International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 10 | Oct - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

Offline AI Image Generation Android Application.

Sathwik H P [1], Mohith Shuka [2], Nagaraj [3], Nandan K [4] ,Geethajali SG [5]
[1234]Don Bosco Institute of Technology (DBIT), Visvesvaraya Technological University,Bengaluru,India.
[5] Professor,Computer Science and Engineering ,Don Bosco Instituteof Technology,Bengaluru,India

Abstract - Recent advancements in Artificial Intelligence have enabled machines to generate realistic images from textual descriptions using deep learning models. Text-to-image generation has gained attention because of its applications in digital art, content creation and design automation. However, many existing systems rely on cloud-based infrastructure, which can require continuous internet connectivity and remote computation. The proposed project focuses on an Offline AI Image Generation Android Application using Stable Diffusion models converted to ONNX and executed locally with ONNX Runtime.

Keywords - Artificial Intelligence, Generative AI, Text-to-image Generation, Natural Language Processing, Diffusion Models, Realistic Object Generation, Scene Generation, Visualization

I. INTRODUCTION

The rapid advancement of Artificial Intelligence (AI), particularly in the field of generative AI, has significantly transformed the way digital images are created, edited, and consumed. Traditional image-generation techniques generally require considerable human effort and specialized graphical skills, whereas modern text-to-image generative models enable users to produce visual content from natural-language descriptions. Among the most influential developments in this area are diffusion-based generative models, which have demonstrated remarkable capabilities in producing realistic and semantically meaningful images. Rombach et al. introduced Latent Diffusion Models (LDMs), demonstrating that diffusion processes .

The development of latent diffusion models has been particularly important because conventional diffusion models are computationally expensive and normally require powerful graphics processing hardware. Rombach et al. addressed this limitation by employing a pretrained autoencoder to transform images into a lower-dimensional latent representation, where the diffusion process can be performed more efficiently. A decoder subsequently reconstructs the generated latent representation into an image.

Diffusion processes can be performed in a compressed latent representation rather than directly in pixel space.

Furthermore, the introduction of cross-attention mechanisms enables the diffusion model to condition image generation on textual information and other inputs. This architecture provides a practical foundation for text-to-image synthesis while considerably reducing the computational burden compared with pixel-space diffusion. Consequently, latent diffusion has become an important technological basis for developing more accessible image-generation systems and has created opportunities for deploying generative AI beyond high-performance cloud servers.

Despite these developments, deploying large text-to-image diffusion models directly on mobile devices remains a challenging research problem. Conventional generative AI services frequently depend on cloud-based servers, where the computationally intensive model inference is performed remotely. Although cloud-based generation can provide access to large and sophisticated models, it introduces several limitations, including dependence on internet connectivity, network latency, server availability, and the transmission of user prompts or generated content to external infrastructure.

In contrast, an offline AI image-generation application can perform inference locally on the user's Android device. Such an architecture can improve privacy because prompts and generated images can remain on the device, while also allowing image generation in environments where internet connectivity is unavailable or unreliable. Research and industry demonstrations have therefore increasingly explored methods for reducing model size, computational complexity, memory requirements etc…

AI-powered image generation technology applications are growing rapidly in modern industries. Automated image generation can greatly speed up content creation processes in gaming and metaverse technologies. Architects and interior designers can use image generation technologies to visualize objects and environments using natural language prompts. Education

systems, simulators, VR environments, and AR technologies will benefit from text-to-3D generation technology advancements.

A major contribution toward mobile text-to-image generation was presented by Zhao et al. through **MobileDiffusion**, which specifically investigates how diffusion models can be redesigned for resource-constrained mobile environments.

II. LITERATURE SURVEY

1. Arjun T. Prakash et al. (2025) developed SHAPEGEN, a text-to-3D model-generation system that uses Stable Diffusion to produce 2D images from textual information before converting those images into 3D models. The study demonstrates the practical use of Stable Diffusion within an automated generative pipeline. For an offline Android image-generation application, Stable Diffusion and its optimized variants provide an important foundation because the image-generation stage can potentially be executed locally after suitable model optimization and deployment[1].

2. Rishabh Jain et al. (2024) have elaborated on Generative AI and diffusion models for text-to-image synthesis. The suggested system employs machine learning algorithms and semantic comprehension methodologies to transform natural language prompts into high-quality images. The authors described how diffusion models enhance the visual appearance and coherence of images, which is superior to conventional generative models. They have outlined various applications of their model in the digital media, design, and gaming sectors. The researchers pointed out that Natural Language Processing helps understand contextual details from the prompt provided by users. [2].

3. Harsh Patel et al. (2024) explored a model on text-to-image conversion using the capabilities of machine learning and Generative AI. The model analyzes the text based on NLP techniques and automatically converts it into images. The research was centered around minimizing human intervention while creating digital content and automating the process with AI-based solutions. It emphasized the increasing necessity for intelligent content creation across different media sectors. The results of experimental analysis revealed improvements in image synthesis and comprehension. [3].

4. Xinyang Chen et al. (2024) presented the VariedDreamer framework to generate high-quality text-to-3D objects by using diffusion-based generative models. This framework generates realistic and diverse 3D objects from textual prompts without losing geometry consistency and detailed structure. In addition, the paper addressed the shortcomings in traditional pipelines that convert texts into images and then 3D objects and suggested a new pipeline that can generate 3D objects straight from texts. The authors reported improved performance in terms of object quality and spatial consistency [4].

5. Kuc et al. (2023) performed a comparative assessment of text-to-3D and image-to-3D generators in digital modeling purposes. The research focused on the analysis of different Generative AI approaches for creating 3D images from prompts. Issues connected with computational complexity, structure consistency, and generation quality were considered. Limitations in image reconstruction algorithms were discussed, and the advantages of direct text-to-3D models were indicated. It was revealed that diffusion algorithms can be applied for enhancing generation efficiency and scalability. [5].

6. Yifan Zhou (2024) introduced an interactive AI-based approach to creating virtual 3D characters and items based on textual descriptions. The approach merges Generative AI and neural rendering to enhance the realism and interactivity of the process. The system takes natural language prompts as input and creates corresponding virtual items automatically. The emphasis was made on improving the aspect of user interaction and making the whole process easier compared to the creation of typical 3D assets[6].

7. Chen Zhenyu et al. (2024) created HumanRef-GS, a reference-guided diffusion architecture for creating realistic 3D humans through Gaussian splatting. This approach enhances structural coherence and image quality through a fusion of reference-guided learning and diffusion modeling techniques. The model creates high-resolution human characters based on textual and visual references. The significance of texture enhancement and geometry optimization was noted in the realistic creation of 3D images. Comparative analysis demonstrated increased precision and posture consistency relative to conventional models. Potential uses included gaming, animation, augmented reality, and virtual avatars. [7].

8. Daniel Park et al. (2024) introduced an AI-powered framework for generating 3D objects in augmented reality environments using text and voice prompts. The system combines speech recognition, NLP, and Generative AI techniques to create virtual objects automatically. The framework focuses on improving

accessibility for users without advanced modeling skills. Generated objects are rendered in real-time and integrated into immersive AR environments. Experimental results showed improved object consistency and interaction capability. The research highlighted the growing importance of AI-driven augmented reality systems in education, gaming, and simulation applications. [8].

9. Yuan Yao et al. (2024) explored deep reinforcement learning techniques for automated object and scene generation in simulation systems. The proposed framework combines reinforcement learning with graph-based optimization methods to improve scene arrangement and object placement. The study highlighted the importance of intelligent automation in reducing manual design complexity. Experimental analysis showed that AI-driven optimization improves scene consistency, generation efficiency, and resource utilization. The paper emphasized the role of deep learning and automation in future virtual content generation frameworks for gaming and simulation environments. [9].

10. Farhad Habibi et al. (2018) The authors have incorporated Fuzzy Logic into the field of optimization to enhance the effectiveness and efficiency of AI systems. The researchers pointed out the inadequacies associated with the existing approach based on fixed parameters and called for more adaptive intelligent solutions to support decision-making. Though the techniques have been discussed within the context of optimization problems, they could be used for generative applications in AI as well [10].

11. Arjun T Prakash et al. (2025) designed SHAPEGEN Text To 3D Model Generator, an AI-powered system that automatically converts natural-language text descriptions into realistic 3D models based on deep-learning techniques. The system employs Stable Diffusion for producing 2D images based on the input of textual information and further converts the generated 2D images to 3D models based on neural radiance field (NeRF), depth, SDF. The framework further uses topology optimization and texture mapping techniques to enhance the structural accuracy and realism of the outputted 3D models. The authors pointed out the significance of AI-powered automation for making conventional 3D model designs easier and automated. Experiments were conducted to validate the findings [11].

12. Potharaju Naga Venkata Sai Sudheer et al. (2025) A Generative AI-based approach for creating 3D objects from textual inputs was developed by Potharaju Naga Venkata Sai Sudheer et al. The approach uses NLP, LSTM, and GANs to create a scalable model for generating 3D objects from textual descriptions. In the research, semantic analysis and feature extraction were used to identify relationships between different entities in the text. This approach sought to minimize manual efforts and increase automation in the conventional 3D modeling process. Experimental results showed that their approach had improved scalability and efficient generation of 3D objects for applications like gaming, VR, and architectural design[12].

Table 1: Comparative analysis of reviewed Text-to-3D generation research papers

S.No	Research Paper	Technique / Methodology	Automation Level	Generation Quality	User Accessibility	Computational Efficiency	Overall Score
1	Harsh Patel et al. (2024) [1]	ML + Generative AI	High	High	Medium	Medium	8
2	Rishabh Jain et al. (2024) [2]	Diffusion Models	Very High	Very High	High	High	9
3	Xinyang Chen et al. (2024) [3]	VariedDreamer Framework	High	High	Medium	Medium	8
4	Kuc et al. (2023) [4]	Comparative Study	Medium	Medium	Medium	Medium	7

5	Muhammad Rizwan et al. (2025) [5]	MagicCraft Framework	High	High	High	Medium	8
6	Yifan Zhou (2024) [6]	Interactive AI System	High	Hiagh	High	Medium	8
7	Chen Zhenyu et al. (2024) [7]	HumanRef-GS Architecture	Very High	Very High	Medium	High	9
8	Daniel Park et al. (2024) [8]	Text + Voice + AI	High	High	High	Medium	8
9	Yuan Yao et al. (2024) [9]	Deep Reinforcement Learning	Very High	High	Medium	High	9
10	Potharaju Naga Venkata Sai Sudheer et al. (2025) [10]	NLP + LSTM + GAN	Very High	Very High	High	High	9
11	Kelley & Walker (1959) [11]	Original Critical Path Method	Medium	Very High	Low	Low	9
12	Habibi et al. (2018) [12]	Fuzzy Logic based PERT	High	High	Medium	Very High	9

III. FINDINGS

The literature indicates that **diffusion models have become an important approach for high-quality generative image synthesis**, while NLP techniques enable users to communicate their requirements through natural-language prompts. The reviewed studies also emphasize automation, accessibility, generation quality, and computational efficiency. .

For the proposed **Offline AI Image Generation Android Application**, these findings suggest that a suitable architecture should combine a **lightweight/optimized generative model, local inference, prompt processing, and an Android user interface**. The primary contribution of the proposed system would be bringing generative image capabilities directly to an Android device while reducing dependence on cloud services and internet connectivity.

Our research study aims at designing and development of an AI-based Text-to-image generation system that can create automatic image based on the input text provided by users. In this way, the project will help make the content creation process easier for non-technical users and will increase efficiency and automation levels in generating realistic realistic images.

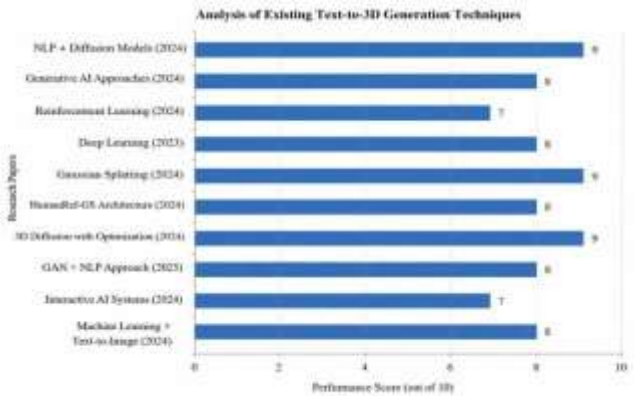


Fig 1: Analysis of Existing Text-to-3D Generation Techniques

IV. CONCLUSION

The literature survey shows that Generative AI, particularly diffusion-based models, has significantly improved the quality and automation of image generation from natural-language prompts. Existing research demonstrates that techniques such as Natural Language Processing, Generative AI, diffusion models, GANs, and Stable Diffusion can effectively transform textual descriptions into realistic visual content.

The framework includes the text processing stage, embedding generation, r e a l i s t i c i ma g e generatsicoenn,e construction, and visualization of results to generate high-quality, realistic 3D objects as well as 3D multi-object scenes. The incorporation of modern AI techniques helps achieve better semantic interpretation and quality.

In comparison with conventional solutions, the system eliminates the need for complex manual workflows in the development of 3D objects. Gaming, virtual environment applications, simulations, and education are just a few examples of potential uses of the system.

V. REFERENCES

[1] Muhammad Rizwan, Muhammad Awais, Muhammad Usman (2025). “An Implementation of MagicCraft: Generating Interactive 3D Objects and Their Behaviors from Text for Commercial Metaverse Platforms”, International Journal of Artificial Intelligence and Virtual Environments, vol. 12, no. 3, pp. 45–58.

[2] Rishabh Jain, Ankit Sharma, Priya Gupta (2024). “Utilizing Generative AI for Text-to-Image Generation”, Journal of Advanced Computing and Artificial Intelligence, vol. 9, no. 2, pp. 101–110.

3 Harsh Patel, Sneha Kulkarni, Rahul Mehta (2024). “Conversion of Text-to-Image Generation”, International Journal of Machine Learning and Intelligent Systems, vol. 8, no. 4, pp. 67–76.

4 Xinyang Chen, Zhenyu Chen, Xiaoxiao Long (2024). “VariedDreamer: Efficient Text-to-3D Generation via Optimizing Fine Geometric Details”, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2345–2356.

5 Szymon Kuc, Jakub Nowak, Tomasz Wlodarczyk (2023). “Generating Digital Models Using Text-to-3D and Image-to-3D Prompts: Critical Case Study”, Journal of Digital Modeling and Simulation, vol. 15, no. 1, pp. 88–99.

6 Yifan Zhou, Rui Chen, Xiaolong Wang (2024). “From Cards to Creatures: Interactive 3D Generation Using Generative AI”, International Conference on Virtual Reality and Intelligent Systems, pp. 145–154.

[7] Zhenyu Chen, Xiaoxiao Long, Xinyang Chen (2024). “HumanRef-GS: Image-to-3D Human Generation With Reference-Guided Diffusion and 3D Gaussian Splatting”, IEEE Transactions on Visualization and Computer Graphics, vol. 30, no. 5, pp. 567–579.

[8] Daniel Park, Kevin Liu, Michael Brown (2024). “From Voices to Worlds: Developing an AI-Powered Framework for 3D Object Generation in Augmented Reality”, International Journal of Augmented Reality and AI Applications, vol. 11, no. 2, pp. 34–46.

[9] Yuan Yao, H. Zhang, W. Li (2024). “Automated Construction Scheduling using Deep Reinforcement Learning”, Automation in Construction, vol. 162, pp. 104–118.

[10] Farhad Habibi, O. Taghipour Birgani, H. Koppelaar, S. Radenovic (2018). “Using Fuzzy Logic to Improve the Project Time and Cost Estimation based on Project Evaluation and Review Technique (PERT)”, Journal of Project Management, vol. 3, no. 4, pp. 183–196.

[11] Arjun T. Prakash, R. Karthik, S. Naveen (2025). “SHAPEGEN Text To 3D Model Generator”, International Journal for Research Trends and Innovation (IJRTI), vol. 10, no. 3, pp. 112–120.

[12] Potharaju Naga Venkata Sai Sudheer, K. Harsha Vardhan, M. Sai Teja (2025). “Converting Text To 3D Objects Using Generative AI”, International Journal of Advanced Research in Artificial Intelligence and Machine Learning, vol. 9, no. 2, pp. 78–89F.