VRCopilot: Authoring 3D Layouts with Generative AI Models in VR

Lei Zhang
University of Michigan
Ann Arbor, MI, USA
raynez@umich.edu

Jin Pan
University of Michigan
Ann Arbor, MI, USA
jhinpan@umich.edu

Jacob Gettig
University of Michigan
Ann Arbor, MI, USA
jgettig@umich.edu

Steve Oney
University of Michigan
Ann Arbor, MI, USA
soney@umich.edu

Anhong Guo
University of Michigan
Ann Arbor, MI, USA
anhong@umich.edu

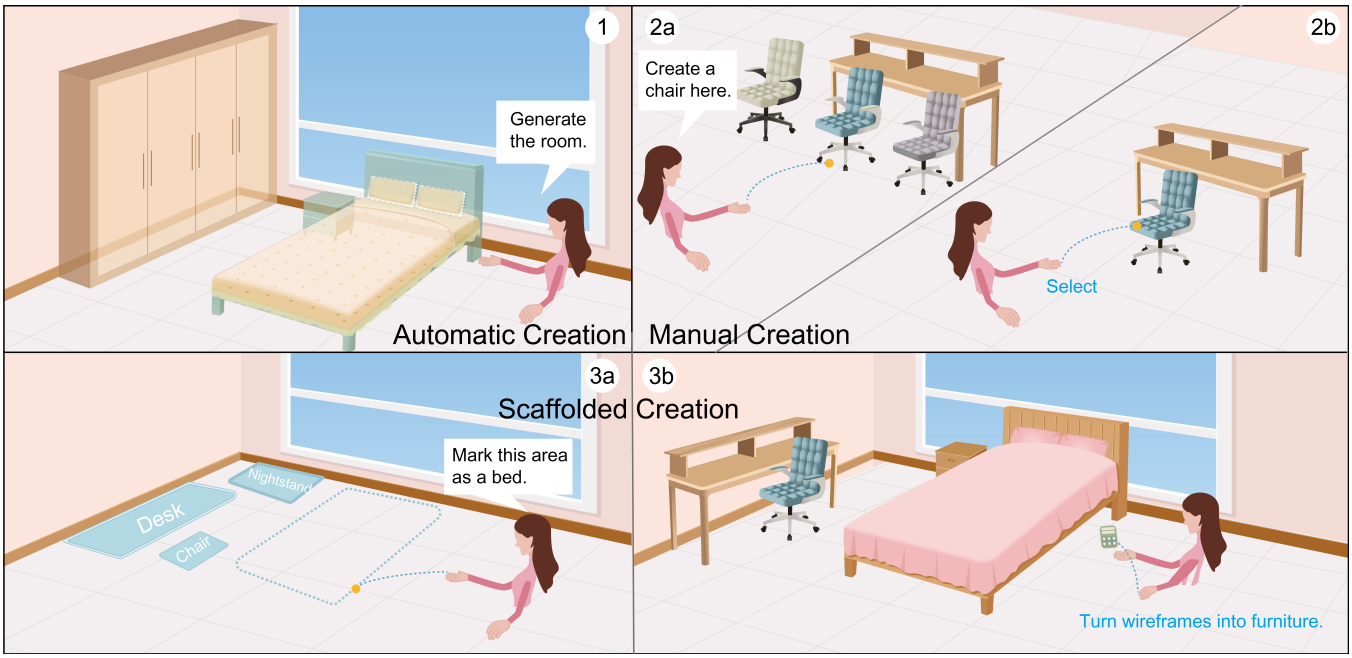


Figure 1: System Overview of VRCopilot. 1) Automatic Creation: Users can use voice commands to ask the generative model to generate a full-room layout based on an empty room. 2) Manual Creation: Users can use multimodal specification by speaking with simultaneous pointing to ask the system to suggest a chair (a); they can select from one of the three suggestions offered by the system (b). 3) Scaffolded Creation: Users can create *wireframes* by drawing on the floor while speaking, in addition to automatically generated wireframes (a); They can then turn the wireframes into specific furniture (b).

ABSTRACT

Immersive authoring provides an intuitive medium for users to create 3D scenes via direct manipulation in Virtual Reality (VR). Recent advances in generative AI have enabled the automatic creation of realistic 3D layouts. However, it is unclear how capabilities of generative AI can be used in immersive authoring to support fluid interactions, user agency, and creativity. We introduce VRCopilot, a mixed-initiative system that integrates pre-trained generative AI models into immersive authoring to facilitate human-AI co-creation in VR. VRCopilot presents multimodal interactions to support rapid prototyping and iterations with AI, and intermediate representations such as wireframes to augment user controllability over the created content. Through a series of user studies, we evaluated the potential and challenges in manual, scaffolded, and automatic creation in immersive authoring. We found that scaffolded creation using wireframes enhanced the user agency compared to automatic creation. We also found that manual creation via multimodal specification offers the highest sense of creativity and agency.

UIST '24, October 13–16, 2024, Pittsburgh, PA, USA

ACM ISBN 979-8-4007-0628-8/24/10…$15.00
https://doi.org/10.1145/3654777.3676451

CCS CONCEPTS

• **Human-centered computing** → **Interactive systems and tools**; **Virtual reality**.

KEYWORDS

Virtual Reality, Generative AI, Human-AI Co-creation

ACM Reference Format:
Lei Zhang, Jin Pan, Jacob Gettig, Steve Oney, and Anhong Guo. 2024. VRCopilot: Authoring 3D Layouts with Generative AI Models in VR. In *The 37th Annual ACM Symposium on User Interface Software and Technology (UIST '24), October 13–16, 2024, Pittsburgh, PA, USA.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3654777.3676451

1 INTRODUCTION

As Virtual Reality (VR) continues to gain momentum across various domains, such as education [67], gaming [56], and spatial design [63], the need for effective tools and techniques to author high-quality 3D scenes becomes increasingly important. Immersive authoring is a paradigm that leverages users' spatial capabilities to enable them to create and evaluate 3D scenes directly while immersed in the virtual environment [1, 18, 23, 30, 41, 63–65]. Prior work in immersive authoring tools has demonstrated benefits of lowering the barrier for end-users with little technical skills to create 3D content [33, 65].

While existing immersive authoring tools make it intuitive for users to visualize their design concepts for 3D scenes in VR, most current 3D layouts such as architectural designs and game scenes are laboriously created through manual placement of 3D models. This manual process is not only tedious and time-consuming, but can also limit the user's ability to explore a diverse range of ideas [31]. In recent years, generative Artificial Intelligence (AI) models have emerged as powerful means for automatic generation of intelligible text [44], photorealistic images [46], videos [2], music [37], and 3D layouts [34, 45, 57]. By leveraging generative models, we can potentially provide users with automatically generated 3D layouts during the process of immersive content creation, enabling users to save time and effort while exploring alternative design possibilities.

Prior work has demonstrated promising results in generating realistic 3D layouts [34, 45, 57] and text-to-layout generation [20, 39]. However, integrating these models into immersive authoring workflows poses unique challenges of how users can collaborate and interact with generative models—specifically, understanding, controlling, and refining model outputs in immersive virtual environments. This difficulty is compounded by generative AI models' well-known issues with transparency, controllability, and user agency. Current generative models for 3D layouts use either room sizes (e.g., [45]) or text captions (e.g., [20]) as prompts. It is difficult for users to define their design objectives, such as requesting layout designs with elements in particular locations or sizes (as seen in Fig. 1.3.b).

In this paper, we introduce VRCopilot, a mixed-initiative system that integrates pre-trained generative models into immersive authoring workflows. VRCopilot is instantiated in the context of layout design for indoor scenes, where users are able to co-create with generative models via requesting, controlling, and refining generative models' outputs in VR. VRCopilot introduces two key interaction techniques: *(1) multimodal specification* and *(2) intermediate representation.* Inspired by multimodal interactions such as "Put-that-there" [4], our system enables users to use speech and simultaneous pointing to specify their creation needs, increasing the naturalness and economy of language description in the immersive environments. For instance, users can point to a location in the room while saying "create a wooden chair here." As a response, the system will offer three options for the user to choose from. Besides, to help users co-create with the generative model in a more transparent and controllable way, VRCopilot proposes the notion of *wireframes* as intermediate representations for the generated outcomes. Inspired by the concept of low-fidelity prototyping in Human-Computer Interaction [7, 47, 50], wireframes are 2D representations of 3D layouts similar to floor plans in interior design. These representations can be hand-drawn by users together with speech specifying the their types, or suggested by generative models. VRCopilot allows users to iteratively refine the design with generative AI by enabling them to convert between intermediate representations and 3D layouts.

Taking the above techniques together, we propose three ways of human-AI co-creation in VR enabled by VRCopilot: (1) manual creation, where users create individual objects to complete a layout design via a catalog menu and multimodal specification; (2) automatic creation, where users request and refine suggestions from generative AI for full-room layouts; and (3) scaffolded creation, where users co-create intermediate representations with generative AI for guiding the final layout design.

To provide an in-depth understanding of the human-AI co-creation process in VR, we conducted two rounds of user studies. Our first study aimed to compare user experiences of creating 3D layouts with and without AI. Specifically, we compared creation without AI using manual placement and creation with AI using generative models. We found that co-creating 3D layouts with generative models is generally more preferable as it could save users' effort while resulting in 3D layouts with more complete *functionality* and diverse *color palette*. However, users struggled with the generative model's non-deterministic output, where the generated results might misalign with the user's design goals due to the lack of controllability of the generative model.

Based on the insights and challenges from the first study, we further evaluated VRCopilot by comparing different levels of AI automation in the creation process including manual creation, automatic creation, and scaffolded creation. We found that users' sense of agency significantly increases in the order of automatic creation, scaffolded creation, and manual creation. Specifically, the design of wireframes in scaffolded creation enhances users' agency by allowing them to define the 3D layout including object types and sizes compared to automatic creation. Manual creation offers the highest agency by enabling additional visualization and control over object styles. We also found that users felt significantly more creative in manual creation, than in scaffolded or automatic creation, with no significant difference found between the latter two. Specifically, having multiple suggestions via multimodal specification in manual creation can make users feel more creative. Users felt less creative in the other two conditions since AI generated outcomes could lead to fixation and prohibit users from creative exploration.

In sum, our paper makes the following contributions: 1) VRCopilot, an immersive authoring system that enables users to interact and co-create with generative AI models in virtual immersive environments; and 2) empirical results gained from two user studies that provide insights on user experiences such as perceived agency and creativity, as well as potential and challenges of human-AI co-creation in immersive authoring workflows.

2 RELATED WORK

VRCopilot draws inspiration from prior literature on 3D scene synthesis using generative models, creativity support with generative design, and interactive interfaces with computational agents.

2.1 Generative Models for 3D Scenes

The demand for automatically generating 3D scenes has never been higher in the domain of gaming, AR & VR, architecture and interior design. In the field of computer vision, this topic named 3D scene synthesis is gaining popularity and prior researchers have explored generating new 3D scenes via various input including images [22, 36], text [16, 62], or room shape [45]. A key line of work is 3D *indoor* scene synthesis, which refers to the task of automatically generating a set of 3D furniture objects along with their positions and orientations, given a room layout [66]. Some of the early work in this space offered suggestions using hardware-accelerated Monte Carlo sampler based on interior design guidelines [40]. Follow up work has been focused on data-driven approaches, given the rise of large 3D object datasets such as SUNCG [51] and 3D-FRONT [21]. The data-driven approaches can be approximately categorized into graph-based [57] and autoregression-based approaches [45, 48, 58, 59]. Graph-based approaches encode 3D layouts as scene graphs, where objects are nodes, and the spatial relationship between objects are edges. This method treats the task of generating 3D scenes as generating directional graphs. The main motivation behind this is to process it with graph convolutional networks. Most notably, Ritchie et al. [48] introduced a CNN-based architecture that operates on a top-down image-based representation of a scene and inserts objects in it sequentially by predicting their category, location, orientation, and size. More recently, autoregression-based approaches have been introduced. Wang et al. introduced SceneFormer [59], a series of transformers that autoregressively add objects in a scene. ATISS [45] simplifies the process by proposing a single model trained end-to-end to predict all attributes. Most notably, ATISS encodes 3D objects' positions, rotations, and scales in transformers for training. More recently, DiffuScene utilizes a denoising diffusion model that is able to generate more plausible and diverse indoor scenes [54].

Our work contributes to the existing literature on 3D scene synthesis by introducing generative models into immersive environments. While prior work has been focused on generating realistic 3D layouts, VRCopilot aims to integrate state-of-the-art generative AI models into immersive authoring and explores the ways of co-creating with generative AI models in VR.

2.2 Creativity Support via Steering Generative Models

The acceleration of AI capabilities has enabled human-AI co-creation in domains such as drawing [15, 19], creative writing [13], video game content creation [25], and music composition [28, 37]. For example, Bach Doodle [28] is able to complete a music composition in the style of J.S. Bach by requiring users to only write a few notes. While recent research has focused on building co-creation experiences in 2D interfaces, there has been relatively little HCI work examining how to design interactions with these state-of-the-art generative models to ensure they are effective for co-creation in the immersive environments. Our research contributes an understanding of how interactions with these AI models can be designed, how they affect the immersive authoring experience, and users' attitudes towards AI co-creation in VR.

Integrating existing generative AI models into creative work presents unique challenges in itself such as adapting actions of AI based on users' preferences [12, 32, 53]. Research has also observed that users desire to take initiative in their partnership with AI, and thus sought to provide steering tools to make AI align with users' creative goals. For example, TaleBrush [12] uses a combination of line sketching and natural language narration to create stories. DreamSketch [32] uses sketches as input for the generative design of 3D models. In the domain of 2D layouts, Scout [53] uses high-level constraints based on design concepts to generate multiple designs. Building on this need, our work investigates how users express their preferences to generative AI through multimodal specification and intermediate representations in VR.

2.3 Interaction Techniques in Immersive Environments

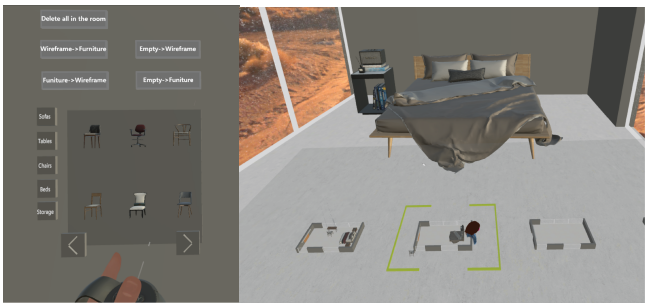
Our proposed interactions are inspired by prior interaction techniques in immersive environments including multimodal interaction, spatial interaction, and world in miniature (WIM). While recently there has been extensive exploration in using natural language interactions with generative AI models to build virtual scenes, using just natural languages alone might be effective for tasks such as referencing [8] in immersive environments. Building on this line of work, our system demonstrates how multimodal interaction can be used for specifying objects to generative AI in the immersive authoring process. Finally, Stoakley et al. introduced the concept of World in Miniature (WIM), which enables both navigation and interaction in a large VR scene [52]. A WIM represents the virtual environment and allows users to manipulate objects offered by the miniature, or rapidly teleport in the virtual environment by selecting locations directly in the miniature. It also has the benefit of allowing users to see a preview of the immersive virtual environment without having to travel back and forth between different views. We built on the WIM technique to enalbe users to design and edit multiple variations of the 3D layouts.

3 VRCOPILOT

VRCopilot is a mixed-initiative immersive authoring system that enables users to co-create 3D layouts with pre-trained generative models in VR. Users can ask generative models to generate full-room layouts or use multimodal specification to create individual objects. They can also manually place objects from a catalog menu or request suggestions from the system using multimodal interactions (i.e., pointing and speaking). VRCopilot further allows users to create *wireframes* — intermediate representations that help guide and refine the layout generation process. We detail our system design in the sections below.

3.1 Scope

We situate our design of VRCopilot in the context of interior design tasks, where users can place pre-made 3D furniture models in a virtual apartment. Interior design requires balancing constraints



(a) Palette Menu. (b) Multi-workspace.

Figure 2: User interfaces in VRCopilot including (a) a palette menu where users can select and furniture from the catalog, and (b) a multi-workspace visualization that allows users to work and switch between multiple versions.

(e.g., functional requirements and space limitations) with aesthetic preferences. It has been the application domain of many prior immersive authoring tools [10, 29, 63] and is a common use case for Mixed Reality. For example, several popular home goods stores, including IKEA[1], integrate features that allow customers to virtually preview furniture arrangements in their own homes before making a purchase. VRCopilot includes 7,302 furniture models from 3D-FRONT [21], a large open-source dataset of furniture objects and textures.

Designing VRCopilot for interior design allows us to evaluate it in a realistic domain with demonstrated utility. However, we believe many of the concepts behind our design could generalize to other spatial design tasks, as the low-level tasks (e.g., object instantiation, customization, and manipulation) and multimodal interactions with generative models are broadly applicable across domains.

3.2 Immersive Authoring Features

VRCopilot is designed as an *immersive authoring* tool, meaning users design a room layout while immersed in that room (in VR).

3.2.1 Importing Models. Users can manually import furniture models into the virtual environment from a catalog menu bound to their non-dominant hand, as seen in Fig. 2a. Using the catalog menu, users are able to choose from different categories of furniture such as tables and chairs from a sub-menu. Each page on the catalog menu contains six furniture items and they can turn pages to navigate more items. After a furniture model is imported, users can manipulate and place the model via direct manipulation using the VR controllers.

3.2.2 Design Exploration. The ability to explore multiple alternatives is crucial to supporting creativity in design tasks [27, 49]. For example, in the realm of interior design, designers typically develop a variety of versions to present to clients or stakeholders. To facilitate the exploration of multiple design variations, VRCopilot offers multiple empty workspaces or templates for users to work on (as seen in Fig. 2b). Users can easily switch between workspaces to work on different versions by navigating a list of miniatures in VR. This creativity support is inspired by the concept of "World in Miniature" (WIM) [52] and recent work on version control in VR [63]. To help users reuse partial layouts across different versions of the designs, VRCopilot also includes a copy & paste feature, shown as additional buttons bound to the handheld menu. This feature allow users to copy multiple objects and paste them either in the same workspace or other workspaces.

3.3 Generative Model in VRCopilot

We used ATISS [45], an open-source generative model for indoor scene synthesis using autoregressive transformers, as our generative model. ATISS is trained using an open-source dataset of 3D models called 3D-FRONT [21], from which we also build our furniture catalog. This model takes room dimensions parameters as prompts and generates reasonable furniture arrangements of the full-room layout. It is also versatile for user inputs such as asking for a suggested placement of a given furniture item, or asking for a suggested furniture item for a given location. We chose this model because it has been used as baseline models for work in the domain of indoor scene synthesis (e.g., [20, 60]).

We integrated ATISS in VRCopilot and can generate suggestions for full-room layouts on demand. In VRCopilot, users can access the generative model via either voice commands or the catalog menu (as shown in Fig. 2). Upon receiving the request, our system can fill the entire room by placing suggested furniture in the user's current workspace. Users can delete the suggestions and also run the generative model repeatedly. Our system supports multiple room sizes, shapes, and types (e.g., bedrooms and living rooms), and can be easily extended to support arbitrary room sizes and shapes (e.g., users can draw the room) and the backend generative model can adapt to these specifications automatically.

3.4 Multimodal Specification

Existing immersive authoring tools enable users to directly manipulate virtual objects similar to how they would manipulate them in the physical world. However, direct manipulation does not suffice for all needs during immersive authoring. One clear weakness of direct manipulation is that it makes it difficult to identify or manipulate a potentially large sets of objects. For example, there is a massive number of objects and styles in our furniture catalog (e.g., the catalog is based on 3D-FRONT that contains 7,302 furniture items with textures). It is difficult for users to specify generating a chair with *minimalist* style via direct manipulation. On the other hand, the inherent ambiguity of natural language instructions makes it difficult to use pronominal reference to objects in the scene [14]. For example, it is hard for the system to understand which location the user is referring to when the user describes "generate a chair here" without additional contextual information.

Inspired by multimodal specifications in graphical interfaces such as "Put-that-there" [4], VRCopilot allows users to use speech and simultaneous pointing to specify their creation needs, increasing the naturalness and utility of language description in the immersive environments. Our system can process users' natural language voice commands and categorize them into several possible intents:

- *Object Generation*: generating individual objects;
- *Object Regeneration*: regenerate individual objects;

[1] https://www.ikea.com/global/en/newsroom/innovation/ikea-launches-ikea-place-a-new-app-that-allows-people-to-virtually-place-furniture-in-their-home-170912/

- *Object Duplication*: duplicate selected objects;
- *Scene Completion*: initiating a request for generating the full-room layout;
- *Wireframe Generation*: initiating a request for generating the (see specifics in Section 3.5);
- *Wireframe Labelling*: assigning an object type to a wireframe (see specifics in Section 3.5);
- *Deletion*: delete selected objects.

Users can point to any location in the scene while verbally requesting that VRCopilot suggest furniture to be placed at the designated point. In their specifications, users can use pointing to specify the *location* and voice to specify the *object type* and its *style* and *material* (e.g., "Generate a minimalist wooden chair here"), as seen in Fig. 3. Since its furniture catalog is built on 3D-FRONT, VRCopilot contains a limited number of 21 object types such as beds and chairs, 19 unique styles such as Modern and Japanese, and 15 unique materials such as Wood and Metal. Our system's language processing currently ignores out-of-range intents, such as indication of color or shape, due to the repository's limitation of not supporting color or shape labels. For example, when users indicated a *red* chair, the system would retrieve a chair of any color. VRCopilot also does not include other natural language intents such as repositioning and object selection, as these operations can be more easily achieved via direct manipulation as observed in our pilot tests. Upon parsing the user's request, three suggested furniture items fitting the user's provided criteria are visualized in front of the user (as seen in Fig. 3), and the user can choose one of the three to become a part of the scene (as seen in Fig. 3).

3.5 Intermediate Representation

One of the key challenges of human-AI co-creation is the lack of transparency, control, and user agency [3, 6]. To help users co-create with the generative model in a more transparent and controllable way, VRCopilot proposes the notion of *wireframes* that is used as intermediate representations for the generated outcomes. We took inspirations from low-fidelity prototyping that is commonly used in Human-Computer Interaction [7, 47, 50]. For example, prior work has explored using low-fidelity prototypes such as paper prototypes to quickly scaffold user interface design [50] or Play-Doh as intermediate representations to represent high-quality 3D models [42]. In VRCopilot, wireframes are designed as 2D representations of 3D layouts such as floor plans in interior design. These representations can be hand-drawn by users together with speech specifying their types. For instance, users can use the cursor of the raycast from the controller as the pen tip. They can place the cursor on the floor and start drawing by pressing a button on the controller, while saying "Mark this area as a bed." Upon the intent is recognized, the system will normalize the drawing into a rectangular plane with a text label of the object type (e.g. "Bed") attached to it. Users can further adjust the placement and dimension of the wireframe using direct manipulation, similar to manipulating furniture models. Users can build up intermediate representation of the full-room layout design by creating multiple wireframes in the room. Alternatively, users can ask the generative model to offer suggestions of wireframes by initiating a request similar to generating full-room layout. The system can then generate the intermediate representation of the full-room layout and visualize all generated wireframes in the room.

In addition, VRCopilot allows users to iteratively refine the design with generative AI by enabling them to convert between intermediate representations and 3D layouts. For example, users can use voice commands or button presses to turn their intermediate representations into actual furniture models. The system can then interpret the labels and populates the scene with detailed furniture pieces corresponding to the *object type*, *size*, and *orientation* as specified using each wireframe. For objects that are not placed on the floor, such as ceiling lamps, users can draw wireframes on the floor similarly to how they create other objects. The system will then automatically set the y attribute, representing the height, for these objects when populating the scene. Users can also switch back to intermediate representations from detailed furniture design, enabling an iterative design that leverages both lo-fi and hi-fi representations of the layout.

3.6 Ways of Human-AI Co-creation in VR

With the above generative model and interaction techniques, VRCopilot supports three ways of human-AI co-creation in VR.

3.6.1 Manual Creation. Manual creation enables users to manually create a 3D layout design by creating each furniture item and its placement one after another. Such creation method uses a bottom-up approach, where users start by creating specific furniture items either via the catalog menu or multimodal specification. Once the central pieces are selected (e.g., beds, sofas), users consider how other components can be arranged within the room including placement of furniture, the flow of circulation, and how spaces will be utilized.

Figure 3 1a-e showcases a typical workflow of how users can create layout designs using manual creation: a) Users first use multimodal specification to ask the system to generate a bed. b) They can then pick from one of the suggestions as a central piece in the room. c) Users use multimodal specification to create a nightstand, and d) pick the one that best matches the style of the bed. e) Then they start using the catalog menu to create other furniture models such as desks to complete the layout design.

3.6.2 Automatic Creation. Automatic creation enables users to ask the generative model to generate full-room layouts. After the suggested layout is given, users can modify the layout design based on their design goals. This could include adjusting the placement of objects to avoid overlapping objects or to remove unwanted objects.

Figure 3 2a-e demonstrates a typical workflow of how users can create layout designs using automatic creation: a) Users start out with no concrete design goals and ideas so decide to ask the system to generate the full-room layout for them. b) After the system processes the request and visualize the layout suggestion to the user, c) they can manipulate furniture models such as deleting the wardrobe to fit their own preferences. d) They decide to move the bed to be further away from the window. e) They also decide to add additional furniture models from the catalog menu that are missing from the generated layout such as additional desks.

3.6.3 Scaffolded Creation. Scaffolded creation enables users to create intermediate representations, i.e., wireframes, to scaffold their

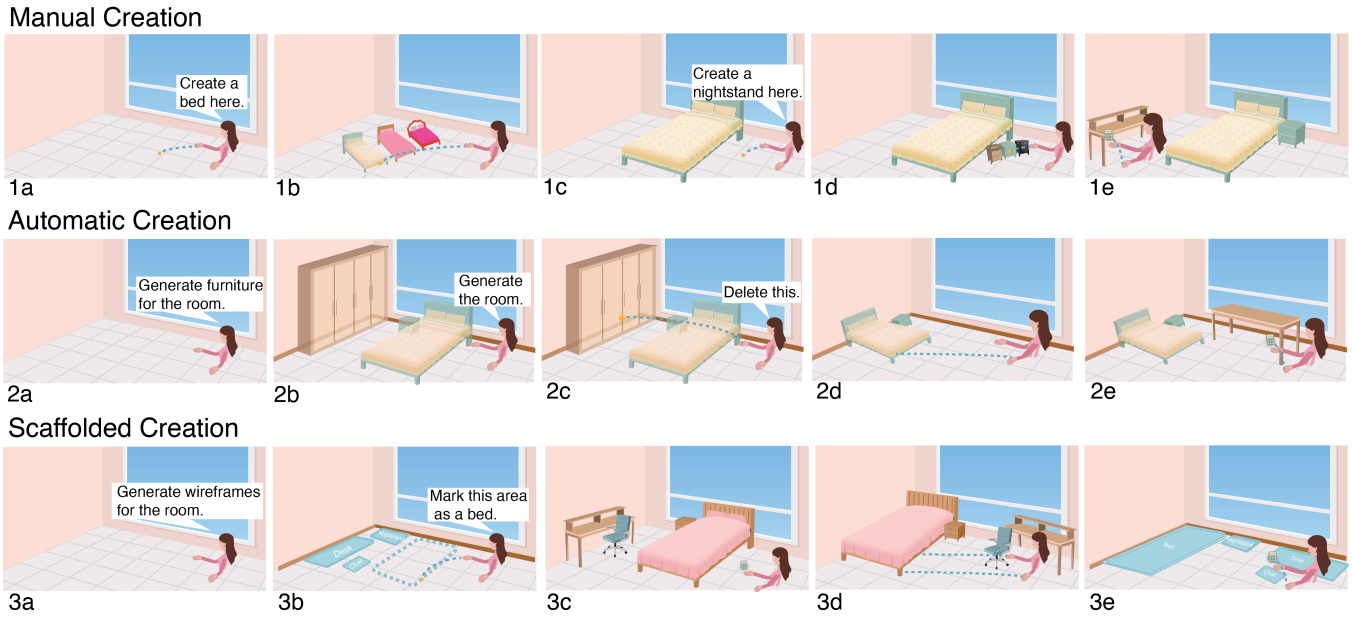


Figure 3: VRCopilot proposes three ways of human-AI co-creation in virtual immersive environments: manual creation (1a-e), automatic creation (2a-e), and scaffolded creation (3a-e).

designs. Such a creation method uses a top-down approach where users begin with a broad, overarching vision of the floor plan by creating wireframes in the immersive environments. They can draw their own wireframes and ask for generated wireframes. They can also modify the placements and sizes of wireframes, and convert between wireframes and furniture layouts.

Figure 3 3a-e a typical workflow of how users can iteratively create layout designs using scaffolded creation: a) Users first ask for generated wireframes from the system. b) Upon getting the results from the generative models, they can draw their own wireframes such as a bed and rearrange the wireframes. c) They can turn the wireframes into furniture layouts via a button press. d) They can then manipulate the furniture models to further fine-tune the design. e) Once they want to explore an alternative design, they can switch back to wireframes for generating another layout option.

3.7 Implementation

VRCopilot is developed using Unity (version 2021.3.20f1) and integrates plugins from Meta Oculus and the Microsoft Mixed Reality Toolkit (MRTK), enabling operation on Meta Quest and Rift VR headsets. The application incorporates advanced voice recognition and response capabilities through integration with the ChatGPT Audio Model (whisper-1) and Chat Model (gpt-4-turbo), with the latter hosted on a dedicated GPU server equipped with an Nvidia RTX 4090 graphics card. A comprehensive system architecture is depicted in Figure 4.

3.7.1 Integration with ChatGPT Models. Interaction with ChatGPT models is facilitated through voice commands. The system captures user voice input via the microphone, converting the audio to an .mp3 format. This file is then translated into text by the ChatGPT SpeechToText model (whisper-1) through an HTTP request. The resulting text is processed by the ChatGPT Chat Model (gpt-4-turbo), which identifies the user's intent from the predefined categories and extracts relevant parameters such as furniture styles or categories. The responses, formatted as JSON, are parsed by the Unity client to execute the corresponding actions. While most actions are deterministic, actions requiring the generation of new items (e.g., "generate a chair in a modern style") involve a selection process from a set of items meeting the specified criteria.

3.7.2 Communication with the Generative Model. For tasks that involve the generation of new furniture, VRCopilot employs socket communication with a generative AI model, ATISS. Furniture attributes (unique ID, position, rotation, scale) are encoded in JSON and sent to the server. Upon completion, the server returns a JSON response with the furniture items that meet the established criteria, which the Unity client then processes and renders in the virtual environment.

3.7.3 Multimodal Feedback Module. To enhance user interaction, VRCopilot integrates a feedback loop through AWS Polly Text-ToSpeech model. After processing an intent, the system generates textual feedback corresponding to the user's request, which is then converted into speech. This multimodal feedback mechanism provides real-time auditory confirmation of actions taken within the virtual environment, enriching the user experience.

4 USER STUDY 1

To understand the effectiveness and challenges of co-creating with generative AI in immersive environments, we fist sought to compare immersive authoring with and without AI. Prior research has provided some insights on how people collaborate with generative

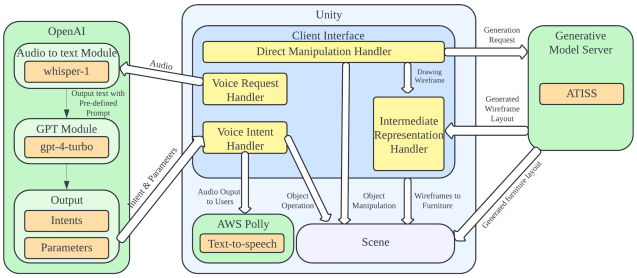


Figure 4: The system architecture of VRCopilot.

AI in creative domains (e.g. music [37] and painting [11]). We extend this line of work by understanding people's behaviors and attitudes when working with generative AI in virtual immersive environments. We conducted a qualitative comparison study between two conditions: 1) immersive authoring using the conventional interfaces (e.g., via direct manipulation and menu selection), 2) immersive authoring with conventional interfaces and generative AI models. We use this study as the first stop to eliciting the challenges that users perceived when co-creating with AI in VR.

4.1 Participants

We recruited 14 participants (10 women and 4 men, age 20—28) from a university through public email lists. All participants had prior experience using VR devices and were compensated with $30 USD Amazon gift cards for two hours of their time.

4.2 Procedure

During the study, users were first given a tutorial of the system that covered individual features of the system including the control of direct manipulation and the usage of the generative model. The tutorial lasted about 30 minutes. Then, participants were asked to design an empty apartment, consisting of two bedrooms and one living room, under two conditions: 1) with conventional immersive authoring interface, 2) with the conventional interface and the generative AI model. The room sizes and types were pre-configured, in order to encourage participants to focus on the co-creation process. The order of the conditions was counterbalanced. Each condition took about 15 minutes to complete. In each condition, participants were asked to aim for finishing three versions of the apartment with at least three items in each room. This instruction was not a strict requirement, but rather a means to encourage participants to design multiple variations of the apartment. After both conditions were finished, we conducted a retrospective interview with participants. Our study protocol was approved by our institution's IRB.

4.3 Analysis

We transcribed and conducted a thematic analysis [5] of the interview data. To assess the creation results from participants, we designed an evaluation that elicits emerging patterns of users' creation through an evaluation workshop. One design expert, a full-time architect with 2 years of working experience, was invited to participate an evaluation workshop with one experimenter that took 90 minutes. The evaluation workshop was held remotely where the experimenters screen shared to the expert. The expert then went through the top down images of the creation results from all participants under each condition. The order of showing the creation results is completely randomized and the expert was not informed of how the design were created under each condition. Then the expert were asked to use an inductive approach to observe the top down images under each condition and use open-coding to elicit emerging patterns in each condition.

4.4 Results and Insights

Below are the insights gained from the qualitative user study and the expert evaluation:

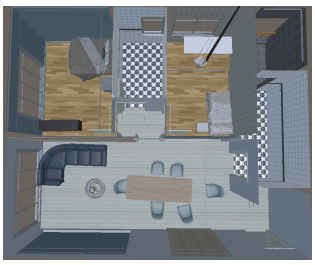
Generative models provide less user agency. Agency refers to the awareness and control over one's action and their results [61]. We found that participants reported feeling less agency over the creation results when co-creating with AI. While the generative AI models could make meaningful layout suggestions that help users explore different ideas, the generated suggestions sometimes misaligned with users preferences in terms of the functionality and other considerations of the layout design. For instance, P2 said "I really have no idea what was going to come out when I did it [using the generative model], like I did not at all expect a bookcase in the middle of the living room, even if it would make sense for that room." P7 also commented on their agency when comparing creating with and without AI: "When there's no AI intervention in the process it is just me thinking about what is the best circulations? What is the best looking furniture to be placed in the room? Those are my primary concern when I was doing it. So I will say I was the most in control when I was doing the first task [without generative AI models]."

Generative models are useful for sparking different ideas. One key dimension of creativity is the ability to explore different ideas [9]. Participants reported that the results from generative AI models can provide inspirations for the layout design that they did not come up with. For example, P8 commented that "It brings up new ideas I hadn't thought about before... also because when I first create the room, I pretty much put in what my favorite idea is for so when I create or generate a new room, it adds more inspiration than what I already had started off with."

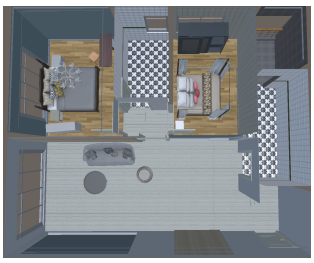
Creating with generative models can lead to more diverse functionality and color palette. Functionality refers to the ability of a space or its components to serve a specific purpose or function effectively and efficiently. We found that creation results with the help of AI encompass more diverse functionalities. Specifically, the expert observed more diverse object types in each room that can support different activities. For example, the bedrooms shown in both Fig. 5c and Fig. 5d include desks (for working), wardrobes (for clothes), and bookshelves (for storage). Color palette refers to the selection of colors used in a design, including primary, secondary, and accent colors, which contribute to the overall mood and atmosphere of a space. We found that creation results generated with AI generally have a richer color palette (seen in Fig. 5), which contributes to the expert commenting the creation "more exciting."

Creating with generative models can lead to poor considerations of circulation and daylighting. Circulation refers to the flow or movement of people within a space. It encompasses the pathways, routes, and patterns that individuals follow as they navigate and move

(a) Creation Results without generative AI (P7).



(b) Creation Results without generative AI (P10).



(c) Creation Results with generative AI. Two bedrooms are created with generative AI while the living room is unfinished (P6).

(d) Creation Results with generative AI. All rooms are created with generative AI (P13).

Figure 5: Exemplary top-down view comparison of participants' creation results with and without the assistance of generative AI in Study 1.

through an interior environment. We found that creation results generated with AI generally have a poorer circulation. For example, one of the bedrooms shown in Fig. 5c includes a nightstand that is blocking the doorway. The dining table in the living/dining room in Fig. 5d does not allow for much movement between the two sides due to its close placement to the walls. This is because our underlying generative model (i.e., ATISS) that we utilize does not take doorway or room of movements into consideration when generating. Daylighting in interior design is a design strategy that focuses on harnessing and optimizing natural daylight to illuminate interior spaces. We found that creation results generated with AI generally have a poorer consideration of daylighting. For example, both bedrooms shown in Fig. 5c have furniture blocking the windows, making it difficult to harness daylight. This is due to the underlying generative model (i.e., ATISS) that we utilize does not take window placement, size and shape into consideration.

Based on these findings around using generative models, our research team investigated further in the second round of study, that was specifically focused on the mitigation of the issue of user agency and the comparison across different ways of human-AI co-creation (as described in Section 3.6). We were also able to design tasks for the second study based on the patterns drawn from the expert evaluation session to further develop our ideas. We describe the second user study in the following section.

5 USER STUDY 2

We conducted a second user study to compare three conditions: 1) manual creation using catalog menus and multimodal specification, 2) scaffolded creation using wireframes, and 3) automatic creation using generative AI. We aimed to compare user perceived effort, creativity, and agency, and to elicit potential and challenges that users perceived when co-creating with AI in VR.

5.1 Participants

We recruited another 15 participants (5 women and 10 men, ages 19—26) through university email lists. All participants had prior experience using VR devices and did not participate in the previous study (Section 4). We labeled the participants as P15-29 below. Each participant was compensated with a $30 USD Amazon gift card for two hours of their time.

5.2 Procedure

We designed a within-subject study where each participant experienced all three conditions during the study. Balanced Latin-Square was used to determine the order of the conditions for each participant. For the study setup, we used the Meta Quest 2 connected to a laptop that was running our system in Unity 3D game engine. Each study session began with an introduction of the study and a tutorial of the system that lasted about 30 minutes. During the introduction, participants were introduced to the study and were informed of all the data that would be collected during the study. Participants were then given a tutorial of individual features of VRCopilot. They were given an atomic task after learning each feature to familiarize themselves with the system.

After the tutorial, participants performed a design task under each condition, where they furnished an empty bedroom in VR. In each condition, they were asked to come up with three design solutions for the same room within 15 minutes. If more than three versions were created, they would be asked to turn in the three versions that they were most satisfied with. The following design goals were given to participants for each condition:

- There should be at least 4 furniture types in the bedroom.
- Make sure the top of the window is not blocked by wardrobes / shelves / bookcases.
- There should be enough space for users to navigate in the room.
- There should be a sofa to accommodate seating.
- Try to make the three versions different in both layouts and appearance.

The design goals were created based on design considerations drawn from the expert evaluation in the previous study (Section 4) including functionality, day-lighting, navigation, seating, and diversity. Participants were encouraged to design multiple variations of the room based on the design goals. They were notified every five minutes during the task. They were also free to ask for time remaining as well as clarification questions related to the task or the system. However, experimenters were not allowed to give any instructions on how to design the room. After finishing each task, participants were asked to fill out a questionnaire while we saved the resulted scenes and the screen recordings from that condition.

After experiencing all conditions, we conducted a semi-structured interview with participants to ask about user experience and their perceptions of each condition. Each interview lasted about 20 minutes. Our study protocol was approved by our institution's IRB.

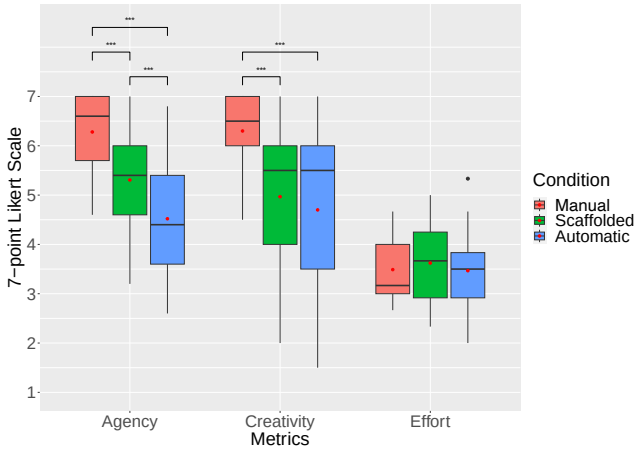


Figure 6: Results from post-task survey comparing three conditions in Study 2.

5.3 Measures and Analysis

We evaluated the following metrics via post-task surveys. Participants rated the items on a 7-point Likert scale (1=Strongly Disagree, to 7=Strongly Agree). To answer the research questions, we measured the following aspects: (1) user perceived **effort** via the NASA-TLX [26] questionnaire; (2) user perceived **creativity** from the Creativity Support Index [9], with an emphasis on how well our system help users explore different ideas; and (3) user perceived **agency** adapted from prior work (e.g., Tapal et al. [55] and Lukoff et al. [38]).

We first performed a Friedman test, to analyze the non-parametric within-subject survey data. For the potential post-hoc analyses, we conducted pairwise comparisons using Conover's test. For qualitative results, we transcribed and conducted a thematic analysis [5] of the interview data.

5.4 Quantitative Results

The results of the post-task questionnaire and the Conover's test are aggregated and shown in summarized in Figure 6.

A Friedman test was conducted to evaluate differences in participants' perceived agency across three conditions. The analysis revealed a statistically significant difference in the sense of agency across the three conditions ($\chi^2(2) = 20.11, df = 2, p < .001$). Post-hoc analyses with Conover's pairwise comparisons were performed with a Bonferroni correction. We found that participants' perceived control was significantly higher in the manual creation condition ($p < .001$ compared to the scaffolded creation condition and $p < .001$ compared to the automatic creation condition). We also found that participants' perceived control was significantly higher in the scaffolded condition compared to the automatic condition ($p < .001$). These results suggest that the design of wireframes was effective in increasing users' sense of control compared to fully automatic generation from AI.

In regards to users' perception of creativity, we found a significant effect of ways of creation on the sense of creativity ($\chi^2(2) = 17.633, df = 2, p < .001$). Further post-hoc analysis revealed that users felt significantly higher sense of creativity in the manual condition compared to both the scaffolded condition ($p < .001$) and the automatic condition ($p < .001$), with no significant difference found between the latter two. These findings show that participants felt they were the most creative when they were working in the manual creation condition, but similarly creative in the scaffolded and automatic creation condition.

For users' perceived effort, we did not find a significant effect of ways of creation on users' perceived effort ($\chi^2(2) = 0.915, df = 2, p = 0.63$). This indicates that users felt similar levels of effort across three conditions.

5.5 Qualitative Results

All participants were able to finish three design variants of the room by the end of the task. To further investigate the reasons behind users' perceptions in subjects such as agency and creativity, we analyzed our interview data and solicited users' qualitative feedback. Overall, our results suggest that users' sense of agency can be enhanced by offering greater control during the human-AI co-creation process, such as control over object types and sizes through wireframes or object styles through multimodal specification. In addition, providing multiple suggestions via multimodal specification can increase users' creativity. We center our findings below around the topics of user agency and creativity in the contexts of human-AI co-creation, interior design, and immersive authoring.

5.5.1 Offering greater control could enhance the sense of agency and ownership. Participants reported having higher agency over the created content in scaffolded creation compared to in automatic creation. Specifically, participants felt that they have control over aspects such as furniture size and placement compared to automatic creation.

> "I felt like I had the most control with the wireframe because in addition to what types of furniture I could also decide what size and how it's positioned. Whereas the others, I think, particularly lost out on the sizing component. Because there were a few times, after I tried the wireframe, that I did try to resize furniture, but then realized that that wouldn't work [in the other conditions]." -P23

In addition, participants reported having higher agency in manual creation since they have additional control over the furniture styles.

> "For example, I can pick, only leather chairs, leather sofas, and then have a bed that matches that style... you just got more control over the style itself, rather than just the layout" -P21

We also found that users felt the least agency in the automatic creation since the generated furniture is already fleshed out and decreases their willingness to control or manipulate things, which further decreased the sense of ownership in the created content.

> "Because it feels like that's already there. So it looks like it already looks pretty good. So I wouldn't want to move it too much, and definitely I have less control with it. Because the furniture and everything were chosen by AI, I feel like it doesn't feel fully like *I* designed it." -P28

5.5.2 Manual creation can spark creativity via multiple suggested options. We found that participants felt the most creative in the manual creation, mostly because the multiple suggestions offered via multimodal specification can give users inspirations. Having the options from the system could inspire participants to keep building the room centering around the piece that they chose from the options.

> "When I saw the bed [from the three suggestions], and it's like bright green, yellow, I was like, 'maybe I can make this the theme of this room.' And I was trying to go with this style when I was choosing the other furniture. Then when I saw a bunk bed, I was like, 'maybe this could be a bunk bed for two children,' and I'm styling the room in that way.... I think the [Manual Creation] condition facilitates creativity a bit more just because you can choose between the three options." -P24

5.5.3 Users tend to follow the designs generated in scaffolded or automatic creation, leading to a reduced sense of creativity. While the scaffolded creation and the automatic creation also suggest furniture to the user, participants reported less creativity mostly because they tended to follow the layout that the generative model suggests to them. We found that users tended to feel fixation when the system generated the full-room layout compared to the system suggested individual furniture items.

> "I think having everything laid out for you already, it decreases your creativity. Because you'll have that bias towards the way that it just puts everything. So it's like the bed's here, I might just keep it there." -P21

5.5.4 Scaffolded creation enables high-level and unbiased design thinking. Participants mentioned that scaffolded creation, specifically the design of wireframe, allowed them to focus on the functionality over the styles and enabled them to think "in the layout sense" (P24).

> "I feel like, by creating all those wireframes, I'm actually doing the job of an interior designer, because I'm not the one who's purchasing the actual furniture for the household. I'm just designing how to maximize the utility of the whole space for this household." -P20

> "I think just where things are and how you move around the room. I think that's very important...if the room is cramped or awkward, it's not going to be as good even if it looks really nice. So wireframe, I think, is very good for that just to see it completely unbiased. Because if I just build using the voice or the menu, I can already see things. Like if it looks good, but it's not really functional, I might be biased just because it looks good, and just go with something that doesn't really work. But wireframe kind of takes away from that. And it really lets me focus on the function. And just making sure that everything flows together nicely." -P19

5.5.5 The design of wireframes in scaffolded creation enables easy manipulation in VR. We found that the design of wireframes in scaffolded creation made it easier for users to navigate the layout and manipulate distant objects due to the reduced occlusion of 2D planes, compared to handling a full layout with 3D furniture.

> "I think it [wireframe] is useful in, getting through the layout, because with all the objects already in the environment, it's been hard to see around and if there's something behind the big cabinet, you can't reach it. But with wireframes, you can see everything at once." -P18

5.5.6 Expectation mismatches with system suggestions reduce user control. We found that users sometimes felt that the system's suggestions, either suggested via multimodal specification or generative AI models, did not match their design expectations, and thus reduced their sense of control. For example, participants were not able to specify the color or the relative size (e.g. big or small) of the objects either through multimodal specification or wireframes. This kind of mismatch is often due to the lack of understanding of the capabilities of the underlying AI models.

> "When I was trying to create a side table to place next to a sofa as a coffee table, either it was not picking up or it was going for more desk or larger-size tables. Even though I switched back and forth between saying small table and side table, it still took a while before it generated something I was happy with." -P23

6 DISCUSSION

In our first study, we found that generative models are helpful for idea inspirations. Through the followup expert elicitation study, we found that when co-creating with generative AI models, users can create 3D layouts with more diverse functionality and color palette, but with poorer consideration of circulation and daylighting. Furthermore, we found that generative models could result in lower user agency when it comes to human-AI co-creation in immersive environments. However, this could be mitigated via the design of wireframes as found in our second study.

Our second study demonstrated that among the three ways of human-AI co-creation, manual creation offers users the most sense of agency and creativity. By visualizing multiple furniture suggestions, manual creation can offer design inspirations. Scaffolded creation offers users higher agency compared to automatic creation. This is because users have additional control over aspects such as furniture size and placement via scaffolded creation. Users also found that scaffolded creation can enable un-biased, higher-level of thinking when designing layouts. In automatic creation, users tended to follow what the system suggested to them and not to make changes, leading to the least sense of creativity and agency among the three conditions.

6.1 Design Implications

Through the lens of creativity and agency, we highlight the opportunities and challenges of human-AI co-creation in immersive virtual environments, and discuss design recommendations drawn from our results.

6.1.1 Offering results of generative AI via intermediate representations. Our design of wireframes offers higher user controllability when working with generative models. Specifically, in the task of creating 3D layouts, users are granted more control over the size and placement of object and think they can view the design in an

unbiased way. Besides, the design of wireframes offers unique affordances in VR by making it easier for users to navigate layouts and manipulate distant objects due to less occlusions compared to handling a fully populated 3D scene. This aligns with prior work that utilizes low fidelity representation when working the generative designs (e.g. [32]). Similarly, there has been also a long-standing body of work in Sketch Based Interfaces for Modeling that utilizes both the coarseness and the expressiveness of sketches to guide the detailed generation of 3D models [43]. This demonstrates the benefits of designing low fidelity representations that can prompt more controllable and sophisticated generated content. We therefore encourage future researchers and designers to consider using more advanced intermediate representations of the generated outcome beyond 2D planes on the floor. These representations should capture richer properties of 3D content, such as color and shape, in the immersive environment while still allowing users to easily manipulate the objects and navigate the scene. The note of intermediate representations could even go beyond immersive environments. The concept of intermediate representations can extend beyond immersive environments. For instance, rather than generating lengthy text, Large Language Models could produce an outline as an intermediate representation, allowing users to make adjustments before finalizing the text. Similarly, other generative models could use intermediate representations like image skeletons for pictures or key frames for videos.

6.1.2 Offering multiple generated suggestions for inspirations. Our study shows that participants felt more creative and more easily inspired when they can choose from multiple generated suggestions. Users tend to get inspirations from suggestions when they don't have a concrete idea in mind or when they don't want to spend too much time on browsing the catalog menu. Contrarily, when given one suggestion in automatic creation, users tended to follow what the system generates, leading to fixation of thinking. Thus, future researchers and practitioners might consider offering user the ability to choose from multiple generated suggestions, in order to enhance users' sense of creativity. For example, generative AI systems can offer parallel comparison by visualizing potentially diverse generated results for users.

6.1.3 Addressing expectation mismatch between users and generative AI. A common challenge across all conditions, based on the study, is the expectation mismatch when unexpected output was generated by AI. Through the expert evaluation, we found that although by co-creating generative AI models users can create 3D layouts that are diverse in aspects such as functionality and color palette, users generally have preferences of the layout design that fall outside of the capabilities of generative AI models. For example, in study 1, layouts co-created with AI showed poorer consideration of circulation and daylighting because the underlying generative AI model was not trained with those criteria in mind. Additionally, users lacked a sufficient understanding of the system's capabilities. This highlights the need for more transparent communication between users and generative AI regarding the system's capabilities and limitations. This aligns with the Explainable AI (XAI) research (e.g., [17, 24, 35]), where researchers aim to provide more transparent explanations of decision-making process of the AI model, with an emphasis on text or images. However, there has been little explorations in the visualization and interaction techniques for making AI models more understandable in the immersive environments. Therefore, future researchers and practitioners should consider designing human-AI systems that can visualize how the generative AI model perceives and completes the user's design.

6.2 Limitations

Our paper explored ways of human-AI co-creation in virtual immersive environments and showed empirical results on the comparison among various ways of creation. However, our work has several limitations. First, both of our studies took place in a lab setting with the participants engaged with the system in a short amount of time. The way that participants design 3D layout with a time constraint in the lab setting could be different from how they would design without time constraint outside the lab. Our studies also had a relatively small sample size, which could reduce the validity of our quantitative results. Second, our system was specifically tailored for interior design tasks and had several technical limitations. For instance, as mentioned in Section 3, users could only draw wireframes on the floor, and for objects not placed on the floor (e.g., ceiling lamps), the system automatically set their heights when converting to furniture. The underlying AI models of VRCopilot also had limitations, such as misidentifying voice intents or not supporting color or size in multimodal specifications. Participants occasionally had to retry their intents or regenerate in a few cases when the system misidentified voice commands. Future work should seek to provide clearer system status and offer alternatives for misidentified or unsupported intents, as well as further extend the model's capabilities and supported attributes. Lastly, the generalizability of our findings to domains or contexts other than interior design necessitates further investigation. Our paper provides insights into user creativity, agency, and strategies in human-AI co-creation in general. Some findings, however, are more specific to interior design or the immersive virtual environments. Future research should evaluate the adaptability and utility of the system across diverse application domains to determine its broader applicability.

7 CONCLUSION

In this paper, we presented a mixed-initiative system named VRCopilot that integrates pre-trained generative models into immersive authoring workflows. We introduce three ways of human-AI co-creation in the immersive virtual environment including Manual Creation, Automatic Creation, and Scaffolded Creation. We conducted two rounds of comparative studies that evaluates the potential and challenges of co-creating with generative AI in VR and user perceived creativity, effort, and agency. Our first study revealed that generative AI could offer design inspirations to users but decrease their sense of agency. Our second study suggested that when users use the wireframes in Scaffolded Creation, they felt higher sense of agency compared to Automatic Creation. Manual Creation offers users the most creativity and agency. We provide insights on the opportunities and challenges around human-AI co-creation in the immersive environments and make recommendations for future research and design.

ACKNOWLEDGMENTS

We would like to thank Bella Palumbi for her help in system implementation. We would also like to thank our participants for their time, and the reviewers for their valuable feedback and suggestions.

REFERENCES

[1] Adobe. 2021. *Adobe Medium.* https://www.adobe.com/products/medium.html
[2] Open AI. 2024. *Sora.* https://openai.com/sora
[3] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N Bennett, Kori Inkpen, et al. 2019. Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–13.
[4] Richard A Bolt. 1980. "Put-that-there" Voice and gesture at the graphics interface. In *Proceedings of the 7th annual conference on Computer graphics and interactive techniques*. 262–270.
[5] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.
[6] Daniel Buschek, Lukas Mecke, Florian Lehmann, and Hai Dang. 2021. Nine potential pitfalls when designing human-ai co-creative systems. *arXiv preprint arXiv:2104.00358* (2021).
[7] Bill Buxton. 2010. *Sketching user experiences: getting the design right and the right design.* Morgan kaufmann.
[8] Jeffrey W Chastine, Kristine Nagel, Ying Zhu, and Luca Yearsovich. 2007. Understanding the design space of referencing in collaborative augmented reality environments. In *Proceedings of graphics interface 2007*. 207–214.
[9] Erin Cherry and Celine Latulipe. 2014. Quantifying the creativity support of digital tools through the creativity support index. *ACM Transactions on Computer-Human Interaction (TOCHI)* 21, 4 (2014), 1–25.
[10] Kevin Chow, Caitlin Coyiuto, Cuong Nguyen, and Dongwook Yoon. 2019. Challenges and design considerations for multimodal asynchronous collaboration in VR. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–24.
[11] John Joon Young Chung and Eytan Adar. 2023. PromptPaint: Steering Text-to-Image Generation Through Paint Medium-like Interactions. *arXiv preprint arXiv:2308.05184* (2023).
[12] John Joon Young Chung, Wooseok Kim, Kang Min Yoo, Hwaran Lee, Eytan Adar, and Minsuk Chang. 2022. TaleBrush: Sketching stories with generative pretrained language models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–19.
[13] Elizabeth Clark, Anne Spencer Ross, Chenhao Tan, Yangfeng Ji, and Noah A Smith. 2018. Creative writing with a machine in the loop: Case studies on slogans and stories. In *23rd International Conference on Intelligent User Interfaces*. 329–340.
[14] Philip R Cohen. 1992. The role of natural language in a multimodal interface. In *Proceedings of the 5th annual ACM symposium on User interface software and technology*. 143–149.
[15] Nicholas Davis, Chih-PIn Hsiao, Kunwar Yashraj Singh, Lisa Li, and Brian Magerko. 2016. Empirically studying participatory sense-making in abstract drawing with a co-creative cognitive agent. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*. 196–207.
[16] Matt Deitke, Ruoshi Liu, Matthew Wallingford, Huong Ngo, Oscar Michel, Aditya Kusupati, Alan Fan, Christian Laforte, Vikram Voleti, Samir Yitzhak Gadre, et al. 2024. Objaverse-xl: A universe of 10m+ 3d objects. *Advances in Neural Information Processing Systems* 36 (2024).
[17] Shipi Dhanorkar, Christine T Wolf, Kun Qian, Anbang Xu, Lucian Popa, and Yunyao Li. 2021. Who needs to know what, when?: Broadening the Explainable AI (XAI) Design Space by Looking at Explanations Across the AI Lifecycle. In *Designing Interactive Systems Conference 2021*. 1591–1602.
[18] Barrett Ens, Fraser Anderson, Tovi Grossman, Michelle Annett, Pourang Irani, and George Fitzmaurice. 2017. Ivy: Exploring spatially situated visual programming for authoring and understanding intelligent environments. In *Proceedings of the 43rd Graphics Interface Conference*. 156–162.
[19] Judith E Fan, Monica Dinculescu, and David Ha. 2019. Collabdraw: an environment for collaborative sketching with an artificial agent. In *Proceedings of the 2019 on Creativity and Cognition*. 556–561.
[20] Weixi Feng, Wanrong Zhu, Tsu-jui Fu, Varun Jampani, Arjun Akula, Xuehai He, Sugato Basu, Xin Eric Wang, and William Yang Wang. 2024. Layoutgpt: Compositional visual planning and generation with large language models. *Advances in Neural Information Processing Systems* 36 (2024).
[21] Huan Fu, Bowen Cai, Lin Gao, Ling-Xiao Zhang, Jiaming Wang, Cao Li, Qixun Zeng, Chengyue Sun, Rongfei Jia, Binqiang Zhao, et al. 2021. 3d-front: 3d furnished rooms with layouts and semantics. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 10933–10942.
[22] Jun Gao, Tianchang Shen, Zian Wang, Wenzheng Chen, Kangxue Yin, Daiqing Li, Or Litany, Zan Gojcic, and Sanja Fidler. 2022. Get3d: A generative model of high quality 3d textured shapes learned from images. *Advances In Neural Information Processing Systems* 35 (2022), 31841–31854.
[23] Google. 2016. *Google Tilt Brush.* https://www.tiltbrush.com/
[24] Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Franco Turini, Fosca Giannotti, and Dino Pedreschi. 2018. A survey of methods for explaining black box models. *ACM computing surveys (CSUR)* 51, 5 (2018), 1–42.
[25] Matthew Guzdial, Nicholas Liao, Jonathan Chen, Shao-Yu Chen, Shukan Shah, Vishwa Shah, Joshua Reno, Gillian Smith, and Mark O Riedl. 2019. Friend, collaborator, student, manager: How design of an ai-driven game level editor affects creators. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–13.
[26] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.
[27] Björn Hartmann, Loren Yu, Abel Allison, Yeonsoo Yang, and Scott R Klemmer. 2008. Design as exploration: creating interface alternatives through parallel authoring and runtime tuning. In *Proceedings of the 21st annual ACM symposium on User interface software and technology*. 91–100.
[28] Cheng-Zhi Anna Huang, Curtis Hawthorne, Adam Roberts, Monica Dinculescu, James Wexler, Leon Hong, and Jacob Howcroft. 2019. The bach doodle: Approachable music composition with machine learning at scale. *arXiv preprint arXiv:1907.06637* (2019).
[29] Hikaru Ibayashi, Yuta Sugiura, Daisuke Sakamoto, Natsuki Miyata, Mitsunori Tada, Takashi Okuma, Takeshi Kurata, Masaaki Mochimaru, and Takeo Igarashi. 2015. Dollhouse vr: a multi-view, multi-user collaborative design workspace with vr technology. In *SIGGRAPH Asia 2015 emerging technologies*. 1–2.
[30] Bret Jackson and Daniel F Keefe. 2016. Lift-off: Using reference imagery and freehand sketching to create 3d models in vr. *IEEE transactions on visualization and computer graphics* 22, 4 (2016), 1442–1451.
[31] David G. Jansson and Steven M. Smith. 1991. Design fixation. *Design Studies* 12, 1 (1991), 3–11. https://doi.org/10.1016/0142-694X(91)90003-F
[32] Rubaiat Habib Kazi, Tovi Grossman, Hyunmin Cheong, Ali Hashemi, and George W Fitzmaurice. 2017. DreamSketch: Early Stage 3D Design Explorations with Sketching and Generative Design.. In *UIST*, Vol. 14. 401–414.
[33] Gun A Lee, Claudia Nelles, Mark Billinghurst, and Gerard Jounghyun Kim. 2004. Immersive authoring of tangible augmented reality applications. In *Third IEEE and ACM international symposium on mixed and augmented reality*. IEEE, 172–181.
[34] Manyi Li, Akshay Gadi Patil, Kai Xu, Siddhartha Chaudhuri, Owais Khan, Ariel Shamir, Changhe Tu, Baoquan Chen, Daniel Cohen-Or, and Hao Zhang. 2019. Grains: Generative recursive autoencoders for indoor scenes. *ACM Transactions on Graphics (TOG)* 38, 2 (2019), 1–16.
[35] Q Vera Liao, Hariharan Subramonyam, Jennifer Wang, and Jennifer Wortman Vaughan. 2023. Designerly understanding: Information needs for model transparency to support design ideation for AI-powered user experience. In *Proceedings of the 2023 CHI conference on human factors in computing systems*. 1–21.
[36] Ruoshi Liu, Rundi Wu, Basile Van Hoorick, Pavel Tokmakov, Sergey Zakharov, and Carl Vondrick. 2023. Zero-1-to-3: Zero-shot one image to 3d object. In *Proceedings of the IEEE/CVF international conference on computer vision*. 9298–9309.
[37] Ryan Louie, Andy Coenen, Cheng Zhi Huang, Michael Terry, and Carrie J Cai. 2020. 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*. 1–13.
[38] Kai Lukoff, Ulrik Lyngs, Himanshu Zade, J Vera Liao, James Choi, Kaiyue Fan, Sean A Munson, and Alexis Hiniker. 2021. How the design of youtube influences user sense of agency. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–17.
[39] Rui Ma, Akshay Gadi Patil, Matthew Fisher, Manyi Li, Sören Pirk, Binh-Son Hua, Sai-Kit Yeung, Xin Tong, Leonidas Guibas, and Hao Zhang. 2018. Language-driven synthesis of 3D scenes from scene databases. *ACM Transactions on Graphics (TOG)* 37, 6 (2018), 1–16.
[40] Paul Merrell, Eric Schkufza, Zeyang Li, Maneesh Agrawala, and Vladlen Koltun. 2011. Interactive furniture layout using interior design guidelines. *ACM transactions on graphics (TOG)* 30, 4 (2011), 1–10.
[41] Mark Mine. 1995. ISAAC: A virtual environment tool for the interactive construction of virtual worlds. (1995).
[42] Michael Nebeling, Janet Nebeling, Ao Yu, and Rob Rumble. 2018. Protoar: Rapid physical-digital prototyping of mobile augmented reality applications. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–12.
[43] Luke Olsen, Faramarz F Samavati, Mario Costa Sousa, and Joaquim A Jorge. 2009. Sketch-based modeling: A survey. *Computers & Graphics* 33, 1 (2009), 85–103.
[44] OpenAI. 2023. *GPT4.* https://openai.com/gpt-4
[45] Despoina Paschalidou, Amlan Kar, Maria Shugrina, Karsten Kreis, Andreas Geiger, and Sanja Fidler. 2021. Atiss: Autoregressive transformers for indoor scene synthesis. *Advances in Neural Information Processing Systems* 34 (2021), 12013–12026.

[46] Scott Reed, Zeynep Akata, Xinchen Yan, Lajanugen Logeswaran, Bernt Schiele, and Honglak Lee. 2016. Generative Adversarial Text to Image Synthesis. In *Proceedings of The 33rd International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 48)*, Maria Florina Balcan and Kilian Q. Weinberger (Eds.). PMLR, New York, New York, USA, 1060–1069. https://proceedings.mlr.press/v48/reed16.html

[47] Marc Rettig. 1994. Prototyping for tiny fingers. *Commun. ACM* 37, 4 (1994), 21–27.

[48] Daniel Ritchie, Kai Wang, and Yu-an Lin. 2019. Fast and flexible indoor scene synthesis via deep convolutional generative models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 6182–6190.

[49] Ben Shneiderman. 2007. Creativity support tools: accelerating discovery and innovation. *Commun. ACM* 50, 12 (2007), 20–32.

[50] Carolyn Snyder. 2003. *Paper prototyping: The fast and easy way to design and refine user interfaces*. Morgan Kaufmann.

[51] Shuran Song, Fisher Yu, Andy Zeng, Angel X Chang, Manolis Savva, and Thomas Funkhouser. 2017. Semantic scene completion from a single depth image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1746–1754.

[52] Richard Stoakley, Matthew J Conway, and Randy Pausch. 1995. Virtual reality on a WIM: interactive worlds in miniature. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 265–272.

[53] Amanda Swearngin, Chenglong Wang, Alannah Oleson, James Fogarty, and Amy J Ko. 2020. Scout: Rapid exploration of interface layout alternatives through high-level design constraints. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[54] Jiapeng Tang, Yinyu Nie, Lev Markhasin, Angela Dai, Justus Thies, and Matthias Nießner. 2024. Diffuscene: Denoising diffusion models for generative indoor scene synthesis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 20507–20518.

[55] Adam Tapal, Ela Oren, Reuven Dar, and Baruch Eitam. 2017. The sense of agency scale: A measure of consciously perceived control over one's mind, body, and the immediate environment. *Frontiers in psychology* 8 (2017), 1552.

[56] Unreal. 2022. *Unreal Editor VR Mode*. https://docs.unrealengine.com/5.0/en-US/vr-mode-in-unreal-editor/

[57] Kai Wang, Yu-An Lin, Ben Weissmann, Manolis Savva, Angel X Chang, and Daniel Ritchie. 2019. Planit: Planning and instantiating indoor scenes with relation graph and spatial prior networks. *ACM Transactions on Graphics (TOG)* 38, 4 (2019), 1–15.

[58] Kai Wang, Manolis Savva, Angel X Chang, and Daniel Ritchie. 2018. Deep convolutional priors for indoor scene synthesis. *ACM Transactions on Graphics (TOG)* 37, 4 (2018), 1–14.

[59] Xinpeng Wang, Chandan Yeshwanth, and Matthias Nießner. 2021. Sceneformer: Indoor scene generation with transformers. In *2021 International Conference on 3D Vision (3DV)*. IEEE, 106–115.

[60] Qiuhong Anna Wei, Sijie Ding, Jeong Joon Park, Rahul Sajnani, Adrien Poulenard, Srinath Sridhar, and Leonidas Guibas. 2023. Lego-net: Learning regular rearrangements of objects in rooms. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 19037–19047.

[61] George Wilson and Samuel Shpall. 2016. The nature of action and agency. *Stanford Encyclopedia of Philosophy* (2016).

[62] Xiaohui Zeng, Arash Vahdat, Francis Williams, Zan Gojcic, Or Litany, Sanja Fidler, and Karsten Kreis. 2022. LION: Latent Point Diffusion Models for 3D Shape Generation. In *Advances in Neural Information Processing Systems (NeurIPS)*.

[63] Lei Zhang, Ashutosh Agrawal, Steve Oney, and Anhong Guo. 2023. VRGit: A Version Control System for Collaborative Content Creation in Virtual Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–14.

[64] Lei Zhang and Steve Oney. 2019. Studying the Benefits and Challenges of Immersive Dataflow Programming. In *2019 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*. IEEE, 223–227.

[65] Lei Zhang and Steve Oney. 2020. Flowmatic: An immersive authoring tool for creating interactive scenes in virtual reality. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 342–353.

[66] Song-Hai Zhang, Shao-Kui Zhang, Yuan Liang, and Peter Hall. 2019. A survey of 3d indoor scene synthesis. *Journal of Computer Science and Technology* 34 (2019), 594–608.

[67] Zhengzhe Zhu, Ziyi Liu, Youyou Zhang, Lijun Zhu, Joey Huang, Ana M Villanueva, Xun Qian, Kylie Peppler, and Karthik Ramani. 2023. LearnIoTVR: An End-to-End Virtual Reality Environment Providing Authentic Learning Experiences for Internet of Things. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.