AI FOR 3D MODEL GENERATION FROM TEXT

[1]Mr SIDDESH K T, [2] SIRAJ J
[1]Assistant Professor, Department of MCA, BIET, Davenagere
[2]Student, Department of MCA, BIET, Davanagere

ABSTRACT

Traditional high-fidelity 3D model creation requires advanced skills, artistic expertise, and significant time investment. Converting abstract ideas into structured volumetric models is mostly manual and resource-intensive. This project proposes an AI-driven framework that generates 3D assets directly from natural language descriptions. The system uses large language models and deep generative techniques to transform text into accurate 3D geometric structures. It predicts object shape, spatial arrangement, textures, and materials based on semantic understanding. Advanced methods like diffusion models and Neural Radiance Fields (NeRFs) are trained on multimodal datasets containing text and 3D data. The platform also supports iterative refinement through prompt-based editing, reducing manual adjustments. By minimizing production time and technical barriers, the system enables fast and accessible 3D content creation for gaming, AR/VR, simulation, and educations.

Keywords *: AI-Driven 3D Generation, Natural Language Processing (NLP), Generative AI, Text-to-3D Modeling.*

I INTRODUCTION

The AI System for 3D Model Generation from Text is designed to simplify one of the most time-consuming stages of digital design: translating conceptual descriptions into usable 3D models. Whether the input is a single sentence a detailed paragraph or a creative writing prompt the system interprets it and produces a structured 3D object that designers developers and creators can immediately refine or use in larger projects.

Traditionally artists and designers move from mental concepts or written briefs to professional 3D tools like Blender Maya or SolidWorks. This process demands specialized knowledge a steep learning curve and many hours of manual modelling. The project aims to solve this gap by offering an intelligent interface that understands the semantic intent behind a text description and reconstructs it as a three-dimensional form through advanced AI-generated geometry.

At the fundamental of the platform is a deep learning engine trained on large datasets pairing textual captions with corresponding 3D shapes. The system learns how nouns adjectives spatial prepositions and stylistic descriptors represent real-world structure. As a result it can infer volume texture proportions and material properties even from abstract or minimal text inputs.

The workflow is simple and linear:

- The user provides a text prompt.
- Back-end will interpret the semantic

meaning using NLP technology.

• The generative model predicts a 3D shape that matches the description.

• The system provides a detailed mesh with an editable topology.

II RELATED WORK

The advancements that have come forth recently related to artificial intelligence have played a vital role in ensuring the availability of technologies used for generating 3D objects. One must note that the traditional method of creating 3D objects requires involvement from professionals, who use specific instruments, thus making the entire process difficult and time-consuming. With the advent of neural networks and deep learning methods, it became possible to create 3D objects using text or imagery.

Some of the earliest technologies used for the generation of 3D objects included the use of neural networks for the representation of meshes or voxels. The use of autoencoders and convolutional neural networks was also common for the creation of low-dimensional vectors representing 3D objects. These technologies were quite resource-intensive since large amounts of data were necessary for this purpose. Some of the modern techniques have made use of Generative Adversarial Networks and Diffusion Models.

The text-to-3D synthesis problem has become more popular due to the introduction of multimodal learning. Existing work in the area utilizes pretrained multimodal architectures to align the meaning of text with geometric shapes in 3D space.

The NeRF-based model and the score distillation sampling technique were introduced to enable generation of 3D scenes from natural language instructions. These methods can create high-quality outputs, but the output is generated using an optimization algorithm that requires a GPU.

Meanwhile, the field of LLMs has made strides in creating structured outputs through techniques such as structured prompting and constrained generation. Natural language can be reliably converted into machine-readable code, databases, and other formats. Most works in the area focus on transforming natural language to code or database queries, however.

Procedural modeling and constructive solid geometry (CSG) techniques have also been widely used for generating 3D structures from predefined rules and parametric descriptions. These methods allow precise control over object composition but require manual scripting and domain expertise. They lack adaptive intelligence to interpret high-level human descriptions automatically. However, despite these advancements, a discrepancy persistsbetween computationally intensive neural rendering methods and manually crafted procedural modeling applications.

Few scholarly papers explore the potential of large language models in procedural modeling and primitive-based 3D scene generation for web rendering purposes. In addition, only a

handful of studies concentrate on the creation of structured 3D scene data based on a pre-defined schema.

The current methodology builds upon previous research but expands its horizon through the inclusion of the capabilities of large language models, along with structured output schemas for generating primitive-based 3D scenes. While the usage of diffusion models and volumetric shapes is based on complex operations, the proposed approach focuses on lightweight geometric modeling utilizing simple shapes like cubes, spheres, and cylinders.

This guarantees that there will be semantic coherence between user queries. The connection between natural language processing and 3D visualization provides a valuable solution to current methods of creating 3D images from text.

III METHODOLOGY

The suggested process utilizes a systematic five-step approach to convert textual input into interactive three-dimensional scenes. This process includes elements such as semantic analysis, structured data generation, geometric representation, real-time rendering, and system optimization.

3.1 Prompt Acquisition and Semantic Interpretation

The system starts by collecting prompts provided by the user in the form of text and voice. These prompts specify the 3D object and the characteristics of its appearance. The large language model analyzes the prompt to identify any semantic relations, object features, spatial information, and descriptive information. This step ensures that the context of the prompt is correctly interpreted by the system.

For reliability reasons, the system adopts strategies for controlled prompting in order to elicit structured responses from the large language model.

3.2 Schema-Constrained Structured Scene Generation

Unlike the generation of arbitrary output responses, the language model utilizes a pre-defined JSON structure to restrict its outputs. The mandatory properties defined in the JSON structure for the scene include object type, position coordinates, rotation, scale, color, metallic surface, roughness, and opacity.

The use of structuring constraints provides uniformity and removes ambiguity from the output data. The output objects are well-defined and can be processed by the rendering engine without further processing.

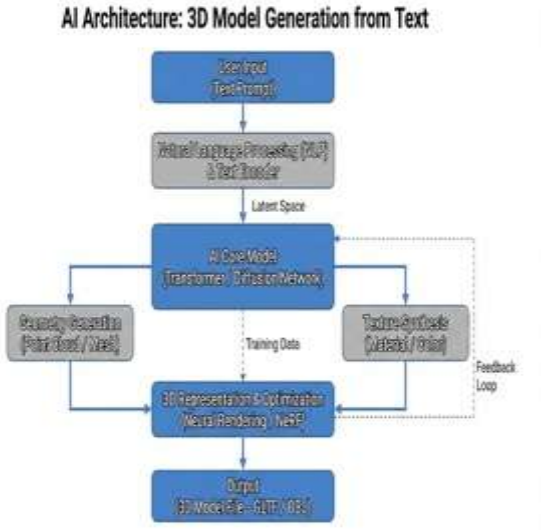

Figure 3.2.1 : Architecture of the Ai for 3D Model Generation From Texts

3.3 Primitive-Based Geometric Decomposition

Natural language descriptions of complex objects are broken down into combinations of basic geometric primitives including cubes, spheres, cylinders, cones, and tori.

The benefit of such a method is that it simplifies the computational process without sacrificing structural information.

The advantage of using primitives in the creation of objects is that they can be assembled like blocks, each of which can be separately placed, oriented, and resized. This technique performs better than neural mesh synthesis and volumetric rendering approaches.

3.4 Real-Time Rendering and Visualization

The organized information about the scene is fed into a rendering engine running on WebGL, which is embedded in the reactive frontend. The geometry primitives are created dynamically, and the required transformations are applied to the primitives using the parameters obtained.

Material properties that are physically realistic are specified, and the lighting and camera parameters are tuned for better visualization. The final rendering is done in an interactive canvas where the object can be rotated, zoomed, and inspected.

3.5 System Optimization and User Interaction Features

To improve usability and scalability, the system introduces extra modules like voice input, scene history, color configuration, and screenshot exports.

The validation process guarantees the accuracy of data prior to the rendering process, avoiding any runtime issues.

With its lean nature, it does not have heavy computational requirements and can be implemented in ordinary web browsers without needing any particular hardware.

3.6 Flowchart

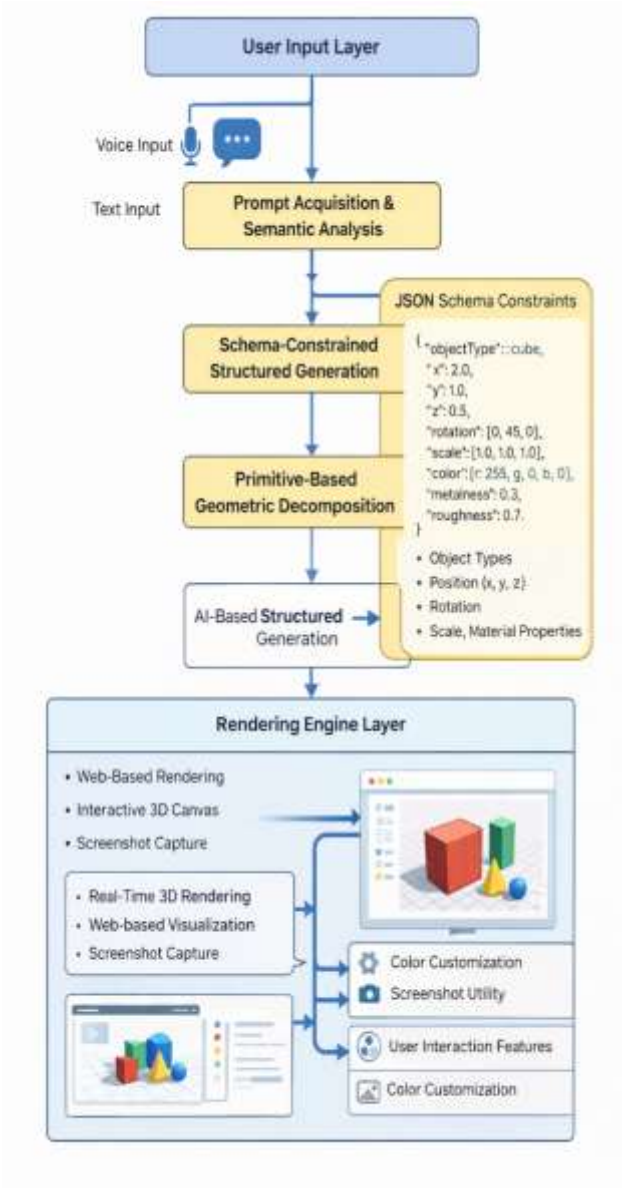

Figure 3.6.1 : Flowchart

IV RESULTS

4.1 INPUT PANEL

Figure 4.1.1 : Prompt Area

Figure 4.1.2 : Processing User Input

4.2 OUTPUT PANEL

Figure 4.2.1 : Generated Output

V CONCLUSION

A lot of advances have been made in the field of digital design due to the development of the three-dimensional model based on textual information, allowing artists, engineers, designers, and even students to design their objects effectively.

The current project demonstrates that it is possible to create a 3D model based on textual inputs containing natural language. Moreover, one may conclude that it is possible to do so without having any technical skills.

By leveraging the power of language modeling and neural rendering, one can reduce the complexity of designing three-dimensional models and save time and energy on doing so.

According to the results obtained during the experiment, the designed system is capable of interpreting the user's intention through textual inputs. Additionally, the system is able to detect important visual elements that will form the basis of the 3D model in question.

VI REFERENCES

Poole B., Jain A., Barron J. T., & Mildenhall B. (2022). DreamFusion: Text-to-3D Using 2D Diffusion. Paper presented at International Conference on Learning Representations (ICLR).

Li C., et al. (2023). A Comprehensive Survey on Generative AI Approaches for Text-to-3D and Image-to-3D Synthesis. arXiv preprint.

Mildenhall B., Srinivasan P. P., Tancik M., Barron J. T., Ramamoorthi R., & Ng R. (2020). Neural Radiance Fields (NeRF).

Lin C.-H., Gao J., Tang L., Takikawa T., Zeng X., Huang X., et al. (2023). Magic3D: Efficient Framework for High-Resolution 3D Generation From Text.

Radford A., Kim J. W., Hallacy C., Ramesh A., Goh G., Agarwal S., et al. (2021). CLIP: Demonstrating the Power of Language to Guide Visual Understanding by Jointly Learning Image and Text Embeddings.

Park J. J., Florence P., Straub J., Newcombe R., & Lovegrove S. (2019). DeepSDF: Learning Continuous Signed Distance Functions For Shape Representation.