Int. J. Sci. R. Tech., 2026 3(11)
A Multidisciplinary peer-reviewed Journal
www.ijsrtjournal.com [ISSN: 2394-7063]

Original Article

A Comparative Study on the Use of Generative AI Tools in Multimedia with Focus on Prompt Writing

Jagdish Kumar*, Dr. Kuldeep Kumar, Garg Palak Parmod

Department of Computer Science and Engineering, Chaudhary Devi Lal University, Sirsa, Haryana, India.

ABSTRACT

This research presents a comparative analysis of generative AI tools used in multimedia content creation with a primary focus on prompt engineering. With the rapid evolution of artificial intelligence technologies such as text-to-image, text-to-video, text-to-audio, and text-to-animation systems, users increasingly depend on prompt quality to achieve optimal output results. This study evaluates major tools including DALL·E, Midjourney, Stable Diffusion, Runway ML, Pictory AI, Speech Synthesis systems, and animation platforms. The research adopts a comparative and exploratory methodology using both qualitative evaluation and user survey analysis (n=100). Results indicate that newer model versions demonstrate superior prompt comprehension, visual coherence, and realism. Furthermore, structured and descriptive prompts significantly improve output quality across all multimedia formats. The findings emphasize that prompt engineering plays a critical role in maximizing generative AI performance in creative industries.

Keywords: Generative AI, Prompt Engineering, Multimedia, Text-to-Image, Text-to-Video, Artificial Intelligence

INTRODUCTION

Generative AI

Generative Artificial Intelligence refers to AI systems capable of generating new content such as text, images, audio, and videos by learning patterns from existing datasets. Unlike traditional AI systems that perform classification or prediction tasks, generative models create original content based on textual instructions known as prompts. With the emergence of tools such as DALL·E, Midjourney, Stable Diffusion, Runway ML, and AI-based speech synthesis models, multimedia production has become highly automated and accessible. However, despite technological improvements, output quality largely depends on how prompts are structured and designed. This study aims to perform a structured comparative analysis of major generative AI tools across multimedia domains and evaluate the influence of prompt engineering on output quality [1].

Literature Review

The chapter presents the review of work done by various researchers, scientists and academicians related to the present study. A literature review provides a context for the research, outlines gaps in previous research, focuses on research questions and tries to identify, select and combine all high-quality research evidence and arguments. The research is related to the comparative analysis of generative AI tools used in multimedia with prompt writing.

Ramesh et al. proposed a transformer-based model for text-to-image generation that processes text and image tokens in a unified framework. The model was trained on a large dataset of image-text pairs and demonstrated the ability to generate high-quality images from textual prompts. This work

laid the foundation for advanced generative models such as DALL·E and further improvements using models like CLIP and diffusion-based approaches [2].

Marcus et al. The emergence of multimodal generative models, particularly OpenAI's DALL·E 2, has sparked significant academic interest regarding their capabilities and limitations. In a foundational study by Marcus, Davis, and Aaronson (2022), the authors conducted a preliminary but systematic analysis of DALL·E 2 to assess its ability to interpret and generate images from complex natural language prompts. Their research involved a series of 14 targeted tests that explored the system's performance across dimensions such as compositionality, spatial reasoning, stylistic accuracy, and commonsense understanding [3].

Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Robin Rombach et al. presented an efficient alternative technique via Latent Diffusion Models, which can generate high quality images even with fewer resources. While Joanna Zilinskas (2022) focused on the artistic and social implications of DALL·E 2, and presented it as a collaborative creative tool. DALL·E 2, where the research team, led by Ramesh and Dhariwal, shifted towards diffusion-based models. This allowed the generation of more realistic, higher-resolution images while improving the model's ability to combine complex concepts, styles, and attributes in a single output (Ramesh et al., 2022) [4].

Yuhao Zhang et al. in which introduced significant advancements have been made in generative AI models, particularly those leveraging transformer architectures. Foundational models such as GPT, BERT, and T5 have demonstrated strong capabilities in text generation and laid the groundwork for prompt-based generation tasks. OpenAI's CLIP (Contrastive Language–Image Pretraining) marked a pivotal development in aligning textual and visual modalities, enabling more coherent image generation from text inputs. Earlier efforts in text-to-image synthesis, including Stack GAN, Attn GAN, and DM-GAN [5].

Dena Magdy Hanna explores the emergence and application of AI-driven art generation tools, particularly Midjourney, in the realms of art and advertising. It discusses how Midjourney, an AI art generator, transforms textual prompts into visual artworks by leveraging extensive datasets of existing art and understanding artistic principles. The review highlights the evolution of AI in creative fields, noting predecessors like DALL-E and the role of GPT-3 in generating images from text descriptions. It emphasizes the shift from traditional, manual design processes to AI-assisted creation, where users without formal artistic training can produce high-quality visuals [6].

P. Logeswari et al. analyzed several AI video generation tools such as Runway, Pictory, Deep Brain AI, Synthesis, Colossians, Hour One, D-ID, Elai.io, Heyen, and In Video. The study compared these tools based on their architecture, features, advantages, and limitations. The results showed that AI video tools make video creation easier and accessible, but challenges such as limited customization, misuse risks, and inconsistent output quality still exist [7].

Weimin Wang et al. introduced MagicVideo-V2, an AI model that integrates text-to-image conversion, video motion generation, reference image embedding, and frame interpolation into an end-to-end pipeline. The model is capable of generating high-resolution, aesthetically pleasing videos with remarkable fidelity and smoothness. User evaluations demonstrate its superior performance over leading text-to-video systems such as Runway, Pika 1.0, Morph, Moon Valley, and Stable Video Diffusion [8].

Tao Yu et al. This study examined the challenges faced by video production professionals in adopting AI video generation tools like Runway and Pictor. It highlights factors such as technical limitations, user experience, and trust in AI-generated content. The research provides empirical

evidence on practitioners' attitudes and acceptance levels, offering insights for developers and industry decision-makers to enhance the effectiveness and adoption of these tools [9].

Serhat Erdem reviewed the role of artificial intelligence in digital storytelling and media editing. The study highlights that AI tools such as Runway ML and Pictory AI simplify video creation using technologies like machine learning and natural language processing. The research also discusses ethical concerns such as content authenticity and the role of human creativity in AI-based media production [10].

Reddy et al. reviewed the evolution of text-to-speech (TTS) systems from traditional methods to modern neural network-based approaches. The study highlights that deep-learning models such as Taco Tron and neural vocoders like Wave Net significantly improve speech naturalness and quality. Performance is commonly evaluated using Mean Opinion Score (MOS) to measure synthesized speech quality [11].

Jinhee and Jeong Jinheon delves into the advancements and applications of AI-driven motion capture technologies, with a particular focus on Deep Motion's solutions. The authors discuss the evolution of motion capture from traditional marker-based systems to contemporary AI-powered, markerless approaches. They highlight the benefits of AI integration, such as enhanced accuracy, reduced human error, and improved processing speeds. The literature review also addresses challenges like multi-person tracking errors and the need for real-time processing capabilities. By analyzing current applications and technological

progress, the study underscores the potential of AI-driven motion capture in various industries, including animation, gaming, and virtual reality [12].

Anna Jaruga et al. (2022) They explained that Midjourney can help create architectural concept sketches when given the right text-prompts. The research results proved that the AI (Midjourney) is able to automatically identify important design features and produce creative results even without being provided with full information. The research also showed that Midjourney can create virtual designs faster than humans, but the final decisions, feasibility and revisions will still have to be done by the human architect. They concluded that Midjourney can become an important supporting tool for speeding up the initial design stage in architecture, although it is still a preliminary concept tool, not a complete architectural plan [13].

Novita Sari et al. discussed the role of AI tools such as Synthesia, Runway ML, Pictory AI, and Lumen5 in multimedia video creation. The study highlights how technologies like NLP, computer vision, and GANs automate video production, while tools like Steve.AI enable script-to-video generation, though research on its performance is still limited [14].

Mahmoud Hamandi et al. proposed Deep Motion, a deep learning system that enables robots to navigate crowded environments by learning human movement patterns from trajectory data. The model improves navigation success and produces smoother movements, though it depends heavily on training data quality [15].

2.2 Research Gaps

- Existing studies do not provide a structured comparison of generative AI tools like DALL·E, Runway ML, and Pictor AI in the context of multimedia creation.
- Most literature emphasizes single-mode outputs (e.g., only text-to-image or only text-to-video), with little attention to tools capable of handling integrated multimedia content.
- Few studies examine how users interact with these tools, including the cognitive load, ease of use, and the impact on user creativity.
- There is limited research on the classification, effectiveness, and strategic use of prompt writing across different generative AI platforms.

3.1 Problem Statement

The problem is entitled as "Comparative Analysis of Generative AI Tools Use in Multimedia with Prompt Writing". In the rapidly evolving digital landscape, the use of generative AI tools for multimedia content creation has become increasingly widespread. However, there exists a significant gap in understanding the comparative strengths and weaknesses of these tools, particularly regarding how prompt writing affects output quality. Without clear, systematic comparisons, users struggle to select appropriate tools and design effective prompts. This research aims to address this problem by providing a structured comparative analysis to guide multimedia creators in making informed choices.

3.2 Objective of Research

- To study generative AI tools used in social media.
- Find the effective solution for prompt writing.
- To study and analyze the effectiveness across different multimedia formats.
- To study the challenges and issues of prompt writing in AI.

RESEARCH METHODOLOGY

Research methodology refers to the systematic, theoretical, and organized approach used to conduct research. It provides a framework for collecting, analysing, and interpreting data in order to achieve the objectives of the study. In the present research, the methodology focuses on the comparative analysis of Generative AI tools used in multimedia content creation with a focus on effective prompt writing. This methodology helps in understanding how different generative AI tools respond to prompts and how prompt engineering techniques influence the quality of generated multimedia outputs such as images, videos, audio, and animations.

4.2 Data Sources

The data and insights for this comparative analysis on the use of Generative AI tools in multimedia with prompt writing were gathered from multiple reliable sources. These sources provided detailed information regarding the functionality, features, and practical

applications of different AI tools used for multimedia content generation.

The following official platforms were referred to during the research:

OpenAI – https://openai.com [16]

Midjourney – https://www.midjourney.com [17]

Fotor AI – https://www.fotor.com/ai-photo-editor/ [18] Art Guru AI – https://www.artguru.ai [19]

Stable Diffusion – https://stability.ai/stable-diffusion [20] Runway ML – https://runwayml.com [21]

Synthesia – https://www.synthesia.io [22]

OpenAI Text-to-Speech – https://platform.openai.com/docs/guides/text-to-speech [23]

Deep Motion – https://www.deepmotion.com [24]

4.3 Tool Used

In this research, different generative AI tools were used to analyse and compare their performance in multimedia content generation based on prompt writing. The tools were selected according to the type of multimedia output they generate.

Text-to-Image Tools: DALL·E (Series), Midjourney, Stable Diffusion

Text-to-Video Tools: Runway ML, Pictory AI, Magic Light AI

Text-to-Audio Tools: OpenAI Text-to-Speech, Speech Synthesis, Cloud Text-to-Speech

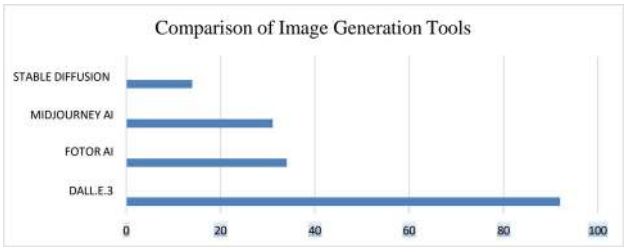
Text-to-Animation Tools: Pika Labs, Deep Motion, Steve.AI

These tools were analysed to evaluate their capabilities, output quality, and effectiveness in generating multimedia content through prompt-based inputs.

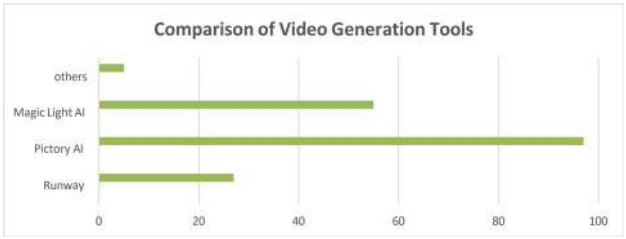
ANALYSIS AND RESULTS

This chapter presents the result obtained from the comparative study of various generative AI tools used in multimedia applications. The analysis is based on parameters such as prompt understanding, image quality, tool performance and application areas. The results are represented using graphical methods for better interpretation. The analysis of generative AI tools shows that each has unique strengths and limitations. Text-to-Image tools such as DALL·E and Midjourney produce highly creative and artistic visuals. Midjourney is widely recognized for producing visually appealing images, while Stable Diffusion provides greater customization and flexibility. Text-to-Video tools like Runway ML allow users to generate cinematic videos using text prompts. Pictory AI focuses on converting scripts into marketing videos quickly and efficiently. Text-to-Audio tools convert written text into natural-sounding speech and are widely used in virtual assistants, audiobooks and video narration. Text-to-animation tools such as Pika Labs and Steve.AI help generate animated videos automatically from scripts, making animation production easier for beginners. The result also indicates that prompt writing significantly affects the quality of generated outputs. Clear and descriptive prompts produce more accurate and relevant results compared to vague or incomplete prompts.

The graphs show the performance of different image generation tools used in this study, including DALL.E, Fotor AI, Midjourney AI and Stable Diffusion. It is observed that DALL.E has the highest score indicating better performance in terms of image quality and prompt understanding. Fotor AI and Midjourney AI show moderate performance, while Stable Diffusion has comparatively lower performance in this analysis.

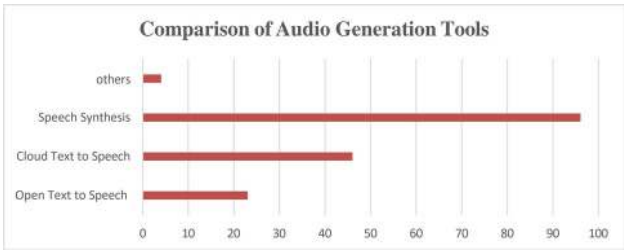


The graph illustrates the performance of various video generation tools, including Runway, Pictory AI, Magic Light AI and others. It is observed that Pictory AI has the highest performance among all tools, indicating its effectiveness in video creation. Magic Light AI shows moderate performance, while Runway demonstrates comparatively lower performance in this analysis. Other tools have minimal contribution.

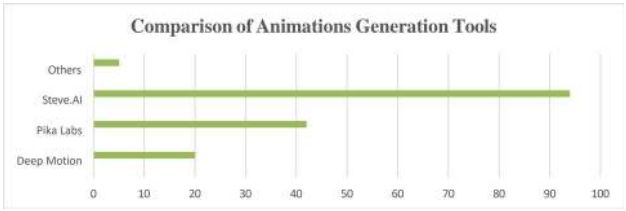


The graph illustrates the performance of various audio generation tools, including Speech Synthesis, Cloud Text to Speech, Open Text to Speech and others. It is

observed that Speech Synthesis achieves the highest performance, indicating superior voice quality and natural speech generation. Cloud Text to Speech and Open Text to Speech also perform well, while other tools contribute moderately to the overall comparison.



The graph illustrates the performance of various animation generation tools, including Steve.AI, Pika Labs, Deep Motion and others. It is observed that Steve.AI achieves the highest performance with 94%, indicating its effectiveness in generating high quality animations with ease of use. Pika Labs show moderate performance with 42% while Deep Motion has comparatively lower performance at 20%. Other tools contribute minimally with only 5%.



DISCUSSION

The results obtained from the analysis, based on data collected through a Google Form survey from generative AI users, show that different generative AI tools perform differently depending on their features and applications. Image generation tools such as DALL·E demonstrate higher performance in terms of image quality and prompt understanding. Video generation tools like Pictory AI show better efficiency in content creation. Similarly, audio tools such as Speech Synthesis provide more realistic voice output, while animation tools like Steve.ai perform effectively in generating animated content.

The findings indicate that no single tool is perfect for all tasks, and the selection of tools depends on the specific requirements of multimedia applications. The survey responses further highlight that users prefer tools that offer a balance between quality, ease of use, and functionality.

CONCLUSION

In conclusion, this study highlights the growing importance of generative AI tools in multimedia applications. The comparative analysis shows that different tools have their own strengths and limitations. Tools like DALL·E.3, Pictory AI, Steve.AI and Speech Synthesis perform better in their respective domains, while others provide ease of use and accessibility. Overall, generative AI tools are transforming the way multimedia content is created, making the process faster, more efficient, and accessible to users.

[8] Future Work

Future work can focus on improving accuracy, quality and usability of generative AI tools. This study was based on data collected through a Google Form survey; therefore, future research can include more users and real time testing for better results. Additionally, integrating multiple AI tools and reducing limitations such as cost and complexity can enhance overall performance.

REFERENCES

1. Pasick, A. (2023). Artificial intelligence glossary: neural networks and other terms explained. *The New York Times*, 27.
2. Ramesh, A., Pavlov et al. (2021, July). Zero-shot text-to-image generation. In *International conference on machine learning* (pp. 8821-8831). Pmlr.
3. Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, 1 (2), 3.
4. Marcus, G., Davis, E., & Aaronson, S. (2022). A very preliminary analysis of DALL-E 2. *arXiv preprint arXiv:2204.13807*.
5. Cilloni, T., Fleming, C., & Walter, C. (2023). Privacy threats in stable diffusion models. *arXiv preprint arXiv:2311.09355*.
6. Hanna, D. M. (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.
7. Logeswari, P., Jebaraj, N. R. S., & BanuPriya, G. (2024). Comparative analysis of AI tools for video production. *Journal of Information Technology Review, DLINE Journals*.

8. Wang, W., Liu, J., Lin, Z., Yan, J., Chen, S., Low, C., ... & Feng, J. (2024). Magicvideo-v2: Multi-stage high-aesthetic video generation. *arXiv preprint arXiv:2401.04468*.
9. Yu, T., Yang, W., Xu, J., & Pan, Y. (2024). Barriers to industry adoption of AI video generation tools: A study based on the perspectives of video production professionals in China. *Applied Sciences*, 14 (13), 5770.
10. Erdem, S. (2025). The synthesis between artificial intelligence and editing stories of the future. In *Transforming cinema with artificial intelligence* (pp. 231-250). IGI Global Scientific Publishing.
11. Ferraro, A., Galli, A., La Gatta, V., & Postiglione, M. (2023). Benchmarking open source and paid services for speech to text: an analysis of quality and input variety. *Frontiers in big Data*, 6, 1210559.
12. Ferraro, A., Galli, A., La Gatta, V., & Postiglione, M. (2023). Benchmarking open source and paid services for speech to text: an analysis of quality and input variety.
13. Xi, C., & Chung, J. (2024). A case study on ai-driven deepmotion motion capture technology. *International Journal of Internet, Broadcasting and Communication*, 87-92.
14. Novita Sari, L. (2024). Developing Audio Visual Media Using Steve. AI for Teaching Writing (Doctoral dissertation, Magister Pendidikan Bahasa Inggris).
15. Hamandi, M., D'Arcy, M., & Fazli, P. (2019, October). Deepmotion: Learning to navigate like humans. In *2019 28th IEEE international conference on robot and human interactive communication (RO-MAN)* (pp. 1-7). IEEE. Web References
16. https://openai.com
17. https://www.midjourney.com
18. https://www.fotor.com/ai-photo-editor/
19. https://www.artguru.ai
20. https://stability.ai/stable-diffusion
21. https://runwayml.com
22. https://www.synthesia.io
23. https://platform.openai.com/docs/guides/text-to-speech
24. https://www.deepmotion.com

HOW TO CITE: Jagdish Kumar, Dr. Kuldeep Kumar, Garg Palak Parmod, A Comparative Study on the Use of Generative AI Tools in Multimedia with Focus on Prompt Writing, Int. J. Sci. R. Tech., 2026, 3 (11), 190-195. https://doi.org/10.5281/zenodo.23942139