Language-To-Scene: Generating Virtual Environments from Natural Language Using LLMs and 3D Tools

Research Proposal

The University of Adelaide

Christopher Hamilton - a1766121

Supervisor: Dr Zhibin Liao

Contents

1 Introduction and Research Problem **3**

2 Dataset **3**

3 Task Selection and Benefit **3**

4 Research Questions **3**

5 Proposed Methodology **4**

6 Ethics Statement **4**

7 Expected outcome and timeline **5**

1 Introduction and Research Problem

Recently, there has been a lot of research and development done in Large Language Models (LLMs) and using them for the generation of text. On top of this research, there has also been research and developments into generation of images and more recently 3D objects using machine learning techniques. This project aims to complete the research necessary to generate a 3D scene from a textual input.

The automatic generation of full 3D scenes could be used in game development or other tasks. In particular, this research aims to generate scenes that are explorable from different angles and perspectives, in contrast to some of the current methods to generate scenes.

2 Dataset

Since this project requires multiple machine learning technologies to be applied together, a single dataset will not be sufficient for training a full pipeline from text input to 3D scene outputs. The project will utilise LLM, 3D object generation, and object placement technologies, and as such, qualitative evaluation will be done on the whole system.

However, different datasets could be used for each part of the pipeline as demonstrated through previous research (Rombach et al. 2021). For example, a labelled dataset of text to 3D objects could be used for the training of the model that completes generation of 3D objects from text. A different labelled dataset would be used for training the text to image part of the pipeline for the background textures. The CLEVR dataset contains descriptions of different scenes with objects laid out inside of them (Johnson et al. 2016). This or a similar dataset could be used for the 3D scene layout, given that the layouts are encoded into a format that could be analysed for training, validation and testing.

3 Task Selection and Benefit

Being able to generate a 3D scene automatically could be of great benefit to game developers and artists, saving time and automating what is usually a manual task. In answering the research questions for the project, the aim is to develop a pipeline which takes a textual description of a scene as an input, generates 3D object and background images as required, and creates a full scene that can be explored. This research may make scene creation more accessible to people who do not have experience as artists or designers in a similar way to how 2D image generation has allowed many people to quickly and easily create images for specific use cases (Enjellina et al. 2023).

This research will focus on evaluating the current methods that have been developed for text generation using LLMs, image and 3D object generation from text, and placement of 3D objects in a scene. If possible, the research will then go into the development of a method to generate a scene, including a background and multiple objects laid out appropriately, utilising existing tools and methods where needed, and implementing or improving the missing artificial intelligence technologies required.

4 Research Questions

The first question that this research aims to answer is: 'Is it possible to generate a 3D scene using a combination of currently researched methods and techniques?'. If the research finds that it is not possible, the focus will shift to the questions of 'Why is it not possible?' and 'Why are researchers not developing methods for this type of generation?'. These questions are designed to guide the research project, with the ideal outcome being an answer to the question of 'How can

AI technologies and other software be constructed in a way to generate a 3D scene from natural language?', given that the research finds it is possible.

There could be challenges that could be faced during the research that will prevent answering the research questions. One such challenge is limited access to open source models and technologies for 3D generation. If previous research is not available for free use in this project, resources that could be used to answer the research questions are limited. Another challenge is limited access to powerful enough compute resources for running some of the tools. This project will be conducted on consumer grade hardware, rather than that which may be designed for running training and inference, which could reduce the effectiveness of the models. By understanding these limitations at the beginning of the project, the research is able to be guided towards tools, technologies and methods that give the best chance of answering the questions posed.

5 Proposed Methodology

In order to conduct this research, qualitative results will be analysed to determine the effectiveness of using a pipeline of machine learning tools for scene generation from text. This research will investigate where existing technologies can fit for each part of the pipeline, and where new tools or models need to be implemented. Large language models will first be utilised to generate some of the objects that will be placed in the scene, as well as any specific details about the layout. 3D object generation models, such as SHAP-E (Jun et al. 2023), will also be utilised to create objects that can be placed in a scene. Finally, Unity will be used as the tool to display and interact with a scene, with all objects being placed appropriately (Unity Technologies 2025). A framework will be built around these tools in order to allow a user to generate a 3D scene from a natural language input and test both if it is possible to generate 3D scenes from text prompts, and the effectiveness of applying the existing tools to this task. While the project does not aim to develop new techniques for the generation of images or 3D models, it does aim to bridge the gaps between where current research ends, providing a full pipeline from text to a 3D scene, with the potential for developing a new machine learning method for the appropriate layout of the objects in the scene.

As part of the upcoming literature review for this project, investigation will be completed into different object layout methods using transformer models to see if they are appropriate for the task, with some preliminary research indicating that they may be effective (Huq et al. 2020).

One issue with the proposed methodology is that the lack of dataset for the entire pipeline makes it difficult to perform quantitative analysis and testing accuracy or error metrics. As such, for this research, it will be important to generate multiple different types of scenes and observe and analyse the layout of the scenes, as well as the objects themselves to effectively perform qualitative analysis. The research should also investigate possible methods for evaluating the scene layout, and how datasets could be used with appropriate values to perform some comparison between the ground truth and the generated layouts, even if these datasets are not yet existing.

6 Ethics Statement

Research in the field of artificial intelligence requires extra ethical considerations to be taken, many of which have been laid out by the Australian Government and the University of Adelaide (National Health and Medical Research Council et al. 2018). While the system that is proposed in this document does not require access to private information, the development of the system should still focus on human, societal and environmental wellbeing, human-centred values, fairness, transparency and explainability, contestability, and accountability (Department of Industry, Science and Resources 2024). When considering these principles, some issues that may be faced during the development of this system are biases in generated scenes or parts of scenes, inappropriate use, and a negative impact on the 3D design industry. To address these issues, this project accepts that currently trained models may have certain biases, and given the use will only be in a

research setting to answer certain research questions, this is acceptable. There is also a possibility of inappropriate use of the scene generation tool, where users may provide inappropriate textual inputs in an attempt to generate an inappropriate scene. While it will not be possible to control how the generated scenes are used, this project will acknowledge that outputs generated by the system could be harmful in certain situations, and make it clear that the intended use of the system is to generate scenes for research only. There is currently debate around how generative artificial intelligence models could negatively impact creative workers like artists and designers, where one side of the debate argues that this kind of automation will take jobs away, and the other believes that it will allow creatives to focus on more complex problems (McKendrick 2024). This research does not attempt to address these issues, and instead the tools that will be developed aim to automate one specific task that artists may do, rather than replace them.

7 Expected outcome and timeline

This project will run until the end of the academic year, with consideration in the timeline given for the assignments that will be required, as well as the specific research tasks to complete the project. One possible timeline for the research has been provided in a Gantt chart in Figure 1. The timeline is split into sections involving the initiation of the project, specific research tasks, meetings, and assignments. As an optimistic expected outcome for this project in the time available, is that the research question of 'How can software be constructed to generate a 3D scene from natural language?' will be answered, with the research focusing on developing the artificial intelligence or machine learning tools to complete this task.

In order to complete the research, the tasks that are expected to need to be completed are: researching existing methods and technologies, setting up a development environment, testing existing methods, developing a new method to bridge gaps in current literature, integrating a full system to generate a 3D scene, and finally refining the machine learning model or method used to improve the system. There are also a total of eight assignments that need to be completed over the rest of the academic year for this research project, including an ongoing project planner, a literature review, analysis presentation, progress presentation, final presentation and final report, which are also important to plan for in the timeline as they are compulsory components with set due dates.

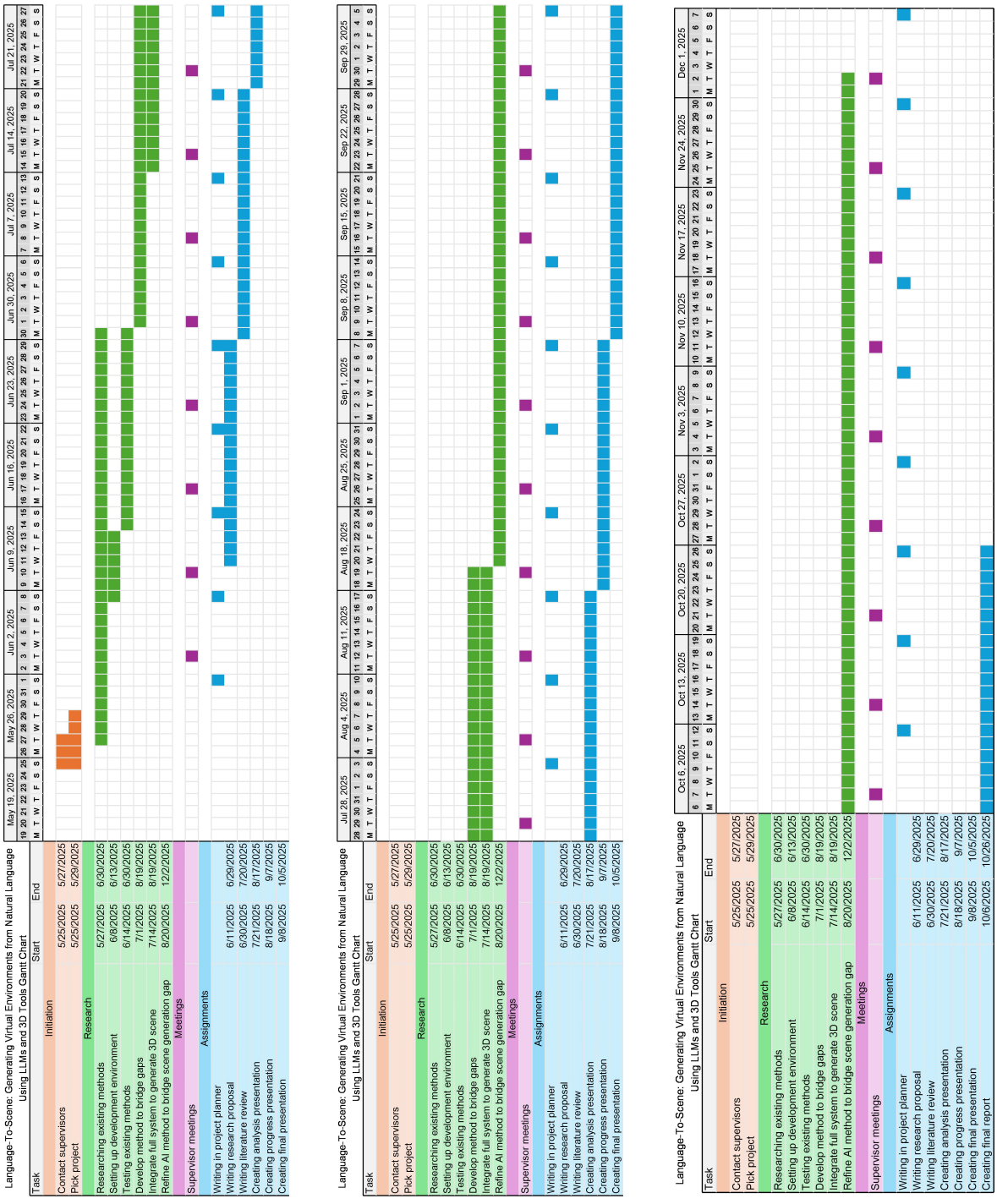

Figure 1: Gantt chart for a possible timeline for the project

References

Department of Industry, Science and Resources, 2024, *Australia's AI Ethics Principles*, viewed, 20 June 2025, URL: https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-principles/australias-ai-ethics-principles.

Enjellina, Beyan, Eleonora, and Rossy, Anastasya, 2023, "Review of AI Image Generator: Influences, Challenges, and Future Prospects for Architectural Field", vol. 2, pp. Pp. 53–65, DOI: 10.24002/jarina.v2i1.6662.

Huq, Faria, Iqbal, Anindya, and Ahmed, Nafees, 2020, "Static and Animated 3D Scene Generation from Free-form Text Descriptions", vol. 1, DOI: 10.48550/arXiv.2010.01549.

Johnson, Justin, Hariharan, Bharath, Maaten, Laurens van der, Fei-Fei, Li, Zitnick, C. Lawrence, and Girshick, Ross B., 2016, "CLEVR: A Diagnostic Dataset for Compositional Language and Elementary Visual Reasoning", vol. 1, DOI: 10.48550/arXiv.1612.06890.

Jun, Heewoo and Nichol, Alex, 2023, "Shap·E: Generating Conditional 3D Implicit Functions", vol. 1, DOI: 10.48550/arXiv.2305.02463.

McKendrick, Joe, 2024, *Generative AI as a Killer of Creative Jobs? Hold That Thought*, viewed, 22 June 2025, URL: https://www.forbes.com/sites/joemckendrick/2024/06/23/generative-ai-as-a-killer-of-creative-jobs-hold-that-thought/.

National Health and Medical Research Council, Australian Research Council, and Universities Australia, 2018, *Australian Code for the Responsible Conduct of Research*, viewed 22 June 2025, URL: https://www.nhmrc.gov.au/research-policy/research-integrity/australian-code-responsible-conduct-research.

Rombach, Robin, Blattmann, Andreas, Lorenz, Dominik, Esser, Patrick, and Ommer, Björn, 2021, "High-Resolution Image Synthesis with Latent Diffusion Models", vol. 1.

Unity Technologies, 2025, *Unity Real-Time Development Platform*, viewed 20 May 2025, URL: https://unity.com/.