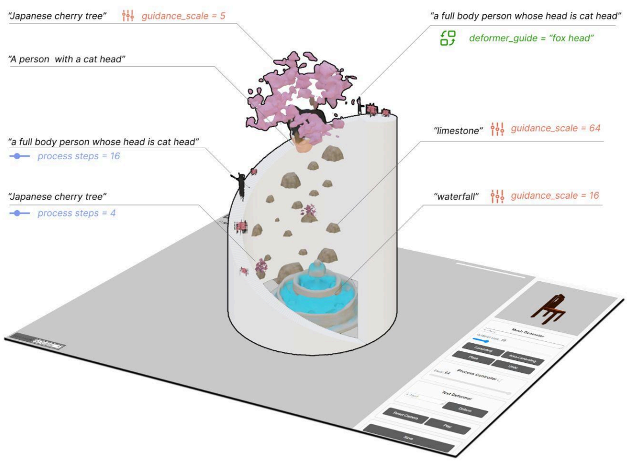
Shap-Explorer: Introducing Manipulable Text-to-3D Generation Into 3D Art Creation



By

Xing Hao

CARNEGIE MELLON UNIVERSITY

School of Architecture

College of Fine Arts

Thesis

Submitted in Partial Fulfillment of the requirements for the degree of

Master of Science in Computational Design

TITLE:

Shap-Explorer: Introducing Manipulable Text-to-3D Generation Into 3D Art Creation

AUTHOR:

Xing Hao

ACCEPTED BY ADVISORY COMMITTEE:

Daragh Byrne	Principal Advisor	DATE: May 8th, 2024

Acknowledgments

I would like to thank Professor Daragh Byrne for being my principal advisor and providing invaluable feedback, insightful recommendations, and patience throughout the research process. Without his suggestions and encouragement, I would not have been able to complete this thesis. I hold the same regard for Professor Daniel Cardoso Llach. His suggestions helped me identify the research direction and make further progress.

I thank my seminar classmates Hongfei and Semina for their feedback and encouragement. Their works also inspire me a lot. Thanks to Steve, Felicia, Zhenfang, Yumeng, Yujeong, and other CD cohorts, the time we spent together at CoDeLab was precious and unforgettable.

I am grateful to my family for their enduring support.

Shap-Explorer: Introducing Manipulable Text-to-3D Generation Into 3D Art Creation

Xing (Zain) Hao MSCD '24

Abstract

Generative Artificial Intelligence (GAI) technologies are increasingly employed by artists for their capability to rapidly generate intricate designs. While text-to-image (T2I) applications have been explored in art creation with constructive outcomes, the integration of text-to-3D (T23D) methodologies remains underexplored, primarily due to the knowledge gap between users and tool-makers. Developers aim to refine models for speed and accuracy, yet artists lack a steerable front-end tool to guide the 3D generation.

To bridge this gap, this research introduces Shap-Explorer, a tool that streamlines the use of T23D, enabling users to iteratively modify the output models through mesh generator and control the generation through process controller and text deformer. Through a preliminary user study, this research examines the affordance of T23D and the impact of enhanced interactions on the generative system, such as providing design alternatives.

Insights include, for example, how users can iteratively refine and adjust the generated 3D models to align with their creative vision into the manipulation capabilities of generative AI tools in the design workflow, offering a step forward in the interactive creation of 3D art.

Keywords: Text-to-3D, 3D art, design interactions, generative models

Committee:
Daragh Byrne

Table of Contents

1. Introduction

"… Our dissatisfactions and visions were related to a further set of boundaries as well, drawn between professional technology design and sites of technologies-in-use. More specifically, as researchers and developers we found ourselves cut off from prospective technology users at the same time that our enterprise was legitimized by them.[21]"

Lucy A. Suchman

1.1 Context

3D models and virtual environments are fundamental in modern art creation, serving multiple roles: as spaces for visualizing and interacting with artifacts, as tools in computer-aided design, and as mediums for simulating and augmenting reality. Their versatility makes them indispensable in the exploration and realization of artistic visions. The landscape of artistic creation is being transformed by Generative Artificial Intelligence (GAI) [1], particularly through its ability to generate multimodal artistic representations. While Text-to-Image (T2I) applications have seen significant exploration and adoption in arts and design [30], leading to substantial academic and practical interest, there remains a wide gap in research focused on other generative models, such as Text-to-3D (T23D). As Lucy A. Suchman suggests, the boundaries between technology design and its use are profound yet often overlooked, creating a division that can obscure rather than illuminate the potential of new tools [21].

Recent advancements in AI transforming 3D, such as the Instruct 3D-to-3D [31] and SKED [32], demonstrate the potential of guiding 3D generative models using additional textual or sketch-based inputs, bridging the gap between human creativity and AI 3D model generation. However, these technologies often emphasize technical precision over the quality of human-AI interaction. They maintain an end-to-end generative approach, utilizing alternative input methods to direct the model's output but often sidelining the artist's intuitive control over the generation process.

1.2 Problem

The integration of machine learning (ML) in art is more than a technological upgrade; it represents a fundamental shift in how art is conceived and created. Artists historically used tools to extend their capabilities. To illustrate, consider the process of sculpting. A sculptor starts with a vision and continuously interacts with the material, adjusting techniques based on the response of the medium. With ML, they now collaborate with systems that learn and generate autonomously [16]. Despite its potential, this shift brings challenges, primarily the opacity of ML algorithms. Artists often find themselves using these tools without a clear understanding of how their inputs are transformed into artistic outputs, leading to a potential disconnect between intention and creation. This gap—highlighted by Suchman's discourse on the boundaries within technology use [21]—underscores a critical need: moving from mere generation to truly creative processes that incorporate iterative, manipulable interactions. Furthermore, limitations in manipulation methods within Text-to-3D (T23D) technology, which heavily rely on language expressiveness and artists' perceptions, may not fully support diverse design intentions or iterative refinement akin to physical artifact creation.

This leads me to my **research question**:
How can text-to-3D be designed to enhance user control and interaction in the process of 3D art creation, and In what ways can manipulatable text-to-3D technology influence and potentially enhance 3D art creation?

1.3 Thesis Vision

My research is inspired by Erik Ulberg's perspective on computational design as a tool: "Computational Design is fundamentally a tool-making enterprise that critically examines the role of technical approaches applied to creative practice [10]." So, this research focuses on workflow and interaction improvement rather than technology development.

This thesis explores how T23D technology can enhance the 3D art creation process and improve interactions between the user and the technology, aiming to transform "generation" into "creation." Inspired by digital sculpting processes, this research provides opportunities for artists to modify artifacts iteratively. Additionally, it explores ways to enhance users' understanding of T23D processes, allowing for manipulation of the generation process.

"Shap-Explorer," the interface introduced in this research, bridges the gap between generative models and artistic creativity. Shap-Explorer offers a direct, user-friendly interface(see Section 4.2 for a full description of the tool) that enables artists to iteratively modify and refine 3D outputs, enhancing the transparency of the generative process and empowering artists to participate actively in the digital creation process. It redefines interactions between artists and T23D technologies by emphasizing manipulation over mere generation. Specifically, it provides *mesh generator* for interactive modification of 3D models and scenes and provides *process controller* and *text deformer* to control and iterate.

A user study is also conducted to test the usability and creativity of Shap-Explorer. From the study, I found that iterative modification and gradual control enhance user control over Text-to-3D generation, highlighting the potential for broader application in other generative tools. I also discuss the influence of generated models on design direction and the challenges users face in integrating these outcomes, emphasizing the need for improved usability. Furthermore, it explores trade-offs between time efficiency and design control, stylistic alignment, and user controllability in T23D generation, as well as the balance between AI autonomy and manual intervention. Limitations related to the web-based modeling tool's functionality, quality, and efficiency concerns in generative outputs, as well as the absence of alternative interaction methods, are also highlighted, suggesting areas for future improvement in generative tool design.

1.4 Outline

This thesis is organized into chapters that form a cohesive narrative of the research process. Beginning with the background chapter, it delves into the historical and technological evolution of 3D art, tracing the transition from traditional sculpting to AI-driven creation while also exploring Text-to-3D (T23D) technologies and Shap-E's pivotal role. Following this, the hypothesis chapter presents the thesis's core premise, proposing that integrating manipulatable T23D technology enhances artistic workflow by granting greater control over model generation and modification. Chapter 4 outlines the methods employed, detailing the development process of Shap-Explorer, a manipulable tool tailored for 3D art creation, covering pipeline integration, interface construction, and outlining plans for a user study. Subsequently, Chapter 5 delves into the preliminary user study, assessing Shap-Explorer's functionality, usability, and impact on creativity through interviews, surveys, and qualitative analysis. Moving forward, the discussion chapter engages in a thorough examination of research findings, focusing on generative tools' controllability, Text-to-3D interactions, design trade-offs, and limitations. The last chapter provides a succinct conclusion, summarizing contributions and proposing potential avenues for future research endeavors.

2. Background

This chapter explores the evolution of 3D art, detailing its journey from traditional sculpting techniques to the sophisticated use of artificial intelligence in art creation. It examines how these technological advancements have transformed artistic processes and practices. Following an overview of the historical and technological context, the chapter delves into the specifics of Text-to-3D technologies, focusing on the technical approaches that guide their development and application. A significant part of this discussion will center on Shap-E, a generative model that has been rigorously tested and applied within this research. The chapter will conclude by examining innovative manipulation methods designed for Text-to-3D systems, which were developed to enhance user control and facilitate the pilot study conducted as part of this thesis.

2.1 3D Art: From Physical to Digital

Three-dimensional art (3D art) refers to those art forms represented in the dimensions of height, breadth and depth [26]. Compared to two-dimensional art (in most cases, it refers to flat painting), 3D art can be perceived in-depth in addition to breadth and height. Although the history of 3D art, which is represented by sculptures, can be traced back thousands of years, significant developments occurred in the late 20th century due to the advent of computer graphics [26].

Before computers' advent, sculptures were the main 3D art form for centuries (Figure 2.1.2). The creation of a sculpture traditionally involves a sequential process, including the initial armature construction, sculpting or modeling, iterative refinement, and final post-processing to achieve the desired aesthetic and structural integrity [40].

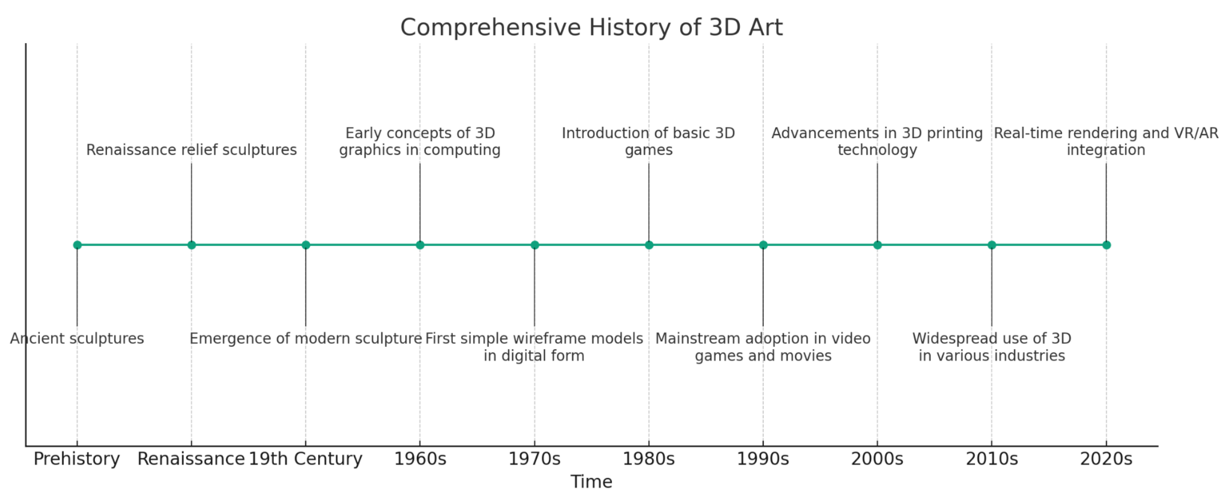


Fig 2.1.1. Timeline of Development of 3D Art. Drawn according to statements of [27].

The emergence of computers heralded a new era for 3D art, marked by the evolution from physical sculpting to digital 3D graphics [27]. Digital sculpting, akin to its traditional counterpart, follows a structured workflow involving blocking out basic forms, refining details, applying textures, rendering the final output, and conducting post-processing enhancements [41].

The integration of computer graphics technologies revolutionized 3D art by providing the ability to iterate rapidly, facilitating experimentation and exploration of creative ideas. The realm of digital 3D modeling and tools has been a focal point of extensive research and innovation. For example, studies have shown that integrating digital sculpting with sketching techniques can significantly enhance conceptual product design processes, fostering creativity and design iteration [28].

This evolution from physical to digital 3D art lays the groundwork for exploring the intersection of Computer-Aided Design (CAD) tools, Artificial Intelligence (AI), and Text-to-3D technologies in contemporary art creation processes, which will be further discussed in subsequent sections.

2.2 3D Art Creation: From Computer-Aided to AI-Powered

Computer-aided design (CAD) has been a transformative force in the realm of 3D art, empowering artists and designers with sophisticated tools to visualize, simulate, and construct intricate structures with precision and efficiency [13]. Initially developed for engineering and architectural applications, CAD technologies swiftly transitioned into the domain of digital art [43], enabling the creation of highly detailed sculptures, complex architectural models, and animated characters that would have been arduous or impractical to craft manually. CAD software offers a comprehensive suite of features, from initial sketching to final rendering, allowing artists to manipulate geometric shapes meticulously and expand the scope of their designs. In industries like animation, CAD systems play a vital role in crafting detailed 3D characters and environments, ensuring that every aspect aligns with the creators' vision. Moreover, the integration of CAD with 3D printing has facilitated the seamless transition of digital models into physical objects, democratizing art production and making it accessible to a wider range of creators.

As CAD technologies continue to evolve and reshape artistic workflows, the integration of Artificial Intelligence (AI) introduces new dimensions of creativity and efficiency to the 3D art creation process. Advancements in AI (See Figure 2.2.1) have significantly impacted visual arts, including 3D art, by augmenting the creation process. AI technologies such as generative models and neural networks have empowered artists to explore new creative avenues and streamline repetitive tasks. These AI-driven tools can assist in generating complex textures, refining designs, and even proposing novel artistic concepts based on learned patterns and data [42].

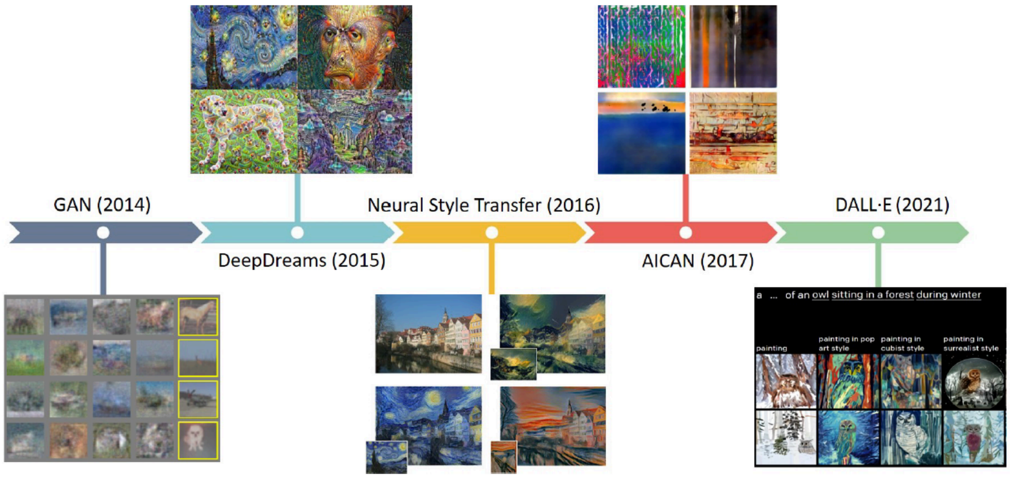


Fig 2.2.1. Illustration of the most important technological milestones that led to the current AI Art production. Image source: [42]

Agre's insights on reflective practice within AI resonate strongly in the design field, emphasizing the importance of considering broader implications, ethical dimensions, and societal impacts alongside technical advancements [6]. Cardoso's observations highlight the need for a balanced approach, encouraging designers to slow down computational processes to allow for interruptions and adjustments [9]. These perspectives underscore the evolving relationship between AI technology and human creativity, prompting practitioners to engage in thoughtful, responsible, and inclusive design practices.

2.3 Text-to-3D

Generative Artificial Intelligence (GAI) has recently gained significant attention in the realm of 3D art creation, particularly due to its multimodal capabilities that bridge various art forms through advanced learning models. Among the technologies emerging from GAI, text-to-3D (T23D) is one of the most practical for direct artistic application. This technology enables users to generate 3D models simply by entering text descriptions and transforming verbal inputs into visual outputs.

2.3.1 Overview of Text-to-3D

Key technologies facilitating the bridge between text-to-image and 3D model generation include Neural Radiance Fields (NeRF) [33], which construct detailed 3D models from 2D images (see Figure 2.3.1); CLIP (Contrastive Language-Image Pre-training) [34] by OpenAI, which learns visual concepts from natural language; and diffusion models [35], often pre-trained on vast datasets to generate detailed images (See Figure 2.3.2) that can be adapted for 3D modeling.

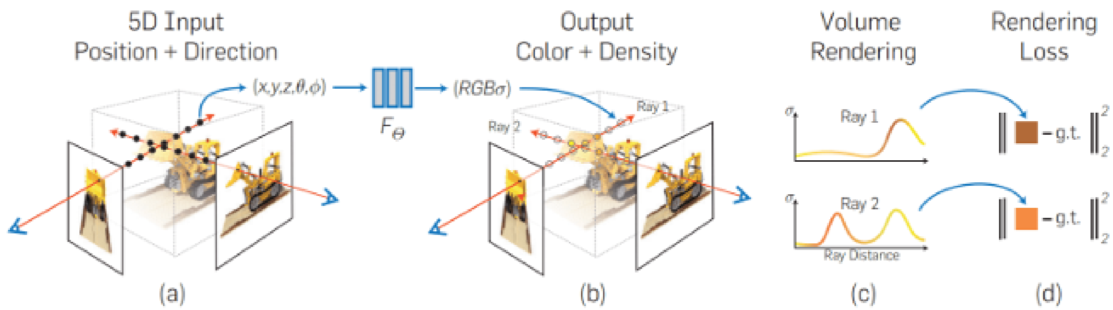


Fig 2.3.1. An overview of neural radiance field scene representation and differentiable rendering procedure. Image source: [33]

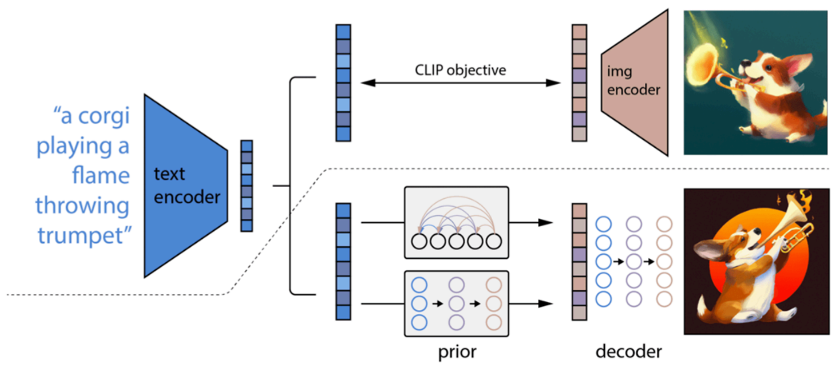


Fig 2.3.2. The architecture of DALLE-2 which applies the diffusion model. Image source: [39]

Recent advancements have highlighted the potential to generate high-quality 3D models through innovative approaches. For example, CLIP-Forge [36], developed by Autodesk AI Lab, utilizes a two-stage training process with an unlabelled shape dataset and a pre-trained image-text network like CLIP. This method bypasses costly inference time optimizations and can generate multiple shapes from a single text input, offering rapid and diverse model generation.

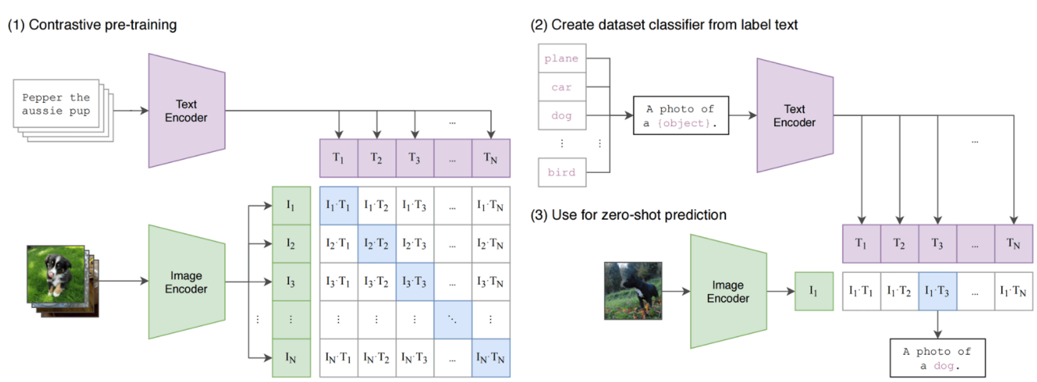


Fig 2.3.3. CLIP Structure. Image Source: [34]

Another notable development is Dream Fields [37], a method created by researchers from UC Berkeley and Google Research, which generates both geometry and color for a wide array of objects without 3D supervision. Dream Fields can create realistic, multi-view, consistent object geometry and color from various natural language descriptions using optimization techniques such as pre-training and geometric priors. Additionally, Text2Mesh is a framework that predicts color and local geometric details to stylize a 3D mesh based on a given text prompt [38]. This method stands out because it operates independently of a pre-trained generative model or a specialized 3D mesh dataset, making it highly versatile for applying diverse styles across different 3D meshes.

Overall, while the trend in developing T23D technology focuses on generating high-quality and large-scale models efficiently and accurately, these generative models still typically adopt a basic end-to-end generation approach. This approach often limits the interaction between the user and the model, which, if enhanced, could significantly improve the creative potential and applicability of T23D technologies.

2.3.2 3D Models Editing with T23D

In addition to generating 3D shapes, related technologies such as Neural Radiance Fields (NeRF) and diffusion models have demonstrated capabilities in facilitating 3D shape transformation. Pilot research works have delved into guiding 3D model generation and editing processes using these technologies.

Depending on the medium of instruction and interaction, there are two approaches to guide the transformation of 3D shapes. The first is text prompt-based guidance, which uses text prompts to modify the shapes. For example, *Instruct 3D-to-3D* utilizes a pre-trained image-to-image diffusion model to realize 3D-to-3D transformations [31]. Similarly, *TextDeformer* employs text prompts to guide the generation of input triangle mesh deformations [45].

Fig 2.3.4. Overview of Instruct-to-3D. Image Source: [31]

Another approach is sketch-based guidance, where images, particularly the user's sketches, are utilized to guide 3D editing processes. SKED (See Figure 2.3.5), for instance, enables users to guide mesh editing using prompts and mesh inputs [32]. Notably, while research on sketch-guided Text-to-Image (T2I) techniques has been extensive and fruitful, yielding solid results, the application of sketch-guided methods in 3D editing is gaining traction. Researchers have introduced highly efficient sketch-guided T2I methods, enabling the production of diverse results corresponding to text prompts and aligning with the spatial layout of sketches [44].

Fig 2.3.5. Overview of SKED. Image source: [32]

2.3.3 Shap-E

In the landscape of text-to-3D generation, Shap-E emerges as a significant and innovative generative model that forms the backbone of our tool [20]. Developed through a synthesis of cutting-edge technologies and research methodologies, Shap-E represents a leap forward in the realm of manipulable 3D model generation.

Shap-E leverages Transformer-based encoding techniques to process textual input and generate corresponding Implicit Neural Representations (INRs) for 3D assets (See Figure 2.3.6). This approach enables an accurate translation of text descriptions into tangible 3D forms.

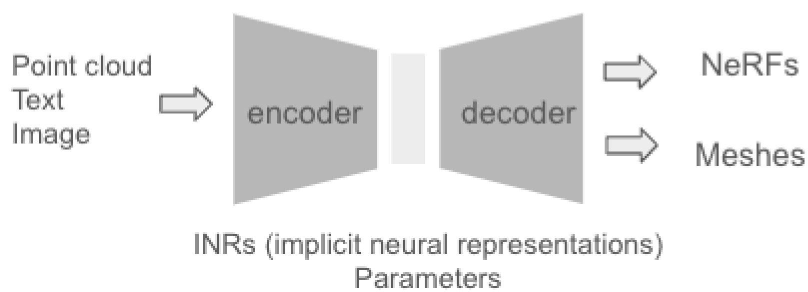


Fig 2.3.6. Workflow of Shap-E. Drawn according to [20].

Unlike traditional models, Shap-E has the unique capability to produce INRs representing both Neural Radiance Fields (NeRFs) and meshes simultaneously. This dual representation offers flexibility in rendering options and enhances compatibility with various 3D applications.

Shap-E empowers users with interactive manipulation capabilities, allowing for real-time adjustments and refinements to generated 3D models. This feature aligns with my tool's vision of providing intuitive and user-friendly experiences in 3D art creation. The following research will use the Shap-E as the base for developing manipulatable T23D.

3. Hypothesis

The research posits that integrating text-to-3D (T23D) technology into the 3D art creation process can significantly enhance the artistic workflow by providing artists with greater control over the generation and modification of 3D models. This hypothesis is founded on several key assumptions.

1. **Enhanced Creative Control**: By employing T23D technology, artists can more intuitively interact with generative systems, potentially leading to a richer, more personal creative output. Shap-Explorer, as a tool, is hypothesized to offer nuanced control mechanisms that allow artists to iteratively refine and alter their creations based on real-time feedback and adjustments.

2. **Increased Efficiency and Flexibility**: T23D is expected to reduce the time and technical barriers traditionally associated with 3D modeling. Artists using Shap-Explorer should be able to generate complex designs more quickly than through conventional methods, enabling rapid prototyping and experimentation.

3. **Demystification of Generative Processes**: By enhancing transparency and user interaction within the T23D technology, Shap-Explorer aims to transform the typically opaque 'black-box' nature of generative AI into a more understandable and manageable tool. This could foster a deeper understanding and trust in AI-driven tools among artists, which in turn could lead to wider adoption and innovation in artistic practices.

4. Methods

This chapter outlines the prototyping process utilized to test the hypothesis stated in the previous chapter. This research aims to develop a manipulable tool, Shap-Explorer, to assist users in 3D art creation. The development process involves two primary stages: 1) Developing a pipeline that integrates text-to-3D (T23D) models into the 3D art creation workflow, guided by insights from an initial pilot study examining the performance and features of Shap-E. 2) Building an interface that enables users to control and guide the T23D models through enhanced interactive functions.

Subsequently, this interface will be employed in a user study (see Chapter 5) to assess its functionality and affordances.

4.1 Prototype Pipeline: From Outside Shap-E to Inside

4.1.1 Pipeline Overview

The Shap-E model serves as the foundational generating model in the development of the primary pipeline for creating 3D models with T23D technology. Using the example code provided by the developers of Shap-E, I adjust the text prompts and render the sample latent representations as mesh files stored in PLY format. Furthermore, I import these files into Blender and display them in vertex mode. To acquire images of the 3D models, I utilize the Microsoft Snipping Tool to capture screenshots.

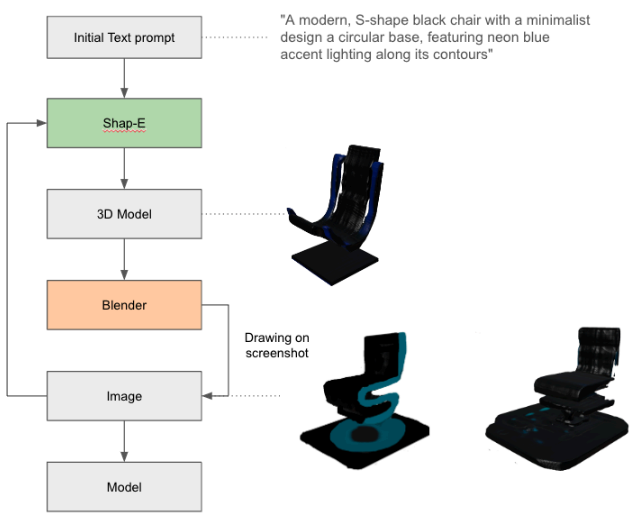


Fig 4.1.1.Text-based, sketch-guided creation pipeline.

4.1.2 Design Experiments

To optimize the pipeline proposed, I conducted three design experiments to test its function and usability. These experiments primarily focus on the users’ control of T23D, interactions, and design trade-offs.

Experiment 1: Design a shark with a horn
To test whether the user can “create” an object, I use the text-based, image-guided pipeline to generate a shark with a horn.

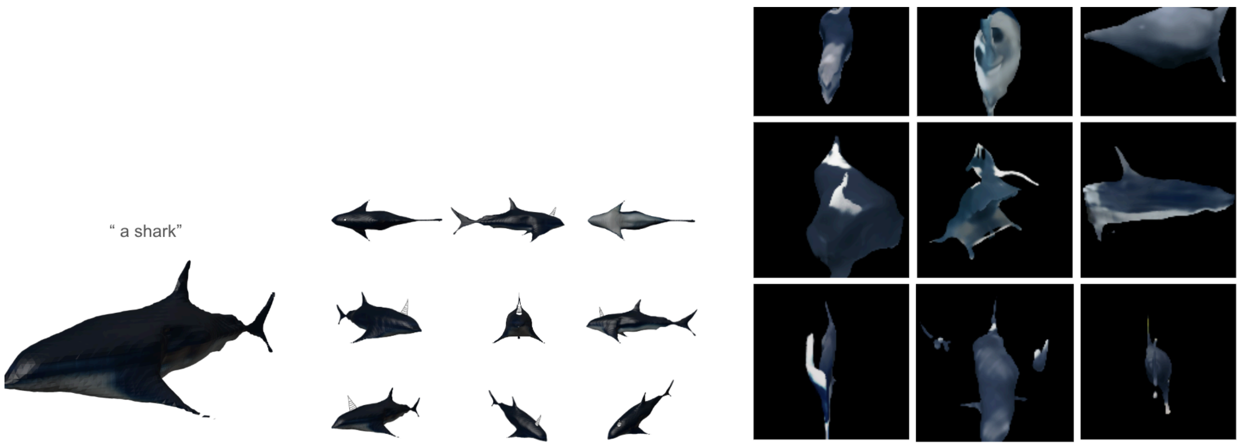


Fig 4.1.2. Overview of experiment 1.

In this experiment, I try to design a shark with a horn in the following steps:

1. Generating mesh with the prompt “a shark.”
2. Get screenshots of 9 different perspectives of the mesh
3. Draw a horn in the head in these screenshots
4. Regenerate the meshes with image input

Experiment 2: Design a fierce Pokémon that can fly as a boss
In this experiment, I tested the function of this pipeline to explore styles:

1. Generating mesh with the prompt “a fierce dragon with a feather wing.”
2. Sketch on the screenshot to create red feathers
3. Regenerate the mesh with Sketch

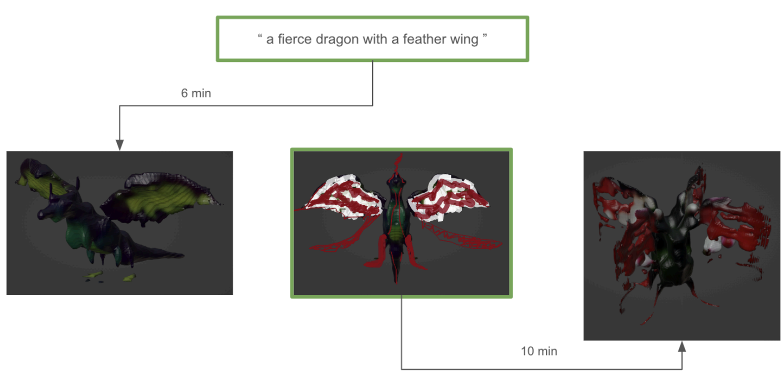


Fig 4.1.3. Overview of experiment 2.

Experiment 3: Design a small pavilion as a cafe

This experiment is conducted to examine whether the pipeline can allow users to complete the design task with strong intention. In this experiment, I assume two artists are engaged in the design process with strong intentions. I select two reference cases for the task and try to modify the design to align with the reference. Here are the steps in this experiment:

1. Choose two different references A and B
2. Generating models with the prompt “ a circular wooden pavilion with a stone roof.”
3. Sketch on the screenshot of step 2’s output mesh according to reference A
4. Regenerating mesh based on image
5. Sketch on the screenshot of the output mesh from the last step according to reference B
6. Regenerating mesh based on image

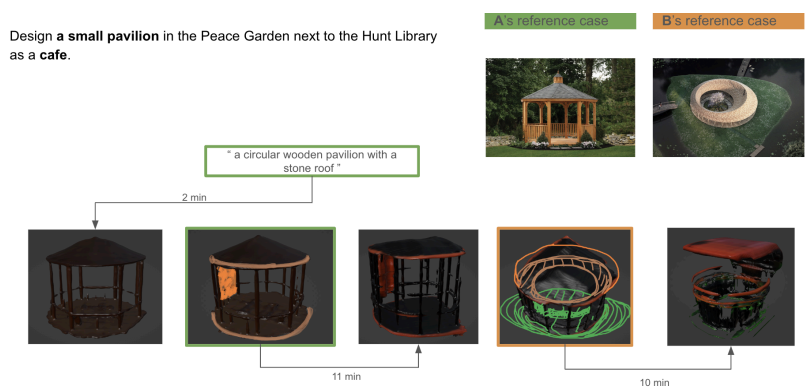


Fig 4.1.4. Overview of experiment 3.

4.1.3 Findings

The insights gained from the initial experiments with the prototype pipeline highlighted critical areas for improvement, particularly in how users interact with and control the generation process of 3D models. While the experiments demonstrated the potential of using T23D technologies in artistic creation, they also revealed significant limitations in user control and model interactivity:

1. Text prompts are not utilized beyond the initial 3D model generation.
2. Users lack control over which aspects of the model to retain or modify.
3. The regenerated model's accuracy heavily depends on the selected image perspective.

These findings underscored the necessity for a more intuitive and manipulative interface that could address these deficiencies. To solve this problem, I will use an interface to integrate the workflow of T23D in art creation. The randomness and uncontrollability of Image-to-3D suggest a need to explore generating models part by part rather than as a whole. Besides, future research should focus on improving user control and interaction within T23D systems.

4.2 Prototype Interface: Shap-Explorer

To address the user pain points identified in the pilot study, I designed the Shap-Explorer interface to enhance interactions with the Shap-E model, facilitating a more intuitive exploration of its capabilities. The primary goals of Shap-Explorer are to:

1. Enhance the user's understanding of Shap-E's inputs and outputs.
2. Provide gradual control and interactions, allowing for a more hands-on approach to 3D model generation.

4.2.1 Interacting with Text-to-3D

Iterative Modification

Text-to-3D (T23D) tools are traditionally designed as end-to-end systems that aim to generate highly detailed meshes from textual prompts[1,20]. While developers often focus on improving the efficiency and accuracy of their AI models, this focus can sometimes overlook the actual needs and preferences of the users. This oversight can lead to tools that, although technically advanced, are not necessarily aligned with the practical requirements of those who use them.

Typically, 3D modelers refine their creations by making incremental adjustments [41], a practice that should ideally be mirrored in how they interact with AI-generated models. However, the current design of most T23D tools does not support such iterative refinements; instead, they deliver final models without the opportunity for user intervention during the generation process.

To solve the problems, Shap-Explorer introduces the idea of modifying the 3D artifacts by iteratively adding parts to the scene.

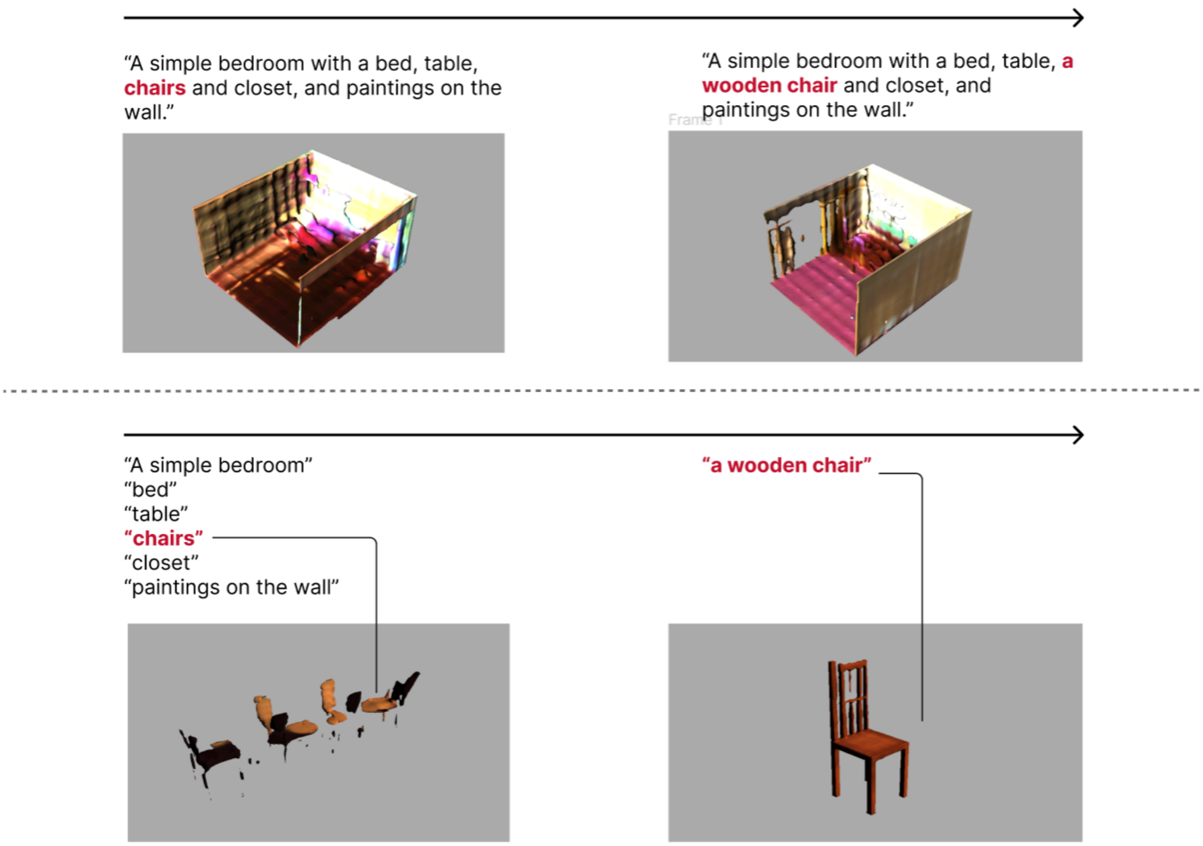


Fig 4.2.1. Iterative modification

Gradual Generation

Unlike traditional digital sculpting, where artists start with basic shapes and adjust details iteratively based on visual feedback, the process with generative models like Shap-E can seem opaque—a "black box." Users input parameters and wait, often without any visibility into the progress beyond a percentage completion bar. This mode of interaction is starkly different from conventional 3D modeling practices, where artists develop a model through a series of deliberate, visible stages.

Inspired by traditional digital sculpting workflows, Shap-Explorer introduces a sliding bar mechanism for a gradual generation. This feature allows users to dynamically adjust the progression of their model's generation, akin to sculpting in clay. This method not only demystifies the generative process but also restores a measure of artistic control by enabling real-time visual feedback and incremental adjustments.

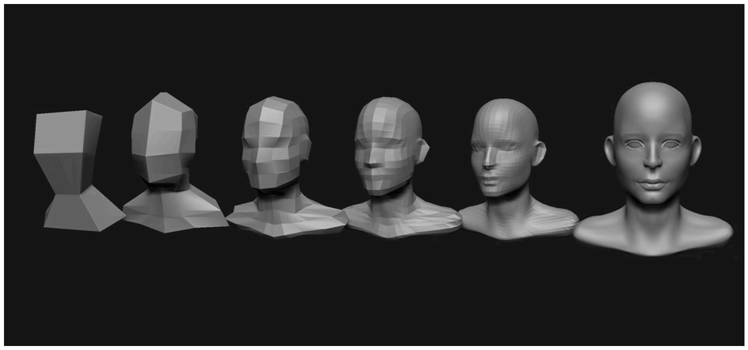


Fig 4.2.2. Regular digital sculpting process. Image source: https://aarne.co.uk/what-we-do/digital-sculpting/

By integrating these features into Shap-Explorer, the tool aims to bridge the gap between traditional artistic methods and modern AI-driven techniques, providing a more intuitive and controllable interface for artists working in 3D.

Besides, to achieve gradual control of the generation process.

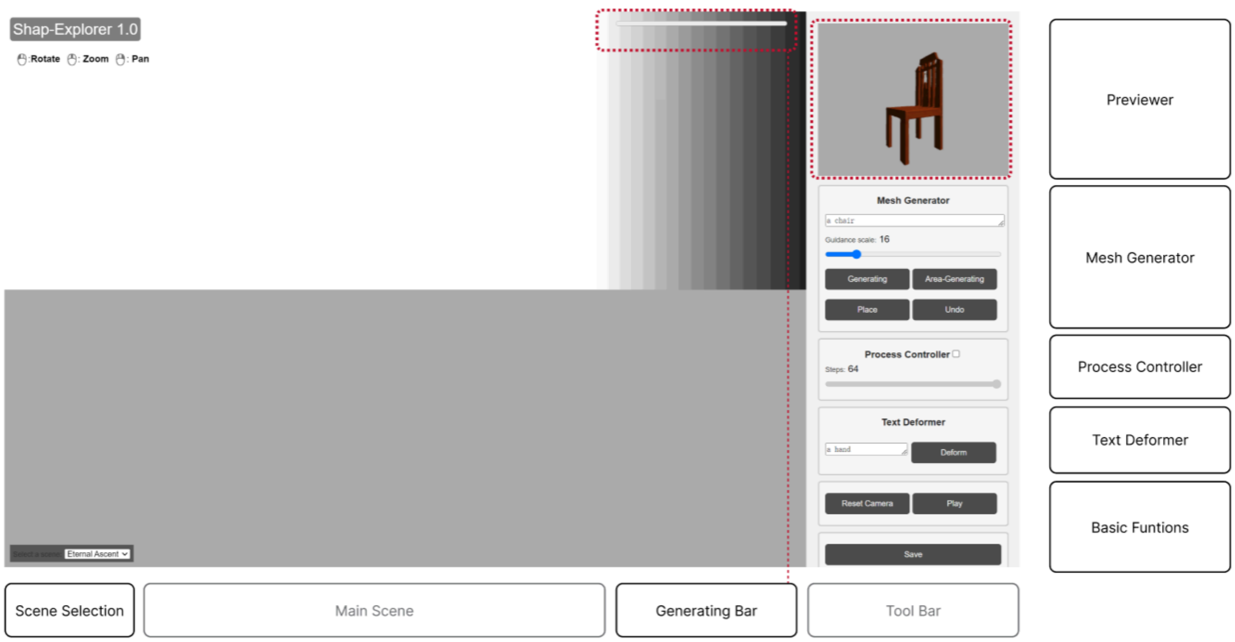


Fig 4.2.3. Shap-Explorer Interface.

4.2.2 Building Interface

Based on the interactions proposed in the last section, I built Shap-Explorer, a web-based 3D modeling tool powered by Shap-E. Shap-Explorer is designed to 1) help users understand and fully use generating models with Mesh Generator. 2) enhance the user's creation by gradually controlling generative models with Process Controller and Text Deformer. To realize this enhancement, Shap-E has the following functions:

3D Scene Setup and Basic Functions

Shap-Explorer uses Three.js to build and render the scene as the space of art creation. Like most 3D modeling software camera controls, the user can rotate, zoom, and pan by pressing and dragging the mouse's left, middle, and right keys. Besides, according to the common 3d modeling operations, I add basic functions to the toolbar like Undo, Resetting the camera, and Saving Files. Additionally, to help users preview the generated mesh and determine the next creation, a previewer scene is designed in the top-right corner of the interface.

Mesh Generator

As the core function of Shap-Explorer, the *mesh generator* sends the user's request to Shap-E, gets the response, and loads the mesh to the previewer. The user can input a text prompt in the text area, adjust the guidance scale, and generate a mesh. Since the generating process costs 30-120 seconds, the user can push several generating tasks, which will be shown in the task queue. When a task is completed, the user can place the mesh with the "Place" button (See Figure 4.2.4)

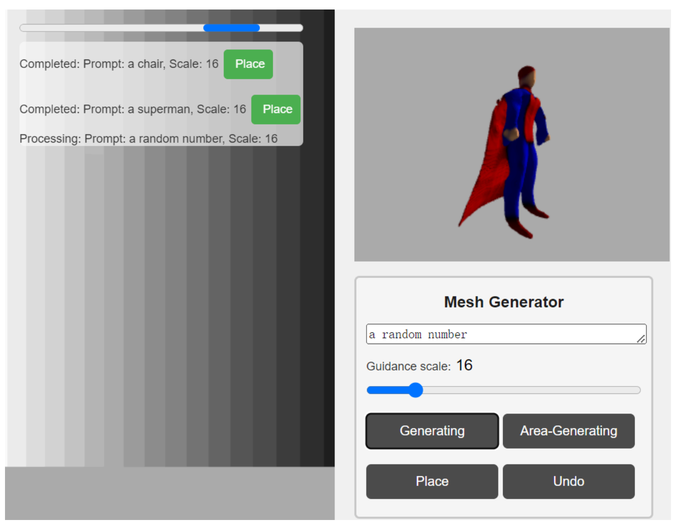


Fig 4.2.4. The task queue can show the task status and generate a place button once it is completed.

After clicking the "Place" button, the user can finish the placing with the mouse step by step: 1) choosing a position in XY plan, 2) choosing a position along the Z axis, 3) Scale mesh, and 4) Rotate it.

The Shap-Explorer also allows the user to control the "guidance_scale" of generative models, which is used to control the degree of influence the conditioning (in a T23D it refers to text prompts) has over the generated output. Figure 4.2.5 shows the influence of guidance_scale: high values result in outputs that adhere more closely to the prompts, while low values make the output less constrained by the input. However, since the Shap-E is trained by point clouds, tighter conditions caused by higher guidance make the resulting model more fragmented and do not allow the vertices to be connected as a whole mesh (in Figure 4.2.5, when the guidance scale is 64 and 100). Considering the limitation of Shap-E, the default value of the guidance scale is set as 16.

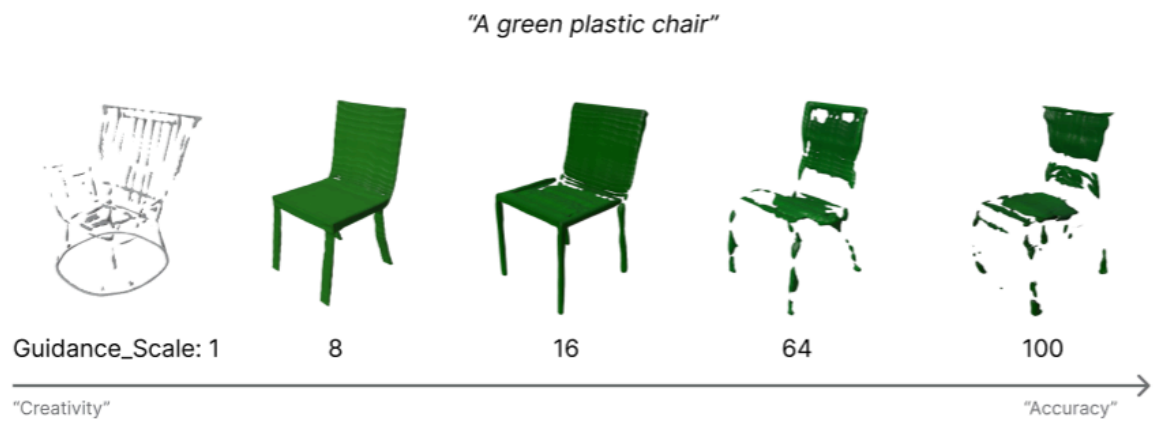


Fig 4.2.5. guidance_scale of Shap-E

Process Controller

With *process controller*, Shap-Explorer provides manipulation in the generating process. When the user enables the Process Controller, the models of mediate stages will be stored and shown in the previewer. (See Figure 4.2.6) The user can trace back the generating process with the "steps" slider.

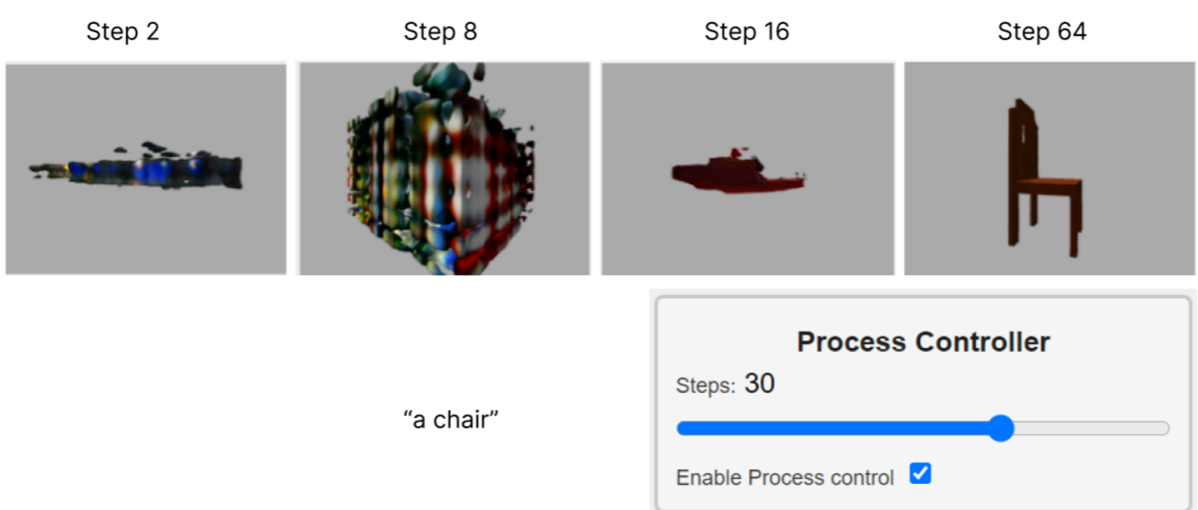


Fig 4.2.6. The process controller

Text Deformer

With *Text deformer,* Shap-Explorer allows the user to explore the forms of shape by adding new "prompt" in generated mesh [45]. The user can input a directing prompt and click "deform" button, then the generative model will reform the base mesh with the directed prompt. For example, if the user wants a hand-like chair, he can generate a chair first, then use "a hand" as a directing prompt to deform the shape(See Figure 4.2.7).

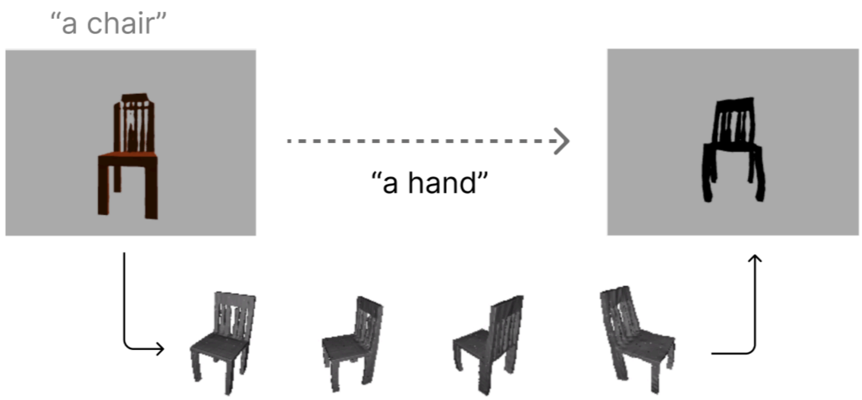


Fig 4.2.7. How the Text Deformer works.

4.2.3 Technical Details

This section explains the technical approach used to develop the Shap-Explorer.

Project Code Implementation

Shap-Explorer is implemented as a web application using HTML, CSS, and JavaScript. To support the generating functionality, I built a WebSocket-based Flask server based on the open-source code of Shap-E[4]. Shap-Explorer has been deployed by Vercel (shown in Appendix).

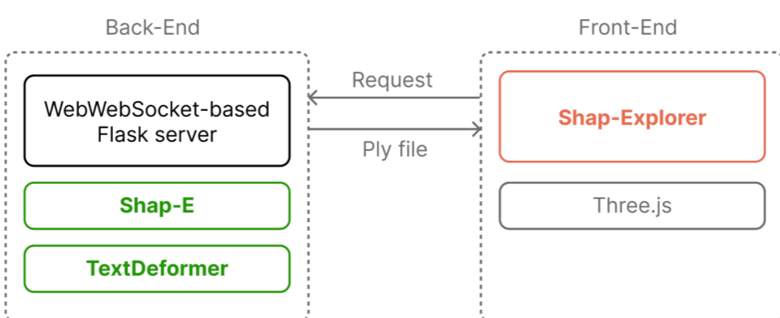


Fig 4.2.8. the architecture of Shap-Explorer's front-end and back-end.

5. User Study

I conducted a preliminary user study to test and evaluate the pipeline and functions of Shap-Explorer. This study focuses on the usability and creativity of this tool. Interviews, pre-surveys, and post-surveys are conducted to determine how the function of Shap-Explorer will influence the user's creation process using text-to-3D generating AI. The findings served as the basis of quantitative and qualitative analysis.

5.1 Participants

I recruited five participants for my user study, comprising one female and four males with ages ranging from 24 to 28 years old. The recruitment process involved reaching out to university communities specializing in computational design and artificial intelligence (AI), as well as tapping into my personal networks within academic and professional circles. Participants were selected based on their proficiency and experience levels in using Generative Artificial Intelligence (GAI) and 3D modeling software. Each participant was offered an Amazon gift card valued at $10 to incentivize participation.

5.2 Procedure

Every participant needs to complete the following tasks:

1. Complete pre-survey (2 mins)
2. Watching the tutorial (5 mins) showing how to use the tool in different templates
3. Finishing the design task within (40 mins)
4. Competing Post-task survey(10 mins)
5. Short interview(10 mins)

I selected the System Usability Scale (SUS) [22] as a tool to gauge the user-friendliness and overall satisfaction with Shap-Explorer. The SUS offers a quantitative measure, providing insights into participants' perceptions of the tool's ease of learning, efficiency, and general usability. By incorporating SUS, I aim to obtain structured feedback that can help pinpoint specific areas for improvement in terms of user experience.

In tandem with SUS, I integrated the Creativity Support Index (CSI) [23] to assess the extent to which Shap-Explorer fosters creativity in 3D model creation. The CSI's multidimensional approach allows for a comprehensive evaluation of creativity support like ideation, exploration, expressiveness, and immersion. This inclusion enables me to gather both quantitative ratings and qualitative insights regarding participants' creative experiences with Shap-Explorer, aiding in a nuanced understanding of its effectiveness in enhancing creativity.

As the most essential part of this user study, the design task is to let the user **create a prototype of [eternal ascent] using Shap-Explorer.** My design task is inspired by the popular 3D creation challenge raised by *pwnisher*, a famous 3d artist and video maker. He gives a template of a 3D scene where a character slowly ascends a rotating staircase and challenges 3D artists to create unique art based on that. Instead of letting user freely "create" anything they want, a base template can be an ideal starting point for ideas without restricting their creativity.

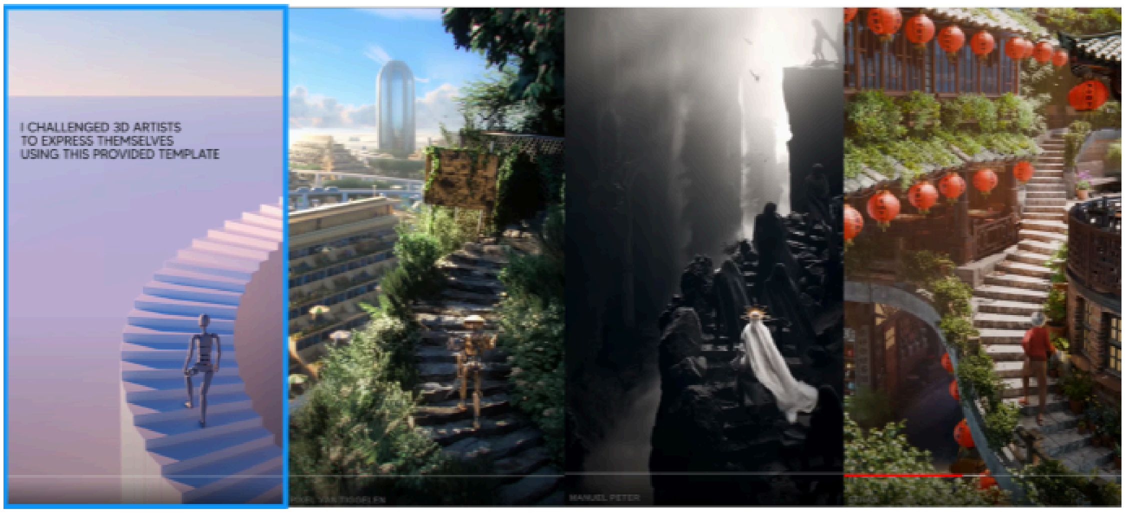


Fig 5.2.1. The template and works of Eternal Ascent.
https://www.youtube.com/watch?v=UNjMSFLkMZA&ab_channel=pwnisher

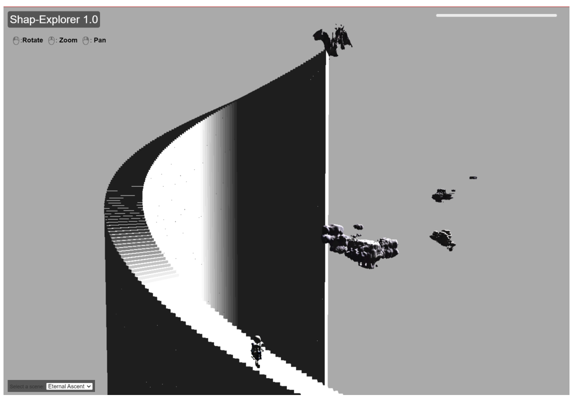
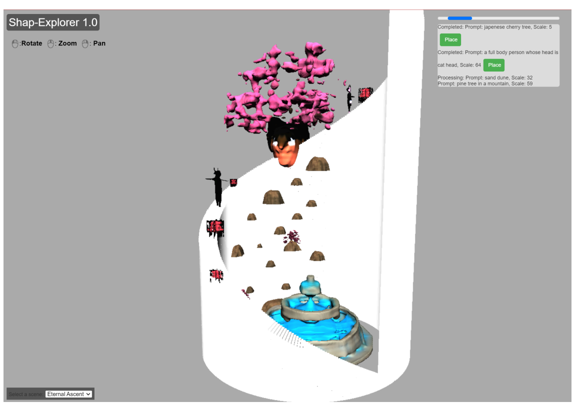
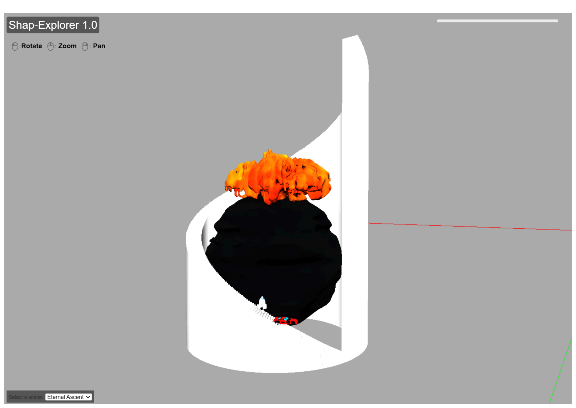
In the post-survey, I asked the participants to give a general evaluation of Shap-Exploration's usability and creativity. Besides, according to the features of GAI, I asked them for a sense of contribution to the outcomes. This survey also asked about the usability of three functions: mesh generator, process controller, and text deformer. After the post-survey, I conducted a short interview during which I asked about the good experience they felt, the challenges they encountered, and the possible improvement suggestions.

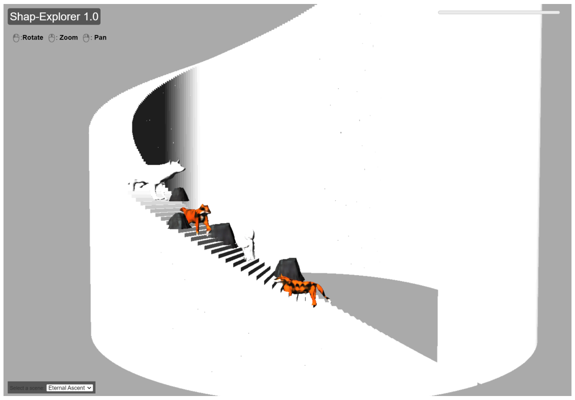
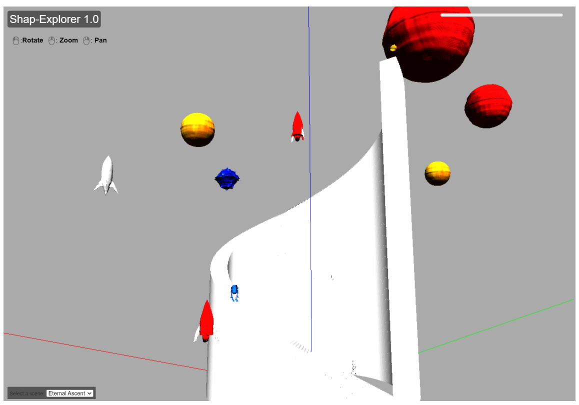
The whole study, including interviews, has to be completed within 70 minutes. The pre-survey, post-survey and interview questions can be checked in the Appendix.

5.3 Result

This section shows the participants' outcomes, as well as the findings from the surveys and interviews.

Table of Design Results

Scene	Prompts used	Function used	Time to complete the task, mins
	*"Golden crown", "Dragon", "A walking humanoid robot", "Clouds", "cloud"*	Mesh generator Process controller	**33**
	*"a full-body person whose head is cat head", "a person with a cat head", "Japanese cherry tree", "Limestone rock",*	Mesh generator Process controller	**41**
	*"Alien", "astronaut", "Black hole", "explorsion", "Sport car",*	Mesh generator Process controller Text deformer	**32**

	*"Tiger", "Big tiger", "Monk", "bucket"*	Mesh generator Process controller Text deformer	**40**
	*"A High Technology Robot", "a transformer", "blue robot", "Cyberman walking", "rocket", "rocket in red" "Stars", "super nova"*	Mesh generator Process controller	**28**

(High-resolution pictures are listed in the Appendix)

5.3.1 Qualitative Analysis

For the qualitative findings of this study, I analyze the think-aloud and interviews and conclude the findings about three functions.

General Evaluation

Participants explained why they felt neutral or negative about the usability of Shap-Explorer. Most of them (N=4) mentioned that the current one-time place-scale-rotation was inconvenient. *"P1: Hard to get the scale I want" "P3: I need a select button, which allows me to reselect the mesh, then edit or delete it."* All of them (N=5) mentioned the waiting time when they were asked about challenges encountered in completing the design task.

Most participants (N=3) perceived positively about the creativity of Shap-Explorer *"P2:. (the output) is very different from my goal but interesting."* However, one participant felt unsatisfied with the editing function *"P5: the tool doesn't work so well, I expect more powerful editing functions."*

Participants also gives some suggestion for improving basic features. One participant mentioned the perspect saving is important for building a 3D scene *“P1: need for saving the perspective based on which I develop my design.”*

The Limitation of Generative Models

In examining the limitations of generative models as reported by the participants, several key points emerged. P1 noted that the output meshes had a points and voxels style, suggesting a lack of smoothness or detail (*P1: "The output meshes look points and voxels style"*). P3 observed inconsistency in the output style, indicating variability or unpredictability in the generated models (*P3: "It is inconsistent in the output style"*). P4 mentioned that while the mesh could match their prompts, it was still too rough, highlighting issues with the quality or fidelity of the generated models (*P4: "The mesh can match my prompts, but too rough"*).

Design Trade-offs

Regarding design trade-offs, P1 raised questions about the differences between the generated models at various steps, indicating a desire for clearer distinctions or iterations in the generative process (*P1: "If possible, I wonder about the difference between dragon of steps 31, 32, 33…"*). They also mentioned that if the generative models were less time-consuming, they would spend more time experimenting to achieve satisfying outcomes, highlighting the trade-off between time efficiency and desired results (*P1: "If the generative models cost less time, I will try more time to get the satisfying outcomes"*) (*P3: "Too time-consuming, but easy to learn"*)

5.3.2 Survey Findings

The usability criteria include satisfaction, usability, integration and consistency. Figure 5.3.1 shows the findings on usability. The horizontal axis represents the extent to which the user agrees with the statement, the higher value(5) indicates that the user perceive that Shap-Explorer support tha criteria. The vertical axis represents the number of users who rated this score.

Generally participants were neutral about the usability of Shap-Explorer and most of them felt it was hard to control the web-based scene, and the waiting time for generation was too long. There was one participant who answered positively about the usability and satisfaction.

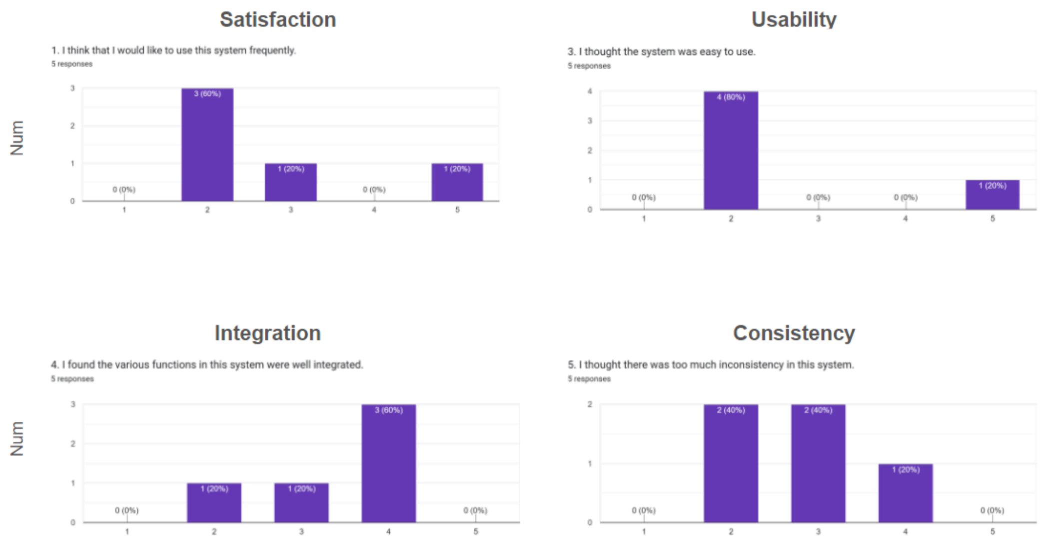


Fig 5.3.1. The histogram of responses on the system usability scale questions.

Figure 5.3.2 shows the creativity criteria include collaboration, enjoyment, exploration, expressiveness, immersion and results worth effort. Compared with SUS questions, participants were more positive about the creativity of Shap-Explorer especially in enjoyment , immersion and expressiveness. However, participants responded neutrally about the exploration.

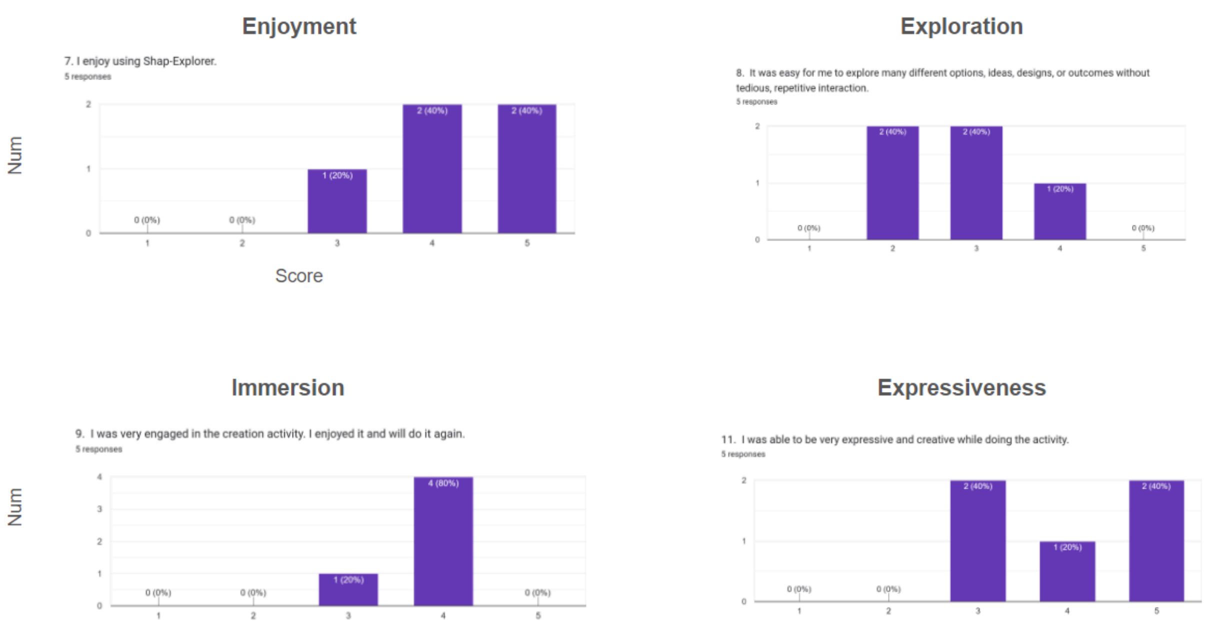


Fig 5.3.2. The histogram of responses on the creativity scale index questions.

Figure 5.3.3 shows the evaluation of individual functions. Overall, participants perceived the functions neutrally. Of the three functions, the participants' attitude towards the process controller was the most positive, affirming its positive effect on control generation.

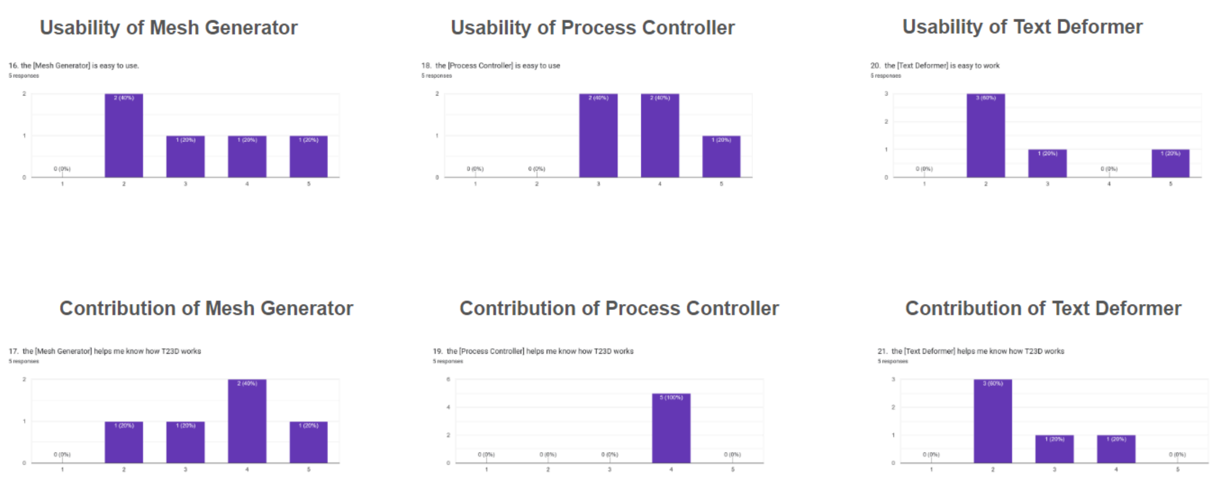


Fig 5.3.3. The histogram of responses on the individual function.

6. Discussion

In this chapter, I will discuss the findings based on the research background, prior work, and Shap-Explorer. To be specific, the 1) controllability of generative tools, 2) interactions for Text-to-3D, 3) design trade-offs while using Text-to-3D, and 4) limitations.

6.1 Controllability of Generative Tools

I hypothesized that implementing iterative modification and gradual control would enhance user control over Text-to-3D (T23D) generation. The pilot study and survey findings largely support this hypothesis. Specifically, participants found the Process Controller and Text Deformer functionalities easy to use. These findings suggest broader implications beyond T23D. Compared to other forms of generated content, 3D models are more likely generated as a design module and then combined following a sequence. Similar approaches could be applied to other generative tools, such as text-to-music systems, to improve user control and interaction.

While Shap-Explorer currently offers limited fine-degree control (e.g., the Process Controller provides predefined steps, and the Text Deformer generates a single intermediate content), it explores the potential to fine-tune factors beyond the initial input parameters. This includes exploring different semantics and control over various stages of generation, hinting at possibilities for future enhancements in generative tools.

6.2 Interactions for Text-to-3D

In the interview, three participants mentioned that generated models would influence them and even change design direction, showcasing the potential for generative tools to inspire creativity and shape artistic outcomes.

However, participants also encountered challenges in integrating these outcomes into their designs, pointing to a need for improved usability and understanding of intermediate results. Although most participants felt “interesting” with unexpected outcomes generated by process controller and text deformer, they still found it difficult to use them in their designs. Learning to use this function and understanding how it works add a lot of burden to them. While these functions were designed to enhance user control over generative models, the steep learning curve and the inherent randomness of these functions hindered users' ability to comprehend and utilize the intermediate results effectively. This highlights the importance of balancing advanced control features with user-friendly design and usability considerations in generative tools.

6.3 Design Trade-offs while Using Text-to-3D

One of the key trade-offs identified in the user study was the balance between time and attempts. The efficiency of model generation within Shap-Explorer posed challenges, with tasks taking considerable time to complete. This impacted users' willingness to explore multiple prompts or iterations.

Another critical consideration was the balance between stylistic alignment and user controllability in Text-to-3D (T23D) generation. T23D often focuses on generating a single model at a time, disregarding the design context and coherence with other generated models. Users faced challenges in maintaining consistent styles across different pieces when integrating them into a scene. Conversely, using lengthy prompts to generate multiple items posed difficulties in controlling their relative positions and scales effectively.

Additionally, the balance between AI autonomy and manual intervention emerged as a crucial consideration. More sliders don't mean better control [26]. The cost of learning should also be considered. For example, in the user study, if participants are satisfied with the previous model, they are unlikely to adjust the default guidance scale of the mesh generator. The additional control can be served as the specialized functions above basic functions for generative tools.

6.4 Limitation

The utilization of a web-based modeling tool posed significant challenges for Shap-Explorer, particularly in terms of functionality and user experience. Three.js, while suitable for web-based 3D rendering, lacked the advanced features and computational capabilities found in professional 3D modeling software. This limitation hindered the tool's ability to offer a comprehensive set of manipulations and interactions, restricting users from performing complex operations or achieving high-fidelity results.

Moreover, Shap-Explorer faced quality and efficiency concerns in its generative outputs. Users encountered issues with the quality of generated 3D meshes, including lower resolution, limited detail, and suboptimal surface smoothness, diminishing the visual appeal and realism of the models. The time-consuming nature of model generation within Shap-Explorer also posed challenges, with tasks often requiring significant computational time. This lack of real-time feedback during the design process impeded iterative workflows and timely adjustments, impacting user productivity and interaction capabilities.

Additionally, the absence of alternative interaction methods, like showing multiple results to select, presented usability challenges, highlighting areas for improvement in future iterations of the tool.

7. Conclusion

This thesis introduces an interface called Shap-Explorer, which interacts with T23D through a more steerable approach. Drawing inspiration from Suchman's perspective on the interplay between professional technology design and technologies in use [21], Shap-Explorer aims to bridge the gap between developers and users by offering intuitive and effective tools for 3D art creation. These are three key steps in my research: 1) Observed the limitation of T23D and built Shap-Explorer, 2) Design three functions: mesh generator, process controller, and text deformer, 3) Conduct a user study and identify how people use the suggested interactions. The user study's findings validate Shap-Explorer's utility and inform the design and development of future generative tools, emphasizing the importance of user-centric approaches in advancing technology for creative endeavors.

Reference

1. Li, Chenghao, et al. "Generative AI meets 3D: A Survey on Text-to-3D in AIGC Era." arXiv preprint arXiv:2305.06131 (2023).
2. de Guevara, Manuel Ladron, et al. "Multimodal Word Sense Disambiguation in Creative Practice." 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE, 2020.
3. Bødker, S., & Grønbæk, K. (1991). Cooperative prototyping: users and designers in mutual activity. International Journal of Man-Machine Studies, 34(3), 453-478.
4. Odom, W., Wakkary, R., Lim, Y. K., Desjardins, A., Hengeveld, B., & Banks, R. (2016, May). From research prototype to research product. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (pp. 2549-2561). https://dl.acm.org/doi/abs/10.1145/2858036.285844
5. Agre, Philip, and Philip E. Agre. Computation and human experience. Cambridge University Press, 1997.
6. Agre, Philip E. "Lessons learned in trying to reform AI." Social science, technical systems, and cooperative work: Beyond the Great Divide 131 (1997).
7. Ferrando, C. (2018). Towards a Machine Learning Framework in Space Syntax (Version 1). Carnegie Mellon University. https://doi.org/10.1184/R1/7178417.v1 ([])
8. Steinfeld, K. (2021). Significant Others: Machine Learning as Actor, Material, and Provocateur in Art and Design. In As, I. and Basu, P. (eds.), The Routledge Companion to Artificial Intelligence in Architecture. Routledge, pp.3–12.
9. Llach, Daniel Cardoso. "Sculpting spaces of possibility: Brief history and prospects of artificial intelligence in design." The Routledge Companion to Artificial Intelligence in Architecture. Routledge, 2021. 13-28.
10. Ulberg, E. (2021). Crafting the Weights of a Convolutional Neural Network to Make a Drawing (Version 1). Carnegie Mellon University. https://doi.org/10.1184/R1/14135663.v1 ([])
11. Hasey, Michael David. The 3D Form Analysis of Regional Architecture Using Deep Learning: A Case Study of Wooden Churches from the Carpathian Mountains. Diss. Carnegie Mellon University, 2022.
12. Rhee, Jinmo (2019): Context-rich Urban Analysis Using Machine Learning: A Case Study in Pittsburgh, PA. Carnegie Mellon University. Thesis. https://doi.org/10.1184/R1/8235593.v1
13. Lu, Yingxiu (2018): Conversational Form-Generation: An Application of Interactive Genetic Algorithm to Architectural Design. Carnegie Mellon University. Thesis. https://doi.org/10.1184/R1/7182263.v1
14. Huang, Zhewei, Wen Heng, and Shuchang Zhou. "Learning to paint with model-based deep reinforcement learning." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.
15. Coates, P. (2010). Rethinking Representation. In Programming.Architecture. Routledge, pp.5–23.
16. Veloso, P. and Krishnamurti, R. (2019). From Black Box to Generative System. In 107th ACSA Annual Meeting Proceedings, pp.525–533.
17. Epstein, Z., Hertzmann, A., Herman, L., Mahari, R., Frank, M.R., Groh, M., Schroeder, H., Smith, A., Akten, M., Fjeld, J., and Farid, H. (2023). Art and the Science of Generative AI: A Deeper Dive. arXiv preprint arXiv:2306.04141.
18. Gozalo-Brizuela, Roberto, and Eduardo C. Garrido-Merchan. "ChatGPT is not all you need. A State of the Art Review of large Generative AI models." arXiv preprint arXiv:2301.04655 (2023).
19. Veloso, P. and Krishnamurti, R. (2021). Mapping Generative Models for Architectural Design. In As, I. and Basu, P. (eds.), The Routledge Companion to Artificial Intelligence in Architecture. Routledge, pp.29–58.
20. Jun, Heewoo, and Alex Nichol. "Shap-E: Generating Conditional 3D Implicit Functions." arXiv preprint arXiv:2305.02463 (2023).
21. Lucy A. Suchman (2002) Practice-Based Design of Information Systems: Notes from the Hyperdeveloped World, The Information Society, 18:2, 139-144, DOI: 10.1080/01972240290075066
22. Jordan, P.W., Thomas, B., McClelland, I.L., & Weerdmeester, B. (Eds.). (1996). Usability Evaluation In Industry (1st ed.). CRC Press. https://doi.org/10.1201/9781498710411
23. Erin Cherry and Celine Latulipe. 2014. Quantifying the Creativity Support of Digital Tools through the Creativity Support Index. ACM Trans. Comput.-Hum. Interact. 21, 4, Article 21 (jun 2014), 25 pages. https://doi.org/10.1145/2617588
24. Yuki Koyama, Issei Sato, and Masataka Goto. 2020. Sequential Gallery for Interactive Visual Design Optimization. ACM Trans. Graph. 39, 4, Article 88 (aug 2020), 12 pages. https://doi.org/10.1145/3386569.3392444
25. Chenlin Meng, Yutong He, Yang Song, Jiaming Song, JiajunWu, Jun-Yan Zhu, and Stefano Ermon. 2022. SDEdit: Guided Image Synthesis and Editing with Stochastic Differential Equations. In International

Conference on Learning Representations. https://openreview.net/forum?id=aBsCjcPu_tE
26. Ahmed, Eman, et al. "A survey on deep learning advances on different 3D data representations." arXiv preprint arXiv:1808.01462 (2018).
27. Charlene, Lewis, "What Is 3D Art? – The Different Aspects of Three-Dimensional Art." Art in Context. November 8, 2022. URL: https://artincontext.org/what-is-3d-art/
28. Daniel Dixon, Manoj Prasad, and Tracy Hammond. 2010. ICanDraw: using sketch recognition and corrective feedback to assist a user in drawing human faces. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10). Association for Computing Machinery, New York, NY, USA, 897–906. https://doi.org/10.1145/1753326.1753459
29. Alcaide-Marzal, Jorge, et al. "An exploratory study on the use of digital sculpting in conceptual product design." Design Studies 34.2 (2013): 264-284.
30. Jonas Oppenlaender. 2022. The Creativity of Text-to-Image Generation. In Proceedings of the 25th International Academic Mindtrek Conference (Academic Mindtrek '22). Association for Computing Machinery, New York, NY, USA, 192–202. https://doi.org/10.1145/3569219.3569352
31. Kamata, H., Sakuma, Y., Hayakawa, A., Ishii, M., & Narihira, T. (2023). Instruct 3d-to-3d: Text instruction guided 3d-to-3d conversion. arXiv preprint arXiv:2303.15780.
32. Mikaeili, A., Perel, O., Safaee, M., Cohen-Or, D., & Mahdavi-Amiri, A. (2023). Sked: Sketch-guided text-based 3d editing. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 14607-14619).
33. Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. 2021. NeRF: representing scenes as neural radiance fields for view synthesis. Commun. ACM 65, 1 (January 2022), 99–106. https://doi.org/10.1145/3503250
34. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G. & Sutskever, I.. (2021). Learning Transferable Visual Models From Natural Language Supervision. <i>Proceedings of the 38th International Conference on Machine Learning</i>, in <i>Proceedings of Machine Learning Research</i> 139:8748-8763 Available from https://proceedings.mlr.press/v139/radford21a.html.
35. Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. Advances in neural information processing systems, 33, 6840-6851.
36. Sanghi, A., Chu, H., Lambourne, J. G., Wang, Y., Cheng, C. Y., Fumero, M., & Malekshan, K. R. (2022). Clip-forge: Towards zero-shot text-to-shape generation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 18603-18613).
37. Ajay Jain, Ben Mildenhall, Jonathan T. Barron, Pieter Abbeel, Ben Poole; Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 867-876
38. Michel, O., Bar-On, R., Liu, R., Benaim, S., & Hanocka, R. (2022). Text2mesh: Text-driven neural stylization for meshes. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 13492-13502).
39. Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). Hierarchical text-conditional image generation with clip latents. arXiv preprint arXiv:2204.06125, 1(2), 3.
40. Rawson, P. (2016). Sculpture. University of Pennsylvania Press.
41. Santoni, C., Calabrese, C., Di Renzo, F., & Pellacini, F. (2016). Sculptstat: Statistical analysis of digital sculpting workflows. arXiv preprint arXiv:1601.07765.
42. Cetinic, E., & She, J. (2022). Understanding and creating art with AI: Review and outlook. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 18(2), 1-22.
43. Tornincasa, S., & Di Monaco, F. (2010, September). The future and the evolution of CAD. In Proceedings of the 14th international research/expert conference: trends in the development of machinery and associated technology (Vol. 1, No. 1, pp. 11-18).
44. Voynov, A., Aberman, K., & Cohen-Or, D. (2023, July). Sketch-guided text-to-image diffusion models. In ACM SIGGRAPH 2023 Conference Proceedings (pp. 1-11).
45. Gao, W., Aigerman, N., Groueix, T., Kim, V., & Hanocka, R. (2023, July). Textdeformer: Geometry manipulation using text guidance. In ACM SIGGRAPH 2023 Conference Proceedings (pp. 1-11).

Appendix

Related Source:

Shap-Explorer site: https://shap-explorer.vercel.app
Source code of Shap-Explorer: https://github.com/MaxWebb96/Shap-Explorer_DOM
Source code of Shap-E: https://github.com/openai/shap-e
Source code of Text-Deformer: https://github.com/threedle/TextDeformer?tab=readme-ov-file

User Study Question:

Pre-task Survey

1. Do you know GAI? Did you use generative AI for creation or design?
2. (If yes) Did you think it would be helpful for your purpose?

Post-task Survey

Rate your agreement with the following statements:

System Usability Scale [22]

1. I think that I would like to use this system frequently.
   1 2 3 4 5
   Highly Disagree Neutral Highly Agree
2. I found the system unnecessarily complex.
   1 2 3 4 5
   Highly Disagree Neutral Highly Agree
3. I thought the system was easy to use.
   1 2 3 4 5
   Highly Disagree Neutral Highly Agree
4. I found the various functions in this system were well integrated.
   1 2 3 4 5
   Highly Disagree Neutral Highly Agree
5. I thought there was too much inconsistency in this system.
   1 2 3 4 5
   Highly Disagree Neutral Highly Agree
6. I felt very confident using the system.

1 2 3 4 5
Highly Disagree Neutral Highly Agree

Creativity Support Index [23]

7. **Enjoyment:** I enjoy using Shap-Explorer.
   1 2 3 4 5
   Highly Disagree Neutral Highly Agree
8. **Exploration**: It was easy for me to explore many different options, ideas, designs, or outcomes without tedious, repetitive interaction.
   1 2 3 4 5
   Highly Disagree Neutral Highly Agree
9. **Engagement**: I was very engaged in the creation activity. I enjoyed it and will do it again.
   1 2 3 4 5
   Highly Disagree Neutral Highly Agree
10. **Effort / Reward Tradeoff**: What I was able to produce was worth the effort I had to exert to produce it.
    1 2 3 4 5
    Highly Disagree Neutral Highly Agree
11. **Expressiveness**: I was able to be very expressive and creative while doing the activity.
    1 2 3 4 5
    Highly Disagree Neutral Highly Agree
12. **Immersion**: I was able to concentrate on the creation process while using Shap-Explorer
    1 2 3 4 5
    Highly Disagree Neutral Highly Agree

Metrics for T23D

13. **Contribution**: The [Eternal Ascent] is my work.
    1 2 3 4 5
    Highly Disagree Neutral Highly Agree
14. **Collaboration**: It was easy to convey my ideas to AI and easy to understand the AI's outcomes.
    1 2 3 4 5
    Highly Disagree Neutral Highly Agree
15. **Controllability**: I led the entire design process, and the result is what I want to create.
    1 2 3 4 5
    Highly Disagree Neutral Highly Agree

Controllability and Inspiration of Functions

16. **Controllability of Mesh Generator**: the [Mesh Generator] is easy to use.
    1 2 3 4 5
    Highly Disagree Neutral Highly Agree
17. **Contribution to understanding T23D**: the [Mesh Generator] helps me know how T23D works
    1 2 3 4 5
    Highly Disagree Neutral Highly Agree
18. **Controllability of Process Controller**: the [Mesh Generator] is easy to use.
    1 2 3 4 5
    Highly Disagree Neutral Highly Agree
19. **Contribution to understanding T23D**: the [Mesh Generator] helps me know how T23D works
    1 2 3 4 5
    Highly Disagree Neutral Highly Agree
20. **Controllability of Text Deformer**: the [Mesh Generator] is easy to use.
    1 2 3 4 5
    Highly Disagree Neutral Highly Agree
21. **Contribution to understanding T23D**: the [Mesh Generator] helps me know how T23D works
    1 2 3 4 5
    Highly Disagree Neutral Highly Agree

Interview Questions:

Experience

1. How do you approach the design task using Shap-Explorer?
2. Are you satisfied with the final design output? Why or why not?
3. Do you think Shap-Explorer works well?
4. (if yes) Which part of it do you think works well?

Challenges

5. Have you encountered any difficulties using Shap-Explorer?

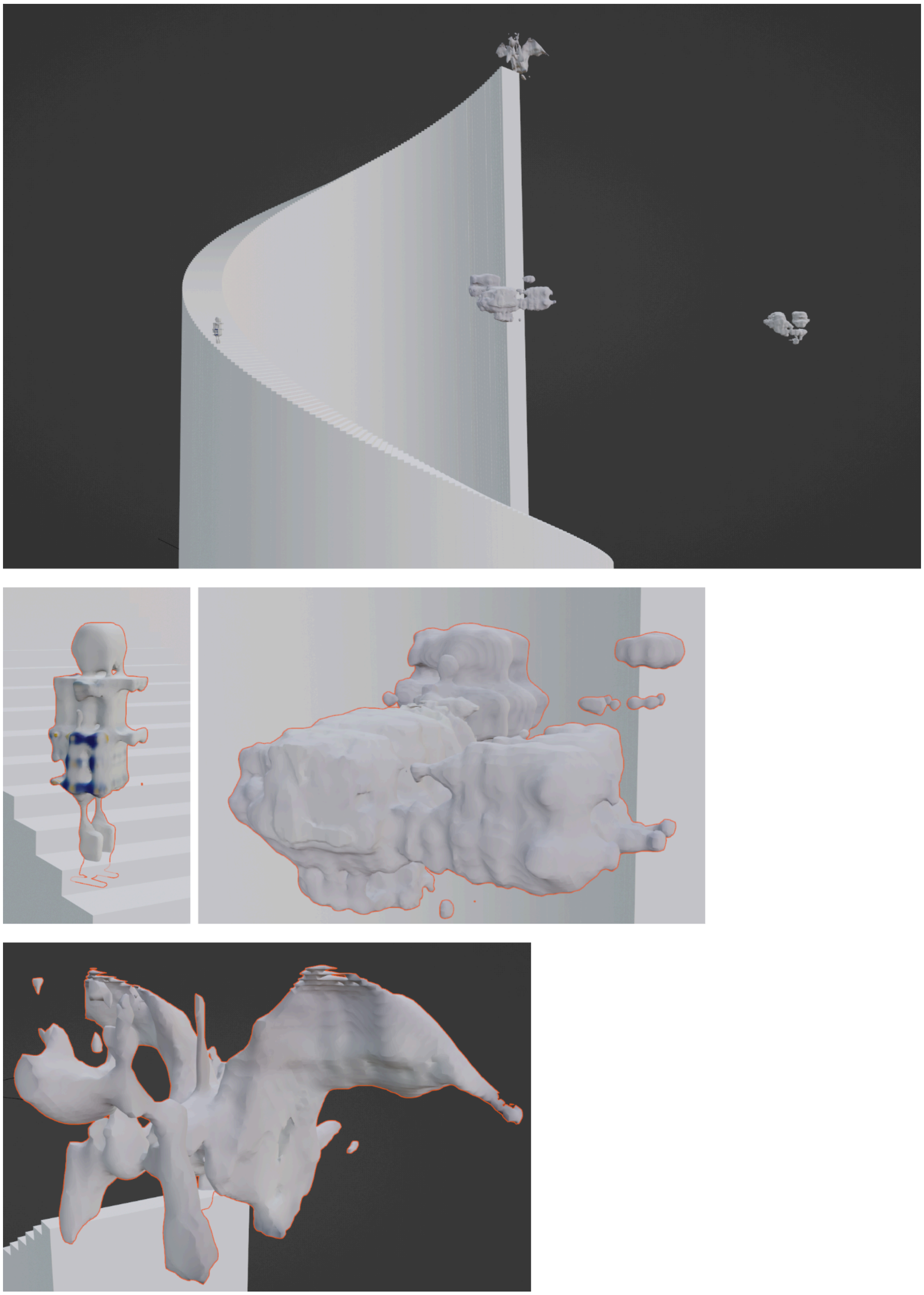
6. （if yes) Which part do you think Shap-Explorer works well, and tell me the challenges you encountered

Improvements

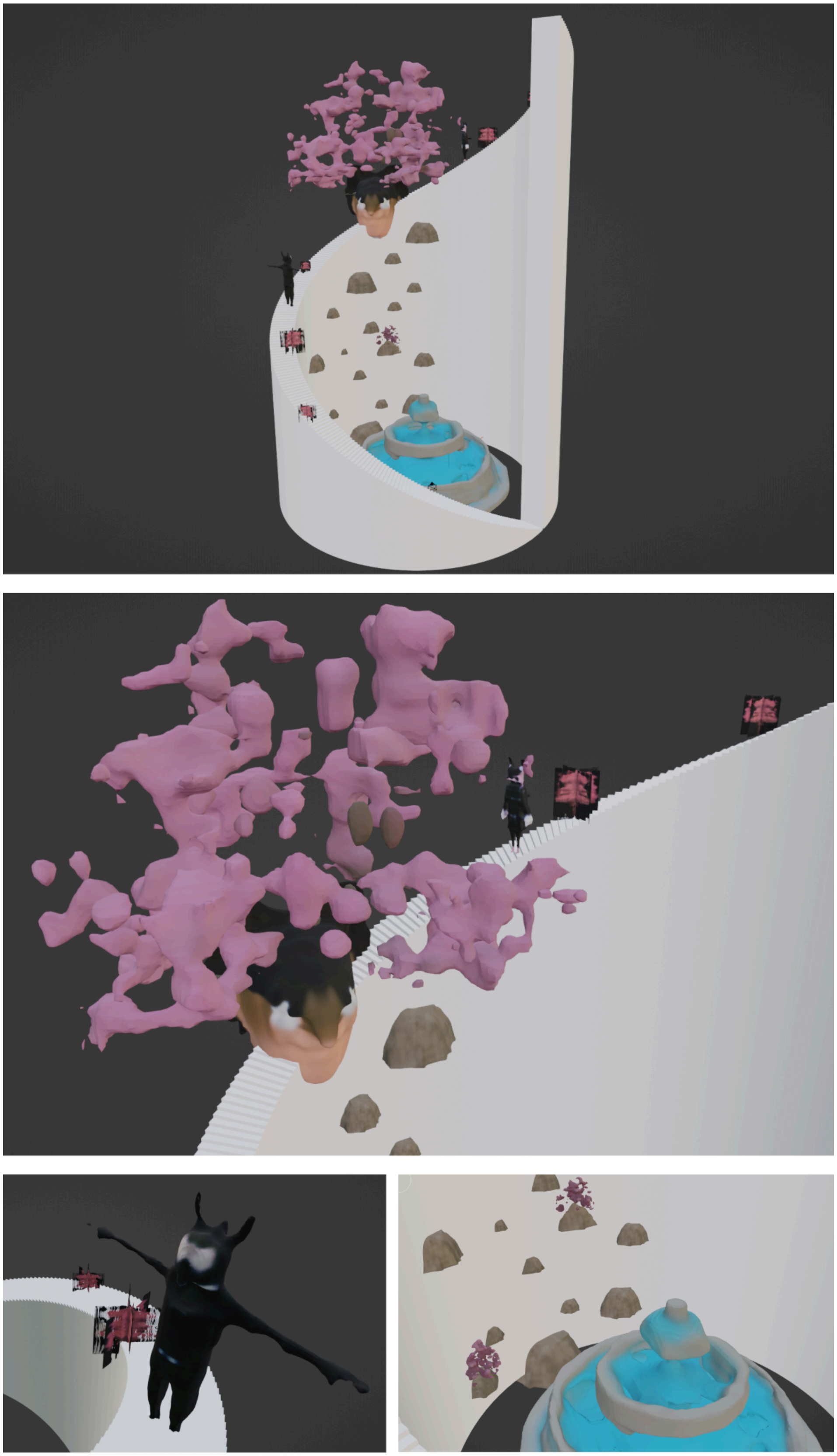
7. Based on the challenges you encountered, which part do you think can be improved? Or any new features can be incorporated?
8. Which part of Shap-Explorer do you think is valuable for inspiration?

Participants’ Design Works

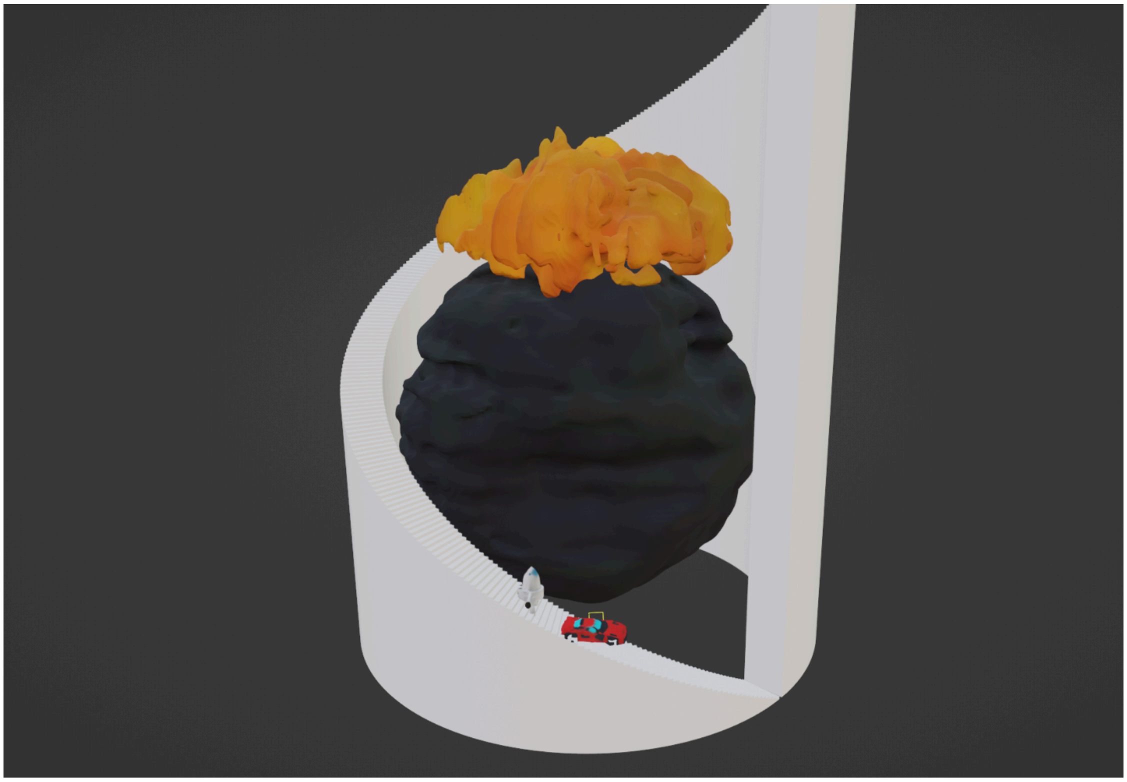
P1:



P2:

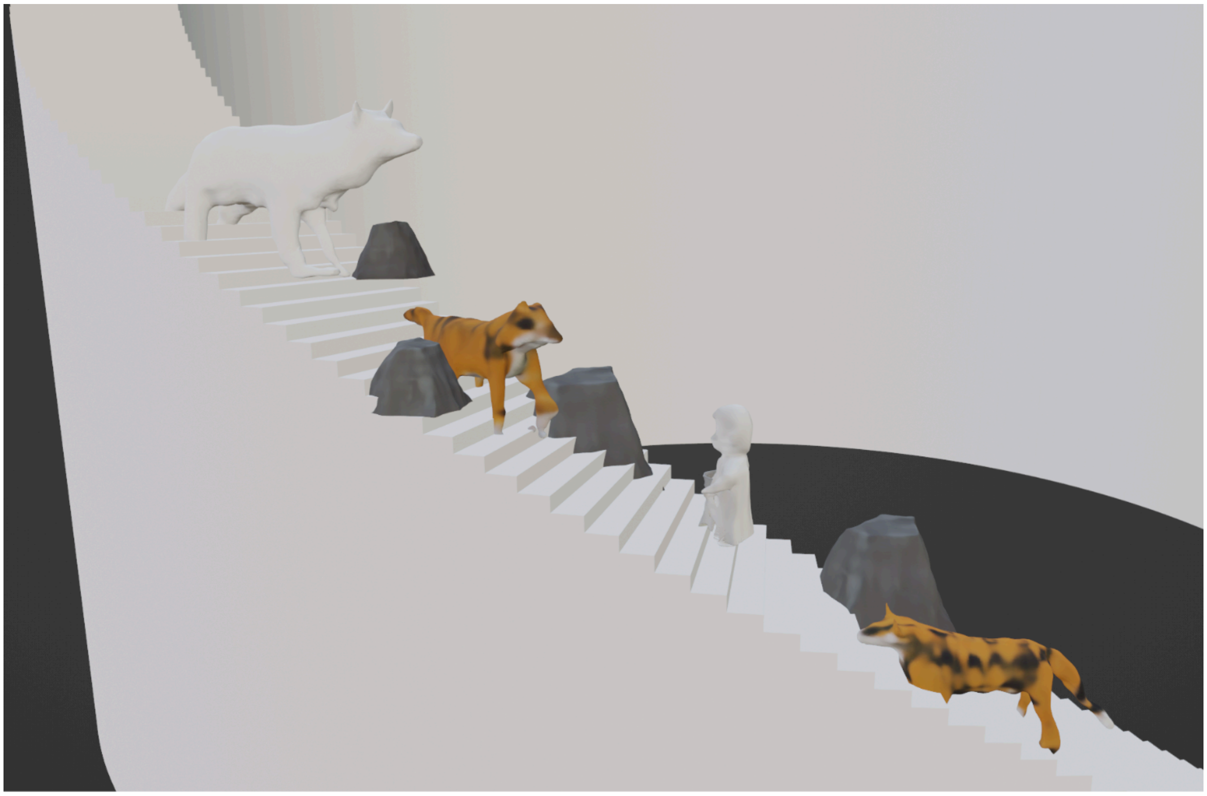


P3:

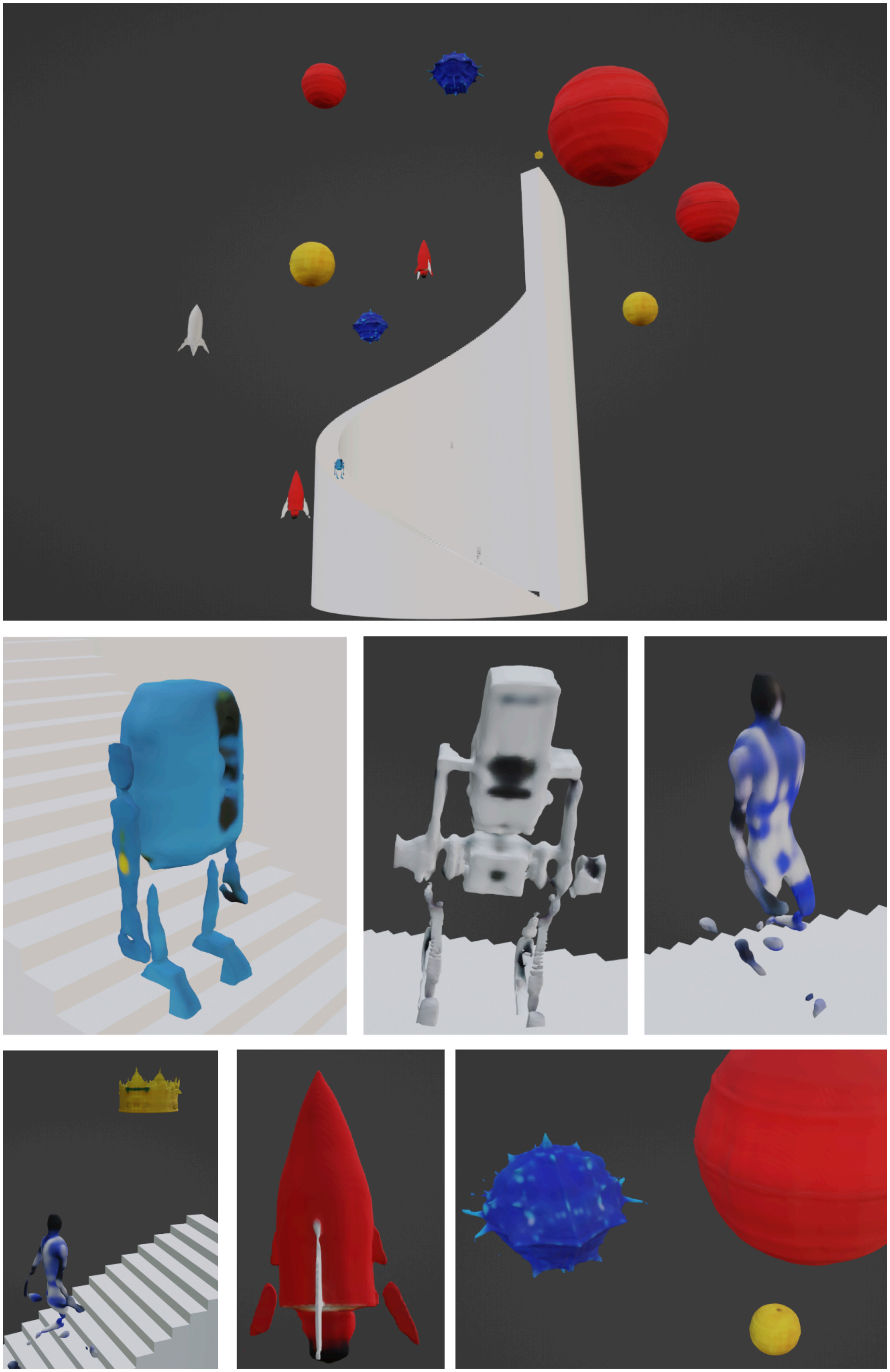


40:

P4:



P5:



42: