, I. Utomo, and W. Mulyono, "PSNR and SSIM Performance Analysis of Schur Decomposition for Imperceptible Steganography," vol. 11, no. 3, pp. 803–810, 2024, doi: 10.15294/sji.v11i3.9561.
- [7] L. Zhao, X. Peng, Y. Tian, M. Kapadia, and D. N. Metaxas, "Towards Image-to-Video Translation : A Structure-Aware Approach via Multi-stage Generative Adversarial Networks," *Int. J. Comput. Vis.*, 2020, doi: 10.1007/s11263-020-01328-9.
- [8] F. Fan, C. Luo, W. Gao, and J. Zhan, "BenchCouncil Transactions on Benchmarks , Standards and Evaluations AIGCBench : Comprehensive evaluation of image-to-video content generated by AI," *BenchCouncil Trans. Benchmarks, Stand. Eval.*, vol. 3, no. 4, p. 100152, 2024, doi: 10.1016/j.tbench.2024.100152.
- [9] F. Syahputra *et al.*, "Evaluasi Efektivitas Ai Generatif Dalam Membantu Guru Menyusun Materi Pembelajaran Di Indonesia," vol. 3, no. 3, pp. 265–271, 2025.
- [10] Y. Tang *et al.*, "Generative AI for Cel-Animation : A Survey," pp. 1–20.
- [11] A. Q. Measurement, O. F. Video, P. Application, and U. Ssim, "PENGUKURAN KUALITAS APLIKASI PEMUTAR VIDEO MENGGUNAKAN METODE," pp. 501–508, 2020.
- [12] I. A. Detection, "Tinjauan Literatur Sistematis tentang Structural Similarity Index Measure untuk Deteksi Anomali Gambar," vol. 7, no. 2, 2020.