

# CreepyCoCreator? Investigating AI Representation Modes for 3D Object Co-Creation in Virtual Reality

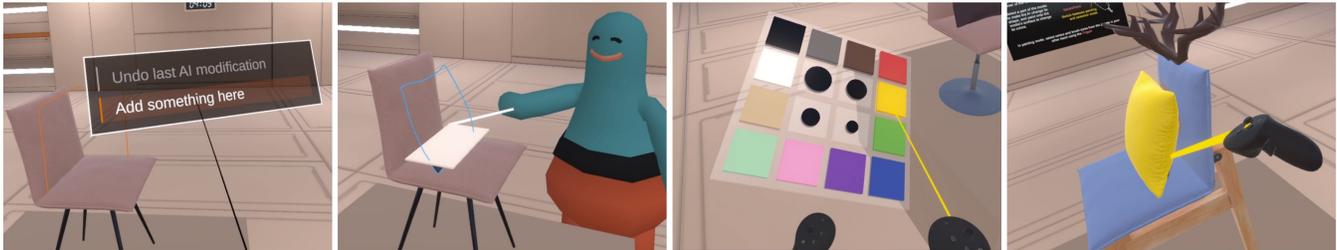
Julian Rasch  
LMU Munich  
Munich, Germany  
julian.rasch@ifi.lmu.de

Julia Töws  
Saarland University  
Saarland Informatics Campus  
Saarbrücken, Germany  
julia\_toews@web.de

Teresa Hirzle  
Department of Computer Science  
University of Copenhagen  
Copenhagen, Denmark  
tehi@di.ku.dk

Florian Müller  
TU Darmstadt  
Darmstadt, Germany  
florian.mueller@tu-darmstadt.de

Martin Schmitz  
Saarland University  
Saarland Informatics Campus  
Saarbrücken, Germany  
Department of Computer Science  
University of Copenhagen  
Copenhagen, Denmark  
mschmitz@cs.uni-saarland.de



**Figure 1:** In this paper, we explore VR interaction with an object-generating co-creative AI via a Wizard-of-Oz study on how the AI communicates its intent to users. The study investigates the impact of different modes of AI representation (modification highlighting, incremental visualization of changes, and the embodiment of the avatar) on user perception.

## Abstract

Generative AI in Virtual Reality offers the potential for collaborative object-building, yet challenges remain in aligning AI contributions with user expectations. In particular, users often struggle to understand and collaborate with AI when its actions are not transparently represented. This paper thus explores the co-creative object-building process through a Wizard-of-Oz study, focusing on how AI can effectively convey its intent to users during object customization in Virtual Reality. Inspired by human-to-human collaboration, we focus on three representation modes: the presence of an embodied avatar, whether the AI's contributions are visualized immediately or incrementally, and whether the areas modified are highlighted in advance. The findings provide insights into how these factors affect user perception and interaction with object-generating AI tools in Virtual Reality as well as satisfaction

and ownership of the created objects. The results offer design implications for co-creative world-building systems, aiming to foster more effective and satisfying collaborations between humans and AI in Virtual Reality.

## CCS Concepts

• **Human-centered computing** → **Human computer interaction (HCI); User studies.**

## Keywords

Co-creative systems; Generative AI; 3D Creation; Virtual Reality; User Studies



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

CHI '25, Yokohama, Japan

© 2025 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-1394-1/25/04

<https://doi.org/10.1145/3706598.3713720>

## ACM Reference Format:

Julian Rasch, Julia Töws, Teresa Hirzle, Florian Müller, and Martin Schmitz. 2025. CreepyCoCreator? Investigating AI Representation Modes for 3D Object Co-Creation in Virtual Reality. In *CHI Conference on Human Factors in Computing Systems (CHI '25)*, April 26–May 01, 2025, Yokohama, Japan. ACM, New York, NY, USA, 14 pages. <https://doi.org/10.1145/3706598.3713720>

## 1 Introduction

Building 3D worlds and objects for Virtual Reality (VR) has become increasingly accessible, enabling users to immerse themselves in detailed environments and actively (co-)create them. Recent advances have also seen the integration of generative Artificial Intelligence (AI) in this process [16], leveraging text prompts and other inputs to generate elaborate 3D objects and even entire 3D worlds. However, these emerging approaches often rely on a "set-it-and-forget-it" approach, where users input a prompt and receive a fully realized 3D object or world with limited opportunities for customization or interaction. While this trend toward AI-driven generation can democratize access to high-quality 3D objects, it challenges human agency and shifts control from humans to AI.

To empower users to get more involved in the design process, a stream of research has started exploring co-creative systems, which enable users to collaborate with AI in a creative process [13]. These systems have been applied across a variety of domains, including music, storytelling, game design, and visual arts [14, 34, 36, 41, 48]. There has been growing interest in extending such co-creative systems to 3D object co-creation in recent years. AI systems can now generate and modify 3D objects based on user inputs such as text prompts or sketches [27, 35, 42–44, 52], and some systems even allow for interactive refinement of AI outputs [38, 62].

Despite these advances, a critical question remains largely unexplored: how should the actions of a co-creative tool for object generation be represented in VR? This question touches on fundamental aspects of co-creative interactions and raises several essential considerations: (1) Human collaborative work often involves asynchronous or gradual changes, making immediate modifications potentially counterintuitive. Should AI-driven changes to 3D objects be presented incrementally, mirroring the rhythm of human collaboration, or is immediate modification more appropriate in co-creation? (2) Drawing from the literature on non-player characters (NPCs), the embodiment is often deemed crucial for effective interaction. However, does this hold for a co-creative system focused on object generation, which may not inherently require an embodied visualization as NPCs do? Does the user perception of the system change from a tool to a collaborator only through embodiment, and if yes, what are the potential downsides? (3) Decades of HCI research and design practices emphasize highlighting focus or activity (e.g., showing all cursors in collaborative text editing tools or speaker indicators in video conferencing). Is visually indicating where and how AI modifications occur sufficient for co-creative object generation systems, or do these systems necessitate new ways of co-creative interaction?

Addressing these questions is pivotal to advancing the design of co-creative systems for 3D objects and world generation. This paper contributes to answering these questions by investigating different representations of AI actions in the context of co-creating 3D objects in VR (see Figure 1). Specifically, we examine the aforementioned three aspects of co-creating a 3D object in VR in a Wizard-of-Oz user study on (1) perception of embodiment, (2) incremental visualization of co-creative contributions, and (3) effect of highlighting the area where the AI is going to perform a modification. Our findings show that co-creating with an embodied AI

significantly influences the perceived AI contribution to the created model. Further, highlighting does not increase predictability or communication with the AI system but decreases users' enjoyment and a perceived partnership with the AI. Finally, users pay more attention to the AI when it uses an incremental generation.

In summary, the contributions of this paper are two-fold: First, we conduct a Wizard-of-Oz user study to examine the influence of highlighting, incremental visualization, and embodiment on user perception, collaboration, overall system appeal, and behavioral engagement. We contribute empirical insights into how the different representation strategies affect the co-creative experience in VR-based object-building. Secondly, based on the study's results, we provide implications for designing future co-creative object-building tools in VR.

## 2 Related Work

In the following, we discuss related work in 3D modeling in VR, AI for 3D object generation, co-creative systems in general, embodied AI, and AI contribution visualization.

### 2.1 3D Modeling in VR

Three-dimensional modeling has been a topic of interest in computer research since as early as the 1970s [11], with HoloSketch [17] being among the earliest contributions to 3D modeling software for immersive environments. In general, there are several ways to approach 3D design in VR. For instance, sketching, though intuitive and quick, suffers from a lack of haptic feedback mid-air. Some systems mitigate inaccuracies by smoothing out hand-drawn lines in post-processing [63], featuring the use of real physical objects in AR as guides [53], using haptic feedback of the controller as assistance [22] or letting the user sketch on 2D surfaces and then transform the lines in 3D space [4, 19, 26, 63]. Similar problems arise with digital sculpting, in which the user shapes desired forms out of primitives and pre-included models using virtual sculpting tools [59]. Other approaches include mesh modeling [9, 53] and assembly from primitives [17, 18, 53]. Additionally, there also exist several commercially available tools for 3D design in VR that allow users to paint 3D artworks [40], design objects by free-hand drawing [59], and modify the mesh directly [61]. Closely related to this paper, GetWild [66] is a VR environment-editing system incorporating AI-generated models into the creation process to speed up the design pipeline and increase accessibility for users with little modeling experience.

While allowing for expressive 3D modeling, direct modeling approaches can be complex for users with little modeling experience. In contrast, this paper investigates systems that co-creatively help users create 3D objects in VR.

### 2.2 AI for 3D Object Generation and Manipulation

While AI research for 2D image synthesis has made notable progress in the past few years, the field of 3D object generation has faced more difficulties due to limitations of the different representation modes of 3D data [33, 58] as well as a lack of text-to-3D training data for deep learning models [33]. Neural Radiance Fields (NeRF)

[45] is a recently developed approach to representing 3D data, initially serving the purpose of 3D image reconstruction. Due to their ability to bridge the gap between 2D and 3D data, NeRFs can be combined with powerful text-to-image generative models to construct advanced text-to-3D generative models [33]. This has led to a recent upsurge in AI tools for 3D object generation, such as DreamFusion [52] and Magic3D [35]. However, outputs of NeRF-based 3D generation are difficult to use in downstream 3D applications since a lot of 3D graphics software requires standard data representations like meshes [33]. Shap-E [27] is a recent work demonstrating object generation based on a text prompt, with a mesh output created within less than two minutes.

Though generative models like the ones mentioned above are already helpful for exploring 3D design ideas, it can be difficult for users to formulate a textual prompt that will produce an output meeting all of the user's envisioned criteria. This one-step approach often forces the user to make the AI generate an entirely new object based on a rephrased prompt. For this reason, efforts have been made to develop AI tools that can modify a given 3D object according to a user's instructions.

SKED [44] uses sketch-based instructions to enable interactive editing of 3D objects. In contrast, ShapeCrafter [23] generates a 3D object from a text-based description and lets the user refine object characteristics by recursively adding descriptive phrases. Similarly, Text2Mesh [43] can change the style of 3D objects: Given an input mesh and a text prompt, it modifies color and geometry details that adhere to the user-specified style. Additionally, SPAGHETTI [25] allows users to select parts of a 3D object and perform rough local transformations on them, based on which the generative model creates an edited version of the same object.

Although previous works include interactive elements similar to co-creative systems, none leverage the intuitive and expressive potential of spatial 3D interaction and presentation offered by VR. Instead, these systems rely on indirect instructions via text or 2D interface interactions. In contrast, this work focuses explicitly on VR environments.

### 2.3 Co-creative Systems

Human-computer co-creativity is a process in which a human and a computer contribute as equals to the same creative process [13]. Such co-creative systems have been commonly developed for various domains [54]. Examples are game design [34, 41] and visual art [14, 36, 48].

A stream of research has also started to investigate co-creative systems for editing 3D objects. For instance, Liu et al. [38] have proposed a system where users can take turns creating and editing 3D objects with an AI. Closely related, Calliope [62] has explored interaction possibilities of using generative adversarial networks as an active collaborator in VR. However, it did not investigate the specific influences of different representation modes of such a collaborator on user perception. Other than Calliope, co-creative systems are typically not studied in VR environments but in physical or digital 2D ones. We want to explore whether findings from those areas also apply to co-creative systems in VR.

### 2.4 Embodied AI

Embodiment can be defined in several ways, though Ziemke [68] has conceptualized the most basic form of an embodied system as a system that can act on and be acted on by its environment. Other notions include physical embodiment, which necessitates a physical body, and organismoid embodiment, which further constrains the physical body to resemble a living being. Guckelsberger et al. [24] extend Ziemke's typology with the concept of virtual embodiment, which requires a virtual body that can act on and be acted on by a virtual environment.

Research has shown that the virtual embodiment of agents can improve motivation, positive attitude, and collaborative experiences in creative systems [6, 14, 55]. Additionally, organismoid embodiment enhances identification, empathy, and perceived creativity and is considered essential for co-creativity due to the unique perspective it provides [24]. Also, Kim et al. [28] investigated the effect of embodiment on the perception of a digital AR assistant. Moruzzi [47] found that AI embodiment increased perceived creativity in a collaborative artistic process, while Rezwana and Maher [56] showed that AI personification influenced users' perceptions of the AI as an independent collaborator and affected their ethical views on ownership of co-created artifacts.

Studies on the effects of embodiment in VR have mainly examined AI agents without co-creativity. In contrast, studies on the embodiment of co-creative AI have mainly been carried out in non-VR environments. We combine both promising directions of previous works by studying embodiment in the context of a co-creative object-generating agent embedded in VR.

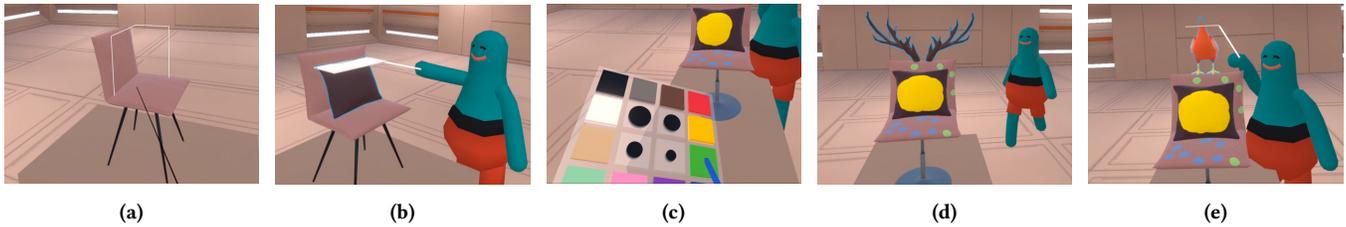
### 2.5 Non-player Characters in Games

In the domain of game development, the design of autonomous human-like agents, so-called Non-Player Characters (NPCs), was addressed in the context of believability [5]. Recent research explores the potential of NPCs in virtual reality [67] and of LLM-based agents with more human-like reasoning, planning, and execution abilities [50]. Moreover, in a systematic literature review, Wittmann and Morschheuser [64] discuss NPC design patterns and how to transfer those to the design of AI systems for human-AI collaboration. The authors point out six relevant focus fields: NPC responsiveness, appearance, communication patterns, emotional aspects, behavioral characteristics, and player-NPC and NPC-NPC team structures.

Unlike NPCs, which are inherently embodied as avatars, co-creative tools for object building do not inherently require such embodiment. This property raises important questions about the role of embodiment in such tools and its interaction with factors like the highlighting and visualization of changes, which are the focus of this paper.

### 2.6 AI Contribution Visualization

Computational creativity can be assessed through the creator, the product, the process, or the environment, but humans often interpret a system's creativity based on appearance, behavior, and output [20, 32, 49, 51, 57]. Observing the creative process may lead to higher evaluations due to empathy and perceived effort, known as the "effort heuristic" [12, 29], and delays in AI output generation



**Figure 2: Steps during the collaboration creation at the example of a chair: (a) The chair at the center of the room on a podium (b) embodied, incremental, highlighted AI adds a new element to the selected position (c) the user paints elements using the palette, (d) AI adds another (creative) element to the chair (e) and modifies the previous addition.**

can increase user engagement and perceived control [39]. However, empirical studies on incremental visualization of AI creativity are lacking [37], prompting the need for research on both AI embodiment and perceptual evidence in co-creative VR systems.

Thus far, no empirical studies have investigated the effects of such an intentionally incremental visualization of AI contributions, especially not separately from the impact of AI embodiment. Therefore, we address this lack of research by contributing a study in which both embodiments and perceptual evidence in the form of incremental contribution visualization are examined in the context of a co-creative VR system.

### 3 Methodology

Our study investigates the collaboration with an object-generating AI, focusing on the often overlooked crucial factor of AI-to-human communication [54]. We conducted a controlled experiment to examine how three representation modes (highlighting, incremental visualization, and embodiment) of the AI’s generative contributions impact the co-creative process. Our study is guided by the following research questions (RQs), each targeting one of the three representation modes:

**RQ1** How does **highlighting** affect the co-creative collaboration experience, in particular regarding the predictability of AI? Based on decades of HCI research and design practices that emphasize the importance of highlighting focus or activity, we hypothesize that highlighting will positively affect perceived communication quality, as it is a well-known mechanism to guide users’ attention. For the same reason, we also expect that with highlighting, users will pay more attention to the AI, and the clarity of the AI intentions will be perceived better.

**RQ2** How does **incremental visualization** of changes affect the co-creative collaboration experience, in particular regarding a user’s perception of the AI’s efficiency and competence compared to a discrete visualization of the AI’s contributions? Inspired by human collaboration, where it takes time for a modification to take shape, we expect that incremental visualization will affect measures concerning those same qualities as above, but also measures on the perceived system efficiency and competence, measures on the perception of AI outputs, perceived alignment of AI outputs with the users’ visions, perceived agency, and the proportion to which contributions to the final artifact are attributed to the AI.

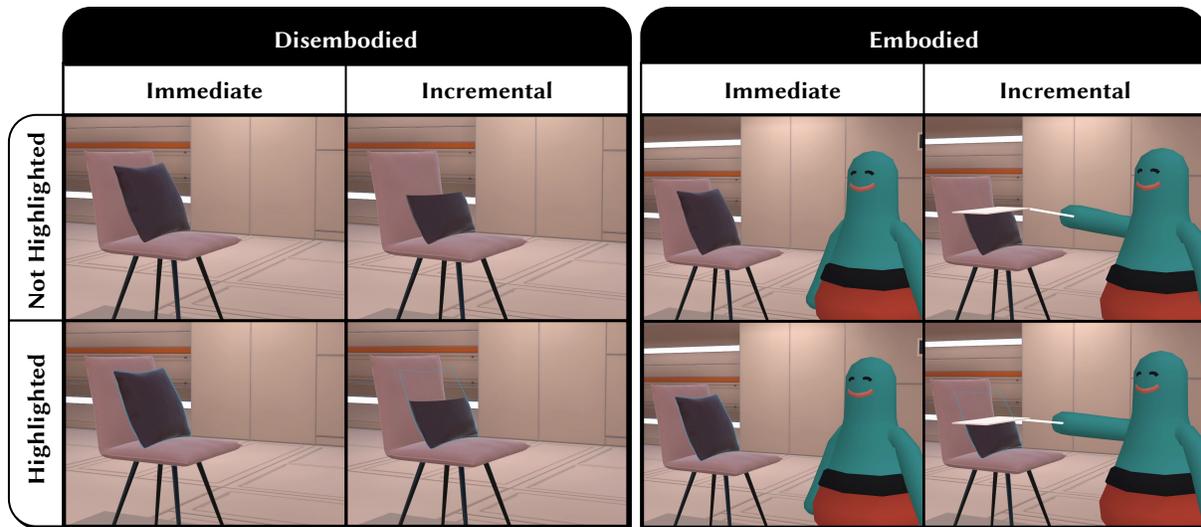
**RQ3** How does the **embodiment** of an AI affect the co-creative collaboration experience, in particular regarding the perceived supportiveness, efficiency, and competence of the AI? Research on NPCs and games highlights embodiment as a critical factor in interaction, making embodiment a potentially promising factor for co-creative tools in VR. We thus expect that embodiment will affect the perceived supportiveness, efficiency, and competence of the AI, the perceived value and creativity of AI outputs, measures concerning collaborative experience, appeal, and behavioral engagement, as well as the previously mentioned contribution attribution and users’ perceived closeness to the AI.

#### 3.1 Task

For our study, we placed participants in a virtual room, which was empty, besides a podium in the center (see Figure 2a) and a timer on the wall. In the embodied AI conditions, the AI avatar is shown additionally (see Figure 2b). During the task, the user and the AI can collaboratively and iteratively modify a given 3D model.

At the beginning of a modeling session, a 3D object appears on the podium, and the timer starts running. As long as the timer was running, the user could modify the object by either painting its texture (see Figure 2c), adding elements to the existing object (see Figure 2d), or instructing the AI to modify (see Figure 2e). We chose painting as the users’ primary design responsibility as it allowed for a consistent engagement with the 3D object during the AI-driven modifications and ensured comparable levels of object alteration and creative expression while avoiding interference. For the modifications, users were always controlling if and when any changes to the object should be executed (including being able to revert the AI’s modification). We simulated the AI agent in a Wizard-of-Oz approach, in the sense that its contributions to the 3D objects were not the results of an actual generative process but instead object additions and replacements created by human 3D artists in advance.

Further, the creative contributions of the AI agent are not designed to align with the contributions of the human partner but instead emerge completely independently of the human’s behavior. This contribution style is typical for what Rezwana and Maher [54] refer to as *provoking agents* (as opposed to *pleasing agents*), which evoke creativity in their human partners by providing divergent ideas and alternative explorations. While the AI was active, the user could not give the AI another modification instruction.



**Figure 3:** We vary three different independent variables: (1) whether the AI has an embodied representation or is disembodied, (2) whether the AI's contributions are shown immediately or build up incrementally, and (3) if the area that the AI is going to modify is highlighted or not.

We provided participants with four starting objects available for modification: A dinosaur, a chair, a pickup truck, and a house (see Figure 7). Each object allowed for seven modifications randomly chosen by the system. The AI modifications either consisted of an object part being replaced by another one (e.g., switching out the legs of a chair with legs of a different type) or an object part being added to the object (e.g., adding a dormer to the roof of the house).

While most modifications matched the object's type, a few unexpected changes were intentionally introduced (e.g., bunny ears on a dinosaur or cheese in a pickup truck). This design allowed us to explore how users perceive AI creativity, novelty, and predictability across different visualization settings.

## 3.2 Design

**3.2.1 Independent Variables.** In our study, we vary three representation modes that change how the AI carries out a modification instruction (see Figure 3):

**HIGHLIGHTING** If highlighting is enabled, instructing the AI to modify a part of the object will cause an outline to appear around the object part in question. If none of the other visualization aspects are enabled, this outline will be displayed for three seconds and then disappear at the same time that the object part is replaced with its modified version. If incremental visualization is enabled, the highlight will be displayed on the object part for exactly as long as the incremental process takes. If embodiment is enabled, the highlight will be displayed on the object part from the moment the AI avatar starts walking toward the object part until it has finished stepping back from the object. We base this concept of highlighting active elements on traditional 2D interface design to guide user attention.

**INCREMENTAL VISUALIZATION** If incremental visualization is enabled, instructing the AI to modify an object part will cause this part to disappear and its modified version to slowly appear slice by slice as if a 3D printer was printing it. If the embodiment is enabled, a "printing layer" will be placed on top of the incrementally growing object part, vaguely resembling the additive manufacturing technique of sheet lamination. The AI avatar will appear to move this layer along the printing direction through a line connecting its right hand to one of the layer's corners.

**EMBODIMENT** If embodiment is enabled, an AI avatar will sit on a chair in the corner of the room at the beginning of a modeling session. Once instructed to modify a part of the 3D object, it will get up and walk toward the object part to perform the modification. If incremental visualization is enabled, it will stand beside the object holding the printing layer until the incremental process is finished. If incremental visualization is not disabled, the avatar will carry out a waving hand gesture reminiscent of a person performing a magic trick, after which the modified version of the object part replaces its previous version. Once the object part is finished being modified, the avatar will walk away a few steps from the object and wait in this location until it is instructed for another modification. After the AI has finished, it walks back to its chair autonomously.

When none of the representation modes were enabled, the modification process solely consisted of the object part being replaced by its modified version instantaneously without highlighting or any additional embodiment.

We varied all three independent variables in a repeated measures design, resulting in a total of  $2 \times 2 \times 2 = 8$  conditions. We counterbalanced the order of all factors in a Balanced Latin Square to

**Table 1: The list of questions participants were asked in the semi-structured interview.**

Overall experience	How was your experience carrying out the given task? Which of the runs did you prefer? Which one did you like the least?
Perception of and satisfaction with the results	What do you think of your results? Were you aiming for these kinds of results?
Perception of AI	How would you describe the relationship between you and the AI in the context of these tasks? Would you have preferred the Ai to do more/less/other things? In an optimal future version of this system, how would the AI behave?
Suggestions	Do you have any other comments or suggestions?

prevent learning effects. Each participant conducted eight modeling trials across eight different conditions. In each session, participants modified one of the four 3D objects (see Figure 6). Each object was presented to each participant twice, with the order randomized for each individual. We chose four objects as a suitable compromise, striking a balance between variety and familiarity. This approach allowed participants to engage their creativity by attempting various design approaches with the same model while also providing a second chance to revisit and revise previous modifications – an opportunity to learn from failure that we argue is valuable in our research context.

**3.2.2 Dependent Variables.** Data collected during the study consisted of the following:

**Before the experiment:** Demographic information was collected through a Google Forms survey.

**During modeling sessions:** A log of user interactions with the system was recorded. This consisted of the type of each interaction with the object and the AI, as well as the time at which it was performed. Additionally, a screencast of the participant’s perspective in VR was recorded in each modeling session.

**Between modeling sessions:** After each modeling session, the user was asked to fill out a questionnaire. We base our questionnaire on several other questionnaires from the HCI and Psychology:

- The Mixed-Initiative Creativity Support Index (MICSI) [31] measures the degree to which a co-creative system supports a user in creative tasks. For the purposes of our questionnaire, we adopted eight items, some of them as slightly modified versions: the two items on *Enjoyment* and one item each on *Alignment* and *Agency* – all measured on 7-point Likert scales – and one item each on *Contribution*, *Satisfaction*, *Surprise*, and *novelty* – all measured on 7-point non-Likert scales.
- Rezwana and Maher [55] modified the Creativity Support Index (CSI) [10], which the MICSI was based on, to include two items evaluating *partnership* and *communication* in human-AI collaboration. We adopted both of those items in our questionnaire.
- The User Experience Questionnaire (UEQ) [30] measures general user experience for interactive systems and consists of items on seven-point adjective scales. We adopted five of those items: One from the *efficiency* scale, two from the *Dependability* scale, and two from the *Stimulation* scale.

- The Inclusion of Other in the Self (IOS) Scale [3] is a pictorial single-item scale measuring *closeness* between the respondent and another person. Though the AI in our co-creative system is not a person, we are nevertheless interested in whether closeness, when applied to a co-creative AI partner, is affected by our visualization aspects. We, therefore, included this item in our questionnaire.
- The Social Presence in Gaming Questionnaire (SPGQ) [15] was developed for digital games and measures players’ awareness of and involvement with their co-players. Though our co-creative system is not a game, we are interested in whether our visualization aspects influence the degree to which users perceive a co-creative AI as a separate presence that influences and can be influenced by the user. This is why we adopted all eight items from the *Behavioral Engagement Scale* of this questionnaire, all of which are answered with 5-point intensity scales.

In addition to the items sourced from the above-mentioned questionnaires, we included three Likert items on AI competence, contribution value, and AI creativity, amounting to 27 questions in total<sup>1</sup>.

**At the end of the experiment:** In a semi-structured interview, participants answered questions about their overall experience with the system and their perception of the finished 3D objects and the AI (see Table 1).

### 3.3 Apparatus

The study was carried out using a Windows 11 computer with an NVIDIA GeForce GTX 1080 graphics card and a Meta Quest Pro head-mounted display (HMD) connected to the computer via Quest Link. The software was run directly from its Unity environment. Participants were given both of the Meta Quest Pro controllers and were free to walk around in a room-scale tracked space. Questionnaire replies were filled out on a separate computer.

### 3.4 Procedure

After welcoming participants and before starting the study, participants filled out an informed consent form and completed a demographic information survey. We then provided an overview of the study procedure and system usage guidelines. Next, we instructed the participants on using the HMD and controllers and let

<sup>1</sup>See the supplementary materials for a list of all items

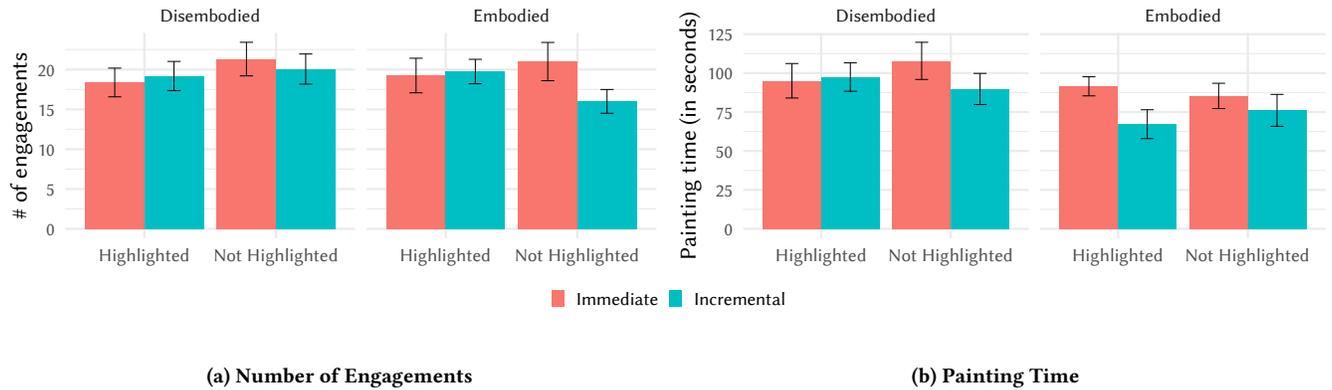


Figure 4: The number of engagements and painting time in our experiment. All error bars depict the standard error.

them familiarize themselves with the VR interactions in a virtual room by testing selection and painting operations on a sphere. Once ready, participants began their first modeling session, tasked to *"be creative and work together to create a new version of this model"*, emphasizing a focus on making the given object look different from its default state but letting participants choose their own design goals. We timed a 5-minute session (visualized by a virtual clock) and automatically disabled all interactions once time ran out. Participants were invited to voice any thoughts they had during the sessions. Participants then filled out questionnaires outside the VR environment before repeating this process for a total of eight conditions. After completing all sessions, we conducted a semi-structured interview with each participant. In total, the study lasted around 90 minutes. All participants were given cake as compensation for their time.

### 3.5 Participants

We recruited 16 participants (7 identified as female, 9 as male, and 0 as non-binary). Ages ranged from 20 to 30, with a mean age of 24 ( $SD = 3$ ). Out of all the participants, five had no prior experience with VR, while nine had tried it before, and two had used it more often. Nine participants had no experience with 3D modeling, three had tried it before, and four had done it more often. Most participants had interacted with AI in the form of Large Language Models before. In contrast, half of the participants had used AI for image generation before, and only three participants had experience with AI for 3D object generation.

## 4 Results

In this section, we report our empirical study's quantitative and qualitative results.

### 4.1 Quantitative Results

We analyzed the influence of our independent variables on the dependent variables using inferential statistics.

**4.1.1 Number of Engagements.** We analyzed the number of engagements to measure the participant's interest in contributing to the model (see Figure 4a). We defined the number of engagements as

the total number of selections, modifications, and undo actions that a user performed. We found results ranging from  $\mu = 16.0$ ,  $\sigma = 6.0$  (with HIGHLIGHTING and INCREMENTAL VISUALIZATION without EMBODIMENT) to  $\mu = 21.3$ ,  $\sigma = 8.4$  (without any ai contribution visualization).

For statistical analysis, we fitted Poisson regression models and applied Type III Wald chi-square tests for significance testing. The analysis indicated a significant ( $\chi^2(1) = 6.29$ ,  $p < .05$ ) main effect of HIGHLIGHTING on the number of engagements that was, however, not supported by post-hoc tests.

Further, we found a significant ( $\chi^2(1) = 6.68$ ,  $p < .01$ ) interaction effect between HIGHLIGHTING and INCREMENTAL VISUALIZATION. Without HIGHLIGHTING, the addition of an INCREMENTAL VISUALIZATION significantly ( $< .05$ ) reduced the number of engagements. With HIGHLIGHTING, however, the addition of INCREMENTAL VISUALIZATION seemingly increased the number of engagements (*n.s.*)

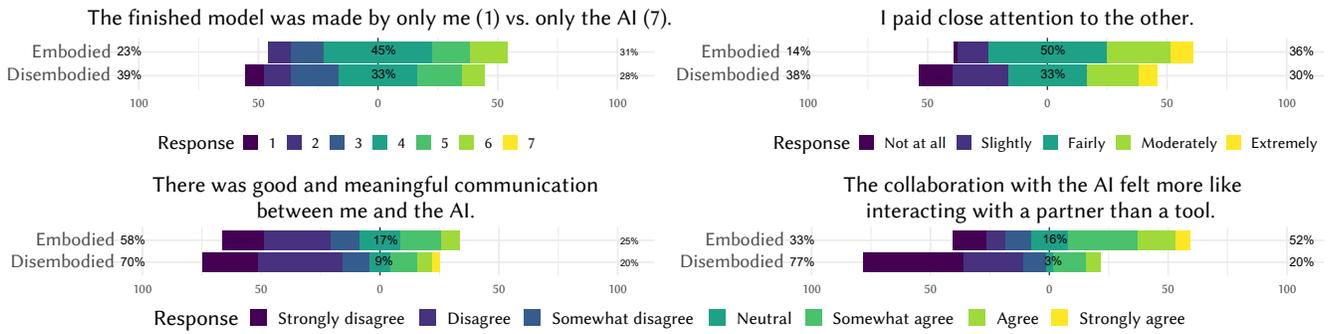
Finally, we found a significant ( $\chi^2(1) = 4.86$ ,  $p < .05$ ) interaction effect between HIGHLIGHTING and EMBODIMENT. Here, we found the same trend where without HIGHLIGHTING, EMBODIMENT reduced the number of engagements while increasing it with HIGHLIGHTING. However, these results were not significant.

**4.1.2 Painting Time.** As another measure for the engagement of the participants, we measured the time participants spent painting parts of the model (see Figure 4a). We found painting times ranging from  $\mu = 67.2s$ ,  $\sigma = 37.1s$  (with all ai contribution visualizations) to  $\mu = 108.0s$ ,  $\sigma = 47.8s$  (without any ai contribution visualization).

We analyzed the data using 3-way repeated measures ANOVA. We tested the data for violations of normality and sphericity assumptions using Shapiro-Wilk's and Mauchly's tests and found no violations.

The analysis indicated a significant ( $F_{1,15} = 4.926$ ,  $p < .05$ ) main effect of the INCREMENTAL VISUALIZATION on the painting time. Post-hoc tests confirmed significantly ( $p < .05$ ) higher painting times without the INCREMENTAL VISUALIZATION.

Further, the analysis indicated a significant ( $F_{1,15} = 9.349$ ,  $p < .01$ ) main effect of the EMBODIMENT on the painting time. Again, post-hoc tests confirmed significantly ( $p < .01$ ) higher painting times without the EMBODIMENT.



**Figure 5: Significant effects of embodiment on attributing contributions to the AI, perceived communication and partnership, and attention paid toward the AI. The plots show the percentages of participants who gave a response to a question, grouped by whether embodiment was enabled in the corresponding modeling session or not.**

## 4.2 Questionnaire Results

We analyzed the responses of our participants using the Aligned-Rank Transform (ART) procedure as proposed by Wobbrock et al. [65] followed by ART-C [21] corrected using Tukey’s method for post-hoc comparisons where appropriate. Figure 5 depicts some of the most interesting effects found in our analysis.

**4.2.1 Perception of AI.** We assess the perception of AI in general by analyzing the questions on predictability, support, efficiency, and competence. None of the visualization aspects significantly affected predictability, efficiency, or competence. However, in conditions with embodiment, participants rated the system as significantly more supportive ( $F(1, 15) = 5.74, p < .05$ ).

**4.2.2 Perception of AI output.** We assess the perception of AI output by analyzing the questions on unexpectedness, unusualness, value, and creativity. There were no significant differences in ratings for unusualness, value, or creativity for any visualization aspects. However, in conditions with incremental visualization, participants rated the finished 3D model as significantly more unexpected ( $F(1, 15) = 7.52, p < .05$ ).

**4.2.3 Collaborative experience.** We assess the collaborative experience by analyzing the partnership, communication, alignment, and agency questions. There were no significant differences in ratings for alignment or agency. However, with embodiment, participants rated communication significantly higher ( $F(1, 15) = 26.19, p < .001$ ). Additionally, participants rated partnership lower when highlighting was enabled ( $F(1, 15) = 6.18, p < .05$ ) and higher when embodiment was enabled ( $F(1, 15) = 23.5, p < .001$ ). We also found a significant interaction of highlighting and embodiment concerning this measure ( $F(1, 15) = 11.02, p < .01$ ). Post-hoc tests do not show significant differences in ratings between the individual variable combinations.

**4.2.4 Appeal.** We assess the overall system appeal by analyzing the questions on excitement, motivation, affinity, and enjoyment. There were no significant differences in ratings for motivation and affinity. However, in conditions with highlighting, participants enjoyed using the system significantly less ( $F(1, 15) = 7.89, p < .05$ ). Additionally, we found a significant interaction of highlighting and

embodiment concerning the measure of excitement ( $F(1, 15) = 5.83, p < .05$ ). Post-hoc tests do not show significant differences in ratings between the individual variable combinations.

**4.2.5 Behavioral engagement.** We assess behavioral engagement by analyzing the questions from the Behavioral Engagement Scale. Only two questions showed significant differences between the enablement of any visualization aspects: The extent to which the AI affected what the participant did was larger when incremental visualization was enabled ( $F(1, 15) = 6.54, p < .05$ ), and the amount of attention participants paid to the AI was more significant with embodiment ( $F(1, 15) = 9.43, p < .01$ ). We also found a three-way interaction of all visualization aspects concerning the latter variable ( $F(1, 15) = 7.87, p < .05$ ); however, post-hoc tests do not show significant differences in ratings between the individual variable combinations.

**4.2.6 Closeness, satisfaction, and contribution.** We assess these factors by analyzing satisfaction, contribution, and closeness questions. There were no significant differences in closeness to the AI as measured by the IOS Scale. However, in conditions with highlighting, participants were significantly less satisfied with the finished 3D model ( $F(1, 15) = 5.06, p < .05$ ). In conditions with embodiment, participants attributed significantly more of the contribution to the finished 3D model to the AI ( $F(1, 15) = 9.19, p < .01$ ).

## 4.3 Qualitative Results

**4.3.1 Data Analysis.** We collected qualitative data in a final semi-structured interview. Our data analysis procedure is based on Braun and Clarke’s [7] step-by-step approach of thematic analysis. While two authors conducted the analysis, the results and intermediate steps were discussed among all authors. The analysis was performed based on the transcripts of the audio-recorded interviews. The interview responses were a mix of English and German. All quotes were translated into English by the authors, who are fluent in English and German. The two authors first familiarized themselves with the data by reading the transcripts. We used Delvetool<sup>2</sup> for coding the data. In the first analysis step, both authors read the

<sup>2</sup>Delvetool for qualitative data analysis <https://app.delvetool.com/>, last accessed: August 30, 2024

same subset of transcripts (8 transcripts/50%) separately and created initial codes inductively. They then discussed the codes and formed themes by grouping them using an online mind map tool<sup>3</sup>. After the initial theme generation, all authors discussed the themes before one of the authors coded the remaining transcripts using the updated codes. Some new codes were merged with the existing themes during the second coding stage. Finally, the two authors met again to discuss the final codes and update and name the themes.

**4.3.2 Codes and Themes.** Coding the first half of the transcripts produced 103 codes. They were then grouped into 21 initial summary codes. Before coding the second half of the transcripts, we discussed and summarised the codes into 73 unique codes. Next, the second half of the transcript was coded, during which we added 15 codes to the coding set and merged them with the existing ones into 20 summary codes. We then formed the following four themes based on the summary codes<sup>4</sup>.

**4.3.3 Task Division and Control. Participants experienced varying levels of control over the AI, ranging from dependency to independence and different hierarchical levels.** One key aspect discussed by participants was how to divide tasks between themselves and the AI, particularly concerning the level of control each would have (or desire to have) over the task. Their comments revealed a relationship axis involving a degree of dependence/independence and, in the case of dependence, a hierarchy of interaction. The participants' comments were quite diverse, showing no standard agreement on the type of control they felt over the AI. Some participants (P13, P15) mentioned they did not feel they were working with the AI. Others noted that they felt like they were co-working with the AI but not collaborating (P8): *"It felt more like ... when you're [...] doing exercise sheets, and you have a partner who never talks to you and just hands something in without you even noticing."* Some participants also stated that they ignored the input from the AI (P15). Although the AI's input was indeed predetermined, some users felt in control of its actions and described a hierarchical relation (P1). They viewed the AI as a tool rather than a collaborator or co-worker. In this context, participants felt that the AI should support the user. The third relationship we observed was that of a partner, where participants felt they were pursuing a common goal with the AI. While many participants did not note any particular influences of embodiment on the relationship, P5 explained that they felt that collaboration was more relevant to them than working with an embodied AI. Finally, some participants described feeling that the AI was in control and that their decisions were overruled by it (P2, P8).

Participants also wished for different types of relationships when asked for improvements or desired changes in the interaction. Some participants suggested a form of task sharing with the AI, where both would choose from a set of options (P4). Other participants desired a more hierarchical way of control, where they wanted to give commands to the AI (P1, P2), while others were okay with the AI working on its own as long as the user could make the final decision about accepting or rejecting its decisions (P2): *So, [the AI] should be able to try out things, but in any case, it should not*

*override my decisions.* These wishes are closely related to the desire for better communication between AI and the user (P3, P6). For example, the AI should indicate whether they understood the user's request and whether the requested action was possible (P3). P2 also indicated they would like to receive feedback from the AI about its perception of the task.

In addition to spoken or verbal feedback, participants indicated that they would like to interact with the AI through different modalities and meta-level instructions, for example, by pointing at things or by making general style requests that are then performed by the AI (P2: *"for example I can write: I want to make a goth house, and then the AI automatically suggests goth windows and so on."*).

**4.3.4 AI as Enabler of New Experiences. Interacting with the AI sparked creativity among participants and enabled new experiences.** We observed that some participants felt the AI enabled them to have more creative experiences. For example, participants liked the AI's suggested modifications (P1). P8 liked that these suggestions guided them towards a satisfying result, and P7 appreciated that the AI suggestions led to new inputs: *[...] so when they [the AI] conjured something new, then, of course, one got new input.* P2 also mentioned that they felt inspired by the AI: *[...] the things from the AI also inspired me during the process.* Some participants even mentioned that they felt the AI helped them to be more creative and open (P7, P8), stating that they were happy with the result, although they, for example, chose a color scheme that they would not have gone for themselves.

**4.3.5 Emotional Experience with AI. Participants had both positive and negative experiences with the AI.** Participants generally found the experience interesting (P2, P6, P7). Their opinions and experiences with the AI were both positive and negative. On the positive side, participants felt that the AI created funny results and a funny experience in general (P3, P4). On the negative side, participants mentioned that they were confused about the AI and its created outcome (P2). They also did not like the slow response time or waiting for the AI to finish their modifications (P4, P5). Finally, some participants felt surprised (P3, P4) by the suggestions or distracted by the AI output (P1, P3).

**4.3.6 Experiences with Embodied AI. The embodied AI evoked a variety of emotions.** The participants experienced quite diverse emotions when interacting with the embodied AI. Some found it annoying because they had to wait for the embodied AI to finish the animation (P5) and disliked it being slow (P3, P4, P5, P7). While some participants liked to watch the AI, others felt it was distracting (P3). However, participants also felt curiosity when working with the embodied AI (P8).

Despite these negative feelings, participants also expressed positive emotions towards the AI. For example, they perceived it to be helpful (P1), cute (P2), or funny (P2). Participants also made comments about the appearance of the embodied AI. Some participants found it uncanny and weird (P3). Some participants specifically criticized the avatar's permanent smiling expression, explaining that *"smiling for no reason is creepy"* (P11) and mentioned that they did not prefer the AI to be human-like (P7). Indeed, several participants noted that they would like to be able to customize the appearance of the embodied AI (P3). P5 found it even scary when seeing the AI

<sup>3</sup>Miro tool for online mind maps: <https://miro.com/>, last accessed: August 30, 2024

<sup>4</sup>See the mind map of the codes in the supplementary materials

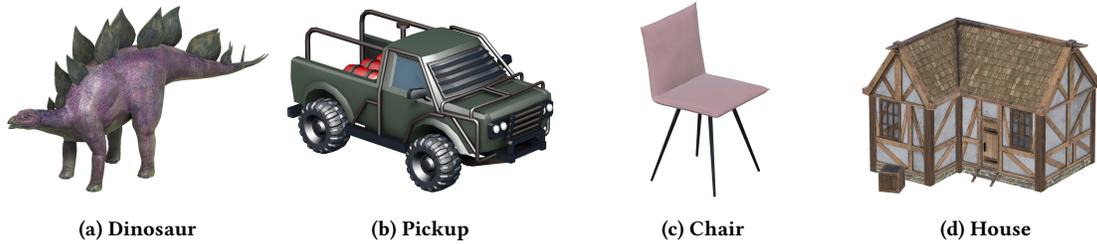


Figure 6: The four modifiable objects in their default form before any modifications.



Figure 7: A selection of finished models designed by participants, each influenced differently by the AI or its contributions.

for the first time: "Yes, I would say that during the iterations, I got used to it, but all in all, I was quite scared when I saw it the first time". However, we also observed positive interaction effects. P1 stated that "changes felt more tangible when done by the embodied AI". P4 felt that social conventions applied to the interaction with the AI: "it was nicer to wait when at least there was this person who made something", and P7 stated that the embodied AI helped not to feel alone. Overall, some participants felt a connection to the embodied AI. P7 even said they missed the embodied AI when it disappeared for some of the iterations. Finally, while some participants did not like the slow version of the continuous interaction mode, they felt it was more acceptable when the AI was embodied versus when it was not.

#### 4.4 Participants' Approach and Task Limitations

This section summarizes how participants generally approached the task. It also describes the limitations introduced by the study task unrelated to the human-AI interaction.

Almost all participants divided their modeling sessions into two stages: First, they searched for AI modifications and let the AI perform all the options they could find, undoing any modifications they were not happy with. The second stage was typically spent entirely on painting. In a few cases, participants were satisfied with the model before the five minutes had elapsed and spent the remaining time mostly just looking at their creations. Most participants did not discover all possible modifications that could be performed on the four models.

Participants often criticized the fact that only limited modifications were possible for the given models and that potential changes did not vary when a model appeared for a second time. Some participants hoped to be able to influence the AI's modifications through

their behavior. P11 voiced the name of the object that he was hoping to appear in a modification spot. At the same time, some participants tried to elicit specific modifications using the colors they painted with, and P8 even wrote a modification prompt on the object's surface itself.

Participants were also bothered that it could not be re-instantiated once they had undone a modification. They moreover criticized several aspects of the painting mechanism, such as the limited amount of colors, the lack of an eraser, and the issue that if a model had repeating textures, painting on one part of a model could make a copy of the painted stroke appear at another location on the model that used the same texture.

Some exemplary images of finished models created by participants during the modeling sessions are depicted in Figure 7.

## 5 Discussion and Implications

In this section, we revisit our quantitative and qualitative analysis results to interpret them and derive implications for the future design of co-creative VR systems. Additionally, we discuss the generalizability and limitations of this work.

### 5.1 Highlighting does not alleviate the confusion caused by AI actions

We found the highlighting to decrease perceived partnership, enjoyment, and satisfaction with the finished model. At the same time, highlighting did not affect the system's predictability, the perceived communication between humans and AI, the users' attention toward the AI, or the clarity of AI intent, which we expected to be most likely affected by highlighting. On the qualitative side, the observational analysis also did not indicate that highlighting object parts caused differences in participant behavior. This result is supported by participants never referring to highlighting during the

modeling sessions or interviews. Our quantitative and qualitative findings suggest that highlighting is not as much of a mitigator against user confusion about AI intentions as expected and might lead to a more negative overall experience. Future research should investigate whether these adverse effects can be replicated or if they occur similarly with different implementations of a highlighting variable.

**Design implication for highlighting** For co-creation use cases where users create one complete object at a time, like in our use case, we recommend not highlighting changes. In particular, in combination with embodied AI avatars, it introduced more negative effects than it contributed to user satisfaction. Future work has to show if highlighting proves more useful for co-creation use cases of dynamic objects.

## 5.2 Incremental visualization increases the attention to AI actions but introduces unexpected side effects

Since most participants watched all incremental visualization processes from beginning to end, it is unsurprising that this variable significantly affected the attention participants paid to the AI. However, we also expected that the visualization aspect would affect the perceived efficiency of the system, which our quantitative analysis did not confirm. Considering that participants were likely aware that the AI was not generating model parts "on the spot," they likely interpreted this process as the loading time of the overall setup. This interpretation also explains why the incremental visualization did not affect predictability, competence, unusualness, value, creativity, or contribution, as we hypothesized.

Further, this lack of impact on the perception of the AI and its outputs could be caused by the specific incremental visualization we used. This incremental visualization was inspired by AI drawing agents that build images stroke-by-stroke. Transferring this concept directly into 3D space resulted in an AI agent that slowly draws individual horizontal lines of different colored pixels, which does not correspond to our experiences in the physical world. A possibly more appropriate 3D equivalent could be the molding or carving of primitives. Future work should investigate whether alternative ways of incrementally visualizing 3D modifications affect observers' perceptions of the system differently.

Another surprising finding is that incremental visualization increased the perceived unexpectedness of AI outputs. This finding contrasts with our hypothesis that incremental visualization would lead to a decrease in unexpectedness rather than an increase, as the longer visualization duration gave participants more time to familiarize themselves with the modified version of an object part. We attribute this reverse effect to participants having had more time to engage with the version of an object part due to the extra time (e.g., to think about what the fully instantiated version of the object part would look like or what kind of change they would have expected instead at this point). As a result, they perceived the AI outputs as more unexpected than in conditions where the output appeared immediately, and participants had no choice but to immediately accept them as the new version of an object part.

Additionally, our interviews revealed that incremental visualization was perceived more positively when an avatar appeared

to carry out the incremental process. We attribute this to a more sympathetic attitude towards an AI with characteristics similar to the user's. An AI that does not have a physical or virtual body might be expected not to be limited by time and space to the same degree as a human and might, therefore, seem more capable of performing actions that would be impossible for a human user. On the other hand, an embodied AI might appear to be constrained by its surroundings in similar ways to a human. It would make sense for a human user to be more sympathetic toward AI, which "has to" slowly work on its task until it is finished, and therefore, to feel more like the AI is equal to the user. One participant also applied this idea to the "tool" that the AI used in the incremental process, which was a line that looked similar to the pointer ray that participants used to interact with the VR environment. It would be interesting for future work to study whether the effects of this similarity of tools only apply when a similarity of processes (i.e., an incremental process) is also given or whether the same effect can be observed given an instantaneous visualization.

**Design implication for incremental visualization** We recommend using an immediate creation of content to co-create complete objects, where the creation process is unimportant. We ground this in our findings where incremental generation did not reduce participants' feelings of unexpectedness, while at the same time, the visualization took longer. However, we recommend incremental visualization for use cases where the process of creating objects is important and where users should observe the generation more closely, and time is no relevant factor.

## 5.3 Embodiment not only affects the relationship to the AI but also to the 3D model created

Embodiment was the visualization aspect that participants had the most substantial feelings about. The divergence in comfort levels regarding the presence of the avatar was unexpected: To avoid an avatar design that created feelings of uncanniness in users [46], we intentionally chose an avatar for our AI that was heavily caricatured and only distantly resembled the form of a human, as well as one that had a friendly, child-like appearance. We do not know whether participants' discomfort with the avatar was indeed a case of the uncanny valley, whether it was only due to the surprise participants felt when they were faced with the avatar for the first time, or whether it can be attributed to characteristics that are specific to co-creative systems.

The divergence in participants' discomfort with the avatar is also a possible explanation for why embodiment did not significantly affect ratings of factors measuring appeal. Nevertheless, our qualitative analysis indicated an improvement in the perceived relationship between users and the AI when the avatar was present, manifested by an increased tangibility of the relationship, the feeling of being less alone, and more empathy expressed towards the AI. This result is consistent with quantitative data pointing to an effect of embodiment on partnership and AI-human communication.

As hypothesized, AI embodiment did affect the extent to which participants felt that the AI contributed to the finished 3D model. This result suggests that factors external to the AI functionality can manipulate users' perceptions of who created an artifact. This

finding is noteworthy considering ongoing discussions around ownership of AI-generated content, as the question of who created an artifact is often one of the most influential in deciding who owns it [56]. Rezwana and Maher [56] found that this decision of ownership, and additionally users' stances on other AI-related moral dilemmas, is also greatly influenced by the perception of a co-creative AI partner as either a collaborator or as a tool. Considering our finding that the embodiment of our AI significantly affected participants' perceptions of the AI as a partner rather than a tool, we argue that such ethical questions should be kept in mind when deciding whether to integrate an AI avatar into a co-creative system. This result is in line with Buschek et al. [8] listing unclarity of ownership as a potential pitfall in the design of co-creative systems.

**Design implication for embodiment** We recommend using an embodied AI for most co-creation use cases, as it increases the perceived support of the system, leads to a higher perceived partnership and communication, and leads to users paying more attention to the creation. However, users also felt reduced ownership of the outcome, which should be considered when designing a system. For use cases where users need a tool and the creative, collaborative aspect is less relevant, the non-embodied AI might be the appropriate choice, as it is faster, and users have higher ownership of the outcome.

## 6 Limitations and Future Work

Our work provides valuable insights and implications for the future design of co-creative systems for object generation in VR. In particular, we are confident that our technology-agnostic approach, which does not rely on current AI capabilities or technologies, will ensure that the results remain relevant for future systems, even if these are not constrained by the same technological limitations.

However, we acknowledge that our experiment focused on **specific AI representation modes**, and other variants or visualizations may yield different results. We selected these specific modes of representation because of their importance for human collaboration to explore their impact on users' perceptions of the co-creative process and to stimulate further research and alternative implementations.

Moreover, the **limited set of models** that participants were asked to modify was an extraneous variable, potentially affecting participants' experience with and perception of the system. Though we tried to minimize the effects of this variable by randomizing the order in which models were presented to each participant, we expect that with a small sample size such as ours, effects could not be avoided.

Also, the **limited number of possible AI modifications**, which remained constant during iterative modifications, could have led participants to the conclusion that their AI partner did not actually generate object parts but retrieved them from previously modeled parts. We hypothesize that this may have had an impact on the perception of AI and its results, as there was nothing about AI's capabilities and how its "generative" process changed between conditions.

Our study also did not explore the potential of a co-creative AI in **more complex, object-building scenarios**, such as simultaneous adjustments to multiple copies of the same object, adding and

manipulating entire environments, or deformation and distortions of the object mesh (cf. [1, 2, 60] on 3D shape editing and deforming). We deliberately opted for a constrained set of predefined actions to focus our contribution on the fundamental questions of co-creative interaction and to provide a better understanding of co-creative tools in VR for object-building that are more than one-shot tools. While we are convinced that this provides a robust foundation for future work, we acknowledge that the constrained set of predefined actions might limit the generalizability of our findings, and future work in this domain is necessary. Exploring how the AI's functionality scales with increased diversity and complexity of tasks is an important direction for future research, particularly for understanding the full potential of AI in co-creative VR systems.

Moreover, the efficiency of the co-creative tool itself was not explicitly evaluated in our study. The results presented thus may be correlated with how effectively the co-creative system executed its tasks, potentially influencing users' perceptions of the system's creativity. As generative AI becomes more advanced and prevalent in co-creative processes, future research should examine the role of AI efficiency and its impact on users' experiences and outcomes in co-creative tasks. Such investigations could provide a deeper understanding of how system performance and responsiveness affect the perceived value and usability of co-creative systems.

Since we performed the study as a controlled experiment, we acknowledge limitations in **external validity** in favor of establishing cause-effect relationships to recognize the impacts of individual representation modes.

There are several interesting directions for **future work**: One important area is exploring how different methods of emphasizing the AI's contribution process might positively impact a user's experience with a co-creative system, including whether such methods can enhance user navigation and coordination during collaboration. Additionally, investigating various ways to incrementally visualize the AI's generation of a 3D object could reveal how these visualizations affect users' perceptions of the AI and its creative process. There is also a need to examine the specific characteristics of avatars for co-creative AI that may cause discomfort to users, as well as how the perceived closeness to AI agents is affected by their embodiment and which other factors might influence this perception. Finally, future research should also explore whether manipulating the visual appearance of an AI's embodiment can alter people's understanding and evaluation of the AI and its creations.

## 7 Conclusion

Research on co-creative systems has predominantly focused on how humans participate in the co-creative process, often overlooking how AI contributions are represented and perceived by users. This paper addresses this gap by examining how an AI's contributions can be effectively represented for co-creative object generation in VR. Through a Wizard-of-Oz study in a VR-based co-creative 3D object-generating environment, we investigated the effects on the co-creative collaboration experience for three key representation modes AI can use to communicate to the user: highlighting, incremental visualization, and embodiment of AI contributions. Finally, we derive implications for the design of tools for co-creative object-building: First, designers should avoid combining highlighting with

embodiment to increase user satisfaction. Second, while an embodiment increases perceived support, it also reduces ownership of the created objects. Third, designers should favor immediate visualization when the creation process is secondary.

## Acknowledgments

We would like to thank all participants and reviewers. This research was supported by the HumanE AI Network from the European Union's Horizon 2020 research and innovation program under grant agreement No 952026, and the Pioneer Centre for AI, DNRF grant number P1.

## References

- [1] Panos Achlioptas, Ian Huang, Minhyuk Sung, Sergey Tulyakov, and Leonidas Guibas. 2023. ShapeTalk: A Language Dataset and Framework for 3D Shape Edits and Deformations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 12685–12694.
- [2] Dimitris Angelis, Prodromos Kolyvakis, Manos Kamarianakis, and George Papiannakis. 2024. Geometric Algebra Meets Large Language Models: Instruction-Based Transformations of Separate Meshes in 3D, Interactive and Controllable Scenes. arXiv:2408.02275 [cs.CV] <https://arxiv.org/abs/2408.02275>
- [3] Arthur Aron, Elaine N. Aron, and Danny Smollan. 1992. Inclusion of Other in the Self Scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology* 63, 4 (1992), 596–612. <https://doi.org/10.1037/0022-3514.63.4.596>
- [4] Rahul Arora, Rubaiat Habib Kazi, Tovi Grossman, George Fitzmaurice, and Karan Singh. 2018. SymbiosisSketch: Combining 2D & 3D Sketching for Designing Detailed 3D Objects in Situ. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–15. <https://doi.org/10.1145/3173574.3173759>
- [5] Joseph Bates. 1994. The role of emotion in believable agents. *Commun. ACM* 37, 7 (July 1994), 122–125. <https://doi.org/10.1145/176789.176803>
- [6] Amy L. Baylor. 2009. Promoting motivation with virtual agents and avatars: role of visual presence and appearance. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364, 1535 (Dec. 2009), 3559–3565. <https://doi.org/10.1098/rstb.2009.0148>
- [7] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. <https://doi.org/10.1191/1478088706qp0630a> arXiv:<https://www.tandfonline.com/doi/pdf/10.1191/1478088706qp0630a>
- [8] Daniel Buschek, Lukas Mecke, Florian Lehmann, and Hai Dang. 2021. Nine Potential Pitfalls when Designing Human-AI Co-Creative Systems. <https://doi.org/10.48550/arXiv.2104.00358> arXiv:2104.00358 [cs].
- [9] Jeff Butterworth, Andrew Davidson, Stephen Hench, and Marc. T. Olano. 1992. 3DM: a three dimensional modeler using a head-mounted display. In *Proceedings of the 1992 symposium on Interactive 3D graphics - SI3D '92*. ACM Press, Cambridge, Massachusetts, United States, 135–138. <https://doi.org/10.1145/147156.147182>
- [10] Erin Cherry and Celine Latulipe. 2014. Quantifying the Creativity Support of Digital Tools through the Creativity Support Index. *ACM Transactions on Computer-Human Interaction* 21, 4 (Aug. 2014), 1–25. <https://doi.org/10.1145/2617588>
- [11] James H. Clark. 1976. Designing surfaces in 3-D. *Commun. ACM* 19, 8 (Aug. 1976), 454–460. <https://doi.org/10.1145/360303.360329>
- [12] Simon Colton. 2012. The Painting Fool: Stories from Building an Automated Painter. In *Computers and Creativity*, Jon McCormack and Mark d'Inverno (Eds.). Springer, Berlin, Heidelberg, 3–38. [https://doi.org/10.1007/978-3-642-31727-9\\_1](https://doi.org/10.1007/978-3-642-31727-9_1)
- [13] Nicholas Davis. 2021. Human-Computer Co-Creativity: Blending Human and Computational Creativity. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment* 9, 6 (Jun. 2021), 9–12. <https://doi.org/10.1609/aiide.v9i6.12603>
- [14] 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*. ACM, Sonoma California USA, 196–207. <https://doi.org/10.1145/2856767.2856795>
- [15] Yvonne A.W. de Kort, Wijnand A. IJsselstein, and Karolien Poels. 2007. Digital games as social presence technology : development of the social presence in gaming questionnaire (SPGQ). In *Proceedings of the 10th Annual International Workshop on Presence, October 25-27, 2007*, L. Moreno (Ed.). Starlab, 195–203.
- [16] Fernanda De La Torre, Cathy Mengying Fang, Han Huang, Andrzej Banburski-Fahey, Judith Amores Fernandez, and Jaron Lanier. 2024. LLMR: Real-time Prompting of Interactive Worlds using Large Language Models. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 600, 22 pages. <https://doi.org/10.1145/3613904.3642579>
- [17] Michael F. Deering. 1995. HoloSketch: a virtual reality sketching/animation tool. *ACM Transactions on Computer-Human Interaction* 2, 3 (Sept. 1995), 220–238. <https://doi.org/10.1145/210079.210087>
- [18] Trien V. Do and Jong-Weon Lee. 2010. 3DARModeler: a 3D Modeling System in Augmented Reality Environment. *International Journal of Mathematical and Computational Sciences* 4, 3 (March 2010), 377–386. <https://publications.waset.org/14381/3darmodeler-a-3d-modeling-system-in-augmented-reality-environment>
- [19] Tomás Dorta, Gokce Kinayoglu, and Michael Hoffmann. 2016. Hyve-3D and the 3D Cursor: Architectural co-design with freedom in Virtual Reality. *International Journal of Architectural Computing* 14, 2 (June 2016), 87–102. <https://doi.org/10.1177/1478077116638921>
- [20] Dina El-Zanfaly, Yiwei Huang, and Yanwen Dong. 2023. Sand-in-the-loop: Investigating embodied co-creation for shared understandings of generative AI. In *Designing Interactive Systems Conference*. ACM, Pittsburgh PA USA, 256–260. <https://doi.org/10.1145/3563703.3596652>
- [21] Lisa A. Elkin, Matthew Kay, James J. Higgins, and Jacob O. Wobbrock. 2021. An Aligned Rank Transform Procedure for Multifactor Contrast Tests. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) (UIST '21). Association for Computing Machinery, New York, NY, USA, 754–768. <https://doi.org/10.1145/3472749.3474784>
- [22] Hesham Elsayed, Mayra Donaji Barrera Machuca, Christian Schaarschmidt, Karola Marky, Florian Müller, Jan Riemann, Andrii Matvienko, Martin Schmitz, Martin Weigel, and Max Mühlhäuser. 2020. VRSketchPen: Unconstrained Haptic Assistance for Sketching in Virtual 3D Environments. In *26th ACM Symposium on Virtual Reality Software and Technology*. ACM, Virtual Event Canada, 1–11. <https://doi.org/10.1145/3385956.3418953>
- [23] Rao Fu, Xiao Zhan, Yiwen Chen, Daniel Ritchie, and Srinath Sridhar. 2023. ShapeCrafter: A Recursive Text-Conditioned 3D Shape Generation Model. <https://doi.org/10.48550/arXiv.2207.09446> arXiv:2207.09446 [cs].
- [24] Christian Guckelsberger, Anna Kantosalo, Santiago Negrete-Yankelevich, and Tapio Takala. 2021. Embodiment and Computational Creativity. <http://arxiv.org/abs/2107.00949> arXiv:2107.00949 [cs].
- [25] Amir Hertz, Or Perel, Raja Giryes, Olga Sorkine-Hornung, and Daniel Cohen-Or. 2022. SPAGHETTI: Editing Implicit Shapes Through Part Aware Generation. <https://doi.org/10.48550/arXiv.2201.13168> arXiv:2201.13168 [cs].
- [26] 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. <https://doi.org/10.1109/TVCG.2016.2518099>
- [27] Heewoo Jun and Alex Nichol. 2023. Shap-E: Generating Conditional 3D Implicit Functions. <http://arxiv.org/abs/2305.02463> arXiv:2305.02463 [cs].
- [28] Kangsoo Kim, Luke Boelling, Steffen Haesler, Jeremy Bailenson, Gerd Bruder, and Greg F. Welch. 2018. Does a Digital Assistant Need a Body? The Influence of Visual Embodiment and Social Behavior on the Perception of Intelligent Virtual Agents in AR. In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 105–114. <https://doi.org/10.1109/ISMAR.2018.00039>
- [29] Justin Kruger, Derrick Wirtz, Leaf Van Boven, and T. William Altermatt. 2004. The effort heuristic. *Journal of Experimental Social Psychology* 40, 1 (Jan. 2004), 91–98. [https://doi.org/10.1016/S0022-1031\(03\)00065-9](https://doi.org/10.1016/S0022-1031(03)00065-9)
- [30] Bettina Laugwitz, Theo Held, and Martin Schreppe. 2008. Construction and Evaluation of a User Experience Questionnaire. In *HCI and Usability for Education and Work*, Andreas Holzinger (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 63–76.
- [31] Tomas Lawton, Francisco J Ibarrola, Dan Ventura, and Kazjon Grace. 2023. Drawing with Reframer: Emergence and Control in Co-Creative AI. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*. ACM, Sydney NSW Australia, 264–277. <https://doi.org/10.1145/3581641.3584095>
- [32] Sunok Lee, Minji Cho, and Sangsu Lee. 2020. What If Conversational Agents Became Invisible?: Comparing Users' Mental Models According to Physical Entity of AI Speaker. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 3 (Sept. 2020), 1–24. <https://doi.org/10.1145/3411840>
- [33] Chenghao Li, Chaoning Zhang, Atish Waghvase, Lik-Hang Lee, Francois Rameau, Yang Yang, Sung-Ho Bae, and Choong Seon Hong. 2023. Generative AI meets 3D: A Survey on Text-to-3D in AIGC Era. <https://doi.org/10.48550/arXiv.2305.06131> arXiv:2305.06131 [cs].
- [34] Antonios Liapis, Georgios N. Yannakakis, and Julian Togelius. 2013. Sentient sketchbook : computer-assisted game level authoring. (May 2013). <https://www.um.edu.mt/library/oar/handle/123456789/29607>
- [35] Chen-Hsuan Lin, Jun Gao, Luming Tang, Towaki Takikawa, Xiaoohui Zeng, Xun Huang, Karsten Kreis, Sanja Fidler, Ming-Yu Liu, and Tsung-Yi Lin. 2023. Magic3D: High-Resolution Text-to-3D Content Creation. <https://doi.org/10.48550/arXiv.2211.10440> arXiv:2211.10440 [cs].
- [36] Yuyu Lin, Jiahao Guo, Yang Chen, Cheng Yao, and Fangtian Ying. 2020. It Is Your Turn: Collaborative Ideation With a Co-Creative Robot through Sketch. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*

- ACM, Honolulu HI USA, 1–14. <https://doi.org/10.1145/3313831.3376258>
- [37] Simo Linkola, Christian Guckelsberger, Tomi Männistö, and Anna Kantosalo. 2022. How Does Embodiment Affect the Human Perception of Computational Creativity? An Experimental Study Framework. <http://arxiv.org/abs/2205.01418> arXiv:2205.01418 [cs].
- [38] Jerry Liu, Fisher Yu, and Thomas Funkhouser. 2018. Interactive 3D Modeling with a Generative Adversarial Network. <https://doi.org/10.48550/arXiv.1706.05170> arXiv:1706.05170 [cs].
- [39] Yiren Liu, Si Chen, Haocong Cheng, Mengxia Yu, Xiao Ran, Andrew Mo, Yiliu Tang, and Yun Huang. 2023. How AI Processing Delays Foster Creativity: Exploring Research Question Co-Creation with an LLM-based Agent. <https://doi.org/10.48550/arXiv.2310.06155> arXiv:2310.06155 [cs].
- [40] Google LLC. 2016. Tilt Brush by Google. <https://www.tiltbrush.com/>
- [41] Pedro Lucas and Carlos Martinho. 2017. Stay Awhile and Listen to 3Buddy, a Co-creative Level Design Support Tool. In *Proceedings of the Eighth International Conference on Computational Creativity, ICC3 2017, Atlanta, Georgia, USA, June 19-23, 2017*, Ashok K. Goel, Anna Jordanous, and Alison Pease (Eds.). Association for Computational Creativity (ACC), 205–212. [http://computationalcreativity.net/icc32017/ICC3\\_17\\_accepted\\_submissions/ICC3-17\\_paper\\_12.pdf](http://computationalcreativity.net/icc32017/ICC3_17_accepted_submissions/ICC3-17_paper_12.pdf)
- [42] Zhaoliang Lun, Matheus Gadelha, Evangelos Kalogerakis, Subhransu Maji, and Rui Wang. 2017. 3D Shape Reconstruction from Sketches via Multi-view Convolutional Networks. <https://doi.org/10.48550/arXiv.1707.06375> arXiv:1707.06375 [cs].
- [43] Oscar Michel, Roi Bar-On, Richard Liu, Sagie Benaim, and Rana Hanocka. 2021. Text2Mesh: Text-Driven Neural Stylization for Meshes. <https://doi.org/10.48550/arXiv.2112.03221> arXiv:2112.03221 [cs].
- [44] Aryan Mikaeili, Or Perel, Mehdi Safaei, Daniel Cohen-Or, and Ali Mahdavi-Amiri. 2023. SKED: Sketch-guided Text-based 3D Editing. <https://doi.org/10.48550/arXiv.2303.10735> arXiv:2303.10735 [cs].
- [45] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. <https://doi.org/10.48550/arXiv.2003.08934> arXiv:2003.08934 [cs].
- [46] Masahiro Mori, Karl F. MacDorman, and Norri Kageki. 2012. The Uncanny Valley [From the Field]. *IEEE Robotics & Automation Magazine* 19, 2 (June 2012), 98–100. <https://doi.org/10.1109/MRA.2012.2192811>
- [47] Caterina Moruzzi. 2022. The (Artificial) Physicality of Creativity : How Embodiment Influences Perceptions of Creativity. In *Proceedings of the 13th International Conference on Computational Creativity, ICC3'22*. [https://computationalcreativity.net/icc322/wp-content/uploads/2022/06/ICC3-2022\\_8L\\_Moruzzi.pdf](https://computationalcreativity.net/icc322/wp-content/uploads/2022/06/ICC3-2022_8L_Moruzzi.pdf)
- [48] Changhoon Oh, Jungwoo Song, Jinhan Choi, Seonghyeon Kim, Sungwoo Lee, and Bongwon Suh. 2018. I Lead, You Help but Only with Enough Details: Understanding User Experience of Co-Creation with Artificial Intelligence. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–13. <https://doi.org/10.1145/3173574.3174223>
- [49] S Parise, S Kiesler, L Sproull, and K Waters. 1999. Cooperating with life-like interface agents. *Computers in Human Behavior* 15, 2 (March 1999), 123–142. [https://doi.org/10.1016/S0747-5632\(98\)00035-1](https://doi.org/10.1016/S0747-5632(98)00035-1)
- [50] Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2023. Generative Agents: Interactive Simulacra of Human Behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) (UIST '23). Association for Computing Machinery, New York, NY, USA, Article 2, 22 pages. <https://doi.org/10.1145/3586183.3606763>
- [51] M. T. Pearce and G. A. Wiggins. 2001. Towards a framework for the evaluation of machine compositions. In *Proceedings of the AISB '01 Symposium on Artificial Intelligence and Creativity in the Arts and Sciences*. AISB (Society for the Study of Artificial Intelligence and the Simulation of Behaviour), 22–32.
- [52] Ben Poole, Ajay Jain, Jonathan T. Barron, and Ben Mildenhall. 2022. DreamFusion: Text-to-3D using 2D Diffusion. <https://doi.org/10.48550/arXiv.2209.14988> arXiv:2209.14988 [cs, stat].
- [53] Patrick Reipschläger and Raimund Dachselt. 2019. DesignAR: Immersive 3D-Modeling Combining Augmented Reality with Interactive Displays. In *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces*. ACM, Daejeon Republic of Korea, 29–41. <https://doi.org/10.1145/3343055.3359718>
- [54] Jeba Rezwana and Mary Lou Maher. 2022. Designing Creative AI Partners with COFI: A Framework for Modeling Interaction in Human-AI Co-Creative Systems. *ACM Transactions on Computer-Human Interaction* (Feb. 2022), 3519026. <https://doi.org/10.1145/3519026>
- [55] Jeba Rezwana and Mary Lou Maher. 2022. Understanding User Perceptions, Collaborative Experience and User Engagement in Different Human-AI Interaction Designs for Co-Creative Systems. In *Creativity and Cognition*. ACM, Venice Italy, 38–48. <https://doi.org/10.1145/3527927.3532789>
- [56] Jeba Rezwana and Mary Lou Maher. 2023. User Perspectives on Ethical Challenges in Human-AI Co-Creativity: A Design Fiction Study. In *Creativity and Cognition*. ACM, Virtual Event USA, 62–74. <https://doi.org/10.1145/3591196.3593364>
- [57] Mel Rhodes. 1961. An Analysis of Creativity. *The Phi Delta Kappan* 42, 7 (1961), 305–310. <http://www.jstor.org/stable/20342603>
- [58] Zifan Shi, Sida Peng, Yinghao Xu, Andreas Geiger, Yiyi Liao, and Yujun Shen. 2023. Deep Generative Models on 3D Representations: A Survey. <https://doi.org/10.48550/arXiv.2210.15663> arXiv:2210.15663 [cs].
- [59] Gravity Sketch. 2020. Transform the Industrial Design Workflow with Gravity Sketch. <https://www.gravitysketch.com/blog/articles/transform-the-industrial-design-workflow-with-gravity-sketch/>
- [60] Habib Slim and Mohamed Elhoseiny. 2024. ShapeWalk: Compositional Shape Editing Through Language-Guided Chains. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 22574–22583.
- [61] Jason Toff. 2017. Blocks: Easily Create 3D Objects in VR. <https://blog.google/products/google-ar-vr/blocks-easily-create-3d-objects-vr/>
- [62] Josh Urban Davis, Fraser Anderson, Merten Stroetzel, Tovi Grossman, and George Fitzmaurice. 2021. Designing Co-Creative AI for Virtual Environments. In *Creativity and Cognition*. ACM, Virtual Event Italy, 1–11. <https://doi.org/10.1145/3450741.3465260>
- [63] Gerold Wesche and Hans-Peter Seidel. 2001. FreeDrawer: a free-form sketching system on the responsive workbench. In *Proceedings of the ACM symposium on Virtual reality software and technology*. ACM, Baniff Alberta Canada, 167–174. <https://doi.org/10.1145/505008.505041>
- [64] Maximilian Wittmann and Benedikt Morschheuser. 2022. What do games teach us about designing effective human-AI cooperation? -A systematic literature review and thematic synthesis on design patterns of non-player characters.
- [65] Jacob O Wobbrock, Leah Findlater, Darren Gergle, and James J Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 143–146. <https://doi.org/10.1145/1978942.1978963>
- [66] Shing Ming Wong, Chien-Wen Chen, Tse-Yu Pan, Hung-Kuo Chu, and Min-Chun Hu. 2022. GetWild: A VR Editing System with AI-Generated 3D Object and Terrain. In *Proceedings of the 30th ACM International Conference on Multimedia*. ACM, Lisboa Portugal, 6988–6990. <https://doi.org/10.1145/3503161.3547733>
- [67] Michael Yin and Robert Xiao. 2024. Press A or Wave: User Expectations for NPC Interactions and Nonverbal Behaviour in Virtual Reality. *Proc. ACM Hum.-Comput. Interact.* 8, CHI PLAY, Article 333 (Oct. 2024), 25 pages. <https://doi.org/10.1145/3677098>
- [68] Tom Ziemke. 2003. What's that Thing Called Embodiment? *Proceedings of the Annual Meeting of the Cognitive Science Society* 25, 25 (2003). <https://escholarship.org/uc/item/60w6v9jz>