Generative AI for Immersive Experiences: Integrating Text-to-Image Models in VR-Mediated Co-design Workflows

BUSSELL, Chris, ABDELSALAM, Ehab, HARTLE-RYAN, Daniel and KAPSALIS, Timo

Generative AI for immersive experiences: integrating text-to-image models in VR-mediated co-design workflows

Chris Bussell[1[0000-0003-0869-9947]] Ahmed Ehab [2[0000-0002-9752-7321]]

Daniel Hartle-Ryan[3[0009-0009-6899-2708]] Timo Kapsalis[4[0000-0003-1598-6426]]

[1,2,3,4] Derby Urban Sustainable Transition (DUST), University of Derby, UK
c.bussell@derby.ac.uk

Abstract. Text-to-image AI models can generate novel images for design inspiration. Yet, their applications for collaborative design (co-design) purposes and interoperability within simulation-based, immersive settings have been scarcely explored. In this paper, we propose a novel, multi-modal approach for interactive public participation in urban design projects. The main objectives of our research are (a) to describe a methodological workflow of integrating text-to-image AI models into VR-mediated co-design workshops, and (b) to investigate the applicability of the proposed workflow through a set of completed and prospective case studies. Both studies are parts of a broader research project, which aims to revitalize the city of Derby, UK through producing a series of sustainable design visions. Through these case studies, we discuss some preliminary results and introduce our future work.

Keywords: Virtual Reality (VR); Text-to-Image; Generative AI; Co-design; Urban Design

1 Introduction

Public participation is a process that involves engaging members of the public in decision-making processes that affect their lives, communities, and environment [11]. In an urban design context, the goal of public participation is to foster greater transparency, accountability, and democratic governance, by ensuring that the voices and perspectives of all stakeholders are taken into account [12]. Collaborative design (or, *co-design*) workshops have been regularly applied by city officials as well as design professionals and researchers to effectively engage local stakeholders in urban processes, such as neighborhood regeneration schemes [13]. A co-design workshop typically involves a series of structured activities to facilitate a shared understanding of the challenge being addressed and generate ideas for potential design solutions that better reflect the needs and preferences of the local population [14]. In many cases, designers have employed tools and techniques from the field of Human Computer Interaction (HCI) – for instance, visual representation in digital forms – to facilitate stakeholders' participation in the process of designing the built environment, thus resulting in participatory and interactive paradigms of urban dialogue [15,16].

1.1 Theoretical background

Virtual Reality (VR) is a computer-generated immersive experience that simulates a three-dimensional environment in which users can interact with digital content in real-time [1]. VR is being increasingly applied in the built environment industry for various purposes, such as design collaboration, design review, performance analysis, and facility management. Specifically, VR enables designers, clients, and other stakeholders to experience and evaluate a building project before construction, facilitating design decision-making and problem-solving [2]. Moreover, VR provides a more efficient and effective way of reviewing design elements than traditional methods, and allows for interactive and immersive training and simulation of building operations and maintenance. Additionally, VR can enhance user experience and engagement in a building project, contributing to greater user satisfaction and ultimately improving the success of the project [5].

VR has emerged as a promising technology for urban design collaboration. It offers immersive and interactive experiences to urban design stakeholders, including citizens, planners, designers, and policymakers, by creating a 3D virtual environment that can be navigated in real-time [3-4]. In recent years, several studies have explored the potential of VR in urban design collaboration, covering a range of topics, such as public participation, decision-making, and stakeholder engagement [5,6,9]. VR has been used for visualizing urban design scenarios, assessing the impact of design decisions on the environment and the quality of life, and enhancing collaboration among stakeholders. These studies have shown that VR can facilitate a more effective and inclusive urban design process, leading to better outcomes in terms of design quality, social equity, and environmental sustainability [6-7].

The creation of an immersive VR model for collaborative urban design poses several challenges that must be addressed in order to ensure its success. These challenges include the design of a VR model that accurately represents the complexities of real-world urban environments, which requires a high level of detail and precision and the integration of a diversity of data sources such as Geographic Information Systems (GIS) and Building Information Models (BIM) [8]. Additionally, the development of a framework that supports multi-user collaboration in a user-friendly and effective manner is crucial and requires careful consideration of the user interface and interaction design. The system must provide the necessary tools for users to manipulate and interact with the VR model, while also being intuitive and straightforward to use. Furthermore, ensuring the accessibility of the VR model to all collaborators, regardless of their technical expertise or equipment, requires the creation of a scalable and flexible system that can accommodate a wide range of hardware and software configurations [4,9,10]. In conclusion, overcoming these challenges is critical for the success of VR-based collaborative urban design initiatives, as it enables urban designers to collaborate effectively and create innovative designs that accurately reflect the complexities of real-world urban environments.

Recent advancements in the field of artificial intelligence (AI), especially with respect to image generation and natural language processing tasks, have led to the increasing application of machine learning to domain-specific aspects of the built environment [17,18]. The proliferation of AI generative models, such as Variational Autoencoders [19] and Generative Adversarial Networks [20], have allowed design professionals and researchers to recreate images of buildings, districts, and cities in a bid to explore design alternative options, integrate contextual data, and improve communication of design proposals. More recently, diffusion models have been developed as a more powerful technique for image generation. Some popular generative models include DALL-E 2 [21], Stable Diffusion [22], Imagen [23], and Make-a-Scene [24], which all utilize natural language processing algorithms to generate images from textual descriptions given by the user. Text-to-image diffusion models simulate the distribution of the corresponding image using a diffusion process that iteratively refines the image over multiple steps [25]. This process enables the model to capture complex visual details and synthesize images that are both diverse and realistic [26,27]. It has been observed that images generated through diffusion processes are more accurate to what humans would perceive to be real, when compared to those produced by other generative models [28,29,30].

Despite the emergence of generative AI, remarkably little research has hitherto focused specifically on possible uses of text-to-image models for urban design purposes [18]. A recent study, which delved into this topic, evaluated the potential for text-to-image models to design realistic images and scenes with a view to supporting urban design processes [31]. After a series of software experiments, researchers concluded that AI-generated imagery could reflect the human perception on various aspects of the built environment, such as design qualities [31]. They thereby suggested that text-to-image models might have the potential to substitute population surveys as a tool for gathering human feedback on urban design matters [31]. Drawing on this claim, researchers from Toretei and SPIN Unit have developed *Urbanist AI*, a generative AI platform for participatory planning [32]. The platform leverages natural language processing tools to enable community members with limited design experience communicate their ideas on urban revitalization schemes and create urban design scenarios using textual prompts [32]. Urbanist AI was recently employed by the City of Helsinki as the main tool for a series of co-design workshops, which intended to study local community's preferences with respect to redesigning certain streets of the city [32]. Workshop participants produced numerous alternative visions, which were later collated and discussed with city officials and planning consultants [32]. Whilst both pieces of research are regarded as encouraging steps towards validating the potentiality of generative AI for co-creating urban design scenarios, neither have dealt with (a) communicating the AI-generated concepts in various formats other than images (e.g., VR environments), and (b) subsequently exploring users' interactions with neoteric and multi-modal visualization techniques. By addressing both aspects, the present research provides a significant opportunity to advance the understanding of text-to-image models, from an HCI perspective.

1.2 Purpose

In this paper, we propose a novel, multi-modal approach for interactive public participation in urban design projects. The main objectives of our research are (a) to describe a methodological workflow of integrating text-to-image AI models into VR-mediated co-design workshops, and (b) to investigate the applicability of the proposed workflow through a series of completed and prospective case studies.

2 Methodology

We propose a workflow of integrating text-to-image AI models into VR-mediated co-design workshops to support urban design interventions. For the prospective co-design workshops, we assume that at least one study participant and one investigator will partake in a co-design session. In an urban design context, members of the community (e.g., residents, local entrepreneurs, and action groups) as well as planning authorities will comprise the group of participants. The proposed workflow consists of five steps, which are analyzed below:

- *Step 1 – simulation of the current urban settings*. Firstly, the investigators create a virtual 3D model of the real-world environment, such as a city street or region, whose design they intend to revision. For our experiments, we have created the models using the Rhino 3D software. The simulated model is typically a three-dimensional mesh of polygons that defines the surface geometry of the object or objects included, such as buildings, landscape elements, and street infrastructure. The model should precisely represent the physical attributes and characteristics of the actual urban environment. As soon as the investigators create the digital replica for the area-of-interest, they need to export a two-dimensional representation (i.e., base image) of the model along with a depth map of the exported image. Investigators can choose among top, side, or axonometric views of the model to export as a base image for the next steps. They can also select multiple base images corresponding to different perspectives of the model to better inform the design ideation process.
- *Step 2 – generation of participants' visual input*. Base images along with their depth maps are then imported into a platform for text-to-image generation. For our experiments, we have selected Stable Diffusion, a pre-trained text-to-image model, as the basic component of the generative platform, which we then custom-built using the Grasshopper3D component of Rhino3D software. The platform will enable workshop participants to generate images from textual descriptions of their alternative visions for the study area. For example, if an instance of a residential building block is presented as a base image, participants can provide

prompts of their preference, such as "a building with a brick façade", to recreate a visualization of the initial building coated in a brick-made frontage. The pre-trained, text-to-image model that underpins the platform will then return a recreated image according to the given prompt. The integration of the base-image depth localizes the generation of alternative scenes for the base image based on participants' prompts. This step can be repeated multiple times until the output image satisfies participant's requirements. In the case that more than one base image has been selected, the step is also repeated for as many times as appropriate.

- *Step 3 – selection of design scenarios.* Next, the investigators collate all images generated by participants' text prompts to identify the most prevalent features on images. Those can refer to natural elements (e.g., trees), built elements (e.g., curb ramps), or design qualities (e.g., cycle infrastructure). The collation process is achieved through an integrative method, which combines recognition of similar patterns in the generated images and counting the frequency of specific words or phrases used by participants to construct their prompts. For instance, if the word "trees" has been found multiple times among participants' prompts and tree-like patterns have been repeatedly detected by the visual analysis, it is very possible that most participants would prioritize trees for the reimagined area. As soon as the investigators identify the most significant elements based on participant generations, they should be able to synthesize appropriate design scenarios for the area of study. The exact number of scenarios produced as well as level of design intervention applied depend on investigators' intentions and requirements of different projects.
- *Step 4 – 3D modelling and animation of the selected scenarios.* Once a design scenario is selected, the first step is to model the design of the scene using 3D modelling software. Autodesk Maya and Houdini were chosen for this task due to their high capabilities in creating a detailed and high-fidelity model. Following the modelling stage, the models are then exported to real-time rendering software such as Unreal Engine 5. This software is utilized to refine the model's lighting, materials, and textures in preparation for the animation phase. The animation phase in Unreal Engine 5 plays a critical role in the development of an immersive virtual reality (VR) model for collaborative urban design projects. The purpose of this phase is to enliven the three-dimensional (3D) representation of the urban environment, thereby creating a realistic and interactive experience for stakeholders. The animation phase involves the integration of interactive elements and animations into the 3D model, including dynamic elements such as traffic, people, and other factors that contribute to the immersive experience.
- *Step 5 – functionality of an interactive VR platform.* The final stage in this workflow involves creating an interactive virtual reality (VR) platform using the Unreal Engine 5 game engine to project the animated scenes and models into an immersive environment. Building an interactive immersive model for VR requires advanced coding skills from investigators and the use of visual scripting languages such as Blueprints or C++. Through this process, investigators can add interactivity to the VR model by incorporating elements such as user-controlled movements, object interactions, and animations. This stage represents a crucial

step in the development of an immersive VR experience, as it determines the extent to which users can engage with and explore the virtual environment. However, it is also a complex and challenging step that requires a high degree of technical expertise and attention to detail to ensure the VR experience is seamless and immersive.

3 Case study

To validate the proposed workflow, we have selected a case study that was implemented as part of a wider research project conducted in the University of Derby, UK. The project for Derby's Urban Sustainable Transition (DUST) is a co-creation initiative that aims to develop innovative design solutions and strategies for the homonymous city. The DUST research team, which comprises of all four co-authors of this paper, collaborates with local stakeholders to reconceptualize Derby as a more sustainable, prosperous, and resilient urban environment, while improving the quality of life for residents. Specifically, we have been preparing design proposals for public realm enhancements across the city and will invite local stakeholders to contribute to the discussion on how certain interventions may affect the cityscape. Through utilizing the workflow proposed in the Methodology section of this paper, we intend to co-create multiple intervention scenarios for various sites in the city to better inform the place-making process (Figure 1). We provide some preliminary results from the application of the proposed workflow for co-design workshops in the context of DUST in the following subsection.

Fig. 1. 3D digital model of the city of Derby created using Unreal Engine 5, source: Author.

3.1 Open workshop

In order to evaluate the effectiveness of the workflow, participants were invited to test the Virtual Reality (VR) platform in open showcase events prior to launching the experiment. These events served as an opportunity for participants to provide feedback and preferences on the VR model, which were then used to fine-tune and optimize the VR platform to meet the needs and requirements of stakeholders. The open showcase events also provided a means to demonstrate the potential of the VR model to a wider audience and generate interest in the urban design project. To meet the hardware requirements, the Oculus Quest 2 was utilized for virtual interactions and connected with a PC and two big screens, one of which showed the participants' real-time interaction in the virtual world, while the other displayed a video animation of the city model. The investigators recorded and analyzed the participants' interactions during the showcase.

The showcase event was hosted at the Derby City Lab, which is based on the concept of Urban Rooms and houses the City Living Room. It features a rolling program of dynamic exhibitions and discussion groups focused on ideas to regenerate Derby. The DUST event took place over three weeks in September 2022, during which both the public and stakeholders, including policy makers and organizations such as Derby City Council, Canal River Trust, Marketing Derby, Down to Earth, Environment Agency, and transportation agencies such as Toyota and Trent Barton, were invited to test the VR model and share their ideas. The feedback gathered during these showcase events contributed to the success of the urban design project and ensured that the VR model met the needs and requirements of stakeholders (Figure 2).

Fig. 2. Participants interactions during the DUST showcase event, Derby City Lab, source: Author.

4 Discussion and future work

The application of the proposed workflow helped us identify three key themes resulting from participants’ interaction with virtual models: sustainability, accessibility, and place making. Participants expressed their ideas and concerns related to humanizing the streets and increasing pedestrianization, the lack of vibrant spaces in the city center, new ideas and means of public and clean transportation, the need for more green spaces and gardens along the river side, and the implementation of 15-minute mixed-use neighborhoods. We conducted a coding analysis to organize and highlight these discussions into themes and subthemes, which required a micro-scale focus on different case studies and areas in Derby. This approach allowed for a comprehensive and detailed understanding of the needs and preferences of the stakeholders, which informed the optimization of the VR platform. Specifically, the feedback gathered during the showcase event facilitated the creation of an interactive environment and different design scenarios that participants could interact with and test during the subsequent stages of the VR experiment.

In the near future, we plan to employ the workflow proposed in the Methodology section of this paper into a co-creation study, which investigates the potentiality of low-traffic, shared streets at Derby's neighborhoods. The overall study design has recently received ethical approval by the College Research Ethics Committee, University of Derby. For this research, the goal is to design a more pedestrian-friendly environment that encourages community engagement and improves the quality of life for residents. To achieve this, we plan to introduce a series of design interventions that encourage social interaction, enhance livability, and help build a stronger sense of community in the area, such as nature-based parklets, outdoor exercising spaces, and cycle infrastructure. Utilizing the proposed workflow, we will be inviting locals to test different design scenarios and solicit their ideas in a series of open workshops. We expect that participants' text-to-image-generated input as well as their interaction with the simulated environments in VR will help us recreate the streetscape and contribute to local planning policies.

Funding

This project has been generously supported by the Osborne Legacy. The financial assistance provided by the legacy has been instrumental in the successful completion of this research effort.

References

1. Portman, M. E., Natapov, A., & Fisher-Gewirtzman, D. (2015). To go where no man has gone before: Virtual reality in architecture, landscape architecture and environmental planning. Computers, Environment and Urban Systems, 54, 376-384.
2. Meenar, M., & Kitson, J. (2020). Using multi-sensory and multi-dimensional immersive virtual reality in participatory planning. Urban Science, 4(3), 34.
3. Zhang, C., Zeng, W., & Liu, L. (2021). UrbanVR: An immersive analytics system for context-aware urban design. Computers & Graphics, 99, 128-138.

4. Safikhani, S., Keller, S., Schweiger, G., & Pirker, J. (2022). Immersive virtual reality for extending the potential of building information modeling in architecture, engineering, and construction sector: systematic review. International Journal of Digital Earth, 15(1), 503-526.
5. Liu, X. (2020). Three-dimensional visualized urban landscape planning and design based on virtual reality technology. IEEE access, 8, 149510-149521.
6. Schrom-Feiertag, H., Stubenschrott, M., Regal, G., Matyus, T., & Seer, S. (2020, May). An interactive and responsive virtual reality environment for participatory urban planning. In Proceedings of the 11th Annual Symposium on Simulation for Architecture and Urban Design (pp. 1-7).
7. Kim, S., Kim, J., & Kim, B. (2020). Immersive virtual reality-aided conjoint analysis of urban square preference by living environment. Sustainability, 12(16), 6440.
8. Yu, R., Gu, N., Lee, G., & Khan, A. (2022). A Systematic Review of Architectural Design Collaboration in Immersive Virtual Environments. Designs, 6(5), 93.
9. Panya, D. S., Kim, T., & Choo, S. (2023). An interactive design change methodology using a BIM-based Virtual Reality and Augmented Reality. Journal of Building Engineering, 106030.
10. Delgado, J. M. D., Oyedele, L., Demian, P., & Beach, T. (2020). A research agenda for augmented and virtual reality in architecture, engineering and construction. Advanced Engineering Informatics, 45, 101122.
11. Michels, A. (2012). Citizen Participation in Local Policy Making: Design and Democracy. International Journal of Public Administration. 35(4), 285–292.
12. Binder, T., & Brandt, E. (2008). The Design:Lab as platform in participatory design research. *CoDesign*, *4*(2), 115–129.
13. Nabatchi, T., & Leighninger, M. (2015). Citizenship, Outside the Public Square. In *Public Participation for 21st Century Democracy* (pp. 1–12).
14. Sanders, E. B.-N., & Stappers, P. J. (2008). Co-creation and the new landscapes of design. *CoDesign*, *4*(1), 5–18.
15. Pillai, A. G., Ahmadpour, N., Yoo, S., Kocaballi, A. B., Pedell, S., Sermuga Pandian, V. P., & Suleri, S. (2020). Communicate, Critique and Co-create (CCC) Future Technologies through Design Fictions in VR Environment. *Companion Publication of the 2020 ACM Designing Interactive Systems Conference*, 413–416.
16. Paasch Knudsen, S., Husted Hansen, H., & Ørngreen, R. (2022). *Exploring the Learning Potentials of Augmented Reality Through Speculative Design* (pp. 156–163).
17. Gu, N., & Amini Behbahani, P. (2021). A Critical Review of Computational Creativity in Built Environment Design. *Buildings*, *11*(1), 29.
18. Yildirim, E. (2022). Text-to-Image Generation - AI in Architecture. *Art and Architecture: Theory, Practice and Experience*, 97.
19. Kingma, D. P., & Welling, M. (2013). Auto-Encoding Variational Bayes. *CoRR*, *abs/1312.6114*.
20. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, *63*(11), 139–144.
21. Open AI. (2022). *DALL·E 2*. https://openai.com/product/dall-e-2.
22. Stability AI. (2023). *Stable Diffusion Public Release*. https://stability.ai/blog/stable-diffusion-public-release.
23. Google Research. (2022). *Imagen: Text-to-Image Diffusion Models*. https://imagen.research.google/.

24. Gafni, O., Polyak, A., Ashual, O., Sheynin, S., Parikh, D., & Taigman, Y. (2022). Make-A-Scene: Scene-Based Text-to-Image Generation with Human Priors. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *13675 LNCS*, 89–106.
25. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-Resolution Image Synthesis with Latent Diffusion Models. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2022-June*, 10674–10685.
26. Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E., Ghasemipour, S. K. S., Ayan, B. K., Mahdavi, S. S., Lopes, R. G., Salimans, T., Ho, J., Fleet, D. J., & Norouzi, M. (2022). *Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding*.
27. Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M., & Aberman, K. (2022). *DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation*.
28. Ho, J., Jain, A., & Abbeel, P. (2020). Denoising Diffusion Probabilistic Models. *Advances in Neural Information Processing Systems*, *33*, 6840–6851.
29. Dhariwal, P., & Nichol, A. (2021). Diffusion Models Beat GANs on Image Synthesis. *Advances in Neural Information Processing Systems*, *11*, 8780–8794.
30. Gu, S., Chen, D., Bao, J., Wen, F., Zhang, B., Chen, D., Yuan, L., & Guo, B. (2022). Vector Quantized Diffusion Model for Text-to-Image Synthesis. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2022-June*, 10686–10696.
31. Seneviratne, S., Senanayake, D., Rasnayaka, S., Vidanaarachchi, R., & Thompson, J. (2022). *DALLE-URBAN: Capturing the urban design expertise of large text to image transformers*.
32. UrbanistAI. (2023). *UrbanistAI*. https://www.urbanistai.com/.