

A Technical and Aesthetic Exploration of Surreal Video Creation through Generative AI : A Focus on Google VEO 3

Hyeyoung Yoo¹

Abstract

This paper investigates Google's Veo3 video generation model as an emerging form of creative agency at the intersection of advanced generative AI and artistic surrealism. Recent advances in generative AI are blurring the boundaries between artistic imagination and technological representation, forging new modes of collaboration between human creativity and machine-driven computation. This study analyzes the technical architecture and training framework of Veo3, examines its relationship to surrealist aesthetics, and critically reviews the visual and temporal language of AI-generated videos through case studies of prompt-based video creation. The central argument is that Veo3 transcends its role as a mere video synthesis tool and instead operates as a nascent creative agent, producing novel and unexpected visual concepts through stochastic generation, cinematic understanding, and temporal coherence. By addressing both the artistic potential and the ethical limitations of AI-based video creation, this research contributes to ongoing discussions on computational creativity, human-AI collaboration, and generative media aesthetics.

Keyword : Generative AI, AI-based Video Creation, Google VEO3, Surrealism, Media Contents

1. Introduction

The recent explosion of AI-generated content has transformed creative media. Breakthrough in large language models (LLMs) and text-to-image generation (e.g. GPT-4, DALL-E, Midjourney) have paved the way for sophisticated video synthesis [1]. In 2024-2025, text-to-video models such as OpenAI's Sora and Google DeepMind's Veo marked a decisive shift from experimental motion synthesis toward coherent, high-definition moving imagery generated directly from natural language prompts. Google describes Veo as 'our most advanced video generation model' capable of high-definition, coherent video [2]. These tools enable novel creative workflows: for instance, Google notes that customers use Veo to streamline video ad production and rapid prototyping in marketing [3]. Google Veo has distinguished by its advertised capabilities for generating high-fidelity, 1080p resolution and higher video clips of extended duration. Crucially, Veo has been developed with a sophisticated understanding of cinematic

¹ Department Media Contents, Seowon University, Cheongju-si, Chungcheongbuk-do, Korea [Professor]
e-mail: hyy345@gmail.com

Received(October 21, 2025), Review Result(1st: November 16, 2025), Accepted(December 12, 2025), Published(December 31, 2025)



© 2025 The Authors. Published by NCISS.
This is an open access article licensed under the Creative Commons Attribution-NonCommercial 4.0 International License.
To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

language, capable of interpreting prompts that include terms like ‘wide shot’, ‘dolly in’, or ‘golden-hour lighting’. This positions Veo not simply as a generator of moving images, but as a potential instrument for deliberate cinematic and artistic expression [4].

However, despite rapid technical advances, the aesthetic implications of AI-generated video—especially its potential to produce surreal, dreamlike imagery grounded in cinematic temporality—remain underexplored. Surrealism in art emphasizes unconscious imagery, uncanny juxtapositions, and narrative disjunction [5]. When combined with generative video systems, this tradition raises new questions about how algorithmic models interpret, reproduce, and transform surreal visual language over time.

This study aims to bridge the technical and artistic domains by examining surreal video creation using Google Veo3. Aesthetically, this paper examines how Veo3 handles surreal prompts, identifying visual motifs and dynamics that echo classical Surrealist art. The ultimate goal is to understand Veo3 as a creative tool: how it augments human imagination, what limitations it has, and what new forms of expression it enables.

And this study seeks to move beyond the simplistic ‘AI as a tool’ metaphor to consider a more nuanced model of human-AI creative collaboration. The creative process with Veo3 is framed as a dialogic interaction between the human director's articulated intent (the prompt) and the model's algorithmic agency. This reframes the artist's role from that of a solitary creator to a curator, provocateur, or collaborator with a complex, non-human system, a shift articulated by creative partners who suggest AI enables a new form of rapid, iterative direction.

2. Theoretical Background

2.1 The Evolution of Generative AI and Video Synthesis

The path to advanced video-making AI like Veo has seen several major technological breakthroughs. Early attempts often used a technology called Generative Adversarial Networks, or GANs, which appeared in 2014. GANs work like a competition between two AIs: one AI, the generator, creates fake images, while the other, the discriminator, tries to spot the fakes. This contest pushes the generator to make incredibly realistic images [6]. While GANs were great for still pictures, they struggled with video. It was hard to keep the video stable and prevent objects from flickering or randomly changing from one moment to the next [7].

A major breakthrough came with the adoption of Transformer-based architectures, which enabled models to capture long-range dependencies and maintain consistency across frames. Building on this

foundation, contemporary systems such as Veo employ diffusion-based approaches. In diffusion models, video generation is framed as an iterative denoising process, in which random noise is progressively transformed into coherent visual sequences guided by textual prompts [8].

To make this process even more efficient, developers created Latent Diffusion Models (LDMs). Instead of working with the massive, high-detail video file directly, LDMs work on a smaller, compressed summary of the video, called a latent space. This saves a huge amount of computing power. To make sure the final video looks smooth and consistent, the AI is built with special temporal layers. These layers help the AI pay attention to how things look and move from one frame to the next, ensuring that a person or object doesn't suddenly change appearance or move illogically. This is the core technology that allows models like Veo to create stable, high-quality videos [9]. [Table 1] compares three major generative AI architectures.

[Table 1] Comparative Analysis of Key Generative AI Architectures

Architecture	Core Principle	Key Strength	Key Limitations	Relevance to Artistic Control
GAN	Adversarial training between a generator and a discriminator.	High degree of realism in generated images.	Training instability, mode collapse, limited temporal coherence for video.	Low; difficult to control output with precision.
Transformer-based	Self-attention mechanism to capture long-range dependencies.	Excellent at understanding global context and relationships in data.	High computational cost, can be less effective at fine-grained texture generation.	Medium; provides structural control but less direct visual manipulation.
Diffusion Model (Veo, Sora)	Iterative denoising of a random signal, conditioned on a prompt.	High-fidelity output, strong prompt adherence, diverse generation.	Slower inference speed compared to GANs, computationally intensive.	High; allows for detailed, nuanced control via natural language prompts.

2.2 Technical Foundations of Google Veo3

Google Veo represents a consolidation of years of research in generative media, building upon predecessors like Imagen Video and VideoPoet. Veo3 is a unified text-to-video and image-to-video model based on latent diffusion, designed to jointly model visual, temporal, and-at a basic level-audio coherence [10].

A primary feature of Veo is its ability to generate video at high resolutions, including 1080p and even 4K, providing a level of visual fidelity suitable for professional and artistic applications. This

moves generative video beyond the realm of low-resolution curiosities. Furthermore, the model has been trained on a vast dataset that includes metadata describing cinematic techniques. This allows creators to use specific directorial language in their prompts such as ‘timelapse’, ‘aerial shot of’, or ‘dolly zoom’ to exert fine-grained control over the final output's cinematography.

Google's official communications emphasize Veo's nuanced understanding of the physical world and its dynamics. The model can generate plausible motion and interactions between objects and characters, suggesting it has developed a sophisticated internal world model. This baseline understanding of reality is a crucial prerequisite for its subsequent artistic subversion. A model must first understand the rules of physics before it can be instructed to break them in a visually compelling way [11].

To achieve its hallmark temporal consistency, Veo relies on a suite of advanced architectural components. Its diffusion transformer framework utilizes spatio-temporal transformers to understand the relationships between different parts of a frame and across different frames over time. Technologies analogous to scene memory networks help maintain the identity and appearance of characters and objects throughout a sequence, preventing the common generative AI problem of subjects inexplicably morphing or changing [12]. This robust mechanism for coherence is precisely what an artist must strategically disrupt to achieve a Surrealist effect.

Veo3 is among the most advanced video AIs. It can create vivid, physically plausible scenes with sound. Veo3 stands as a professional-grade video AI with cutting-edge fidelity. Its closest peer, Sora 2, shares many capabilities, but differs in presentation and community features. Runway and Pika provide more turnkey, creative-platform experiences. This [Table 2] summarizes the key features and limitations of four leading video generation models.

[Table 2] Comparison with Leading Video Generation Models

Model	Key Features	Limitations
OpenAI Sora	Long sequences and realistic physics representation	Closed research model
Runway Gen-2	User-friendly interface	Limited resolution
Pika	Focus on web-based prompt generation	Limited artistic control
Veo3	High-quality spatio-temporal coherence	Limited model access

2.3 Surrealism in Visual Media

Surrealism is an artistic movement (circa 1920s-1960s) that seeks to unlock the unconscious and defy rational perception. The Surrealists believed that the rational mind represses imagination, and so art should reveal the omnipotence of dreams and hidden meanings.

Visually, Surrealist works often juxtapose incongruous elements, create impossible scenarios, and employ dream-logic. For example, Salvador Dalí painted melting clocks and fantastical landscapes, while René Magritte depicted everyday objects in uncanny contexts. In film, Surrealists like Luis Buñuel embraced narrative disjunction: in *Un Chien Andalou* (1929), scenes cut abruptly and illogically, emphasizing shock and subconscious metaphor rather than linear story [13]. René Magritte's surrealist painting *Decalcomania* (1966) [Fig. 1] uses a distinctive cut-out silhouette to explore the boundaries between presence and absence, challenging the viewer's perception of reality.



[Fig. 1] René Magritte, *Decalcomania* (1966)

Within this tradition, cinematic surrealism relies not only on visual juxtaposition but also on temporal rupture-abrupt cuts, irrational continuity, and affective dissonance over time. Veo3's generative architecture enables a contemporary form of this practice. Prompt engineering functions as a digital analogue to Surrealist automatism and *dépaysement*, allowing creators to force the synthesis of incompatible concepts while embracing stochastic variation as a creative resource [14].

Advanced operations such as video-to-video translation, inpainting, and outpainting further support the creation of fluid, dreamlike transformations that would be prohibitively complex using traditional filmmaking techniques. In this sense, Veo3 operates as a technical substrate through which Surrealist aesthetics can be rearticulated in motion-based, AI-mediated form.

3. Case Studies in Surreal Video Creation with VEO 3

3.1 Case Study 1: Dreamlike Landscapes

Prompt: “A wide shot of a twilight forest. Giant bioluminescent mushrooms tower amid mist, and translucent geometric cubes float among the trees. The sky is a deep purple with two moons. Cinematic warm lighting, low angle shot, ethereal atmosphere.”



[Fig. 2] Image Sequence of Dreamlike Landscapes (2025)

Veo3 produced an 8-second clip at 1080p as shown in [Fig. 2]. The scene opens with a slow pan across a dark forest. Oversized glowing mushrooms emit light. Semi-transparent cubes drift lazily in midair, rotating slowly. The dual moons are visible in the sky, casting a soft violet glow. The camera dollies forward among the foliage. Subtle ambient audio plays: soft wind and an otherworldly hum.

The resolution is sharp and stable. Frame-to-frame continuity is good; no obvious flicker. The floating cubes maintain their shape but occasionally their textures blur briefly. The lighting is rendered smoothly. All prompt elements appear: the mushrooms, cubes, double moons, purple sky. The only artifact was a rare moment where a cube's edge looked slightly wavy.

The mood is clearly dreamlike. The lighting and colors feel otherworldly. Floating cubes and enormous mushrooms are classic surreal tropes, reminiscent of Dalí's strange plants and Magritte's incongruous shapes. The scene has narrative ambiguity: nothing happens beyond camera movement. This aligns with Surrealist imagery which often depicts static, uncanny scenes. The lack of logical camera path, it glides through objects, adds to the unreal quality.

3.2 Case Study 2: Reality-Bending Motion

Prompt: “A street scene in broad daylight. Gravity is inverted: cars and bicycles float a few feet above the road, and pedestrians walk upside-down on the sky. Clouds move below their feet. Medium shot, handheld camera style, natural colors, subtle background music.”



[Fig. 3] Image Sequence of Reality-Bending Motion (2025)

The 8-second clip starts with a slightly shaky, low-angle shot of an urban street. Cars are mid-air, gently bobbing. People appear upside-down against the sky, their shoes facing up. Clouds drift past as though the ground were above. Ambient city sounds mixed with a whimsical tune as shown in [Fig. 3].

The resolution is high; motion is mostly smooth. Some distortion occurs at the edges of floating cars. The spatial inversion is consistent throughout. The prompt's elements floating vehicles, inverted people are all present, demonstrating strong prompt fidelity.

This clip exemplifies bent reality. The visuals are both comedic and uncanny. The surreal effect comes from seeing ordinary objects obeying no gravity. This recalls Magritte's knack for twisting context. Emotionally, it elicits surprise and amusement. There is a faint narrative tension: one wonders how daily life proceeds when physics is reversed. Yet the video makes no attempt to explain it. It simply is, echoing Surrealism's favoring of mystery over rationale.

4. Discussion

4.1 Aesthetic Analysis

The visual language of these AI-generated surreal videos blends photographic realism with fantastical elements. Key features include vivid color contrast, dreamlike lighting, and seamless blending of normal and abnormal elements. For instance, the floating cubes in Example 1 are rendered with realistic transparency and shadows, making the impossible seem almost natural. This approach is reminiscent of painterly surrealism: the more realistically the scene is depicted, the more jarring the surreal element becomes. Veo3 often accentuates textures, glowing fungi, crisp city details to heighten immersion.

Notably, AI outputs introduced novel surreal motifs that do not directly mimic any one artist. For example, the combination of bioluminescent flora with geometric shapes is not a classic Surrealist trope but fits the same spirit of juxtaposition. In this sense, the AI is not copying Dalí or Magritte, but synthesizing a new hybrid style from its vast training data. The result is a quasi-surreal visual grammar: ordinary scenes plus one or two violations of reality. This often yields an uncanny but aesthetically cohesive image.

In contrast to prior scholarship on AI-generated art, which has largely focused on static images, algorithmic style transfer, or questions of authorship in computational creativity, this study emphasizes video-based surrealism as a distinct aesthetic domain. Existing studies on AI art often examine visual novelty, imitation, or posthuman authorship at the level of single images. By comparison, the present research highlights how generative video models such as Veo3 introduce temporal coherence, cinematic language, and motion-based affect as central aesthetic variables. Surrealism in this context unfolds not only through visual juxtaposition, but through the continuity of movement, camera behavior, and audiovisual rhythm. This temporal dimension allows surreal effects to emerge dynamically over time, positioning AI-generated video as a qualitatively different medium from image-based AI art and extending discussions of computational creativity into the realm of cinematic and time-based expression.

Building on this distinction, the concept of ‘creative agency’ proposed in this study is not understood as autonomous authorship on the part of the AI system. Rather, creative agency is defined here as a relational and process-based capacity that emerges from the interaction between human intentionality—such as prompt formulation, iterative refinement, selection, and curatorial judgment—and the model’s stochastic, generative, and associative mechanisms. In this sense, Veo3 does not replace human creativity but participates in a dialogic creative process, where meaning and aesthetic direction are co-constructed across human and machine actions.

4.2 Ethical Considerations

The ability to generate photorealistic video raises significant ethical questions. For example, Veo could

in theory generate a convincing news clip of an event that never happened, or a deepfake of a public figure delivering a surreal speech. Google attempts to mitigate this: Veo and other models embed an invisible watermark (SynthID) to help detect AI-generated content. They also filter training data to exclude hate speech or violent extremism.

This study, focusing on surreal content, did not trigger hateful or sensitive themes, but any practical deployment of such models must include safety review. Moreover, authorship and originality raise concerns. While AI can generate surreal visuals, it also raises questions about plagiarism: Veo has been trained on copyrighted movies and videos. Does a Veo-generated clip inadvertently borrow a protected visual element? Google's policy includes copyright indemnity for generative AI customers, but the legal landscape is still evolving. Users and creators should use these tools ethically, giving credit to original inspiration and avoiding misuse of likenesses.

5. Conclusion

This study has established that Google Veo is a technologically potent and conceptually resonant medium for the creation of 21st-century Surrealist video. Veo3 is a powerful diffusion-transformer system: it jointly generates video and audio from text prompts, producing coherent motion and synchronized sound.

On the aesthetic side, Veo-generated clips readily embody Surrealist characteristics. The case studies showed that with dreamlike prompts, Veo can produce vivid fantasy scenes that mirror traditional surreal art such as floating objects, inverted physics, and uncanny juxtapositions. These videos stimulate viewer imagination similarly to human-made surreal films and paintings. However, the narratives remain fragmentary, in keeping with Surrealism's emphasis on mystery and the subconscious. In summary, Veo3 allows creators to explore a digital Surrealism, blending human intent with AI's generative power.

The advent of powerful generative video models like Veo3 signals a profound shift for the future of cinema and the visual arts. These technologies have the potential to democratize filmmaking on an unprecedented scale, enabling individual creators to produce visually complex narratives that were once the exclusive domain of large studios. This may lead to an explosion of new genres and forms of storytelling, blurring the lines between director, animator, visual effects artist, and writer. The ability to rapidly visualize ideas will change pre-production, storyboarding, and the very process of creative discovery.

This study has several limitations that point toward directions for future research. First, the prompt set

employed in this analysis was intentionally limited and illustrative rather than exhaustive; broader and more systematically varied prompt experiments may yield additional insights into the expressive range and constraints of AI-generated surrealism. Second, this study examined Veo3 at a specific moment in late 2025, and subsequent model updates may significantly alter its technical capabilities and aesthetic tendencies. Third, although qualitative comparisons were made with models such as Sora and Runway, controlled cross-model experiments were not feasible due to access limitations; future studies could benefit from comparative benchmarks under standardized conditions.

In conclusion, Google Veo3 represents a significant step toward AI-driven filmmaking. By studying its use in crafting surreal imagery, we gain insights into the evolving role of machines in visual art. The fusion of technical innovation and human imagination promises exciting new genres of expression, if developed thoughtfully and inclusively.

References

- [1] S. Rui, et al., “From Sora What We Can See: A Survey of Text-to-Video Generation”, IEEE Fellows, May 2024, pp. 1-20.
- [2] Google DeepMind, “Veo”, deepmind.google, <https://deepmind.google/models/veo>, (accessed October 3, 2025).
- [3] W. Barkley, “Introducing Veo and Imagen 3 on Vertex AI”, cloud.google.com, <https://cloud.google.com/blog/products/ai-machine-learning/introducing-veo-and-imagen-3-on-vertex-ai?hl=en>, (accessed October 2, 2025).
- [4] A. Wang “What Is Google Veo? The Plain English Guide to Google’s AI Video Generation Model”, deepmind.google, <https://deepmind.google/models/veo>, (accessed October 3, 2025).
- [5] The Art Story Contributors, “Surrealism”, theartstory.org, <https://www.theartstory.org/movement/surrealism>, (accessed October 3, 2025).
- [6] A. Islam, “A History of Generative AI: From GAN to GPT-4”, marktechpost.com, <https://www.marktechpost.com/2023/03/21/a-history-of-generative-ai-from-gan-to-gpt-4>, (accessed October 4, 2025).
- [7] J. Song, et al., “From Sora What We Can See: A Survey of Text-to-Video Generation”, arXiv, February 2025, doi.: 10.48550/arXiv.2502.04116.
- [8] B. Sharma, “Video Generation: Evolution from VDM to Veo2 and SORA”, learnopencv.com, <https://learnopencv.com/video-generation-models>, (accessed October 4, 2025).
- [9] J. Bengochea, “Diffusion models for video generation”, blog.marvik.ai, <https://blog.marvik.ai/2024/01/30/diffusion-models-for-video-generation>, (accessed October 4, 2025).
- [10] Google DeepMind, “Veo 3 Technical Report”, storage.googleapis.com, <https://storage.googleapis.com/deepmind-media/veo/Veo-3-Tech-Report.pdf>, (accessed October, 5 2025).
- [11] Google DeepMind, “Veo”, deepmind.google, <https://deepmind.google/models/veo>, (accessed October 5, 2025).

- [12] Y, Zhiyu, et al., “Sora: A Review on Background, Technology, Limitations, and Opportunities of Large Vision Models”, arXiv, February 2025, <https://arxiv.org/html/2502.17863v1>.
- [13] The Art Story, “Surrealism Movement Overview”, TheArtStory.org, <https://www.theartstory.org/movement/surrealism>, (accessed October 6, 2025).
- [14] E. Collins, “Fuel your creativity with new generative media models and tools”, blog.google, <https://blog.google/technology/ai/generative-media-models-io-2025/#veo-3>, (accessed October 6, 2025).