Exploring the Potential of Generative AI in Prototyping XR Applications

Mohammad Javad Sahebnasi[1,*], Mahdi Farrokhimaleki[1], Nanjia Wang[1], Richard Zhao[1] and Frank Maurer[1]

[1]University of Calgary, Calgary, Alberta, Canada

Abstract

This paper presents the initial stage of our research to develop a novel approach to streamline the prototyping of Extended Reality applications using generative AI models. We introduce a tool that leverages state-of-the-art generative AI techniques to facilitate the prototyping process, including 3D asset generation and scene composition. The tool allows users to verbally articulate their prototypes, which are then generated by an AI model. We aim to make the development of XR applications more efficient by empowering the designers to gather early feedback from users through rapidly developed prototypes.

Keywords

Extended Reality, Prototyping, Generative Artificial Intelligence

1. Introduction

The field of Extended Reality (XR), which encompasses Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR) [1], has seen a significant rise in recent years, particularly with the introduction of modern headsets like the Apple Vision Pro [2] and Meta Quest 3 [3]. These devices have made XR more accessible and opened new possibilities for immersive experiences in various domains, including gaming, education, healthcare, and more. However, the development of XR applications remains complex and challenging. Creating immersive and interactive experiences requires technical expertise and is a time-consuming process. Given the complexity involved, prototyping can play a crucial role in mitigating these challenges. Prototyping allows developers to explore design concepts, iterate rapidly, and gather user feedback early in the development cycle [4]. This iterative process not only helps to refine the design, but also reduces the overall effort and costs associated with the development of XR applications. By prototyping XR applications, designers and developers can better understand the user experience, identify potential issues, and make informed decisions that ultimately lead to more polished and successful XR experiences.

We have witnessed a significant rise in the utilization of generative Artificial Intelligence, particularly following the introduction of large language models [5] such as ChatGPT [6]. Today, there are various generative AI models that can synthesize new text [6, 7, 8], images [9, 10], music [11], or even videos [12]. This capability has led to a wide range of applications across

RealXR: Prototyping and Developing Real-World Applications for Extended Reality, June 4, 2024, Arenzano (Genoa), Italy
mohammadjavad.sahebn@ucalgary.ca (M. J. Sahebnasi)
https://seriousxr.ca/ (M. J. Sahebnasi)

various domains. One of the key potential benefits of generative AI is its ability to automate and enhance the creative process. This can speed up the design process and lead to more innovative and diverse designs. We aim to investigate the integration of generative AI into the prototyping process of Extended Reality applications.

In this paper, we will introduce a novel tool that utilizes generative AI models to streamline the prototyping of XR applications. We will also describe the methods we plan to use for the evaluation of this tool.

2. Related Work

In many cases, designers and developers use expert tools, such as game engines, that are time-intensive in order to prototype XR systems [13]. These tools enable the creation of detailed, high-quality results. Some of the most commonly used tools in this category for developing XR applications are Unity 3D [14] and the Unreal Engine [15]. These tools provide rich development environments with various toolkits. However, using these expert tools is a challenge for non-experts and an obstacle for rapid prototyping, especially due to the need for technical knowledge and programming skills [16, 13]. Therefore, there have been efforts to create no-code tools to facilitate rapid prototyping of different kinds of XR applications, including VR [17, 18], AR [19, 20, 21], and CR (Cross-Reality) [13]. There are also a number of no-code prototyping and authoring tools developed and used in industry, such as [22, 23, 24]. Most of these tools allow users to create basic objects manually, or import existing object files manually.

Recent efforts have been made to utilize the potential of generative AI across various areas of design and prototyping, including mobile applications and websites, such as [25]. We aim to explore the utilization of generative AI for prototyping XR applications to empower designers to create complex XR scenes and objects rapidly.

3. Methodology

Our research aims to explore the potential of generative AI in prototyping Extended Reality applications. The methodology for this research includes multiple phases, combining tool development, auto-ethnography, and case studies.

3.1. Tool Development

We are developing a generative AI-powered prototyping tool. The tool leverages state-of-the-art generative AI techniques, and is designed to facilitate the prototyping of XR applications, including 3D asset generation and scene composition.

In the initial version of a software tool that we have developed, users can verbally articulate what they want to prototype, as depicted in Figure 1. If they wish to prototype a complex scene that potentially contains several objects, they will see a list of objects that can usually be found in such a scene. Then the user can select any number of the suggested objects from the list. Then, the generative AI model that we use will generate the selected objects, and then the objects will be placed in the scene. Users can move the generated objects inside the scene or

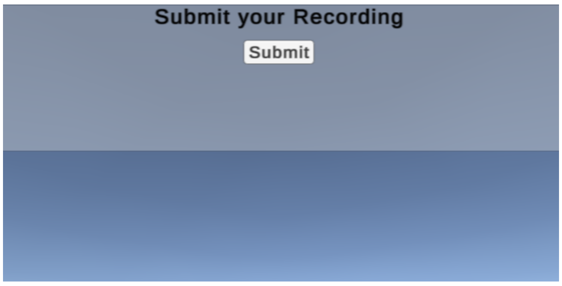
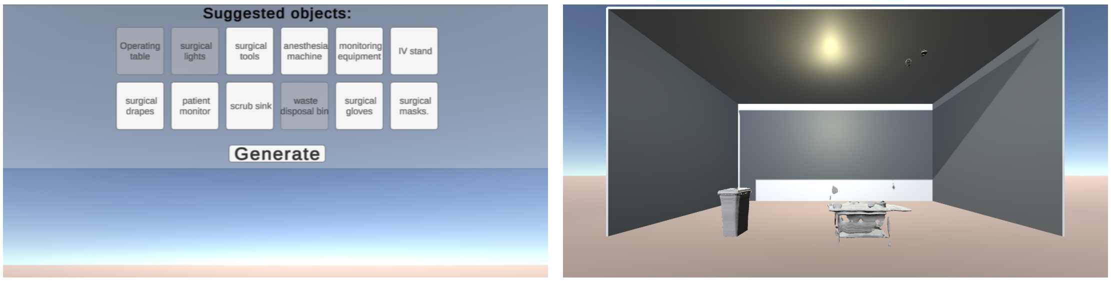


Figure 1: Users records their audio, in which they say what they want to prototype.



(a) A list of related objects are suggested. In this case, some objects one can find in a typical *surgery room* are listed.

(b) The selected objects, which in this case are *operating table*, *surgical lights*, and *waste disposal bin*, are generated.

Figure 2: A sample scenario: user wants to prototype a *surgery room*.

remove them from the scene. In a sample scenario that is depicted in Figure 2, the user wants to prototype a surgery room and then gets a list of objects that one can find in a typical surgery room, such as an operating table, surgical tools, an IV stand, a scrub sink, monitoring equipment, and more. In this case, the user selects the operating table, surgical lights, and waste disposal bin. These objects are then created and placed on the scene. As another example, in Figure 3, the user intends to create a tree. Since it is an object rather than a complex scene, in the list there is only one suggested object, which is tree, and when the user selects that, a tree will be generated and added to the scene. Also, the system will detect if the scene the user wants to create is indoors or outdoors. If it is indoors, a cubic room will automatically be added to the scene, and the generated objects will be placed inside that.

The technologies used in the implementation of the tool include Whisper API [26] for transcribing users' recorded audio, GPT 3.5 API [26] for processing the transcripts, Shap-E [27] for generating 3D objects, and Unity 3D [14] for creating the environment, displaying and supporting interactions with the objects and scenes.

3.2. Planned Evaluation

The development and prototyping of Extended Reality applications are inherently time-consuming processes. Therefore, short, controlled user studies may not adequately evaluate

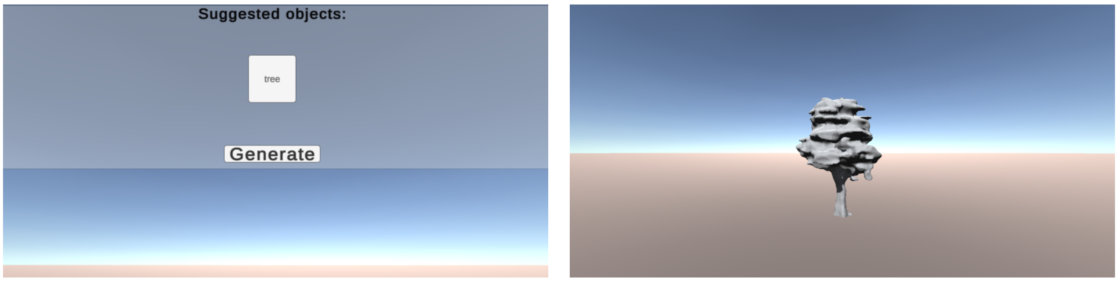


Figure 3: A sample scenario: user wants to prototype a *tree*.

the efficacy of a prototyping tool due to the limited exposure time and controlled nature of the study. To gain a deeper understanding and evaluate the tool properly, users must work with it for an extended period, ideally several weeks. However, this is not feasible within the constraints of controlled user studies. Therefore, we plan to evaluate our work in the following phases.

The first step will be to use auto-ethnography. After developing the first version of the tool, we will use it in a real-world XR project and then evaluate it using the auto-ethnography method. Auto-ethnography is a method where the author creates a detailed study of themselves, plays a dual role as both the subject and the researcher, and analyzes personal behavior and experiences to gain insight into larger contexts [28]. This method goes beyond simple storytelling, aiming for objectivity in interpreting one's own thoughts and actions, while still acknowledging the personal perspective involved [28]. Based on the outcomes of this phase and the feedback we anticipate receiving upon publication, we will enhance our work in preparation for the second phase.

The second phase will be conducting a case study evaluation. Following the auto-ethnography evaluation, we will invite a select group of developers and designers to use the tool for an extended period, ideally several weeks. we will then conduct case studies, which include semi-structured interviews with the participants. Case studies provide an in-depth examination of how the tool is used in practice, allowing a detailed exploration of its strengths, weaknesses, and potential improvements. Semi-structured interviews will allow participants to share their experiences, feedback, and suggestions for the tool, providing valuable insights for further refinement and development.

By employing these evaluation methods, we aim to gain a comprehensive understanding of the tool's impact, effectiveness, and usability in real-world XR development scenarios. The auto-ethnography evaluation will provide rich, qualitative insights into the tool's influence on the researcher's personal experiences and practices, while the case studies will offer broader perspectives from a diverse group of users. Together, these evaluation methods will inform iterative improvements and refinements of the tool, ultimately enhancing its utility and value for XR developers and designers.

4. Conclusion and Future Work

In this research, we aim to make the development of XR applications more efficient by empowering developers and designers to gather early feedback from users through rapidly developed prototypes. We present the initial version of our tool, which allows users to verbally articulate the scene or object that they want to prototype and then generates the objects using a state-of-the-art generative AI model. The next steps for this research are completing the development of the tool by adding more functionalities and evaluating the tool through the methods that we described in this paper.

References

[1] K. M. Stanney, H. Nye, S. Haddad, K. S. Hale, C. K. Padron, J. V. Cohn, Extended reality (xr) environments, 2021. URL: http://dx.doi.org/10.1002/9781119636113.ch30. doi:10.1002/9781119636113.ch30.

[2] Apple vision pro, https://www.apple.com/apple-vision-pro, 2024. Accessed: 2024-03-27.

[3] Meta quest3, https://www.meta.com/ca/quest/quest-3, 2023. Accessed: 2024-03-27.

[4] A. M. Davis, Software Prototyping, Elsevier, 1995, p. 39–63. URL: http://dx.doi.org/10.1016/s0065-2458(08)60544-6. doi:10.1016/s0065-2458(08)60544-6.

[5] M. U. Hadi, q. a. tashi, R. Qureshi, A. Shah, a. muneer, M. Irfan, A. Zafar, M. B. Shaikh, N. Akhtar, J. Wu, S. Mirjalili, A survey on large language models: Applications, challenges, limitations, and practical usage (2023). URL: http://dx.doi.org/10.36227/techrxiv.23589741.v1. doi:10.36227/techrxiv.23589741.v1.

[6] Chatgpt, https://chat.openai.com/, 2022. Accessed: 2024-03-27.

[7] Grok, https://grok.x.ai/, 2024. Accessed: 2024-03-27.

[8] Claude, https://claude.ai/, 2023. Accessed: 2024-03-27.

[9] J. Betker, G. Goh, L. Jing, TimBrooks, J. Wang, L. Li, LongOuyang, JuntangZhuang, JoyceLee, YufeiGuo, WesamManassra, PrafullaDhariwal, CaseyChu, YunxinJiao, A. Ramesh, Improving image generation with better captions, 2023. URL: https://api.semanticscholar.org/CorpusID:264403242.

[10] Midjourney, https://www.midjourney.com/home, 2022. Accessed: 2024-03-27.

[11] R. Louie, A. Coenen, C. Z. Huang, M. Terry, C. J. Cai, Novice-ai music co-creation via ai-steering tools for deep generative models, in: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI '20, Association for Computing Machinery, New York, NY, USA, 2020, p. 1–13. URL: https://doi.org/10.1145/3313831.3376739. doi:10.1145/3313831.3376739.

[12] Sora, https://openai.com/research/video-generation-models-as-world-simulators, 2024. Accessed: 2024-03-28.

[13] U. Gruenefeld, J. Auda, F. Mathis, S. Schneegass, M. Khamis, J. Gugenheimer, S. Mayer, Vrception: Rapid prototyping of cross-reality systems in virtual reality, in: CHI Conference on Human Factors in Computing Systems, CHI '22, ACM, 2022. URL: http://dx.doi.org/10.1145/3491102.3501821. doi:10.1145/3491102.3501821.

[14] Unity, https://unity.com/, 2005. Accessed: 2024-03-28.

[15] Unreal engine, https://www.unrealengine.com/, 1998. Accessed: 2024-03-28.

[16] N. Ashtari, A. Bunt, J. McGrenere, M. Nebeling, P. K. Chilana, Creating augmented and virtual reality applications: Current practices, challenges, and opportunities, in: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI '20, Association for Computing Machinery, New York, NY, USA, 2020, p. 1–13. URL: https://doi.org/10.1145/3313831.3376722. doi:10.1145/3313831.3376722.

[17] M. Nebeling, K. Lewis, Y.-C. Chang, L. Zhu, M. Chung, P. Wang, J. Nebeling, Xrdirector: A role-based collaborative immersive authoring system, in: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI '20, Association for Computing Machinery, New York, NY, USA, 2020, p. 1–12. URL: https://doi.org/10.1145/3313831.3376637. doi:10.1145/3313831.3376637.

[18] M. Nebeling, K. Madier, 360proto: Making interactive virtual reality & augmented reality prototypes from paper, in: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI '19, Association for Computing Machinery, New York, NY, USA, 2019, p. 1–13. URL: https://doi.org/10.1145/3290605.3300826. doi:10.1145/3290605.3300826.

[19] G. Freitas, M. S. Pinho, M. S. Silveira, F. Maurer, A systematic review of rapid prototyping tools for augmented reality, in: 2020 22nd Symposium on Virtual and Augmented Reality (SVR), 2020, pp. 199–209. doi:10.1109/SVR51698.2020.00041.

[20] M. Nebeling, J. Nebeling, A. Yu, R. Rumble, Protoar: Rapid physical-digital prototyping of mobile augmented reality applications, in: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18, Association for Computing Machinery, New York, NY, USA, 2018, p. 1–12. URL: https://doi.org/10.1145/3173574.3173927. doi:10.1145/3173574.3173927.

[21] M. Speicher, B. D. Hall, A. Yu, B. Zhang, H. Zhang, J. Nebeling, M. Nebeling, Xd-ar: Challenges and opportunities in cross-device augmented reality application development, Proc. ACM Hum.-Comput. Interact. 2 (2018). URL: https://doi.org/10.1145/3229089. doi:10.1145/3229089.

[22] Microsoft maquette, https://learn.microsoft.com/en-us/windows/mixed-reality/design/maquette, 2019. Accessed: 2024-03-28.

[23] Shapesxr, https://www.shapesxr.com/, 2023. Accessed: 2024-03-27.

[24] Bezi, https://hq.bezi.com/, 2024. Accessed: 2024-03-27.

[25] Uizard, https://uizard.io/, 2024. Accessed: 2024-03-28.

[26] Introducing chatgpt and whisper apis, https://openai.com/blog/introducing-chatgpt-and-whisper-apis, 2023. Accessed: 2024-03-28.

[27] H. Jun, A. Nichol, Shap-e: Generating conditional 3d implicit functions, arXiv preprint arXiv:2305.02463 (2023).

[28] S. J. Cunningham, M. Jones, Autoethnography: a tool for practice and education, in: Proceedings of the 6th ACM SIGCHI New Zealand Chapter's International Conference on Computer-Human Interaction: Making CHI Natural, CHINZ '05, Association for Computing Machinery, New York, NY, USA, 2005, p. 1–8. URL: https://doi.org/10.1145/1073943.1073944. doi:10.1145/1073943.1073944.