

## Analisis dan Perbandingan Akurasi *Image Generative AI DALL-E 3* dan *Midjourney* Menggunakan Metode *Frechet Inception Distance (FID)*

**Allan Nugraha<sup>1</sup>, Fathoni<sup>2</sup>, Alif Baidhawi<sup>3</sup>, Deni Agus Hendrawan<sup>4</sup>, M Bintang Naufal Riansyah<sup>5</sup>, Ali Ibrahim<sup>6</sup>**

Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Sriwijaya, Indonesia  
[09031282227130@student.unsri.ac.id](mailto:09031282227130@student.unsri.ac.id), [fathoni@unsri.ac.id](mailto:fathoni@unsri.ac.id), [09031282227122@student.unsri.ac.id](mailto:09031282227122@student.unsri.ac.id),  
[09031282227116@student.unsri.ac.id](mailto:09031282227116@student.unsri.ac.id), [09031282227118@student.unsri.ac.id](mailto:09031282227118@student.unsri.ac.id),  
[aliibrahim@unsri.ac.id](mailto:aliibrahim@unsri.ac.id)

**Abstrak:** Perkembangan teknologi *Generative AI* telah membawa kemajuan pesat dalam pembuatan gambar berbasis teks, dengan DALL-E 3 dan Midjourney sebagai dua model terdepan. Penelitian ini bertujuan untuk menganalisis dan membandingkan akurasi visual hasil gambar dari kedua model menggunakan metrik *Frechet Inception Distance (FID)*. Lima jenis *prompt* teks dipilih secara sistematis berdasarkan kategori anjing ras tertentu, dan setiap model menghasilkan 50 gambar yang dibandingkan dengan gambar acuan dari *dataset* Stanford Dogs. Hasil penelitian menunjukkan bahwa DALL-E 3 memiliki rata-rata skor FID sebesar 19.85, sedangkan Midjourney sebesar 28.42, yang berarti DALL-E 3 menghasilkan gambar yang lebih mendekati visual nyata. Uji statistik menggunakan *Independent Sample t-Test* menunjukkan adanya perbedaan signifikan antara kedua model. Dengan demikian, DALL-E 3 lebih cocok untuk kebutuhan yang menuntut realisme visual, sedangkan Midjourney lebih unggul dalam eksplorasi artistik.

**Kata Kunci:** DALL-E 3, Midjourney, *Frechet Inception Distance*, *Generative AI*, Perbandingan Akurasi

**Abstract:** The rapid development of *Generative AI* has led to significant advancements in text-to-image generation, with DALL-E 3 and Midjourney emerging as two of the leading models. This study aims to analyze and compare the visual accuracy of images generated by both models using the *Frechet Inception Distance (FID)* metric. Five systematically designed text prompts were used, based on different dog breeds, and each model generated 50 images, which were then compared against real reference images from the Stanford Dogs dataset. The results show that DALL-E 3 achieved an average FID score of 19.85, while Midjourney scored 28.42, indicating that DALL-E 3 produced more realistic images. A statistical analysis using the *Independent Sample t-Test* confirmed a significant difference between the two models. Therefore, DALL-E 3 is more suitable for tasks requiring high visual realism, while Midjourney excels in artistic exploration.

**Keywords:** DALL-E 3, Midjourney, *Frechet Inception Distance*, *Generative AI* accuracy comparison

### 1. Pendahuluan

Teknologi *Generative Artificial Intelligence (AI)* telah berkembang pesat dan membawa inovasi signifikan dalam pembuatan konten visual berbasis teks (*text-to-image*). Teknologi ini memungkinkan pengguna untuk menghasilkan berbagai konten visual melalui instruksi atau perintah teks (*prompt*), di mana AI beroperasi berdasarkan interaksi manusia untuk mengarahkan hasil yang diinginkan (Chiu, 2024); (Vartiainen & Tedre, 2024). Perkembangan model bahasa yang besar dan teknik difusi juga telah meningkatkan kemampuan interpretasi deskripsi pengguna secara drastis (Saharia et al., 2022). Fenomena ini menjadikan topik model generatif sangat ramai diperbincangkan,

terutama karena kemampuannya yang semakin mendekati kualitas gambar profesional. Persaingan antar pengembang model AI pun kian meningkat untuk menciptakan model yang mampu menghasilkan gambar dengan tingkat akurasi tinggi, realistik, dan memiliki estetika yang menarik.

Dalam lanskap ini, DALL-E 3 dari OpenAI dan Midjourney menjadi dua *platform* terdepan yang mendominasi industri kreatif dan akademik. DALL-E 3 dikenal dengan kemampuannya mengubah perintah teks menjadi gambar melalui antarmuka percakapan yang logis, sedangkan Midjourney, yang berbasis *diffusion model* dan beroperasi melalui Discord, diakui karena gaya artistiknya yang unik sehingga disukai seniman digital (Hanna, 2023; Jovanovic & Campbell, 2022; (Zhang & Liu, 2024). Kedua teknologi ini telah dimanfaatkan secara luas di berbagai sektor, mulai dari seni, desain, arsitektur, pendidikan, hingga kesehatan mental untuk terapi visual, karena kemampuannya mempercepat penciptaan ide tanpa memerlukan keterampilan menggambar manual (Derevyanko & Zalevska, 2023; Enjellina et al., 2023; Flathers et al., 2024; Lyu et al., 2022).

Meskipun kedua model tersebut memberikan hasil yang impresif, tantangan terkait akurasi dan konsistensi gambar terhadap prompt masih menjadi isu yang perlu diteliti lebih dalam (Jamal et al., 2024). Penelitian sebelumnya oleh (Borji, 2023) telah membahas kualitas gambar generatif secara umum, di mana ditemukan bahwa DALL-E cenderung unggul dalam kesesuaian struktur teks, sementara Midjourney lebih unggul dalam elemen artistik namun terkadang kurang realistik (Yildirim, 2023). Namun, sebagian besar penelitian sebelumnya belum melakukan perbandingan langsung menggunakan dataset pengujian yang sama, sehingga hasilnya cenderung bervariasi. Evaluasi yang lebih objektif terhadap efisiensi pemrosesan dan kualitas akhir sangat diperlukan untuk menilai kedua model ini secara adil (Zhou & Nabus, 2023).

Untuk mengatasi subjektivitas dalam penilaian kualitas gambar, penelitian ini menggunakan *Frechet Inception Distance* (FID) sebagai metrik evaluasi utama. FID dipilih karena keunggulannya dalam mengukur perbedaan distribusi fitur antara gambar yang dihasilkan AI dan gambar nyata dengan cara yang lebih selaras dengan persepsi visual manusia dibandingkan metrik tradisional seperti *Structural Similarity Index* (SSIM) atau *Peak Signal-to-Noise Ratio* (PSNR) (Jamal et al., 2024). Jika SSIM dan PSNR lebih berfokus pada perbandingan *pixel-to-pixel* yang kaku, FID mampu menangkap kesamaan statistik dan kualitas perceptual secara lebih komprehensif (Song et al., 2022). Oleh karena itu, skor FID yang lebih rendah menjadi indikator kuat bahwa gambar yang dihasilkan memiliki kualitas tinggi dan terlihat lebih realistik.

Penelitian ini bertujuan untuk mengevaluasi perbedaan akurasi gambar yang dihasilkan oleh DALL-E 3 dan Midjourney dengan menggunakan metrik objektif FID. Penelitian ini diharapkan mampu mengidentifikasi model yang lebih unggul dalam menghasilkan gambar berdasarkan kualitas visual dan kecocokan terhadap teks yang diberikan. Selain itu, penelitian ini juga bertujuan untuk memberikan rekomendasi bagi pengembang serta pengguna dalam memilih model AI yang paling sesuai dengan kebutuhan mereka, baik dalam bidang seni, desain, maupun keperluan profesional lainnya. Hasil studi ini diharapkan dapat memberikan wawasan lebih luas mengenai evaluasi model generatif dan berkontribusi bagi perkembangan teknologi AI di masa depan.

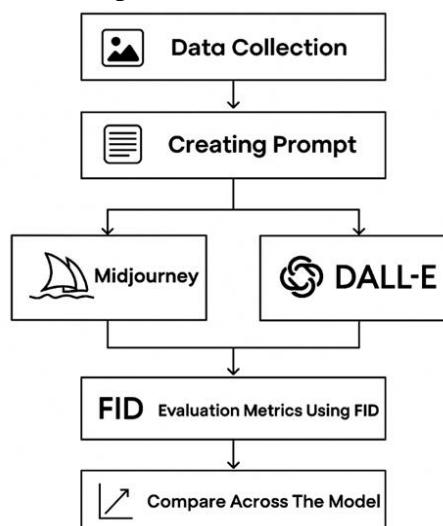
## 2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen untuk mengukur dan membandingkan akurasi gambar yang dihasilkan oleh DALL-E 3 dan Midjourney. Evaluasi dilakukan menggunakan metrik *Frechet Inception Distance*

(FID), yang mengukur kesamaan distribusi fitur antara gambar hasil generasi AI dengan *dataset* referensi.

Metode pengumpulan data yang digunakan dalam penelitian ini adalah eksperimen, *dataset* referensi, dan tinjauan pustaka. Pada metode eksperimen, penelitian dilakukan dengan menyusun total 5 *prompt* teks yang diklasifikasikan secara sistematis berdasarkan kategori visual lima ras anjing (*Border Collie, Rottweiler, Chihuahua, Border Terrier*, dan *Japanese Spaniel*) untuk menghindari bias pada satu jenis karakteristik visual. Setiap *prompt* diinput ke dalam kedua model untuk menghasilkan 10 gambar per kategori, sehingga diperoleh total sampel sebanyak 100 gambar generatif (50 gambar dari DALL-E 3 dan 50 gambar dari Midjourney). Seluruh hasil gambar tersebut kemudian dianalisis tingkat realismenya menggunakan metrik FID dengan pembanding dari *dataset* referensi.

*Dataset* referensi dalam penelitian ini diperoleh dari sumber publik, yaitu *Stanford Dogs Dataset*, yang memiliki karakteristik visual realistik dengan variasi pose, pencahayaan, dan latar belakang yang kompleks (Khosla et al., 2011). Dataset ini dipilih karena menyediakan citra dunia nyata dengan kualitas anotasi yang baik, sehingga relevan untuk mengevaluasi kualitas visual gambar yang dihasilkan oleh model AI generatif. Gambar yang dihasilkan oleh AI dibandingkan dengan *dataset* ini untuk mendapatkan skor FID. Selain itu, penelitian ini juga menggunakan tinjauan pustaka, di mana literatur sebelumnya terkait dengan AI generatif dan evaluasi kualitas gambar digunakan sebagai acuan dalam penelitian ini (Vartiainen & Tedre, 2024).



Gambar 1. Diagram alur Metode keseluruhan

Evaluasi terhadap hasil gambar dilakukan dengan menggunakan metode *Fréchet Inception Distance* (FID), sebuah metrik evaluasi yang mengukur kedekatan distribusi statistik fitur antara gambar hasil AI dengan gambar nyata. Perhitungan FID dilakukan melalui pemrograman Python dengan menggunakan pustaka seperti *TensorFlow*, *NumPy*, dan *Scikit-Learn*, serta model *pretrained Inception-v3* untuk ekstraksi fitur gambar. Rumus FID yang digunakan dalam penelitian ini adalah sebagai berikut:

$$FID = \|\mu^1 - \mu^2\| + \text{Tr}(\Sigma_1 - \Sigma_2 - 2\sqrt{\Sigma_1 \Sigma_2})$$

Keterangan

$\mu^1$  =Mean dari gambar asli

$\mu^2$  =Mean dari gambar fake

$$\Sigma_1 = \text{Matriks kovarian dari fitur gambar nyata}$$
$$\Sigma_2 = \text{Matriks kovarian dari fitur gambar fake}$$

Nilai FID dihitung berdasarkan distribusi kumpulan gambar yang dihasilkan untuk setiap kategori pengujian. Skor FID yang diperoleh dari masing-masing kategori tersebut selanjutnya dianalisis menggunakan uji statistik *Independent Sample t-Test* untuk menentukan apakah terdapat perbedaan yang signifikan secara statistik antara performa DALL-E 3 dan Midjourney secara keseluruhan. Nilai signifikansi (alpha) ditetapkan sebesar 0,05. Jika *p-value* dari hasil uji lebih kecil dari nilai alpha, maka perbedaan rata-rata kinerja antara kedua model dinyatakan signifikan. Hasil analisis disajikan dalam bentuk tabel statistik deskriptif dan visualisasi grafik batang, guna memberikan gambaran yang lebih jelas mengenai perbandingan kualitas distribusi visual dari masing-masing model AI.

### 3. Hasil dan Pembahasan

Tahapan awal penelitian ini meliputi penyusunan lima *prompt* teks secara sistematis berdasarkan kategori visual lima jenis anjing (Border Collie, Rottweiler, Chihuahua, Border Terrier, dan Japanese Spaniel) guna menguji kapabilitas interpretasi variasi deskriptif model. Sebagai acuan pembanding objektif dalam evaluasi, satu gambar nyata (*real image*) yang relevan dengan setiap *prompt* dikumpulkan dari *Stanford Dogs Dataset*, sebagaimana ditunjukkan pada Gambar 2.



**Gambar 2. Foto asli 5 jenis anjing, diambil dari Stanford Dogs Dataset**

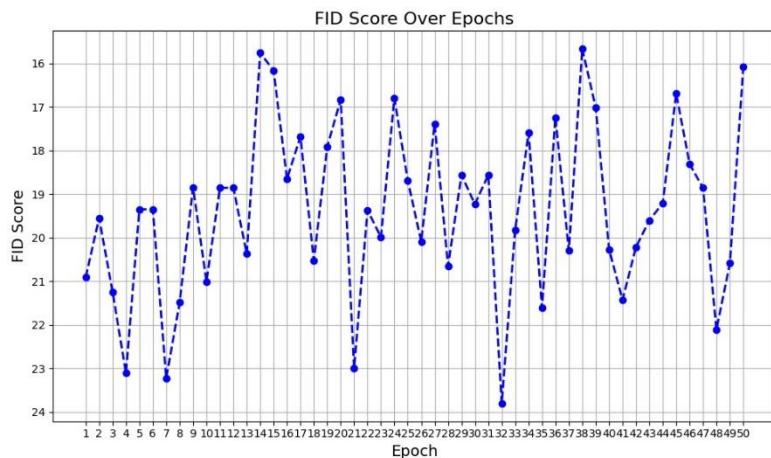
Setelah *prompt* dan gambar acuan dikumpulkan, tahap berikutnya adalah menghasilkan gambar menggunakan dua platform AI generatif, yaitu DALL-E 3 dan Midjourney. Proses generasi gambar dari DALL-E 3 dilakukan melalui antarmuka ChatGPT Plus, sedangkan Midjourney digunakan melalui sistem berbasis Discord. Masing-masing *prompt* digunakan untuk menghasilkan 10 gambar dari setiap *prompt* pada tiap model, sehingga total diperoleh 50 gambar dari DALL-E 3 dan 50 gambar dari Midjourney. Dengan tambahan masing masing *prompt* memiliki 10 gambar nyata sebagai acuan, jumlah keseluruhan gambar yang dikaji dalam penelitian ini adalah 150 gambar, gambar hasil generasi Midjourney ditunjukkan pada gambar 3, dan generasi DALL-E 3 pada gambar 4.



Gambar 3. Hasil Generasi dari model AI Midjourney



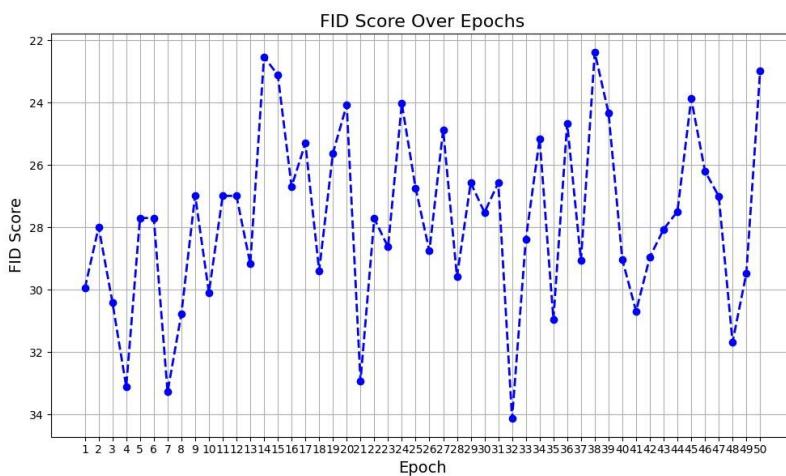
Gambar 4. Hasil Generasi dari model AI Dall-E 3



**Gambar 5. Grafik skor FID Dall-E**

Model DALL-E 3 menunjukkan performa yang sangat baik dalam menghasilkan gambar yang sesuai dengan deskripsi teks. Berdasarkan hasil evaluasi menggunakan metrik *Frechet Inception Distance* (FID), nilai rata-rata FID untuk DALL-E 3 adalah sebesar 19.3682 dengan standar deviasi 1.997. Nilai ini menunjukkan bahwa distribusi fitur visual dari gambar yang dihasilkan oleh DALL-E 3 sangat dekat dengan gambar nyata dari *dataset* referensi. Hasil tersebut menunjukkan bahwa DALL-E 3 mampu menangkap detail visual secara presisi dan realistik, terutama pada objek-objek konkret seperti hewan.

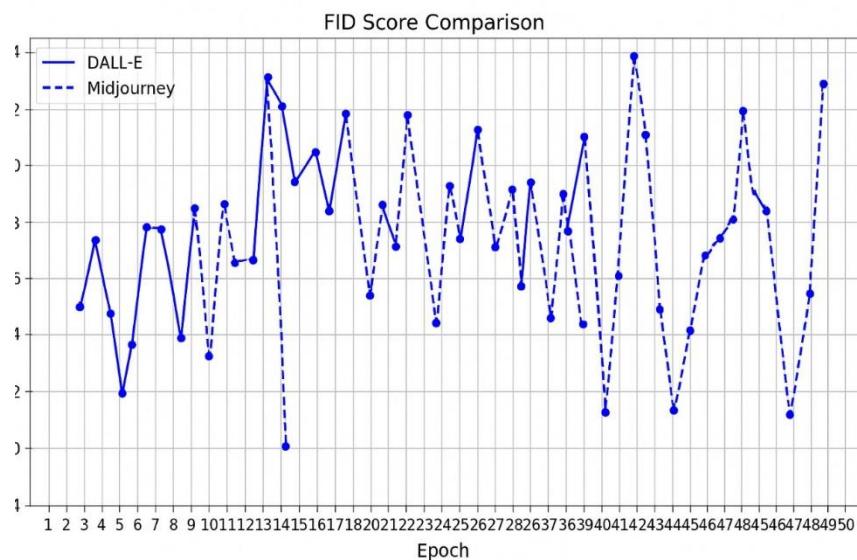
Secara visual, gambar-gambar dari DALL-E 3 memiliki komposisi yang terstruktur, proporsi anatomi objek yang akurat, serta latar belakang yang relevan dengan isi *prompt*. Misalnya, pada *prompt* “*a Border Collie sitting on a green field*,” DALL-E 3 berhasil menghasilkan gambar dengan proporsi tubuh anjing yang tepat, ekspresi natural, serta latar padang rumput yang realistik. Hal ini menunjukkan bahwa kekuatan utama dari DALL-E 3 terletak pada kemampuannya dalam memahami konteks deskriptif dari teks dan menerjemahkannya menjadi gambar yang faktual.



**Gambar 6. Grafik skor FID Midjourney**

Sebaliknya, Midjourney memiliki pendekatan visual yang lebih artistik dan ekspresif. Hasil gambar dari Midjourney menunjukkan rata-rata nilai FID sebesar

27.725 dengan standar deviasi 2.875. Nilai ini lebih tinggi dibandingkan dengan DALL-E 3, yang berarti gambar-gambar Midjourney secara statistik memiliki tingkat kesamaan visual yang lebih rendah terhadap gambar nyata dalam *dataset* referensi. Hal ini menunjukkan bahwa, dari sisi objektivitas visual berdasarkan metrik FID, Midjourney kurang mendekati realita, tetapi secara estetika gambar-gambar Midjourney memiliki karakter visual yang kuat. Gaya artistik yang dihasilkan cenderung kompleks, bertekstur kaya, dan memiliki nuansa warna yang dramatis. Pada *prompt* yang sama, yaitu “a Border Collie sitting on a green field,” Midjourney menampilkan hasil gambar dengan suasana sinematik dan sentuhan imajinatif, meskipun kadang terdapat ketidaksesuaian dalam struktur anatomi anjing atau latar belakang yang tidak relevan. Hal ini menegaskan bahwa kekuatan Midjourney lebih terletak pada kebebasan berkreasi visual daripada ketepatan terhadap isi *prompt*.



Gambar 7. Perbandingan grafik skor FID DALL-E dan Midjourney

Perbandingan komparatif antara DALL-E 3 dan Midjourney memperlihatkan adanya perbedaan fundamental dalam pendekatan pembangunan visual dari teks. DALL-E 3 menunjukkan keunggulan yang konsisten dari segi akurasi struktural dan kesesuaian terhadap deskripsi *prompt*, yang terbukti secara kuantitatif melalui capaian skor rata-rata FID yang lebih rendah serta rentang variasi skor yang lebih stabil antar kategori. Hal ini menandakan bahwa DALL-E 3 memprioritaskan fidelitas terhadap realisme referensi, menjadikannya lebih dapat diandalkan untuk menghasilkan visual yang anatomis dan logis. Di sisi lain, meskipun Midjourney memiliki nilai FID yang lebih tinggi—yang mengindikasikan adanya deviasi statistik dari citra referensi nyata—model ini menawarkan pengalaman visual yang lebih imajinatif dan kaya secara gaya. Tingginya skor pada Midjourney tidak serta-merta menunjukkan penurunan kualitas visual, melainkan mencerminkan kecenderungan model ini untuk melakukan eksplorasi artistik yang ekspresif.

Temuan dari penelitian ini, yang menunjukkan bahwa DALL-E 3 memiliki skor FID lebih rendah (lebih realistik) dibandingkan Midjourney, sejalan dengan pola yang ditemukan dalam studi sebelumnya oleh Borji (2022). Dalam penelitiannya

mengenai *generated faces*, Borji menemukan bahwa Midjourney memiliki skor FID yang paling tinggi dibandingkan kompetitornya (seperti Stable Diffusion dan DALL-E 2) karena kecenderungannya menghasilkan gambar yang bersifat surreal dan artistik, bukan fotorealistik. Meskipun penelitian Borji berfokus pada wajah manusia dan menggunakan DALL-E versi sebelumnya (DALL-E 2), hasil penelitian ini mengonfirmasi bahwa karakteristik fundamental Midjourney yang memprioritaskan estetika di atas realisme statistik tetap konsisten pada DALL-E 3 dan pada domain visual hewan (anjing). Kesenjangan skor FID dalam penelitian ini menegaskan bahwa DALL-E 3 telah mengalami peningkatan signifikan dalam menghasilkan *fidelity* (kesetiaan) terhadap referensi nyata dibandingkan pendahulunya, sementara Midjourney tetap mempertahankan identitas artistiknya yang kuat.

Selain itu, keunggulan DALL-E 3 dalam akurasi visual yang ditemukan melalui metrik FID ini juga mendukung analisis kualitatif yang dilakukan oleh Yildirim (2023) dalam konteks perancangan kota masa depan. Yildirim menyimpulkan bahwa DALL-E 3 adalah alat yang paling praktis untuk menghasilkan elemen desain yang jelas, koheren, dan logis, sedangkan Midjourney lebih unggul dalam menciptakan penggambaran yang mencolok secara visual namun sering kali dilebih-lebihkan (*exaggerated*) dan kurang presisi secara struktural. Kesamaan hasil antara penelitian ini dengan temuan Yildirim menunjukkan bahwa keunggulan DALL-E 3 dalam menghasilkan struktur yang akurat tidak hanya berlaku pada objek arsitektural yang kompleks, tetapi juga pada representasi makhluk hidup. Hal ini memperkuat implikasi bahwa tingginya skor FID pada Midjourney bukan semata-mata menunjukkan "kualitas buruk", melainkan mencerminkan perbedaan filosofi model generatif: DALL-E 3 mengejar presisi interpretatif, sementara Midjourney mengejar eksplorasi artistik.

Dari hasil tersebut, dapat disimpulkan bahwa DALL-E 3 lebih direkomendasikan untuk pengguna yang membutuhkan visual yang akurat, informatif, dan mendekati gambar nyata, seperti dalam bidang edukasi, perancangan produk, atau dokumentasi visual. Sementara itu, Midjourney lebih cocok digunakan untuk keperluan kreatif dan artistik seperti ilustrasi, seni digital, dan eksplorasi desain yang tidak terlalu bergantung pada realisme.

#### 4. Kesimpulan dan Saran

Berdasarkan hasil penelitian dan pembahasan, dapat disimpulkan bahwa terdapat perbedaan yang signifikan dalam kesamaan distribusi fitur visual antara model DALL-E 3 dan Midjourney. Hal ini didasarkan pada pengukuran FID yang menilai kedekatan statistik antara distribusi fitur gambar generatif dengan gambar nyata, bukan semata-mata pada kebenaran semantik konten. Dalam pengujian ini, DALL-E 3 menunjukkan performa lebih tinggi dalam menghasilkan gambar yang mendekati kenyataan berdasarkan skor FID yang lebih rendah, serta menunjukkan konsistensi yang lebih baik terhadap deskripsi *prompt*. Namun, perlu ditegaskan bahwa temuan ini berlaku dalam konteks eksperimen terbatas yang menggunakan 5 *prompt* pada domain visual spesifik (anjing) dengan *dataset* referensi berskala kecil. Oleh karena itu, hasil keunggulan skor FID ini tidak serta-merta dapat digeneralisasi ke seluruh jenis konten visual atau skenario penggunaan yang lebih kompleks. Di sisi lain, meskipun Midjourney mencatatkan skor FID yang lebih tinggi, model ini menghasilkan gambar dengan nilai estetika dan gaya artistik yang menonjol,

menjadikannya lebih unggul untuk keperluan seni dan desain yang mengutamakan kreativitas visual daripada realisme murni.

Secara keseluruhan, pemilihan model generatif sebaiknya disesuaikan dengan kebutuhan spesifik pengguna. Untuk tujuan edukasi, dokumentasi, atau pembuatan gambar yang membutuhkan kedekatan struktural dengan realitas, DALL-E 3 lebih direkomendasikan. Sementara itu, untuk kebutuhan kreatif, ilustrasi, atau karya seni digital yang ekspresif, Midjourney menawarkan fleksibilitas visual yang lebih luas. Penelitian ini berkontribusi dalam menyediakan kerangka evaluasi objektif awal bagi pemanfaatan teknologi generatif. Prospek penelitian ke depan perlu mencakup pengujian terhadap variasi *prompt* yang lebih luas dan domain visual yang beragam, penggunaan *dataset* referensi yang lebih besar, serta integrasi dengan metrik evaluasi tambahan guna menghasilkan penilaian yang lebih komprehensif.

## Daftar Pustaka

- Borji, A. (2023). *Generated Faces in the Wild: Quantitative Comparison of Stable Diffusion, Midjourney and DALL-E 2*. <http://arxiv.org/abs/2210.00586>
- Chiu, T. K. F. (2024). The impact of Generative AI (GenAI) on practices, policies and research direction in education: a case of ChatGPT and Midjourney. *Interactive Learning Environments*, 32(10), 6187–6203. <https://doi.org/10.1080/10494820.2023.2253861>
- Derevyanko, N., & Zalevska, O. (2023). Comparative analysis of neural networks Midjourney, Stable Diffusion, and DALL-E and ways of their implementation in the educational process of students of design specialities. *Scientific Bulletin of Mukachevo State University Series “Pedagogy and Psychology,”* 9(3), 36–44. <https://doi.org/10.52534/msu-pp3.2023.36>
- Flathers, M., Smith, G., Wagner, E., Fisher, C. E., & Torous, J. (2024). AI depictions of psychiatric diagnoses: a preliminary study of generative image outputs in Midjourney V.6 and DALL-E 3. *BMJ Mental Health*, 27(1). <https://doi.org/10.1136/bmjment-2024-301298>
- Hanna, D. (2023). The Use of Artificial Intelligence Art Generator “Midjourney” in Artistic and Advertising Creativity. *Journal of Design Sciences and Applied Arts*, 4(2), 42–58. <https://doi.org/10.21608/jdsaa.2023.169144.1231>
- Jamal, S., Wimmer, H., & Rebman, C. M. (2024). Perception and evaluation of text-to-image generative AI models: a comparative study of DALL-E, Google Imagen, GROK, and Stable diffusion. *Issues in Information Systems*, 25(2), 277–292. [https://doi.org/10.48009/2\\_iis\\_2024\\_123](https://doi.org/10.48009/2_iis_2024_123)
- Jovanovic, M., & Campbell, M. (2022). Generative Artificial Intelligence: Trends and Prospects. *Computer*, 55(10), 107–112. <https://doi.org/10.1109/MC.2022.3192720>
- Khosla, A., Jayadevaprakash, N., Yao, B., & Li, F.-F. (n.d.). *Novel Dataset for Fine-Grained Image Categorization: Stanford Dogs*. <http://vision.stanford.edu/aditya86/StanfordDogs/>
- Lyu, Y., Wang, X., Lin, R., & Wu, J. (2022). Communication in Human–AI Co-Creation: Perceptual Analysis of Paintings Generated by Text-to-Image System. *Applied Sciences (Switzerland)*, 12(22). <https://doi.org/10.3390/app122211312>
- Song, J., Meng, C., & Ermon, S. (2022). *Denoising Diffusion Implicit Models*. <http://arxiv.org/abs/2010.02502>
- Vartiainen, H., & Tedre, M. (2024). How Text-to-Image Generative AI Is Transforming Mediated Action. *IEEE Computer Graphics and Applications*, 44(2), 12–22. <https://doi.org/10.1109/MCG.2024.3355808>

- Enjellina, Eleonora Vilgia Putri Beyan, & Anastasya Gisela Cinintya Rossy. (2023). *JARINA-Journal of Artificial Intelligence in Architecture A Review of AI Image Generator: Influences, Challenges, and Future Prospects for Architectural Field.*
- Yildirim, E. (2023). COMPARATIVE ANALYSIS OF LEONARDO AI, MIDJOURNEY, AND DALL-E: AI'S PERSPECTIVE ON FUTURE CITIES. *Urbanizm*, 82–96.  
<https://doi.org/10.58225/urbanizm.2023-28-82-96>
- Zhang, Y., & Liu, C. (2024). Unlocking the Potential of Artificial Intelligence in Fashion Design and E-Commerce Applications: The Case of Midjourney. *Journal of Theoretical and Applied Electronic Commerce Research*, 19(1), 654–670.  
<https://doi.org/10.3390/jtaer19010035>
- Zhou, K. Q., & Nabus, H. (2023). The Ethical Implications of DALL-E: Opportunities and Challenges. *Mesopotamian Journal of Computer Science*, 2023, 16–21.  
<https://doi.org/10.58496/MJCSC/2023/003>