Vol. 16 (2026): Proceedings of the 2026 International Conference on AI and Technological Innovation Applications (AITIA 2026)
E-ISSN: 3105-0522 P-ISSN: 3105-0514
DOI: https://doi.org/10.70267/aitia.2026427434

A Comparative Study on the Practicality of AI Video Generation Tools in Short Video Creation

Ruolin Wang*

Shandong Normal University, Daxue Road, Jinan City, Shandong Province, China

**Corresponding author: Ruolin Wang.*

Abstract

With the rapid advancement of artificial intelligence, AI-generated content (AIGC) is experiencing explosive growth on short-form video platforms. It is increasingly permeating diverse fields, including lifestyle entertainment, gaming, and anime. Currently, the mainstream AI video-generation technologies primarily include three paradigms: text-to-video (T2V), image-to-video (I2V), and video-to-video (V2V). These three types of technologies differ significantly in their generation mechanisms, underlying principles, and output quality. Empirical studies reveal that the quality of AI-generated videos exhibits considerable heterogeneity. Some works achieve a remarkable level of realism, closely resembling real-world scenarios, while other low-quality content displays obvious flaws that make it easily identifiable. In-depth analysis shows that T2V technology excels in following user instructions and generating high-quality videos, while V2V technology also delivers impressive visual transformation results. By contrast, I2V technology is constrained by the limited information in the source images and insufficient guidance from prompts, which often makes it difficult to produce satisfactory dynamic videos. Moreover, influenced by the algorithmic recommendation mechanisms of short-video platforms, different platforms adopt varying strategies for distributing AI-generated content. Based on these observations, this paper aims to explore how to optimize the AI video creation process and, in conjunction with platform rules, develop effective publishing strategies to enhance the dissemination and impact of AI-generated content within the short-video ecosystem.

Keywords

AI video generation, text-to-video, short-video platforms, sora, runway, Seedance

1. Introduction

Today, AI-generated content is continuously improving. On short-video platforms like Douyin, we can see numerous AI-generated video ads across categories such as online games, commerce, and tourism. Based on consumer-grade AI video-generation tools, the length of videos that could be produced has evolved dramatically—from just 3-4 seconds in 2022 to breakthroughs allowing videos lasting from 10 seconds to 2 minutes by 2026. AI-generated videos are undergoing constant development and innovation. Compared to the earliest AI video-generation technologies, scientists have now developed many sophisticated models.

The use of text-to-video, Temporal Blocks, Frame Interpolation Architecture, and Video Autoencoder can help improve the "building blocks" of model architectures, thereby enhancing both video generation quality and generation speed [1]. In our study of five AI models for text-to-video [2], we found that text2Video-Zero exhibits superior alignment and perception, making it a promising model for T2V applications.

If you want to use image-to-video for video creation and don't have any prompting assistance, you can rely on just a few images or even a single one. When creating videos, AI may generate images that it interprets on its own—images that don't align with our intended ideas. To address this issue, scientists have developed Controllable Image-to-Video Motion Transfer [3], which does not rely on vague text descriptions but instead extracts explicit motion guidance directly from the driving video. Furthermore, by employing a cross-attention mechanism, this approach combines image content with semantic trajectories, enabling precise and flexible control over the generated motions.

As for video-to-video conversion, its primary uses are enhancing image quality, adding and correcting frames, and making videos appear smoother and more vivid in color. At the same time, this method also produces relatively high-quality results.

In addition, researchers conducted a questionnaire survey on social media fatigue related to AI-generated videos among 495 TikTok users in Vietnam, introducing the concept of "perceived AI content overload" [4]. This finding highlights that AI-generated videos are becoming increasingly prevalent on short-form video platforms, leading to varying degrees of acceptance across different videos and circumstances. This paper aims to compare current mainstream AI video-generation tools in terms of generation modes, content quality, creation efficiency, cost, and platform compliance, and to analyze their applicability to short-video creation, thereby producing AI-generated videos that viewers more readily accept.

2. Explanation

This article selects currently popular or previously trending AI-generated video software—Sora, Veo3, Pika, RunwayGen3, and Seedance2.0—and compares them based on their respective functionalities, image quality, resolution, frame rate, video duration, and video style. The data and literature sources are drawn from Web of Science or official authoritative websites; no actual testing was conducted.

2.1 Sora

As a once-leading-edge generative AI (GAI), Sora's outstanding text-to-video capability has been widely adopted by both companies and individuals. Particularly in the fields of advertising and marketing [5], people frequently encounter AI-generated ad videos on TikTok, which indirectly highlight Sora's cutting-edge nature and practical applicability in AI-powered video generation. Meanwhile, researchers' study data demonstrate Sora's high text-video consistency, realism, and technical quality [6]. Moreover, Sora boasts relatively long video durations and a high frame rate of 30 frames per second, indicating that it has reached the most advanced and mature stage in video generation—making it our top choice for text-to-video tasks. However, Sora's capabilities are optimized specifically for text-to-video tasks; it is less adept at image-to-video and video-to-video generation. Therefore, when faced with tasks involving creating videos from images or existing videos, we can turn to other AI software or online platforms to accomplish them.

However, the development of Sora has been accompanied by certain limitations. When presented with specific themes—such as issues related to colonialism—Sora's generated outputs closely align with the dominant visual conventions of settler-colonialism [7]. Rather than offering disruptive alternative perspectives, Sora tends to reinforce the hegemonic framing of memory that it has learned from biased training data. Thus, although Sora excels in video quality, it requires more stringent prompt control and human oversight when dealing with content related to history, politics, identity, and cultural narratives.

As the number of Sora users continues to grow, researchers have begun studying Sora's societal safety implications [8]. Analyzing 292 comments from the X platform, we can see that in the early stages of AI development, people held high expectations for AI. However, as Sora continues to improve and evolve, its creations may raise concerns regarding potential infringements on artists' rights and possible conflicts with others' interests. The issues are gradually emerging. Therefore, when using Sora, lawful usage is particularly

important. However, due to profitability concerns, the developers of Sora will completely shut down the service in 2026.4.26 [9].

2.2 Veo3

Google DeepMind (Veo3) excels in text-to-video generation and natively synchronizes audio, giving it a significant advantage in integrated generation: it can directly generate character dialogue, lip-sync alignment, ambient sounds, sound effects, and basic background music [10]. Therefore, it offers substantial benefits when creating videos featuring human characters. Additionally, it supports high resolutions (1080p, 720p) and incorporates physical principles, ensuring greater stability when handling complex movements. Beyond that, it can process both text-to-video and image-to-video inputs, making it highly functional and practical.

One step further, Veo 3 employs latent diffusion technology [11], applying the diffusion process simultaneously to both temporal audio and spatiotemporal video latent representations. Video and audio are each compressed into latent representations via their respective autoencoders, which significantly enhances the model's learning efficiency compared to directly processing raw pixels or waveforms. During the training phase, the model optimizes a Transformer-based denoising network designed to remove noise from noisy latent vectors. When generating samples, this network is iteratively applied to input Gaussian noise, ultimately producing the final video. In complex tasks such as virtual try-on, the two-stage latent diffusion system enables step-by-step processing, ensuring both geometric alignment between the human body and the garment and high-fidelity texture rendering. At the same time, this technology improves audio synchronization, enhancing video realism and making it better suited for applications in the AI industry.

Researchers convened a group of food vendors in Indonesia and conducted a Veo3 advertising workshop [12]. The study found that when microentrepreneurs use the AI tool Veo3 to safeguard their cultural identity and promote products that resonate with their target audiences, this approach effectively fosters computational empowerment. This phenomenon underscores how AI is once again empowering the evolution of advertising and driving societal progress, while also enhancing the development of individual business skills. It also indirectly reflects Veo3's inclusive integration into mainstream society.

Despite Veo's continued impressive strides in video generation, producing videos with natural and consistent speech audio—especially for shorter clips—remains an active area of development.

In terms of price, Veo3 has a relatively high cost (US$2.07–2.66 per 5 seconds at 1080p), making it suitable for enterprise or corporate applications.

2.3 Pika

As a frequent guest on short-video platforms, Pika has gained fame for its unique advantages. Through in-depth research, it offers features such as one-click special effects (Pikaffects), customizable first frames (Pikaframes), and AI lip-syncing (Pikaformance). These capabilities enable users to quickly create AI-generated videos from short clips, with highly distinctive visual styles. However, the data from Pika 1.0 [6] isn't particularly impressive; research suggests that the newly released Pika 2.2 can only meet the needs of everyday tasks.

In addition to text prompts, PikaAI also supports generating videos from images [13]. Users can upload static images or illustrations, and the platform will animate them with dynamic effects—for example, moving clouds across a landscape or altering facial expressions in a portrait. This feature brings static visuals to life, making it an invaluable tool for marketers, educators, and artists looking to add dynamism and depth to their content.

However, when researchers explored the application of AI technology to theological aesthetics and religious experience [14], they used Pika to transform static images into videos. They found that AI-generated images tend to remain superficial, failing to capture the "darkness of the cross"—the mystical experience of weakness, vulnerability, and failure that lies at the heart of faith. The authors acknowledge the value of AI as a tool but caution against its "mystification" and ideological pitfalls. Thus, this highlights Pika's limitations in expressing complex spiritual or abstract themes. We can use Pika for simple tasks or create easily understandable short videos, avoiding overcomplication that might lead to significant deviations between the generated videos and our intended outcomes. From a consumer perspective, Pika offers a low-usage price

($0.49 per 5 seconds at 1080p), giving it a significant advantage in cost-effectiveness and making generative AI more accessible and easier for the average person to try and use.

2.4 RunwayGen3

Runway Gen3 features two main functions: Alpha, which primarily corresponds to T2V, and Turbo, which primarily corresponds to I2V (Need to upload three pictures) [15]. Runway Gen3 is renowned for its film-quality lighting and textures, as well as its high level of professionalism and consistency. Moreover, it offers higher frame rates and resolutions than other conventional AI software. Not only does it deliver outstanding performance, but it also offers a broad range of functional services, covering T2V, I2V, and V2V. Additionally, it supports the video expansion [16]: Alpha can expand videos up to 3 times (maximum length: 40 seconds), while Turbo can expand videos up to 3 times (maximum length: 34 seconds). This feature is particularly useful for mid-length video production. According to research data [6], Runway Gen3 demonstrates cutting-edge performance across various aspects, exhibiting high overall comprehensiveness and strong versatility, making it ideally suited for both corporate and commercial applications.

Runway has introduced the Autoregressive-to-Diffusion (A2D) visual-language model [17], which achieves faster parallel generation by adapting existing autoregressive VLMs to diffusion-based decoding. The A2D-VL outperforms previous diffusion-based VLMs in visual question answering, while significantly reducing the computational resources required for training. Our novel adaptation technique is crucial for preserving the model's capabilities, ultimately enabling a state-of-the-art conversion from autoregressive VLMs to diffusion-based models with minimal impact on quality. This ensures the original quality remains unchanged, drastically shortening video generation time and allowing us to meet people's needs more efficiently.

In terms of price, Runway Gen3 offers mid-range pricing (0.64 USD per 5 seconds at 1080p), making it suitable for individual creators or those seeking higher-quality productions.

2.5 Seedance 2.0

As an AI video-generation tool independently developed by China's ByteDance, Seedance 2.0 boasts unique advantages in this field. Seedance 2.0 demonstrates higher usability in complex interactive and motion scenarios, with significantly enhanced physical accuracy, realism, and controllability, making it better suited to the demands of industrial-grade creative workflows [18]. Moreover, Seedance 2.0 supports versatile multimodal references, allowing users to combine inputs from diverse sources, including text, images, videos, and audio. The model can precisely interpret multimodal input and generate videos in accordance with specified instructions, taking into account elements such as visual composition, cinematic language, action pacing, and sound effects. It can even reference textual storyboard content directly, greatly boosting creative freedom. In testing Seedance 2.0 for text-to-video and image-to-video generation, researchers found that it exhibits greater advancement and practicality than Sora Pro and Veo 3.1. [19]

As shown in Figure 1, Seedance 2.0 delivers industry-leading overall performance. The model covers a more comprehensive set of reference tasks and supports various creative scenarios, including multimodal reference generation, video editing, and video continuation. At the same time, it boasts superior depth of understanding and precision in responding to reference content. In editing tasks, compared to other models, Seedance 2.0 provides more complete instruction responses and generates more realistic images. In terms of consistency, the model performs relatively well at reproducing the subject's appearance and voice, particularly excelling at maintaining consistency across reference in action logic, special effects styles, and narrative storytelling.

However, as shown in Figure 2, the model still has room for improvement in terms of multi-agent consistency, text-to-video accuracy, and the quality of complex editing effects. When using Seedance 2.0 to create dialogue- or interaction-based videos, you should pay close attention to lip-syncing and certain motion details of the video's characters. Additionally, be mindful of how you use prompts—avoid words or sentences that are difficult for the AI to understand—to minimize instances of inaccurate video.

Figure 1: [18]

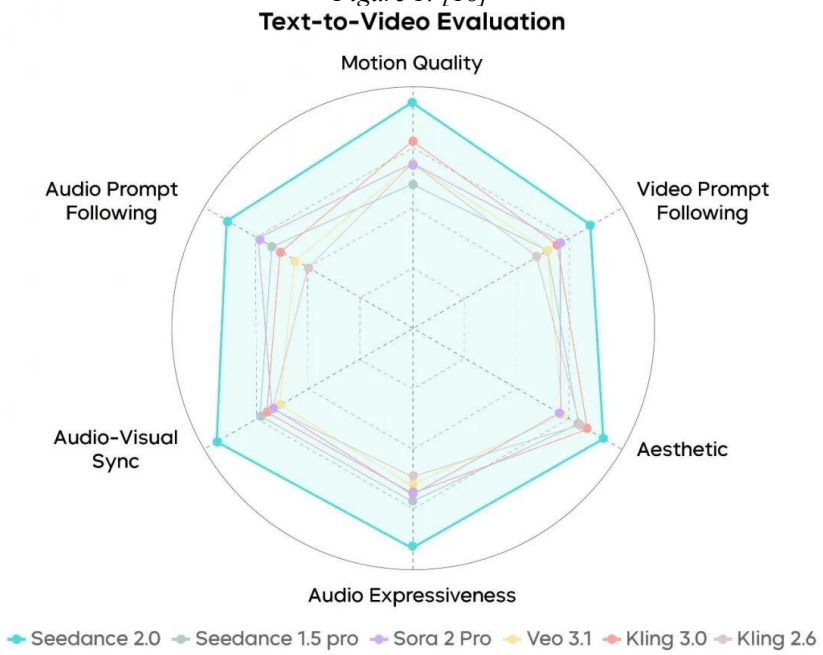
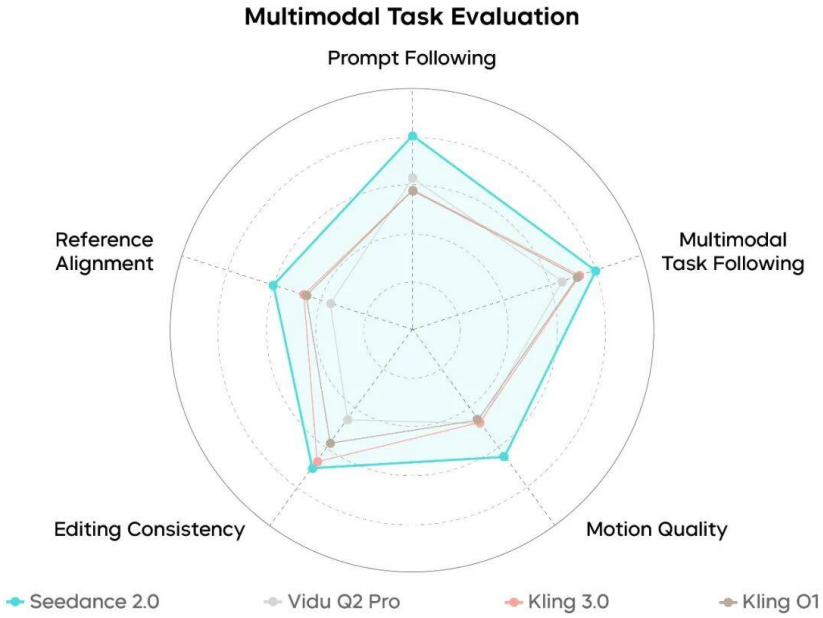


Figure 2: [18]



In terms of pricing, Seedance 2.0 is affordable for most people (US$0.3–0.59 per 5 seconds at 1080p). As an outstanding generative AI platform from China, it's widely used by individuals and businesses across the country.

3. Discussion

Today, AI faces three main types of risks: copyright and portrait rights risks, authenticity and misinformation risks, and platform compliance risks. Regarding authenticity and the risk of misrepresentation, most video platforms have strict guidelines governing the use of AI-generated video content. On short-video platforms such as Douyin, AI-generated content must include dual labeling: explicit and implicit labeling fields [20]. The explicit labeling involves displaying text below the video to alert viewers.

In terms of copyright, right of publicity, and platform rules, users should pay close attention to keyword inclusion during use to avoid duplication and right-of-publicity infringement. According to TikTok's usage guidelines [21], it is prohibited to use generative AI technologies to create or publish infringing content, including, but not limited to, content that violates rights of publicity and intellectual property rights. Additionally, it is forbidden to use generative AI to create or publish content that contradicts scientific common sense, involves falsification, or spreads rumors. The use of AI-based face-swapping, voice imitation, and the creation of false or infringing content are strictly prohibited. Any content that is vulgar, violent, or maliciously mocks sensitive topics will be dealt with rigorously.

Moreover, these AI-generated models each have their own advantages and uses. If individuals wish to apply these AIs, they should try them one by one and determine through practical experience which AI best meets their specific needs. Today, AI is advancing rapidly; Codex can now connect to the APIs of most AI-generated video platforms on the market, making it even more capable of meeting people's needs in specific fields—and more autonomous, faster, and more convenient [22]. Meanwhile, regarding video-generating AI, researchers have also developed a new model—the audio-to-video synthesis model (EchoVid) [23]. We believe that in the future, there will be further developments in generating videos from audio. Today, an increasing number of people are using generative AI to create videos, a trend that spans the advertising, entertainment, and tourism industries. Text-to-video generation is transforming the advertising industry by enabling the creation of visually compelling ad content with minimal time and cost [24]. Moreover, travel companies are leveraging AI-generated promotional videos tailored to different audiences to influence consumer decision-making. This includes shaping audience attitudes and behavioral responses, ultimately increasing tourist numbers [25].

4. Conclusion

In short, today's prevailing trends are driving AI development in many different areas. No single AI video tool is inherently superior or inferior; rather, the choice should be based on factors such as creative purpose, budget, quality requirements, compliance risks, and platform rules. Even if Sora were to shut down [9], there would still be demand for T2V (Text-to-Video) content, and we could opt for any of the other four available tools. If you're looking to create videos featuring human subjects and have a relatively ample budget, Veo3—with its strengths in dialogue synchronization, lip-sync accuracy, ambient sound, sound effects, and basic background music—could be your top pick. If funds are limited, Pika can be an excellent alternative, given its impressive AI-driven lip-syncing capability. Moreover, Pika's ability to animate static images also brings significant value to the advertising industry. For creating videos showcasing landscapes, RunwayGen3 is an outstanding choice, thanks to its cinematic-quality lighting, textures, and high resolution, which produce exceptionally detailed and vivid images.

Additionally, its high frame rate ensures that the resulting videos are smooth and seamless. If you're aiming to produce action-oriented videos—such as dance performances, cycling sequences, or basketball games—Seedance 2.0 offers substantial advantages. It boasts advanced physical accuracy and realism, generating visuals that feel remarkably lifelike, and its models perform particularly well at faithfully reproducing both the subjects' visual appearance and audio.

References

[1] V. Arkhipkin *et al.*, "ImproveYourVideos: Architectural Improvements for Text-to-Video Generation Pipeline," *IEEE Access*, vol. 13, pp. 1986–2003, Jan. 2025, doi: 10.1109/access.2024.3522510.

[2] I. Chivileva, P. Lynch, T. E. Ward, and A. F. Smeaton, "A dataset of text prompts, videos and video quality metrics from generative text-to-video AI models," *Data Brief*, vol. 54, p. 110514, Jun. 2024, doi: 10.1016/j.dib.2024.110514.

[3] Y. Chen, W. Chu, Y. Wu, J. Yang, X. Mao, and W. Liu, "COTA-motion: Controllable image-to-video synthesis with dense semantic trajectories," *Neurocomputing*, vol. 657, p. 131671, Dec. 2025, doi: 10.1016/j.neucom.2025.131671.

[4] G. T. T. Nguyen and H. T. M. Nguyen, "AI Content, User Fatigue, and Churn on TikTok: Testing an Integrated Stressor Model," *Hum. Behav. Emerg. Technol.*, vol. 2026, no. 1, p. 9977742, 2026, doi: 10.1155/hbe2/9977742.

[5] P. Bijalwan, A. Gupta, A. Johri, M. Wasiq, and S. K. Wani, "Unveiling sora open AI's impact: a review of transformative shifts in marketing and advertising employment," *Cogent Bus. Manag.*, vol. 12, no. 1, p. 2440640, Dec. 2025, doi: 10.1080/23311975.2024.2440640.

[6] Z. Qi *et al.*, "T2VEval: Benchmark dataset and objective evaluation method for T2V-generated videos," *Displays*, vol. 91, p. 103178, Jan. 2026, doi: 10.1016/j.displa.2025.103178.

[7] O. Hellmann, "Colonial Bias in AI Training Data: Prompting Sora to Generate Images of Aotearoa New Zealand's Historical Past," *Kotuitui-N. Z. J. Soc. Sci. Online*, vol. 21, no. 1, p. e70016, Feb. 2026, doi: 10.1002/kot2.70016.

[8] K. Z. Zhou, A. Choudhry, E. Gumusel, and M. R. Sanfilippo, "'Sora is incredible and scary': public perceptions and governance challenges of text-to-video generative AI models," *Inf. Res.- Int. Electron. J.*, vol. 30, pp. 508–522, Mar. 2025, doi: 10.47989/ir30iConf47290.

[9] "Sora Service Discontinuation Notice," OpenAI Help Center. Accessed: Jun. 27, 2026. [Online]. Available: https://help.openai.com/zh-hans-cn/articles/20001152-what-to-know-about-the-sora-discontinuation?f_link_type=f_linkinlinenote&show_loading=0&flow_extra=eyJpbmxpbmVfZGlzcGxheV9wb3NpdGlvbiI6MCwiZG9jX3Bvc2l0aW9uIjowLCJkb2NfaWQiOiIzNDZiYTlmMWVjNThjYzkzLTY1MjZmZjFmYzAyMTBhYTIifQ%3D%3D&webview_progress_bar=1&push_animated=1&theme=light

[10] V. Video, "Veo 3 - Google DeepMind," Vexa Video. Accessed: Jun. 27, 2026. [Online]. Available: https://vexavideo.com/zh/veo3

[11] "Veo: a text-to-video generation system." deepmind google. [Online]. Available: https://storage.googleapis.com/deepmind-media/veo/Veo-3-Tech-Report.pdf

[12] A. Budiman, N. A. Rakhmawati, and D. Purwitasari, "Cultivating Participatory Learning Ecologies: Social Network Analysis of Peer-Driven Learning Network," *J. Inf. Technol. Educ.-Res.*, vol. 25, p. 6, 2026, doi: 10.28945/5719.

[13] "Free Pika AI Video Generator | Powered by Pika Art Labs for Efficient Creative Expression," Pika AI. Accessed: Jun. 27, 2026. [Online]. Available: https://pikaai.org/blog/detail/Pika-AI-by-Pika-Lab%3A-Revolutionizing-Video-Creation-with-AI-Technology-c6772b05b14a/

[14] M. Kawka, "Theological imaginal reflections with AI artworks: revealing and obscuring divine knowledge in aesthetic experience," *Pract. Theol.*, vol. 18, no. 2, pp. 152–167, Mar. 2025, doi: 10.1080/1756073X.2025.2474318.

[15] "Creating with Keyframes on Gen-3," Runway. Accessed: Jun. 27, 2026. [Online]. Available: https://help.runwayml.com/hc/en-us/articles/34170748696595-Creating-with-Keyframes-on-Gen-3

[16] "Product Updates & Changelog | Runway AI." Accessed: Jun. 27, 2026. [Online]. Available: https://runwayml.com/changelog?show_loading=0&webview_progress_bar=1&flow_extra=eyJpbmxpbmVfZGlzcGxheV9wb3NpdGlvbiI6MCwiZG9jX3Bvc2l0aW9uIjowLCJkb2NfaWQiOiI3Nzg0MzFjYzU3Yjg4NzE4LTU0NWFmMDkyMmU5ZDMzZTYifQ%3D%3D&f_link_type=f_linkinlinenote&push_animated=1&theme=light

[17] "Runway Research | Autoregressive-to-Diffusion Vision Language Models." Accessed: Jun. 27, 2026. [Online]. Available: https://runwayml.com/research/autoregressive-to-diffusion-vlms

[18] "Team Dynamics - ByteDance Seed." Accessed: Jun. 27, 2026. [Online]. Available: https://seed.bytedance.com/zh/blog/seedance-2-0-official-launch?view_from=content_recommend

[19] "Seedance 2.0." Accessed: Jun. 27, 2026. [Online]. Available: https://seed.bytedance.com/zh/seedance2_0

[20] "Douyin Rule Center." Accessed: Jun. 27, 2026. [Online]. Available: https://api.amemv.com/magic/eco/runtime/release/6458b8f26a0a4e036e02ee15?appType=douyin&magic_page_no=1&magic_source=minebanner

[21] "Announcement on the Governance of the Misuse of AI-Generated Virtual Characters." Accessed: Jun. 27, 2026. [Online]. Available: https://api.amemv.com/magic/eco/runtime/release/6603900cbc1ea9062e329273?appType=douyin&hide_nav_bar=1&auto_play_bgm=1&container_bg_color=%23ffffff&loading_bg_color=%23ffffff&loading_duration=1000&_pia_=1&loader_name=forest

[22] "Stop calling it GPT! Codex is the true 'engineer' of AI programming—a comprehensive guide to breaking down advantages, risks, and implementation." Accessed: Jun. 27, 2026. [Online]. Available: https://juejin.cn/post/7642725839920463882?webview_progress_bar=1&push_animated=1&show_loading=0&flow_extra=eyJpbmxpbmVfZGlzcGxheV9wb3NpdGlvbiI6MCwiZG9jX3Bvc2l0aW9uIjowLCJkb2NfaWQiOiIyNzQzZjRmMDFjYjg1MmZkLWQxYjg3N2UyOGU2OTkzN2EifQ%3D%3D&f_link_type=f_linkinlinenote&theme=light

[23] D. Dharrao *et al.*, "AI-driven audio-to-video generation for dynamic content creation via stable diffusion and CNN-augmented transformers," *Sci. Rep.*, vol. 16, no. 1, p. 10295, Feb. 2026, doi: 10.1038/s41598-026-38758-3.

[24] A. El Assadi, "Ai vs. human creativity: the impact of text-to-image and text-to-video ads on customer engagement," *Electron. Commer. Res.*, Feb. 2026, doi: 10.1007/s10660-026-10103-w.

[25] I. T. Seo, H. Liu, H. Li, and J.-S. Lee, "AI-infused video marketing: Exploring the influence of AI-generated tourism videos on tourist decision-making," *Tour. Manag.*, vol. 110, p. 105182, Oct. 2025, doi: 10.1016/j.tourman.2025.105182.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Open Access