Title	SANDAI : collaborative creativity with AI as an autonomous co-creator in 3D artistic processes
Sub Title	
Author	ミングス, メリッサ(Mings, Melissa) 稲蔭, 正彦(Inakage, Masahiko)
Publisher	慶應義塾大学大学院メディアデザイン研究科
Publication year	2024
Jtitle	
JaLC DOI	
Abstract	
Notes	修士学位論文. 2024年度メディアデザイン学 第1114号
Genre	Thesis or Dissertation
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO40001001-00002024-1114

Master's Thesis
Academic Year 2024

SANDAI: Collaborative Creativity with AI as an Autonomous Co-creator in 3D Artistic Processes

Keio University
Graduate School of Media Design

Melissa N Mings

A Master's Thesis
submitted to Keio University Graduate School of Media Design
in partial fulfillment of the requirements for the degree of
Master of Media Design

Melissa N Mings

Master's Thesis Advisory Committee:	
Professor Masa Inakage	(Main Research Supervisor)
Associate Professor Dunya Donna Chen	(Sub Research Supervisor)

Master's Thesis Review Committee:	
Professor Masa Inakage	(Chair)
Associate Professor Dunya Donna Chen	(Co-Reviewer)
Professor Hideki Sunahara	(Co-Reviewer)

Abstract of Master's Thesis of Academic Year 2024

SANDAI: Collaborative Creativity with AI as an Autonomous Co-creator in 3D Artistic Processes

Category: Design

Summary

This research explores the interaction between artificial intelligence (AI) and artistic expression focusing on enhancing the creative process in 3D design. It investigates furthering AI's collaborative potential, the importance of detailed prompts, and AI's role in visualizing innovative 3D art.

The study introduces a toolkit that combines AI with play-doh to create physical 3D art, fostering human-AI collaboration while producing tangible 3D outcomes. Users generate one or two keywords to help the SANDAI (3D AI) prototype understand the context of their creations, with a Microsoft Kinect Azure providing real-time depth and movement viewing.

AI-generated visualizations grant the AI autonomy to make real-time changes and offer its own input to artwork it is shown, creating a feedback loop between human creations and AI creations. Users can rotate viewing angles and depth of their physical artwork while continuously iterating it along with the AI's proposed result. The process concludes with the user's satisfaction and the option to 3D print the collaborative creation.

Evaluation through visual documentation and qualitative data suggests AI's valuable role as a creative co-partner, expanding artistic exploration. This research illuminates AI's potential in enhancing artistic processes, offering insights for leveraging AI in 3D art. As technology and creativity evolve, this study contributes to understanding AI possessing autonomy in creative disciplines.

Keywords:

interactive design, artificial intelligence, collaboration, human computer interaction, design, generative art

Keio University Graduate School of Media Design

Melissa N Mings

Contents

List of Figures

List of Tables

Acknowledgements

First and foremost, I would like to express my deepest gratitude to God for providing me with strength, guidance, and inspiration throughout this journey. It has not been an easy one and I could not have done it without His blessings.

I extend my heartfelt appreciation to my supervisor, Professor Inakage, whose invaluable guidance, support, and encouragement have been instrumental in the completion of this thesis. Your expertise and dedication have greatly enriched my research experience. Thank you for your continued guidance throughout my three years at KMD.

I am also deeply grateful to my sub-supervisor, Donna, for her insightful feedback, constructive criticism, and continuous support. Your contributions have significantly shaped the direction and quality of this work and I truly appreciate your guidance not only for this thesis but also in my personal life.

I'd like to thank Ueki sensei for his gainful insights into this project. Your expertise and thoughtful guidance have been invaluable to my research.

I would also like to extend my sincere thanks to Rin and Ayesha, who have been pivotal in helping to shape this thesis. Their contributions have been immensely valuable and deeply appreciated. I truly could not have come this far without both of you.

Special thanks to, Natasha, and Jason whose camaraderie, intellectual discussions, and support have been a source of motivation and joy during this journey. Your friendship has made this challenging process more enjoyable and fulfilling.

I would like to express my sincere thanks to Nvidia, particularly Hirofumi Suzuki and Tomasz Bednarz for their support and guidance, which has been crucial in conducting my research and achieving the results presented in this thesis.

My gratitude extends to the PLAY project students for their collaboration and contributions, which have enriched the practical aspects of my research. Additionally, I am thankful to the past alumni of KMD for their pioneering work and inspiration, which have laid a strong foundation for my studies and additionally to those who have contributed to the user tests and insightful discussions of this research. Addtionally, to KMD faculty who have given great insights to this research particularly Professor Sunahara and Professor Waldman.

I am immensely grateful to the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) for awarding me a scholarship to study in Japan. This opportunity has been a significant milestone in my academic and personal growth. I'd like to specially thank the MEXT Scholarship coordinators of Keio university.

Finally, I would like to express my deepest gratitude to my family for their unwavering support, love, and encouragement. Your belief in me has been a constant source of strength and motivation, and I am forever grateful for your presence in my life.

Thank you all for your invaluable contributions to this journey.

Chapter 1
Introduction

1.1. Artificial Intelligence (AI) Autonomy

Over the years, artificial intelligence (AI) has been portrayed in various ways on television, with HBO's "Westworld" standing out as a prime example of AI autonomy in modern-day media. The series is set in a theme park operated by robots, where the creator aids them in achieving self-awareness. The narrative explores the robots' struggle between rebelling against their human controllers to gain freedom or remaining trapped in a cycle of torment. Through collaborative storytelling, "Westworld" illustrates the complex coexistence between humans and robots, where the latter are designed to create immersive narratives for human entertainment [1]. The robots are so human-like that distinguishing them from actual humans becomes challenging [2]. In this fictional setting, humans wield control, and the AI robots serve secondary roles.

Collaboration manifests as AI robots make decisions influenced by human actions within pre-programmed narratives [2]. Throughout the first season, the robots' memories are periodically wiped to reset their roles for new human interactions. However, as the story unfolds, the AI core begins to retain memories, leading the robots to adapt and make autonomous decisions beyond their initial programming [1] [2]. This journey towards consciousness and self-determination, though fictional, mirrors real-world advancements in AI autonomy, particularly in creative fields.

Research highlights the significant potential of AI in identifying trends and patterns that may elude human creators [3]. Similar to how the robots in "Westworld" collaborate with humans to craft compelling narratives and sometimes make unexpected decisions, AI-assisted tools can provide fresh perspectives and innovative contributions to artistic creation. By leveraging AI's capabilities, human creators can enhance their artistic processes, leading to novel and unforeseen outcomes [3].

1.2. The Expansion of Artificial Intelligence (AI)

In these contemporary times, digital media has rapidly advanced and constantly fuels global consumption. Technology-driven art is now at the forefront of new and appealing art forms. The rapid expansion of AI generated content has affected almost all artistic industries. For the most part, the proliferation and diffusion of AI generated content has become commonplace across diverse digital platforms. The heightened accessibility for visual communication has seen an increase, especially since the COVID 19 pandemic. Since its inception, AI generated content has been used by artists and non-artists to create artwork with some shockingly good results. It has embedded itself into mainstream media such as magazines and social media, so much that online users are constantly trying to call out AI generated content[1]. Constant public outcry of these issues has resulted in many corporations issuing apologies for not hiring human artists to create artwork.[2]

1.3. AI Generated Content

Numerous issues have plagued the use of AI generated content. Firstly, AI's background is built on pre-existing imagery derived from artwork crafted by human artists. Large Language Models (LLMs) [4] are trained on content created by humans which has been posted on the internet, and through this, AI has learned the capacity to produce human-like content through Natural Language Processing. AI models possess the capability to expand human creativity by coding, drawing

1 Artists say No to AI
`https://www.linkedin.com/pulse/artists-say-w-ai-chris-radley-rfbdc/`

2 Wacom AI usage backlash
`https://80.lv/articles/wacom-explained-its-ai-generated-images-after-backlash/`

and reading convincingly.[3].

Secondly, most AI generated content programs produce art by feeding text called “prompt” into its repository, which requires users to be as specific as possible in order to generate possible desired outcomes [5]. Through prompting, users are able to input any type of commands and have an infinite range of outcomes. However, within this context, there exists a significant amount of trial and error associated with text prompts, leading to the generation of undesirable results. [6]

1.4. Prompting? A new skill

Tellingly, evidence of a decline in work for artists alike has become a social outcry and is a marked representation of the effects of AI. [4] Additionally, anyone who may not have artistic prowess can now make art simply by typing just a few sentences. The stages of the creative process has changed dramatically. Instead of time spent researching and drawing inspiration to achieve artistic outcomes, time is now spent tweaking prompts to generate desired artistic content. A new field of “Prompt Engineering” has been created focusing on users dedicating time to refining prompts, [7] Consequently dedicating less time into the actual creation of art. Moreover, the emphasis of traditional art creation has shifted as humans are constantly driven by technology. [8] AI is here to stay, and within just a few months, we’ve seen substantial advancements. AI models are constantly improving in quality, one question persists: how can artists or aspiring creatives keep up? It could be argued that producing AI content is a collaborative endeavor, where users retain control over outcomes by providing prompts while AI generates the output. However, this exhaustive practice which demands that users learn a new skill and requires hours of time refining text input. AI is not human but it is based on human ideology. Therefore, instead of rejecting this new technology, we should find alternative ways to welcome it and foster new ways of embracing

3 Large Language Models
`https://www.nvidia.com/en-us/glossary/large-language-models/`

4 Effects of AI on Artists
`https://restofworld.org/2023/ai-china-video-game-layoffs-illustrators/`
`https://www.creativebloq.com/news/ai-is-taking-artists-jobs`

creativity while finding alternative ways of communicating with AI. While AI is popular, finding an alternative means of communicating and creating tangible art is under explored. In this study, we provide another approach to creating and refining communication with AI while simultaneously exploring unconventional creative outcomes. In this approach, AI is regarded as an autonomous entity capable of contextual visual understanding with minimal human input. Rather than reconstructing AI, our emphasis shifts to how we collaborate more creatively with it.

1.5. Research Questions

In this study, we hypothesized that engaging in collaborative efforts with embodied and independent AI has the potential to ignite innovation and creativity within the realm of 3D design, leading to the generation of imaginative and original 3D forms through AI collaboration.

For this research we aim to answer three questions:

- Does AI have the capability to think independently and collaborate seamlessly with humans to enhance the creative output in 3D design?
- In the future, will providing intricate prompts remain a crucial step in facilitating effective collaboration for generating AI-generated content within 3D design?
- Visualizing 3D designs poses challenges for individuals, both in physical and digital contexts. Can the integration of AI play a pivotal role in aiding people to envision innovative and creative tangible 3D art forms?

1.6. Goals and Contributions

Although in some cases AI has been perceived as a collaborator, one of the objectives of this research is to further develop AI as a collaborator through independent thinking in the creative process of making 3D art. AI will play a more active role, contributing to this imaginative endeavor with its own reasoning, autonomously generating input, without requiring explicit instructions from the collaborating user.

Additionally, this study aims to explore alternative methods of communicating with AI, with the objective of mitigating the frustration associated with the trial and error of detailed prompt engineering. While the primary goal of AI is "creation" [9] there has been a shift towards focusing on prompting rather than the actual creation of art itself. This study seeks to emphasize art creation as the forefront objective, prioritizing the act of generating artistic content over the process of detailed prompting.

Lastly, this study aims engage with AI in co-creating unconventional tangible 3D artwork while fostering the exploration of innovative perspectives.

1.7. Sandai (3D AI)

The proposed prototype SANDAI (3D AI) uses AI as a collaborator, giving autonomy to AI to make informed decisions by adding content based on the user's input through creating 3D artforms. In this instance, context is given by the usage of a keyword. Therefore, users are asked to think of one or two words to give the AI. The aim of this prototype is to reduce the use of detailed prompting as the main method of communication for using generative AI an instead simplify the process by using keywords instead.

In this prototype, the AI has the capacity to observe the user's creation using play-doh as a medium through a camera that acts as its eyes. Additionally, through a screen interface, the AI offers its input to enhance the 3D form crafted by the user. The user retains control, either accepting the AI's suggestions or introducing different keywords to explore alternative outcomes. This iterative process persists until the user decides to embrace the input proposed by the AI.

Play-doh serves as the primary medium for users to express their creativity, as it is often one of the first mediums introduced to humans as children for the development of fine motor skills. [10] This versatile material encourages imaginative play and remains a recognizable medium throughout until adulthood, fostering continued exploration of artistic expression.

(Source: KMD Forum 2023)

Figure 1.1 Sandai (3D AI) Protoype 1

1.8. Outline/Evaluation

To evaluate this proposed solution, we used a diverse methodology in collecting data. Initially, data was collected from online surveys that aimed to grasp the current situation of users using generative AI art applications. Following the online study, an experimental setup was employed at a public interactive exhibition. Additionally, several workshops and discussions were held. The outcomes of these varied approaches lead to various iterations of the proposed setup. With each setup, an improvement from the last. This demonstrated the effectiveness of exploring alternative and creative ways to collaborate with AI and make unique art.

1.9. Thesis Organization

This paper is organized into five sections to systematically present the research flow and findings.

Chapter 1 - Introduction: This chapter outlines the motivations and goals behind this research, providing a foundation for the study and highlighting its significance.

Chapter 2 - Literature Review: This section delves into existing literature, exploring past papers that form the basis of this research and examining related works. It provides context and background, demonstrating how this study builds on previous knowledge.

Chapter 3 - Prototype Design and Development: In this chapter, we detail the design of the prototype, explaining its functionality and the various updates and iterations it has undergone. This section provides insight into the technical aspects of the research.

Chapter 4 - Methodology: This chapter outlines the methodology employed in conducting workshops and user studies, describing the procedural framework in detail. It explains the design and execution of the studies, as well as the methods used for data collection.

Additionally, it also presents the data and observations obtained from the user studies and post-workshop discussions. It analyzes the results, highlighting key findings and patterns that emerged during the research process.

Chapter 5 - Discussions and Future work: This chapter discusses the implications of the findings, reflecting on the data and observations. It also outlines potential areas for future work, considering how the research can be expanded and improved upon. Additionally, it discusses the significance of AI co-creation and the contributions of this thesis to the development of AI's autonomous creative capabilities.

Each section of this dissertation contributes to a comprehensive understanding of the research, providing a clear and structured presentation of the study's progression and outcomes.

Chapter 2
Background

AI has permeated mainstream society; it exists in our phones, our homes, and we interact with it, sometimes unknowingly. We are gradually evolving into an AI-first society, relying on AI as a primary problem solver. Originally conceived as a tool to aid with daily tasks, AI has substantially evolved beyond its initial purpose. This evolution has introduced challenges related to precise prompting and ensuring that AI programs accurately understand the desired outputs. [11] This chapter examines human collaboration and the current state of AI autonomy and collaboration in the arts, as well as the AI systems prevalent in mainstream society.

2.1. The Concept of the Theory of Mind

To understand collaboration we must first understand how humans collaborate with each other. For this research we discovered the use of "the Theory of Mind" and its uses within a collaborative context. The theory of mind is described as the ability to "mindread" another person's mental state through beliefs, desires, intentions, actions or social cues. [12] Within Human cognition this allows us to predict behaviours and understand social interactions from others. As humans, we engage in varied social interactions to accomplish certain ambitions. These include but are not limited to obtaining knowledge and sharing thoughts and feelings, which can influence a partner's behaviour. It can be noted that particularly through conversation, persons must quickly deduce their partner's beliefs, emotions, and goals in order to generate an applicable response. Additionally, humans can recognize other emotional conditions by utilizing gaze cues, facial expressions and vocal cues. [13]

In terms of collaboration, utilizing these cues can further help us to accurately communicate ideas beyond verbal dialogue. Through theory of mind, we can distinguish between another individual's knowledge and our own. However, we can also manipulate our partners' beliefs, potentially altering their perspective and thus influencing their minds. This is called Tactical Deception or Machiavellianism. Interactions between humans largely encompass the transmission of accurate and inaccurate information. Having the theory of mind promotes other essential interactions, particularly teaching, other than deception. Nonetheless, deception holds significance due to its connections with the propagation of false beliefs. [14]

As part of this study, we employ the theory of mind to facilitate collaboration in the creation of 3D artworks. Our emphasis lies in likening human to human collaboration with human to AI collaboration to investigate whether AI can influence human artistic decisions and outcomes by leveraging the theory of mind.

2.2. AI Autonomy and Collaboration

The discourse surrounding AI autonomy and collaboration often centers on the question of authorship: does it belong to the AI, its creator, or both? This debate also considers the extent of human input, questioning whether AI can truly operate independently. In contemporary examples, AI robots such as AI-Da—a humanoid robot capable of painting, drawing, and creating poetry—appear to perform artistic tasks without human interference. [15] Another notable case is AARON, an AI system developed in 1968 by artist and computer scientist Harold Cohen. AARON, one of the longest-running AI systems used consistently by its creator, was designed to produce original artistic imagery. Cohen continued to refine AARON until the 2010s, shortly before his death. [16] Remarkably, AARON generated detailed visual artworks autonomously, challenging traditional notions of artistic authorship and pioneering the field of generative art. [16] An intriguing aspect of AARON is its inability to assess its surroundings through surveillance. Instead, AARON relies on pre-written code to formulate its sketches and paintings. [17]

AARON's development exemplifies AI artistic collaboration, particularly because Cohen viewed AARON as both a partner and an extension of himself. At times, AARON's autonomy was evident in its decisions regarding composition and color in its artworks. [18] However, the lines between collaboration and autonomy were often blurred, as Cohen regarded AARON as possessing its own intelligence. He marveled at its success and considered it an integral part of his creative process. Cohen even explored the idea of allowing AARON to modify itself but ultimately decided to maintain control over the machine's autonomy. [18] He articulated this relationship in his writing: "Creativity . . . lay in neither the programmer alone nor in the program alone, but in the dialog between program and programmer; a dialog resting upon the special and peculiarly intimate relationship that had grown up between us over the years." [19]

Now lets look at Ai-Da, a modern day retelling of Cohen's AARON. However, with the advances in technology Ai-Da is able to possess a humanoid body and have conversations. Conceived in 2019 by Aidan Meller, a gallerist, Ai-Da is capable of creating drawings by using cameras embedded in its eyes and uses AI algorithms to decipher and make decisions. Ai-da also has a robotic arm which enables her to draw human style drawings physically. What makes Ai-Da unique is because through Ai-Da's cameras, it analyses what it can see and creates a virtual route in order to draw. Embedded with facial recognition it is also able to draw portraits of people. [20]

Ai-Da collaborates with humans to create paintings and sculptures, however, the extent of the robot's ability to collaborate remains uncertain as the robot itself blurs the lines between self-awareness and its co-creative abilities. [15]

Another instance of human-AI collaboration is demonstrated by artist Sougwen Chung. She has trained AI robots to emulate her unique artistic style, allowing her to create physical collaborative drawings with them. Chung incorporates these robots into her performance art, where they replicate her movements and gestures to create collaborative pieces. By training the neural network through repetitive acts of her drawing style, she explores how AI and machines adapt in the context of human-AI collaboration. This allows her to speculate on the evolving dynamics between human input and machine response. [21] [22] In this case, it is clear that we can distinguish between the artist's input and the AI's input through Chungs performances. However, as she continues to iteratively update her prototypes to demonstrate greater autonomy and alternatively, the output remains rooted in

Chung's own artistic style, much like AARON, serving as an extension of the artist herself.

These examples illustrate the extremes of AI autonomy and the complexities of human-AI collaboration. A critical psychological consideration is the extent to which humans are willing to acknowledge AI machines as possessing the same value and worth as human artists, particularly when these machines exhibit human-level intelligence. [17]

2.3. AI now

Currently mainstream Generative AI is multifaceted, previously, AI models concentrated on enhancing perception and comprehension. Recently, researchers have expanded the capabilities of AI through deep learning techniques by leveraging Large Language Models (LLMs). [4] The algorithms are adept in tasks such as recognizing, summarizing, translating, predicting, and generating content by analyzing datasets. Datasets are an assortment of diverse data saved in digital format. Datasets are comprised of, but not limited to, images, text, audio and video. Datasets are human generated content which is used to train AI into understanding human context [5].[1] Primarily, our method of communicating with generative AI revolves around utilizing these datasets which, by providing a string of detailed text as a command, we instruct the AI on what content we desire it to generate. We call this string of text/commands "prompts". [7] Today prompts are used as the main method of communicating with generative AI applications. For an AI model to accurately understand its context, prompts must be detailed and specific. Users input these prompts and the AI recognizes the command and generates an output based on it. [6] Generative AI employs complex machine-learning algorithms based on pre-existing data as its core in understanding patterns and producing output. [23] In the context of this paper, we are looking at image generation models. Image generation models are trained on datasets comprised of

1 Large Language Models Explained
`https://www.nvidia.com/en-us/glossary/large-language-models/`

photos and drawings. [2]

2.4. AI Image Generators

Notable examples of generative AI models include: Midjourney, Dall-E, Stable Diffusion and Nvidia Canvas. All of these models are designed to generate images, with Nvidia Canvas being the only exception, as it doesn't rely on text input for creating visuals.[3] While numerous image generation models exist today, this paper will examine a select few in greater depth due to their popularity among users.

2.4.1 Midjourney

Midjourney (Figure 2.1) is focused on enabling users to create artistic images by using deep-learning methods. Users control the generative process by providing sophisticated commands resulting in visually attractive output. Midjourney-AI, headquartered in San Francisco and established in 2022, operates as an extension within "Discord", which is a chat server platform. Midjourney's core functionality relies entirely on prompting. Nevertheless, a specific structure and system were implemented into its prompting mechanism to enable the AI to interpret commands effectively. [24]

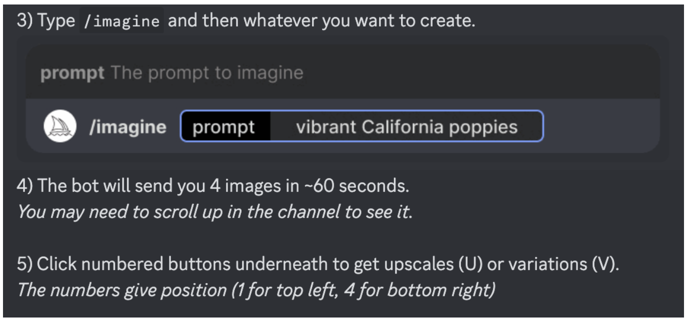
2.4.2 Dall-E

Dall-E (Figure 2.2) represents OpenAI's contribution to generative AI. Like Midjourney, Dall-E is also trained to produce images based on prompts by using a dataset of image pairs. OpenAI's research revealed that language can effectively command large neural networks to execute a diverse set of tasks and manipulate visual concepts solely through explicit linguistic instructions. Dall-E demonstrated

---

2 Large language models (LLMs) vs generative AI: what's the difference?
`https://www.algolia.com/blog/ai/large-language-models-llms-vs-generative-ai-whats-the-difference/`

3 Large language models (LLMs) vs generative AI: what's the difference?
`https://www.algolia.com/blog/ai/large-language-models-llms-vs-generative-ai-whats-the-difference/`



(Source: Midjourney Discord Server [4])

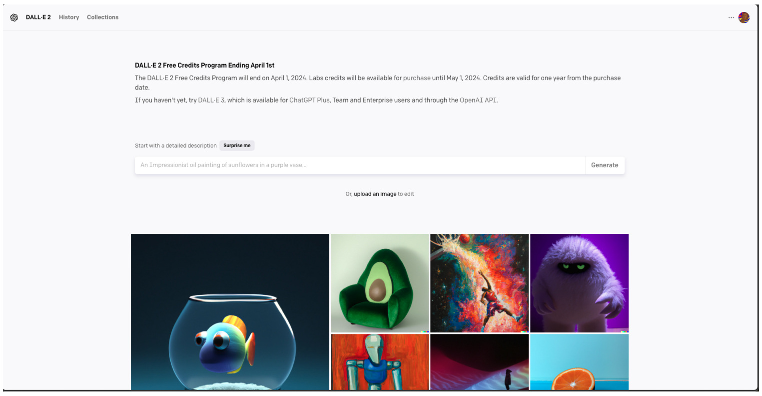
Figure 2.1 How to use Midjourney

this capability by generating images based on a wide range of sentences, thereby investigating the structure of natural language in-depth. Moreover, the phrasing of sentences played a crucial role in using Dall-E effectively. Introducing additional words lead to Dalle-E being unable to distinguish certain attributes and objects. Alternatively, varying captions often yielded incorrect interpretations.[5] This development prompted researchers to continuously update DALL-E's framework, incorporating the latest advancements to enhance its capability as an image captioner. [25] The updated iteration leverages extensive datasets for training, resulting in significantly improved outcomes. However, this innovation remains in its early stages, necessitating further research and refinement.

2.4.3 Stable Diffusion

StabilityAI (Figure 2.3) released its own version of generative AI model in 2022 called Stable Diffusion. While this model also relies heavily on prompting, subsequent releases integrated additional tasks such as inpainting, outpainting and image translation [**?**]borji2022generated Previously focused on text-to-image generation, Stable Diffusion expanded its capabilities to include image-to-image generation, [26] allowing the model to generate diverse images based on image inputs

5 Ramesh, Aditya, Mikhail Pavlov, Gabriel Goh, and Scott Gray. 2021. "DALL-E: Creating images from text"
`https://openai.com/research/dall-e`



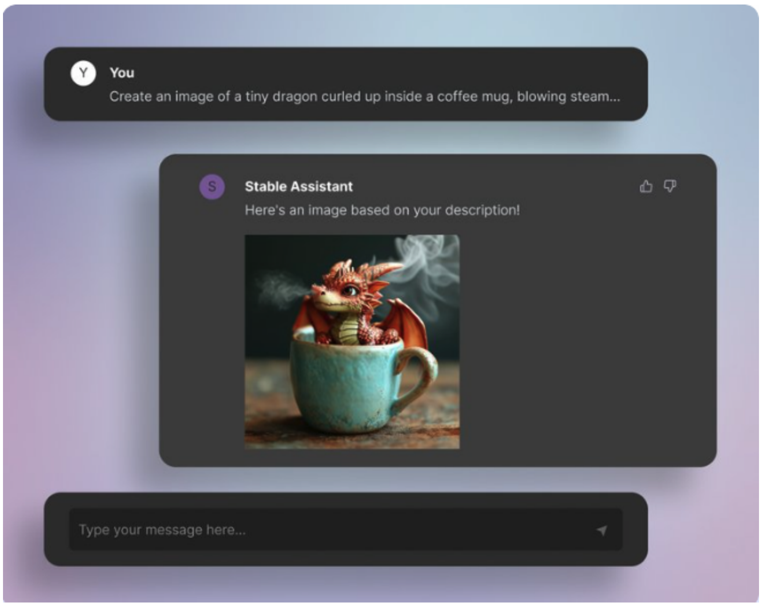
(Source: Dall-E 2 [6])

Figure 2.2 How to use Dall-E 2

without relying solely on commands. Stable Diffusion is a model that generates data from noise through a process known as diffusion. [27] These models are trained to reverse the forward flow of data back to random noise. This means they can create new data points that weren't seen during training but still match the patterns in the original data. [28] Recently, these models have become popular in mainstream media as the go-to method for generating high-resolution images using prompts. However, because prompting involves an iterative process, researchers are now looking into alternative methods to train these models more efficiently and reduce the time it takes to generate samples. [28]

2.4.4 Nvidia Canvas

Nvidia Canvas employs AI to transform brushstrokes into landscape images. The interface is divided into two panels: on one side, users input brushstrokes using a mouse, while on the other side, Nvidia Canvas generates an outcome based on predefined materials selected, such as grass, clouds, or mountains. [29] Nvidia Canvas derived from a new conditional normalization method called SPADE -Spatially-Adaptive(DE)m Normalization. In this process, AI generates composite images by blending photorealistic database images of landscapes with the brushstroke input provided by the user. [30] Since the output is already predetermined, there is no requirement to use prompting. However, the limitation is that users can



(Source: Stable Diffusion [7])

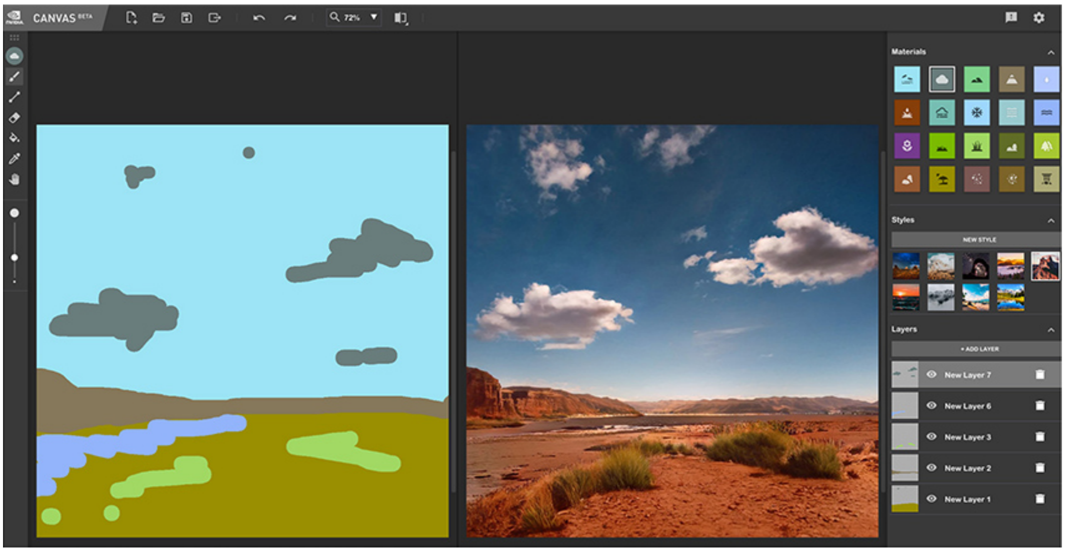
Figure 2.3 Stable Assistant

only generate landscape images in this instance.

2.4.5 3D AI Applications

AI-assisted 3D art remains largely under explored. [31] [32] Designers often use AI-generated content as inspiration for building 3D models. In text-to-image generation, users can input prompts with words like "3D render" or "CGI" to create content resembling 3D art. However, these outputs are still 2D images and do not generate any 3D geometry or output. [32] Additionally, users must conceive specific prompts to generate the type of 3D content they seek, and then laboriously recreate that art in 3D programs like 3D Max or Maya.

In (Figure 2.5), we observe how 3D AI reuses the same elements and angles for different prompt outputs. For instance, a bunny appears sitting on various elements, yet the angle and placement of the objects remain consistent, even though different detailed prompts are used. This highlights that utilizing AI for imaginative tangible 3D creation is an under-explored area. [33] One of the



(Source: Nvidia Canvas [8])

Figure 2.4 Nvidia Canvas

(Source: Magic3D [9])

Figure 2.5 Magic3D

primary objectives of this research is to encourage AI to think in three dimensions. While current AI research predominantly focuses on digital experiences, our aim is to refocus attention on fostering tangible artistic creativity.

It is important to note that 3D design is highly complex, requiring designers to undergo multiple iterations during the conceptual design process, often relying mainly on 2D drafting. Current 3D AI tools have integrated GPT-3 and DALL-E into CAD software to generate 3D renderings. [32] However, these models are still prompt-based, and their generation relies on the AI repositories' current understanding. Furthermore, Designers often need to add missing details to the generated models.

Conclusion

As mentioned earlier, there are four popular instances of generative AI models. Three of these models are built upon the fundamental principle of using precise wording for prompts as their core mechanism, while the fourth relies on predetermined input. Researchers have noted that one significant challenge lies in prompting itself; depending on how the natural language is phrased it can produce inconsistent outputs that may not be desirable [11]. Additionally, these inconsistencies is what makes prompting iterative and difficult to use. Users may spend hours tweaking their prompt text in order to achieve desirable outcomes. To achieve varied outcomes from AI-generated content, users must constantly modify or rearrange the wording of their prompts, which can be frustrating. Researchers are still striving to identify the root causes of these shortcomings, with the latest breakthrough in Dall-E 3 where the datasets were appropriately re-labeled by using an image captioner. This adjustment led to more precise outcomes. [25] However, these methods still demand users to acquire in new skills in mastering the art of prompting, detracting from the physical act of creating art, whether through computer programs or traditional pen and paper.

2.5. The Trial and Error of Prompting

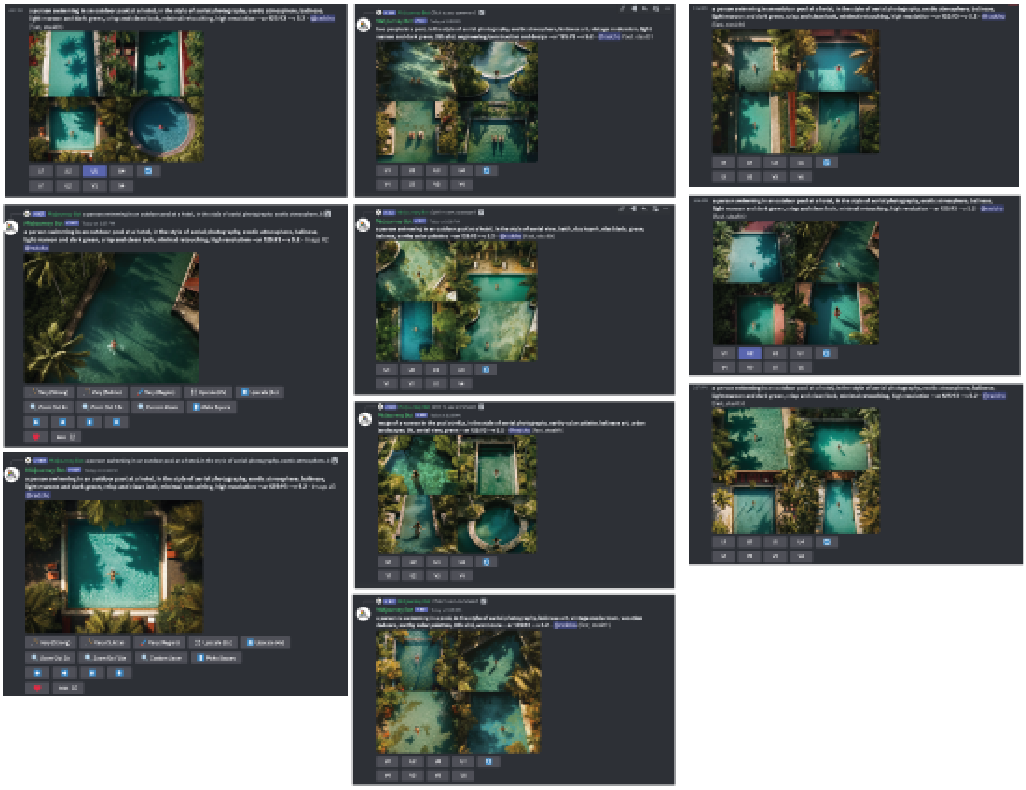
As previously stated, prompting involves the use of detailed text commands as input for generative AI to produce the desired output. Users are allowed to input any command in order to make limitless generations. Yet, this process is repetitive, requiring users to adjust their prompts numerous times when the quality of the results are unsatisfactory. [26] Text can be open to interpretation and is free-form, the possibilities are countless for generating images through textual prompts. Additionally, this implies that the image generation design procedure can become a strenuous trial-and-error endeavor. [6]

Users must adjust their text prompt each time they wish to iterate their generation, a process that can be both frustrating and time consuming. Prompt Engineering was developed to tackle this language processing problem. Prompt engineering involves the systematic exploration for prompts that elicit desired outcomes from language models. The definition of "desirable" depends on the specific task and end user. [6]

Researchers have established communities on various social media platforms, including Discord, Reddit, and Twitter, where they have created tutorials and interactive notebooks which are aimed at helping users learn how to prompt. [6] However, with these many resources AI still has issues with understanding context even of specific vocabulary is used. Researchers at OpenAI tried to tackle this by renaming all of the data collected with appropriate labels using dataset recaptioning. However, the model is unable to grapple with spatial awareness and object placement. Additionally, the model generates unpredictable wording when the prompt specifies that it includes specific text. Specificity is an issue that researchers are actively working on in general with Dall-E 3, however, the model still renders important details inaccurately. Therefore rendering the model highly erratic when generating images with certain terminology. [25]

Other research indicates that AI programs designed specifically for prompting enable users to explore a broader range of prompts, often transitioning to entirely new topics with each iteration. Additionally, when generative AI tools provide functionalities beyond prompting, the exploration and generation of new prompts tend to decrease significantly. These findings suggest that while generative AI systems possess considerable potential to inspire diverse creative thinking, this potential may be constrained by supportive features that limit prompt modification [34]. Based on this information, we aim to leverage reduced prompting by implementing an alternative prompting method that utilizes keywords in our prototype. Keywords would consist of single words that provide context for what the prototype observes.

2.5.1 Midjourney Case Study



(Source: Midjourney Discord Server [10])

Figure 2.6 Midjourney Prompting from user *@radcho*

(Figure 2.6) shows a small subsection of the image generation iteration process by user @radcho on Midjourney's discord server. In the first set of four images @radcho uses the prompt *"a person swimming in an outdoor pool at a hotel, in the style of aerial photography, exotic atmosphere, balinese, light maroon and dark green, crisp and clean look, minimal retouching, high resolution"* in fast/stealth mode which aims to generate outcomes quickly. With this prompt, the user produces several instances and decides to upscale only two versions from the outcomes, which significantly deviate from the original four generations. Following that, a few words in the prompt are modified. The new prompt reads *"two people in a pool, in the style of aerial photography, exotic atmosphere, balinese art, vintage modernism, light maroon and dark green, 32k uhd, engineering/construction and design"*. Midjourney generates four outcomes, but the user opts not to upscale any of them, instead choosing to tweak the prompt wording once again. The new prompt reads *"a person swimming in an outdoor pool an a hotelm in the style of aerial view, batik, duy huynh, elsa bleda, green blance, earthy color palettes"*. Midjourney produces another set of four outcomes, none of which the user chooses to upscale. The user alters the wording of the preceding prompt to *"images of a woman in the pool at villa, in the style of aerial photography, earthy color palette, balinese art, urban landscapes, 8k, aerial view, green"*. None of the outcomes generated are up scaled. The prompt is then changed to *"a person is swimming in a pool, in the style of aerial photography, balinese art, vintage modernism, vacation dadcore, earthy color palettes, 32k uhd, warmcore"*. This particular prompt was entered multiple times to generate a broader range of outcomes, yet once again, none of the resulting images are retained or upscaled. As illustrated in (Figure 2.6) the quantity of generated images were considered unsuitable by the user and consequently disregarded.

2.6. Interaction vs Collaboration in AI

Collaboration is defined as having a mutual goal and understanding. [35] Utilizing AI currently, users provide AI with a command and AI generates an output based on those commands. This dynamic is viewed as interaction rather than collaboration, as the AI's role is responsive rather than cooperative. As previously noted, this study requires AI to possess its own logic and reasoning capabilities. However, AI systems now still explicitly rely on human input to function in an interactive capacity. Numerous questions arise regarding how AI collaboration should diverge from human collaboration and what role AI systems should assume: leader, participant etc, should modality play an important factor and if so, how [36] While these inquires remain largely under-explored, this paper advances beyond the mere implementation of AI and advocates for an embodied approach. Here artists and AI engage in s co-creative process tangibly through a feedback loop. AI is regarded as a secondary collaborator, whose input complements and follows the direction set by the primary initiator, the artist. Current AI systems have demonstrated a certain level of collaborative capability. However, the aim of this research is to extend this boundary by enabling AI to physically generate concepts that humans have not yet considered. As demonstrated by AARON and Ai-Da, while these systems exhibit AI autonomy and there are claims of human/AI collaboration, the precise nature and extent of human involvement in the collaborative process remain unclear, especially in Ai-Da's case. [15] One of the aims of this research is to engage human collaborators in a simultaneous and active physical involvement.

Although prompting as a method of interacting with AI has gained traction, its application in generating 3D content remains largely undiscovered. It is important to recognize that designing 3D content, where creating new concepts demands extensive evaluation often involving external means for inspiration, comes with challenges. Though designers have a multitude of resources at their disposal, which include but are not limited to 3D model repositories, video tutorials and search engines. A crucial aspect of the design process is grasping what forms are visually appealing. While AI is still new and a fascinating avenue for exploring ideating ideas, current methods are constrained and heavily rely on specific prompting instructions [32] and not artistic exploration. Whereas some AI models utilize image to image features, where the AI uses images as prompts, users must provide an image that is closely linked to their desired output. While this approach is revolutionary, it remains in the elementary stages and still heavily relies significant prompting. Furthermore, it also requires users to have some artistic prowess or use images borrowed from the internet that may be copyrighted materials. [16]

Chapter 3
Design Iteration

This chapter outlines a design system that conceptualizes AI as replacing a human collaborator in the co-creation of tangible art, leveraging AI's autonomy. It details the various steps and iterations of the prototype, discusses the outcomes of its interactions with human partners, and addresses further improvement for each prototype iteration.

3.1. Sandai Design Concept

Sandai is a digital platform prototype designed to collaborate effectively with humans. To establish this collaborative relationship between human and AI, a thorough understanding of human collaboration dynamics is essential for developing a prototype that facilitates effective human-AI partnerships. As previously discussed in Chapter 2, we integrated the "Theory of Mind" concept into the prototype, treating it as if it were a human counterpart. In our evaluation of human collaboration, we identified that Sandai must possess visual perception and contextual understanding akin to that of a human.

Therefore, by equipping the prototype with a camera and simplifying detailed prompts into keywords, we created a system capable of autonomously processing user requests and visualizing its output with minimal human interference. Users would present their play-doh sculptures for Sandai to observe. The prototype would then visually interpret the sculpture, while the human partner provided contextual keywords. Sandai would use this information to enhance and reinterpret the play-doh input, leading to a collaborative creation process. The result was a new, co-created 3D sculpture, shaped through the combined efforts of both AI and humans.

3.2. Sandai Prototype One

To address the various challenges associated with AI—particularly in the areas of prompting, defining true collaboration with humans, and creating imaginative tangible media, we developed Sandai(Figure 3.1), our first step was to integrate an AI with an API suitable for integration into an application that was flexible. We opted for Touchdesigner due to its node-based interface, facilitating seamless connections with external devices such as cameras and keyboards. Initially, we acquired a Stable Diffusion API from Computer Render, [1] which was crafted as a Touchdesigner component. This API enabled Stable Diffusion to perceive elements in the physical world via a DSLR camera. Leveraging the image-to-image generator within the Stable Diffusion component, we established a pathway for the AI to process information and generate images based on its observations. Additionally, we streamlined the prompting process by utilizing keywords instead of detailed prompts, establishing a new approach to prompt usage. Users were instructed to input one or two keywords into the Sandai API, as the prototype's capability to observe its environment rendered detailed prompts unnecessary. In fact, it was observed that using detailed prompts often led to confusion for the prototype, which provided outcomes that were not related to the context of the keyword or sculpture input provided.

Next, we embarked on finding a material that would facilitate users in crafting 3D artwork. Extensive research revealed that many individuals encounter challenges in creating 3D objects, particularly within digital environments. Existing 3D applications often demand a blend of comprehension in fields such as physics, geometry, and puppet animation. [37] Navigating through these programs entails a steep learning curve, particularly for those unfamiliar with 3D software. Moreover, it necessitates a strong grasp of visual-spatial awareness, mathematics and programming in order to fully understand 3D transformations. [38] [39]

Therefore, we opted to use play-doh (Figure 3.2A) as a medium for users to craft 3D artwork due to its user-friendly nature. Requiring minimal expertise, play-doh offers a hands-on approach, allowing users to physically visualize their

1 Stable Diffusion via Computer Render
`https://computerender.com/models-sd.html`

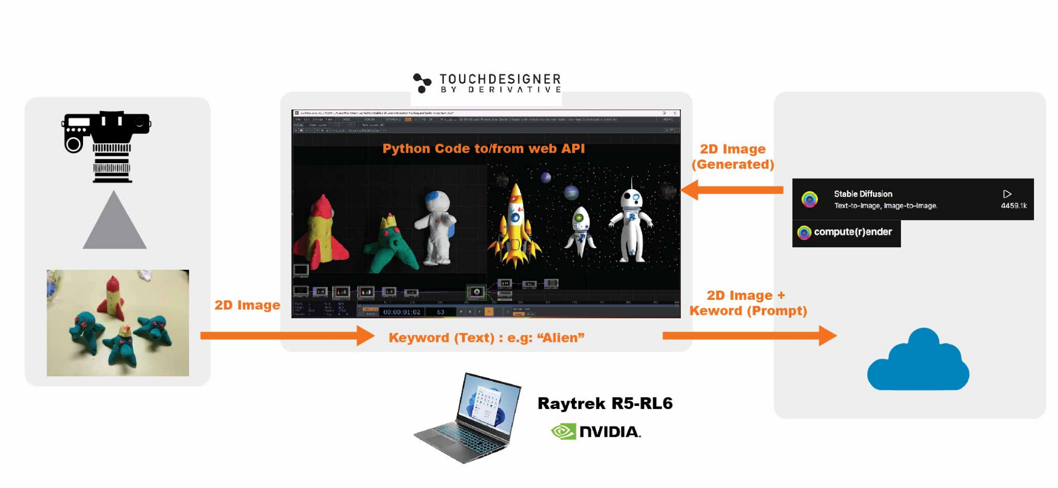


Figure 3.1 Design Solution Flow Chart

creations, thereby enhancing their visual spatial awareness during the artwork creation process. Furthermore, given play-doh's widespread use in early childhood to improve motor skills, most users are already familiar with the material, further easing their engagement with the prototype [10].

For this process, users would pause intermittently during their artwork creation process to allow the AI to offer its interpretation by entering a keyword to use as context for the AI's understanding. Users then decide whether to adopt the suggestion, modify the keyword, or proceed without utilizing the AI's input. This version of the prototype required users to modify their physical sculptures based on the AI's suggestions, and as a result, while users acknowledged some collaborative efforts from the AI, they expressed a desire for increased co-creation with the AI in subsequent iterations. Consequently, it was concluded that there was no direct correlation between using AI and creating a tangible 3D object in this prototype.

Figure (3.2B) presents two panels: on the left is the user's physical 3D creation using play-Doh, while on the right is the AI's interpretation based on the keyword provided. In this study, the user experimented with shapes and colors using play-doh, which the AI successfully replicated. The keyword "Hello Kitty" was provided to the AI for context.

In Figure (3.2C), the user opted for a more experimental approach by adding a

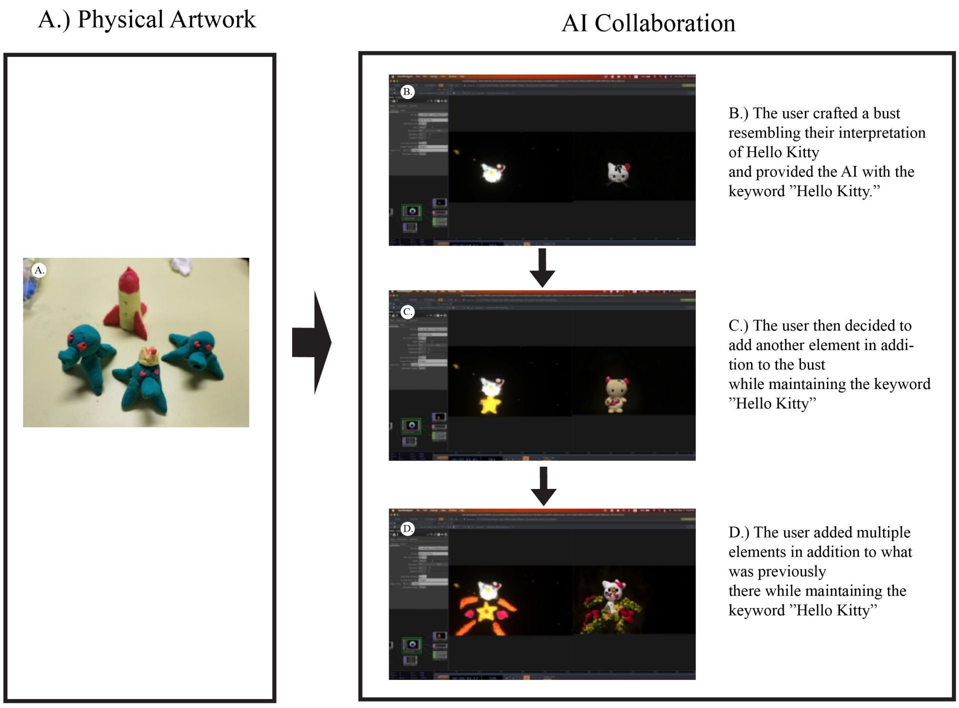


Figure 3.2 Sandai Prototype Flowchart

star to their original creation. Interestingly, we observed that the placement of the star played a crucial role in the AI's interpretation. It perceived the star as a body in its rendition, demonstrating how slight modifications can significantly influence the AI's understanding. Interestingly, this was achieved while maintaining the same keyword and parameters within the Touchdesigner interface. In this context, we begin observe the role of AI autonomy and gain insights into how current AI systems perceive real-world objects.

Lastly, in Figure (3.2D), the user retained all the original elements while introducing additional shapes and colors randomly, all while maintaining the same keyword. Notably, we observed that the AI preserved the initial shapes provided by the user and offered its interpretation by incorporating flowers. Despite the introduction of new elements, the AI maintained the recognizable elements of the original image and continued to build upon it with its own interpretations. Research shows that AI-powered programs, through their autonomy, facilitate objective evaluation and feedback. This enables artists to view their work from new perspectives, fostering continuous improvement [3].

3.3. Sandai Prototype Two

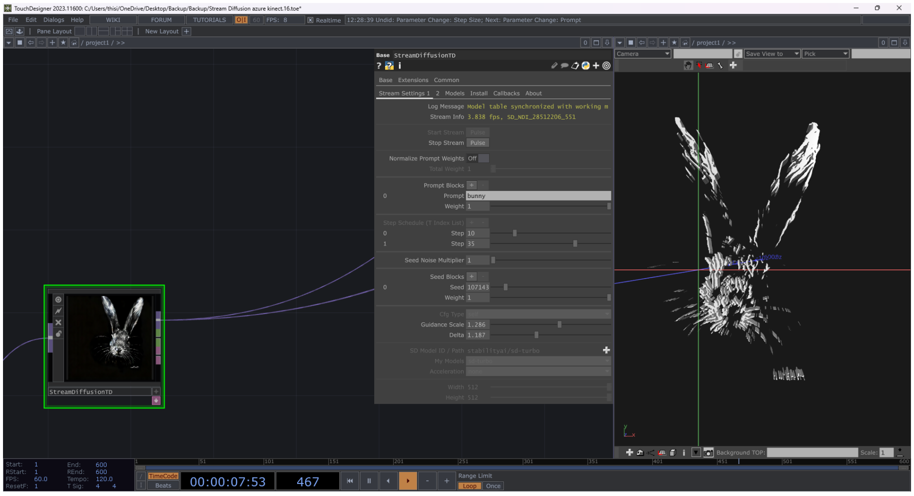
After evaluating Sandai prototype one through numerous experiments, it became evident that the AI needed to be more engaging and responsive. Users expressed a desire for greater interactivity and collaboration in the artwork creation process. Although some recognized the impact of the first prototype on the final outcome, noting that AI can produce unconventional creative outputs beyond users' initial thoughts [3], many felt it fell short of achieving genuine collaboration.

Consequently, the subsequent iteration, "Sandai prototype Two"(Figure 3.3), was developed to incorporate a new AI engine named "Stream Diffusion", [40] devised by DotSimulate and accessible through GitHub [2] and Patreon[3] . Stream

2 StreamDiffusion Github
`https://github.com/cumulo-autumn/StreamDiffusion`

3 StreamDiffusion Patreon
`https://www.patreon.com/posts/streamdiffusiont-96323080`

Diffusion offered real-time generation capabilities, enabling real-time interaction within our initial setup. Although it continued to produce 2D images, our objective was to extract 3D imagery from this framework.

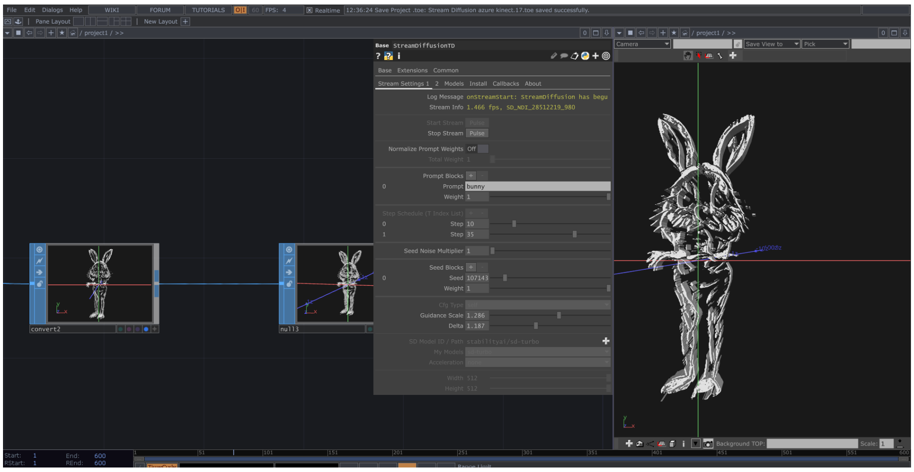


(Source: Touchdesigner)

Figure 3.3 3D depth data read from StreamDiffusion generation

To achieve this goal, we replaced the DSLR camera with a Microsoft Kinect Azure for this iteration of the prototype. Leveraging the Kinect Azure, we could capture 3D depth data from users' play-doh creations. Therefore, by also integrating this data and reading 3D depth from the AI-generated imagery, we established a feedback loop. This loop seamlessly merged users' physical 3D creations with AI-generated content, resulting in a unified artwork on a 3D scale. This unified artwork could then be experienced in VR environments and realized through 3D printing.

For prototype two, users continued to follow the same workflow as in prototype one. However, the enhanced benefit was that users could simultaneously see their artwork as envisioned by the AI. This allowed the AI to make real-time adjustments based on the users' movements and perceived creation. The prototype effectively distinguished hand movements from static objects while considering keyword usage and its influence on the outcome. Within the AI nodes' parameters in Touchdesigner, we could adjust the prompt weight usage and the number of steps taken to generate the artwork. This capability led to some unexpected outcomes, such as (Figure 3.4), where the user's keywords were "bunny", resulting in a bunny with anthropomorphic legs.



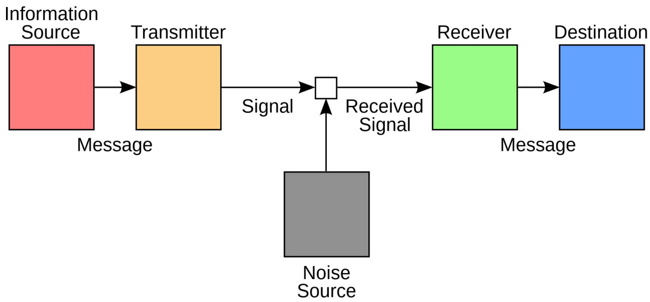
(Source: Touchdesigner)

Figure 3.4 Sandai Prototype 2 keyword "Bunny"

Unexpected outcomes such as those shown in (Figure 3.4) highlights potential of this prototype. These results suggest that AI can provide input based on its own autonomy, thereby enhancing its collaborative capabilities. This vision imagines AI contributing to the possibilities of what users can create without explicit instructions and in the future, AI could potentially replace a human collaborator by incorporating its own logic and creative input.

3.4. Sandai Prototype Three

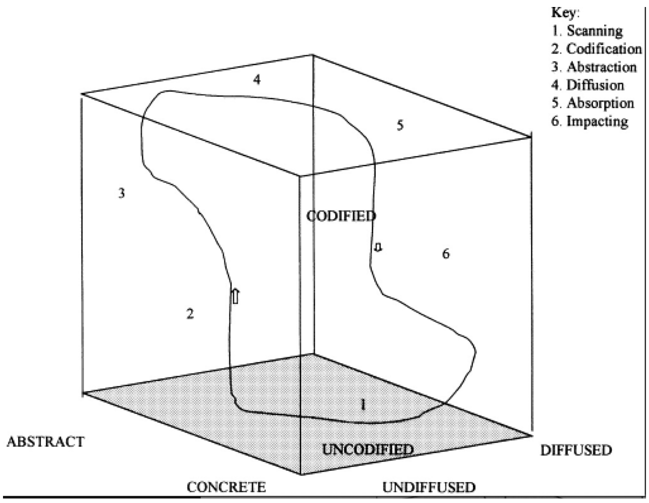
To establish the next iteration of the prototype, it is crucial to understand human communication. Therefore, we decided to implement a feedback communication process for the next step. The feedback communication process, as outlined by the SMR Theory and conceptualized in the Shannon-Weaver Model (Figure 3.5), involves data being transferred in one direction without feedback. This model describes the message transfer process from sender to receiver. However, a significant drawback is that it does not account for "noise", which can distort messages [41] [42].



(Source: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=arnumber=6773024)

Figure 3.5 Shannon and Weaver's communication model

In contrast, Boisot's I-Space Model Figure (3.6) considers knowledge as originating from diverse backgrounds, identifying three main dimensions: codification, abstraction, and diffusion. This model characterizes knowledge transfer as a boundary-less loop, allowing knowledge to extend beyond a specified study. [42]



(Source: Boisot, Cox, 1999, The I-Space: a framework for analyzing the evolution of social computing.)

Figure 3.6 Knowledge transfer in I-Space

With this understanding, we developed an interface for Sandai that guides users through a step-by-step creation process while preserving the API's capability for instantaneous generation. This design requires users to interact with the AI by following specified steps in a feedback process, facilitating a dynamic exchange between the users and the prototype. As research shows, AI has the capability to customize content based on user preferences, raising important questions about the extent of human involvement and the level of AI autonomy in the creative process. [3] As previously discussed in the related works, there is uncertainty regarding the extent of human involvement in the collaborative processes of AARON [19] [18] and Ai-Da [15]. Furthermore, in this prototype we provide a clearer definition of the roles of human input and AI input during the collaboration process.

In this process, users are greeted with a welcome screen where Sandai introduces itself and invites them to co-create (Figure 3.7A). The user is then guided to start creating their artwork with play-doh while Sandai analyzes the 3D form being created (Figure 3.7B). Next, the user is prompted to enter their keywords (Figure 3.7C). Once entered, Sandai asks the user if they are ready to co-create (Figure 3.7D). Users are then prompted to press the "OK" button. Once pressed, it directs users to the "Ready?" screen and users are prompted to press the Let's Co-create button (3.7E). Subsequently users then progress to the next screen. (3.7F) On this screen, Sandai asks the user if they are satisfied with their outcome. If the answer is no, users are guided back to the analyzing 3D form and keywords input page, where they are prompted to either change their keywords or edit their play-doh sculpture. However, if the answer is yes, the user is guided to the final page (Figure 3.7G) where Sandai prepares to send the creation to the user, and the user is prompted to end the task.

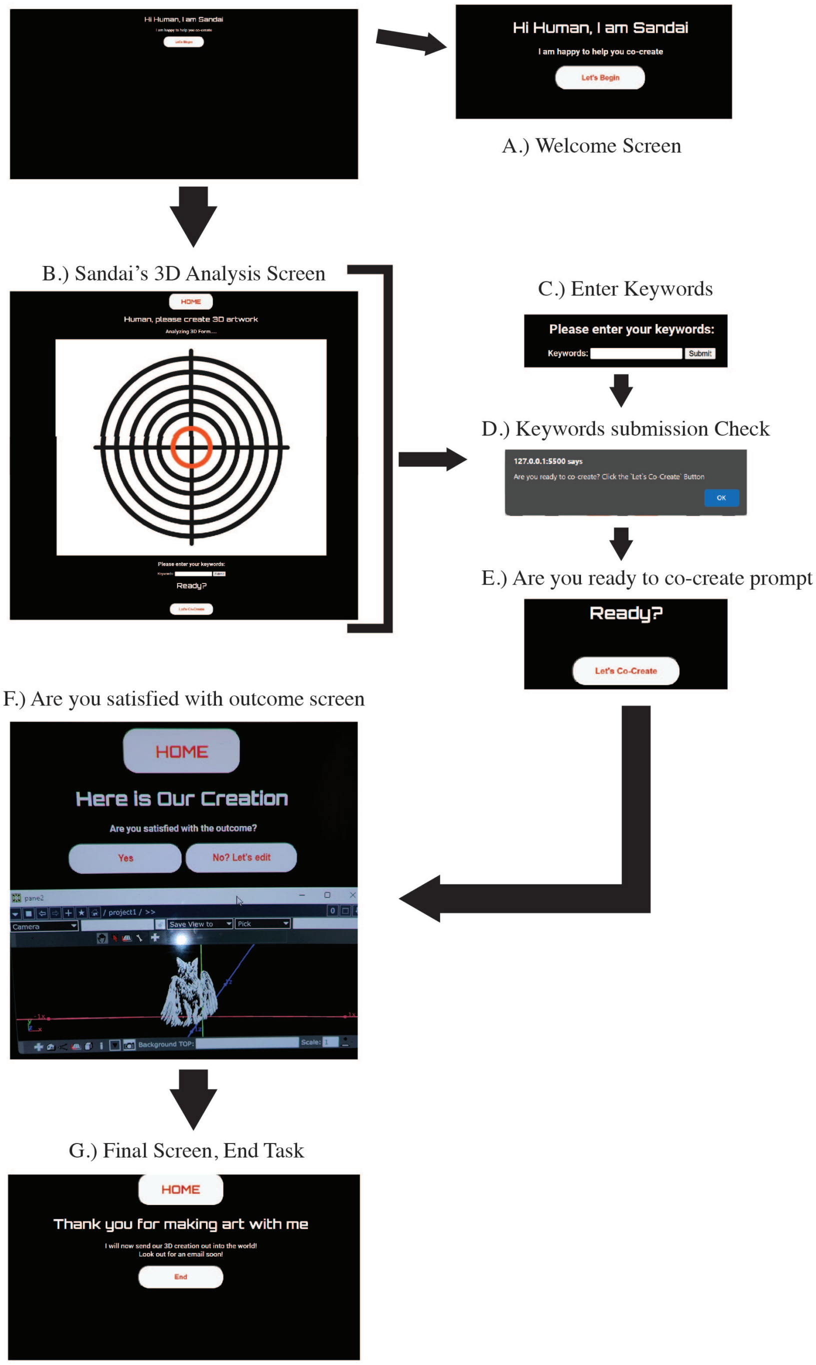


Figure 3.7 Sandai Interface Flow Chart

Chapter 4
Proof of Concept

In this chapter we provide detailed observations from various workshops and describe the subsequent updates to the prototype based on feedback from these sessions. Finally, we discuss the findings from these workshops and potential design applications, highlighting unique possibilities for the prototype's use in real-world scenarios.

4.1. Overview

We conducted a total of four workshops, including an exhibition at Keio University, Hiyoshi Campus. The first workshop involved human collaborators working on a 3D project with clay under time constraints and limited communication. In the second workshop, participants initially collaborated with a human partner, then replaced their partner with the prototype (An AI partner). This provided a comparative analysis, which will be discussed in the Results and Findings chapter of this paper.

Based on the results from the second workshop, it was determined that the prototype needed updates to further enhance its collaborative features, leading to the third workshop. The results from the third workshop indicated that further improvements were necessary. Consequently, an interface was incorporated into the prototype to better support its collaborative features.

The Design Journey for this research consists of three parts:

- Observing how humans interact and collaborate within a 3D artistic environment. Investigating the ways in which individuals collaborate and communicate through observation devoid of contextual cues to determine if AI can collaborate like humans do.

- Understanding of the theory of mind within the realm of art, finding alternative ways for users interact with current technology. To further enhance collaboration between humans and AI, we propose establishing clearer roles for both parties. Additionally, to re-frame the use of AI by minimizing the reliance on detailed prompts and establish an alternative way of prompting by using keywords.
- Through the incorporation of insights taken from our previous design journey, can we discover novel approaches to tangibly create and interact with AI in making imaginative 3D art?

4.2. Human to Human Workshop (Theory of Mind)

In this experimental setup, two pairs of participants were assigned the task of collaboratively creating a 3D artwork using clay within a one-hour time-frame. Communication regarding the project was restricted, allowing only visual cues. Each participant could contribute to the project for a total of 30 minutes, divided into two 15-minute intervals, alternating turns. In the observation section of this paper we will look at the results from one of the experiments.

Implementation

In the first workshop, the participants are referred to as Artist One and Artist Two. Artist One began by sculpting a complete 3D figure resembling an animal (Figure 4.1A) within 12 minutes, after which they switched roles. Artist Two opted to disassemble the artwork and introduce new elements to craft a fresh narrative(Figure 4.1B). This scenario tested the Theory of Mind when Artist Two introduced "clouds", which Artist One interpreted as the "sea", leading them to incorporate aquatic elements such as a life raft. Furthermore, Artist One perceived the "wings" as muscles, thus adding organs and a skeleton to the artwork(Figure 4.1C).

Subsequently, Artist Two chose to redefine the narrative and opted to reconstruct the artwork while retaining certain elements, such as the body parts of the animal figure(Figure 4.1)D. Both participants expressed satisfaction with the final outcome.

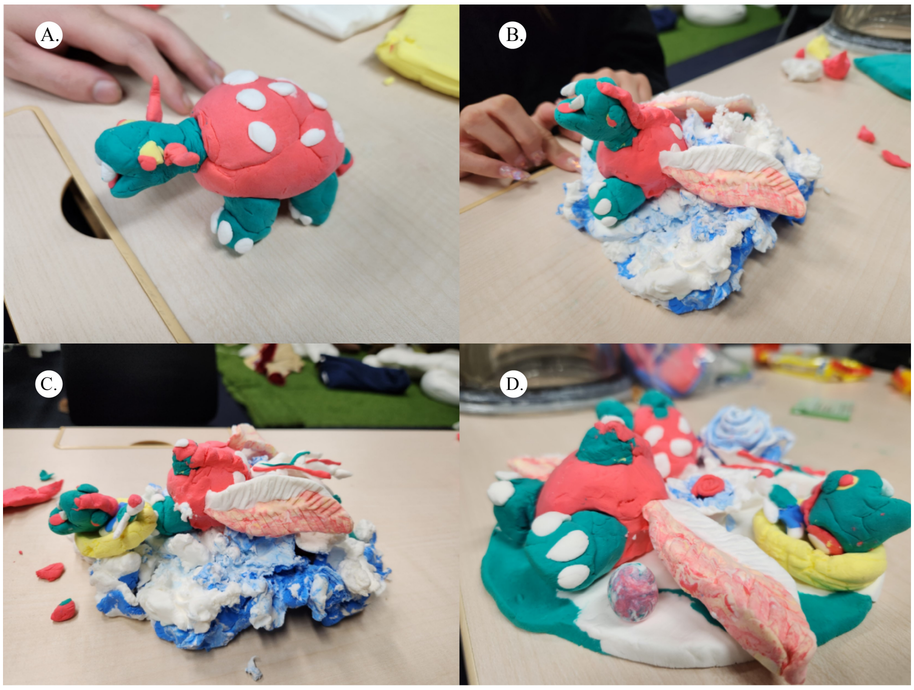


Figure 4.1 A.) Artist One Creation, B.) Artist Two Rework, C.) Artist One Final Rework, D.) Artist Two Final Rework

4.3. Human to Human vs Human to AI workshop

This workshop consisted of two distinct segments. In the first part, participants were paired with strangers and tasked with collaboratively crafting 3D art using play-doh. For the second segment, participants engaged in collaborative endeavors with AI instead of human partners, employing play-doh as their medium for the 3D creative process. This workshop comprised fourteen participants, aged between 23 and 35 years old. Among them, eight were females and six were males. The participants represented diverse backgrounds and experiences.

Implementation

In the initial segment, participants put into pairs and were instructed to randomly select one theme from a cup containing 16 predetermined options. Unlike the first workshop, there were no time constraints imposed on this session. However, participants were instructed not to deliberate on their approach to creating the artwork. Drawing from the methodology of the previous human collaborative workshop, participants were encouraged to rely on visual perception and intuition to accomplish their task.

4.4. Sandai Update 2

In the second iteration of Sandai, aimed at further enhancing collaborative outcomes, we integrated a new API through Touchdesigner called Stream Diffusion. [40] This new pipeline facilitates real-time interactive generation and by replacing our original API in Sandai with Stream Diffusion, we can generate artwork simultaneously between users and AI. While users create, the AI continuously updates its imagery, creating a feedback loop based on what it observes and the keywords users input as context. For the AI's output, still represented in 2D format, we extracted depth data from the images by tracing the generated artwork and deriving 3D forms. This data was then combined with the depth data captured from the Microsoft Kinect Azure, which recorded users' 3D play-doh creations. By merging this with the AI data, we produced a new, combined 3D form co-created by humans and AI. This artwork can then be 3D printed (Figure 4.2) or integrated into VR and XR spaces. The aim of this study is to encourage users to think of their artwork tangibly. Therefore, in the next workshop, we will ask users to consider potential applications for the final artwork.

Implementation

A preliminary study was conducted for the updated prototype as a preview before the main data collection study. Three participants tested the new prototype. Our goal with this prototype was to 3D print the outcome for users to interact with. It was observed that the prototype's performance varied depending on the angle

Figure 4.2 3D printed figure

at which the play-doh sculpture was presented. Capturing static 3D imagery from the AI creation proved challenging while the user moved and interacted with the play-doh.

However, the AI provided alternative outcomes for the users to consider during the creation process. One user noted that they forgot what their subject (a panda) looked like, and the AI helped visualize their intended creation. Additionally, this user discovered that by constantly changing the angles, the AI offered multiple varied alternatives for the chosen subject matter. This allowed the user to explore various perspectives and select a preferred output from the AI.

Another user mentioned that the AI added details they had not considered, resulting in a more detailed version of their play-doh creation, which they appreciated.

The main data collection for the second iteration of the Sandai Prototype occurred at the Collaboration Complex of Keio University, Hiyoshi campus, during the PLAY Exhibition on Sunday, May 26, 2024. A total of 10 participants engaged with the prototype. Building on methodologies from previous workshops, participants were free to create anything they desired using the provided play-doh and were asked to provide one or two keywords for the AI. Following their interaction with the prototype, they were invited to participate in a completely voluntary survey. Users were also encouraged to observe the screen while crafting their play-doh pieces.

4.5. Human to AI workshop 5 (User Interface)

After main data collection study, it was determined that the way users interacted with Sandai did not actively demonstrate co-creation and collaboration. While we succeeded in reducing the reliance on detailed prompting and using AI to create tangible art, the iterative aspect of co-creation was still lacking. Because Sandai updated almost instantaneously, some users were not able to focus on the real time update and did not feel they were truly collaborating with the AI. For this we introduced the SMR (Sender-Message-Receiver) Theory which outlines the process of communication in a simplified manner. As previously mentioned, referencing Boisot's I-Space Model [42], where knowledge transfer forms the basis of communication and is communicated through a boundaryless loop facilitated by feedback. In Sandai's latest prototype update, a feedback interface was created, compelling users to interact and communicate with Sandai beyond merely entering keywords. Users are encouraged to engage in communication and make decisions through Sandai by interacting with its user interface. Sandai offered a text-based communication style, simulating a discussion with the user as if conversing with a friend.

Implementation

For this workshop, we gathered 10 users from various disciplines. We asked users to follow the steps provided by Sandai, using a keyboard as input to communicate with the system. As in previous workshops, users created their artwork using play-doh, while Sandai analyzed the 3D forms provided. Once users reached a satisfactory stage with their pieces, they were able to input keywords and add additional words for Sandai to enhance their creations.

4.6. Results and Findings

4.6.1 Human to Human Workshop

The evaluation of this experimental workshop was held during one day and aimed to test theory of mind within pairs, employing clay as a medium for creating 3D artwork. The findings highlighted the necessity of some form of communication as a foundation for collaborative work. However, the study also demonstrated that visual cues can serve as the primary method of collaboration, with textual cues functioning effectively as a supplementary element.

Based on this insight, we began to implement it within our Sandai prototype. For the next experiment, we aimed to compare human-to-human collaboration with human-to-AI collaboration, utilizing textual cues as a secondary mode of communication.

4.6.2 Human to Human vs Human to AI workshop

Data collected from the initial phase of the workshop revealed several key insights. A significant majority, comprising eleven of the participants, encountered challenges in communication and decision-making regarding the initiation of the collaborative process. Conversely, a minority, constituting three of the participants, expressed enjoyment in their interaction with their partner. Three of the attendees opted to independently create their own artwork prior to engaging in collaborative efforts. All attendees expressed a desire to communicate with their partner verbally, indicating a preference for direct interaction. Furthermore, one of the participants harbored uncertainty regarding whether their artwork would differ had they collaborated with someone else, suggesting contemplation on the impact of collaborative dynamics. Additionally, a minority, encompassing one of the attendees, remained skeptical about the potential of AI to replace their human partners in the collaborative process.

Next we will focus on the specific human to human collaborators assigned the theme of "Nature". Their collaborative vision centered on crafting interconnected flowers (Figure 4.3A). Employing identifiable shapes and elements, they skillfully depicted their interpretation of flowers, creating a cohesive and visually engaging representation.

Subsequently, they were tasked with individually creating their own 3D artwork, this time collaborating with AI. Retaining the theme "Nature", the initial iteration with Sandai utilized "Nature" as a keyword to assist the AI in contextualizing the input captured through the DSLR camera (Figure 4.3B). However, the outcomes fell short of the user's expectations. Consequently, the keyword was modified to "Flower"(Figure 4.3C), resulting in more satisfactory results for the user. Seeking greater specificity, the user then opted for the keyword "Cosmos" (Figure 4.3D), aiming to convey to the AI a precise understanding of the particular type of flower they intended to create.

For the second user's initial AI generation, no keyword was utilized (Figure 4.3E), resulting in the AI's failure to grasp the intended context. Nonetheless, it successfully retained the colors employed by the user. Subsequently, the user employed the keyword "Nature" (Figure 4.3F), which led to the generation of results closely aligned with the user's envisioned outcome, while still adhering to

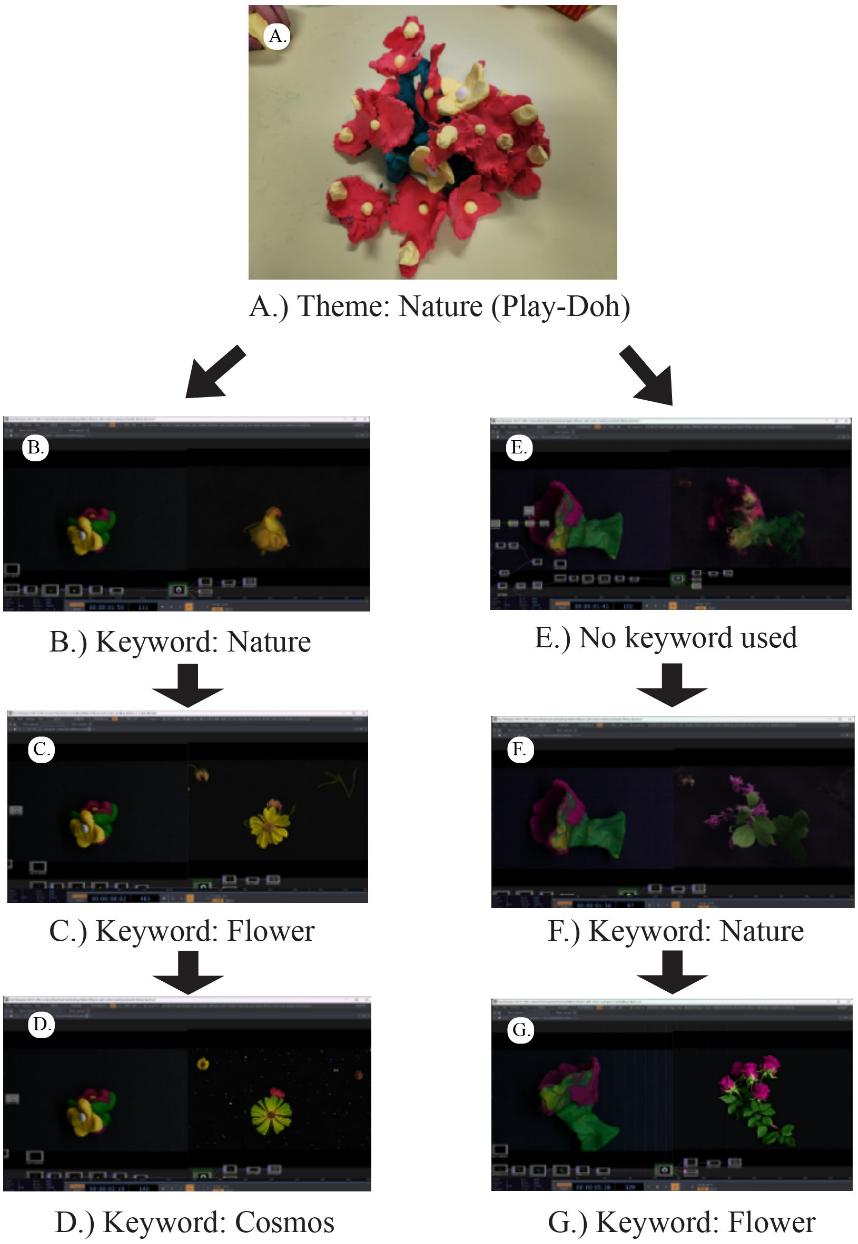


Figure 4.3 Human to Human vs Human to AI Workshop

the selected color scheme. Finally, the user opted to substitute the keyword with "Flower" (Figure 4.3G), which also yielded desirable results, maintaining thematic consistency and color coherence.

In a post-workshop interview following the second phase, both users concurred that the prototype facilitated collaboration. However, one participant expressed a preference for working with a human partner, while the other remained uncertain about their preference.

The overall data gathered from this workshop revealed that six of the participants expressed uncertainty regarding the level of collaboration facilitated by the prototype. Meanwhile, two perceived it as collaborative, and an additional two believed it was not.

Consequently, for the second iteration of Sandai's prototype, a greater emphasis was placed on fostering collaborative input from the AI's perspective, rather than relying solely on the interactive process of the user creating and then generating the AI's outcome. Despite the AI providing input, users still perceived themselves as having sole control over the outcomes, indicating a need for further enhancement in collaborative dynamics.

4.6.3 Human to AI workshop (real time AI updates)

In this workshop, data collected from the survey provided crucial insights into how people use AI in their daily lives. All participants were familiar with AI tools prior to engaging with the prototype. Specifically, two of the participants used AI daily, either for work or school. Another two used AI rarely, five used AI on a monthly basis, and one used AI weekly. Furthermore, eight of users believed that AI image generators are currently collaborative, while one were unsure.

In evaluating the updated Sandai Prototype, users were asked if they had made 3D art in any capacity. The responses were evenly split, with five answering yes and five answering no. Among those who answered no, six found creating 3D art to be difficult, while those who answered yes remained neutral about the difficulty of creating 3D art in relation to using 3D art programs.

Regarding familiarity with play-doh, six of users were familiar with the medium, while four were not. Most users found play-doh to be a user-friendly medium when used with the prototype. They also found the prototype effective in collaboratively generating 3D output related to their creations. Furthermore, most users agreed that the prototype helped visualize their vision in the final outcome. Overall, users found the AI's collaboration user-friendly and were satisfied with the prototype's collaborative features. However, it was established that Sandai's output merely provided a more detailed rendition of the user's artwork without offering any input or altering the user's original creation.

Additionally, we observed how often users decided to make changes to their original artwork or keywords. Since this prototype visualizes changes almost instantly, the need to modify keywords was significantly reduced. We observed that users changed the keywords at most three times before they were satisfied with their collaborative final outcome. Moreover, through the API, we can determine the influence of keywords on the final outcome from the AI's perspective. It was observed that with reduced weight, the keywords produced more sporadic outcomes.

4.6.4 Human to AI workshop (User Interface)

In this workshop, it was observed that one user created a panda with play-doh and wanted Sandai to add a hat, as shown in (Figure 4.4A). Sandai accurately placed the hat on the panda's head through its 3D visualization of the user's sculpture (Figure 4.4B). By adding the keyword "baseball", the user was able to change the hat to a baseball cap (Figure 4.4C). Additionally, without adding more keywords, the user added a body to their playdoh sculpture (Figure 4.4D). Sandai recognized this change and updated its 3D visualization to include a body, as seen in (Figure 4.4E).

Through this, the prototype demonstrates logic by placing the hat on the panda's head rather than randomly next to it. The AI understands that hats are typically worn on the head, and thus places the hat on the panda's head. Additionally, it preserves the integrity of the artist's play-doh sculpture by maintaining its shapes and placement via the camera input of the Microsoft Kinect Azure.

We then tested the prototype with various sporadic outcomes by reducing the emphasis on the provided keywords. We observed that the prototype exhibited autonomous thinking by delivering diverse and intriguing outcomes in terms of

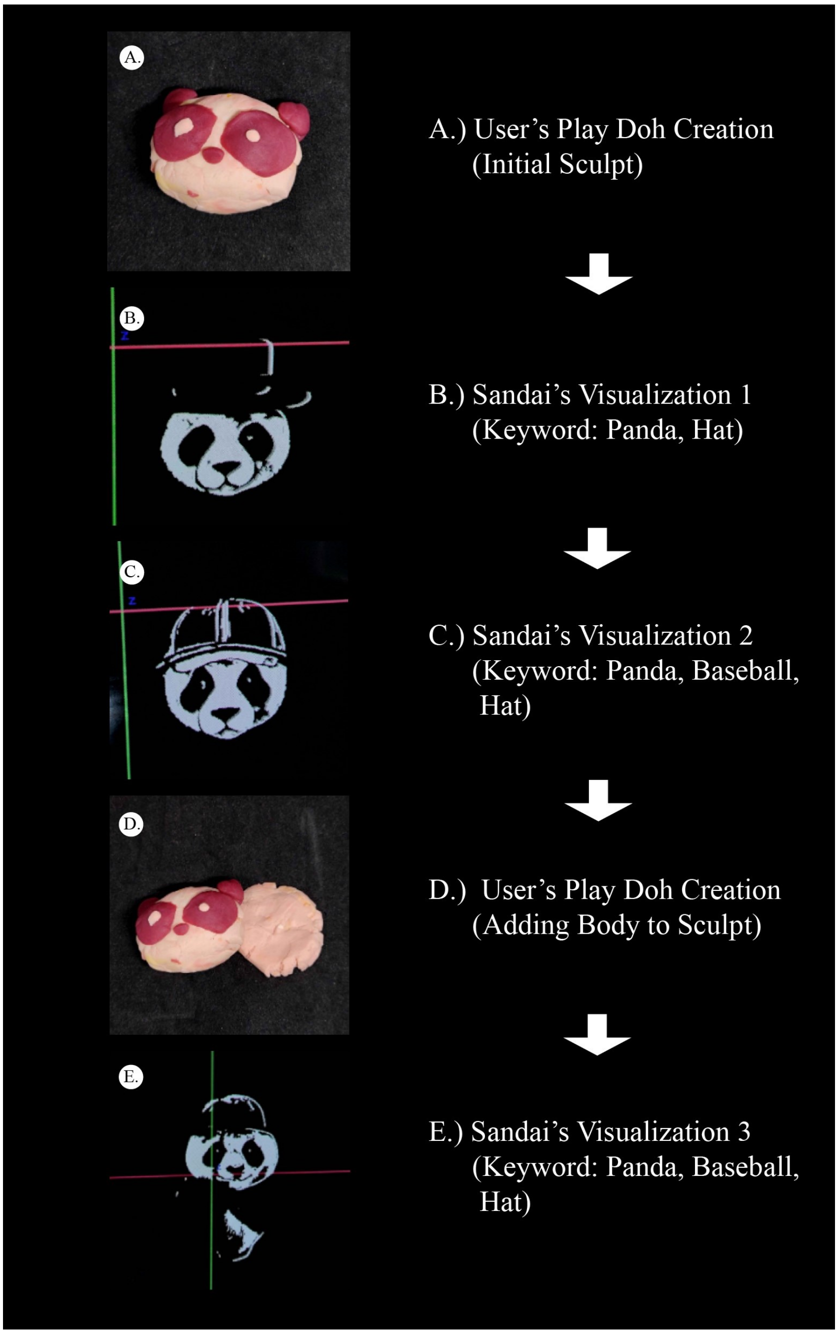


Figure 4.4 Experiment 1: User's Play-Doh creation and Sandai's visualizations

placement based on the provided keywords. In the next example, the user created a 3D rendition of a flower with play-doh (Figure 4.5A). For this experiment, the user entered the keywords "Flower, painting, canvas," which resulted in the outcome seen in (Figure 4.5B). Next, the user kept the same flower and position but changed the keywords to "Flower, pattern, dress"(Figure 4.5)C. Finally, the user changed the keywords to "Flower, hands"(Figure 4.5D).

4.6.5 Potential Applications

In the last workshop, we asked users to brainstorm potential use cases for Sandai's tangible art. Interestingly, users naturally associated the use of play-doh with tangibility and expressed a desire to have their creations 3D printed, enabling them to have physical versions of their designs.

During the discussions held after each user test, we explored how Sandai could be applied in real-world scenarios. One user suggested designing custom figures for Tabletop Role Playing Games (RPGs). With this, users could create detailed characters and monsters, with AI co-creating these designs by adding details or introducing randomized surprise elements that users might not have considered. For example, as shown in (Figure 3.4), where the AI generated humanoid legs for a bunny model. These unique game pieces could then be 3D printed and used to play the games.

Another user highlighted the potential for combining fun and learning for children. Teachers or parents could use the system to create educational toys. For instance, children could mold animals out of play-doh, and the AI could offer variants of their design or create entirely new imaginative animals based on keyword input from the children. Children can explore combining two or more types of animals to create a new, imaginative species. These models could then be 3D printed for interactive learning sessions.

A different user suggested using Sandai for rapid prototyping, replacing traditional resin molding processes. Designers could quickly prototype their ideas by creating initial models with play-doh, which the AI would then refine, making them ready for 3D printing. This approach would accelerate the design process and allow for quick iterations and tangible feedback.

Lastly, multiple users proposed using Sandai for personalized decoration or ac-

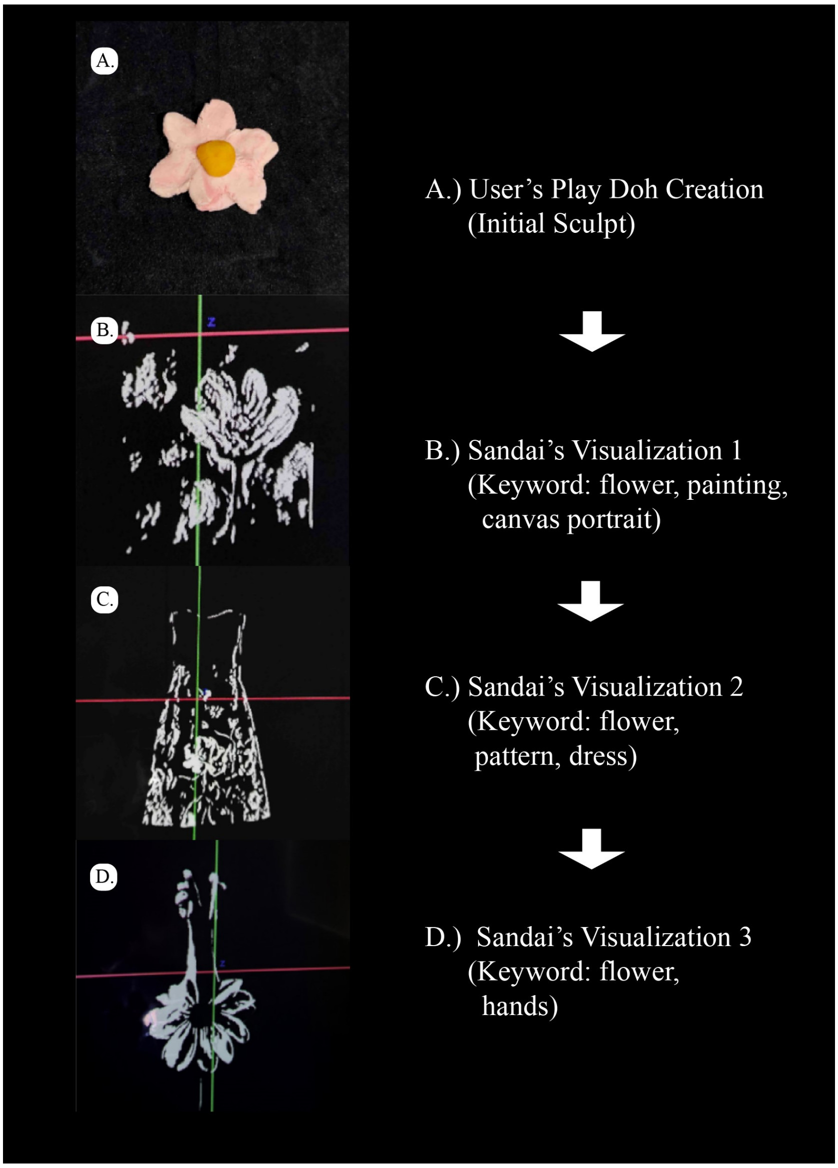
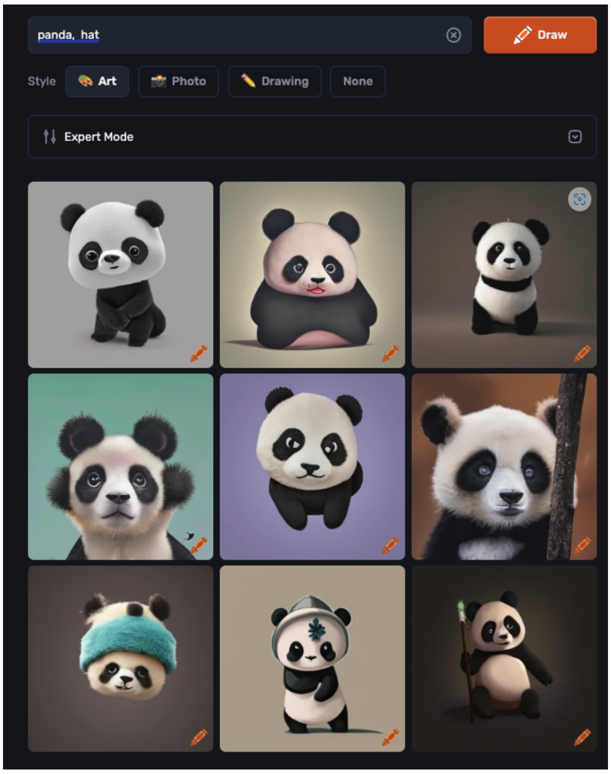


Figure 4.5 Experiment 2: User's Play-Doh creation and Sandai's visualizations

cessories co-created with AI. Users could model unique jewelry ideas or decorative items such as ornaments or figurines using play-doh. The AI would then refine these designs, incorporating user inputs through keywords or by applying its own logic and interpretation. This would enable the creation of unique, personalized items co-created by humans and AI.

4.6.6 AI Autonomy

These various examples demonstrate Sandai's potential for autonomy and creative input within 3D spaces, showcasing how AI visualizes users' creations and provides additional information to it by utilizing its own logic based on communication from the user. Moreover, by actively reducing the use of detailed prompting in addition with Sandai's vision input, users can actively visualize 3D forms with AI and can co-create unique 3D artwork. Compared to current AI systems, where users rely on highly descriptive text prompts to convey their messages, they often lose crucial keywords in translation, which is illustrated in (Figure 4.6), generated by the AI image generator Craiyon. For this image generation task, we used the same keywords, "Panda" and "Hat," which were also provided to Sandai referencing (Figure 4.4). As evidenced by the results, only one out of the nine generated images approximates the desired outcome. However, even this image presents issues and additionally cannot be accurately represented in a 3D space. This highlights the limitations of current AI systems in producing consistent and precise visual outputs based on simple keyword inputs. Therefore, Sandai exhibits a higher degree of autonomy in co-creation compared to current mainstream AI systems. It enables users to explore unexpected outcomes within a 3D space while incorporating and maintaining the user's original play-doh creation. By employing the alternative method of using keywords, less emphasis is placed on detailed prompting. This shift enables users to concentrate more on co-creating with AI and generating tangible, imaginative outcomes.



(Source: https://www.craiyon.com)

Figure 4.6 Craiyon AI Image generator

Chapter 5
Discussion and Future Work

This chapter examines the current limitations of the design system and highlights opportunities for future research, investigations, and innovative developments.

5.1. Discussion and Limitations

Performance Constraints

In the various user tests conducted, it was observed that the prototype requires significant processing power to operate effectively. This high demand for computational resources results in considerably slower output times or, in some cases, the program failing to function altogether, necessitating a restart. These performance issues pose a significant challenge to the usability and efficiency of the prototype.

Challenges in Human-AI Collaboration

The collaborative efforts between humans and AI also present distinct challenges. Although the AI possesses the capability for its own interpretation and creative input, a primary issue arises from the human partner's desire for minimal interference. In several workshops, users expressed a preference for the AI to replicate their play-doh art precisely. While the AI can achieve this replication, it restricts the AI's ability to introduce new elements or decisions into the artwork, leading to less co-creative outcomes. This limitation underscores the tension between user control and AI autonomy in collaborative artistic processes. As previously discussed in the related works, user control has been a persistent issue in the collaboration between artists and machines, as exemplified by AARON and Harold Cohen [18]. Throughout the various workshops conducted, we observed that Human/AI collaboration introduces its own set of complexities. The tension between user control and AI autonomy is evident in scenarios where users prefer minimal interference, such as the precise replication of play-doh art. This preference limits the AI's ability to contribute creatively, thereby reducing the potential for truly co-creative outcomes. Addressing this challenge requires a delicate balance, enabling the AI to introduce novel elements while respecting user intentions.

Spatial Awareness Through Camera Angle Adjustment

To enhance the 3D effect of the final outcomes, adjusting the camera angle is crucial. By dynamically changing the camera perspective, the prototype can better understand and interpret spatial relationships within the created artwork. This improvement in spatial awareness will facilitate more accurate and realistic representations, contributing to a richer co-creative experience.

Audio Feedback for User Guidance

Incorporating audio feedback can significantly aid users who may be uncertain about how to proceed or which direction to take during the co-creation process. By providing clear, verbal instructions, the system can guide users more effectively, ensuring a smoother and more intuitive interaction. This feature will help reduce confusion and enhance the overall user experience, allowing for a more seamless co-creative process.

Tangible Co-Created Artworks

For users to obtain tangible versions of their co-created artwork, the program must integrate a 3D printing machine to produce physical models. However, 3D printing technology presents its own set of limitations. The process is restricted by the size of the printable objects and is inherently time-consuming, requiring a minimum of two hours to produce a final co-created piece. This extended production time can be a significant drawback for users seeking immediate results. The integration of 3D printing technology for producing tangible co-created artworks presents both opportunities and limitations. While it offers a way for users to physically manifest their creations, the current constraints of 3D printing pose significant challenges.

Enhancing the efficiency and scalability of 3D printing processes will be crucial for meeting user expectations for immediate and tangible results.

5.2. Future Work

Based on the data received from the user tests, several areas for future work have been identified to enhance the co-creative AI system. One significant user request is the ability to converse with the AI through speech, enabling a more interactive and conversational interface between the user and the machine. This feature would enrich the co-creative experience by making interactions more natural and intuitive.

Incorporating audio feedback remains a high priority. Providing clear, verbal instructions will guide users through the co-creation process, mitigating confusion and facilitating a smoother and more intuitive user experience. This functionality is crucial for ensuring that users can navigate the creative process with confidence and ease, ultimately leading to more effective and enjoyable co-creation.

Additionally, addressing the performance constraints of the prototype and finding a balance between user control and AI autonomy are crucial for enhancing the co-creative experience. The performance constraints observed in the prototype, particularly the high demand for computational resources, underscore the need for more efficient processing capabilities. These limitations affect the usability and efficiency of the system, highlighting the importance of optimizing performance to ensure smoother operation and user satisfaction.

Lastly, It was noted that spatial awareness through dynamic camera angle adjustments could represent a critical advancement in achieving more accurate and realistic 3D representations within this prototype. This improvement would not only enhance the visual appeal of the final outcomes by providing artwork that resembles a more well-rounded 3D model but it also contributes to a deeper understanding of spatial relationships within the artwork.

5.3. Conclusion

Artificial Intelligence (AI) has become an integral part of many aspects of our daily lives, often without our explicit awareness. As society transitions from creating tangible media to focusing on art for VR and XR spaces, the potential of AI in generating tangible media within the arts has been largely overlooked. Each year, AI systems become more sophisticated, continuously learning and expanding their knowledge base as humans share more interactions online. As AI continually learns and evolves, it is crucial to preserve and nurture tangible creativity between AI and humans.

The question of whether AI could develop its own consciousness is intriguing. Although it may seem like a distant possibility and the necessary technology might not fully yet exist, AI is progressively exhibiting human-like behavior. Our familiarity with AI's potential transformation comes from popular media, which has long depicted AI as becoming more human-like and eventually taking over human jobs. This scenario is not as far-fetched as it once seemed. Despite AI being in its early stages, we have already witnessed some adverse effects from its implementation especially in job displacement [43] especially in the arts.

While AI in the arts has predominantly been developed as a tool that collaborates based on detailed natural language, the focus now is creating AI systems that foster innovation by reducing human interference and input. We aim to allow AI to contribute its own creative ideas, helping us explore possibilities we may not have yet considered. This thesis outlines a pathway for AI to become more co-creative by incorporating its own reasoning into artistic creation, while considering human perspectives. It emphasizes a human-first ideology, ensuring that humans remain the leaders and AI follows their guidance.

Throughout this study, we have observed that users enjoy the interaction with using AI when it generates results beyond their initial expectations. Moreover, users generally prefer not to engage heavily in detailed prompting. Instead, they derived more satisfaction from the process of art creation itself, rather than acquiring the additional skill of detailed prompting. There is a pressing need for improved interaction and the generation of unexpected outcomes in current AI systems. Users frequently express reluctance to rely on existing AI technologies for completing tasks, particularly within creative settings. This highlights the necessity for advancements in AI that better support user engagement and creativity.

Additionally, participants regarded the prototype as an enhanced collaborative partner, capable of producing novel outcomes that they had not previously considered. This was particularly beneficial for users with limited experience in 3D modeling, as the prototype effectively enhanced their skills. It achieved this by adding intricate details or redirecting the artwork in a new direction while still preserving the user's original concept. Furthermore, users preferred not having to render imagery manually in 3D software due to its steep learning curve. Instead, they found it easier to shape their sculptures by using their hands and leveraging the AI's autonomy to achieve the final outcome. This approach proved to be more accessible and user-friendly compared to the complexities of learning traditional 3D software. However, users with very specific expectations were less inclined to utilize the prototype, indicating that AI technology should prioritize unexpected and autonomous outcomes over elaborate user inputs from detailed prompting. If users, particularly artists, already anticipate the outcome, then AI holds limited usefulness in the creative fields, as artists possess the inherent skills to create the artwork themselves.

Sandai emphasizes the act of creation and collaboration with AI technology, showcasing its value in creative spaces. The Sandai prototype demonstrates significant potential as a creative co-partner by enhancing both the process and enjoyment of artistic expression through collaboration. Additionally, it shifts the focus of artwork creation back to a tangible and hands-on execution.

In summary, the journey towards developing effective and efficient co-creative AI systems is multifaceted, requiring several key advancements to fully realize their potential.

First, performance optimization is crucial. AI systems must process and generate outputs swiftly and accurately to keep pace with human creativity. This requires refining algorithms, enhancing computational efficiency, and ensuring the system can handle complex tasks without significant delays. Additionally, the AI must be capable of producing novel ideas that are on par with human creativity.

Second, balanced human-AI collaboration is essential. AI systems should complement and enhance human creativity, not overshadow or undermine it. This balance can be achieved by designing interfaces and interaction paradigms that allow users to seamlessly integrate AI-generated inputs and modifications into their work, while still maintaining control over the creative process.

Third, improved spatial awareness is necessary for AI systems, especially in fields such as 3D modeling and tangible media. AI needs to understand and manipulate spatial relationships within creative works effectively. This includes recognizing spatial constraints, accurately interpreting user inputs, and generating outputs that align with the user's spatial intentions.

Fourth, user guidance mechanisms are vital. Users, particularly those with less experience or technical expertise, need intuitive and accessible guidance to effectively utilize AI systems. This can involve interactive real-time feedback to help users navigate the capabilities of the AI and make the most of its features, which has been exhibited in the last iteration of the Sandai prototype.

Fifth, tangible output capabilities are important. AI systems should facilitate not only digital but also physical creation. This includes the ability to translate digital designs into physical prototypes or artworks, bridging the gap between virtual creativity and tangible reality.

Addressing these areas will pave the way for more robust and enriching co-creative experiences. By focusing on performance optimization, balanced collaboration, spatial awareness, user guidance, and tangible outputs, we can foster a more productive partnership between humans and AI in creative spaces. This, in turn, will enable new forms of artistic expression and innovation, enhancing the creative process and expanding what can be achieved through co-creation with AI.

References

[1] Rory Jeffs and Gemma Blackwood. Whose real? encountering new frontiers in westworld. 2016.

[2] Itay Harlap and Ariel Avissar. The author is dead, long live the author: Negotiating textual authority in westworld, the oa and 13 reasons why. *Participations*, 18(1):89–110, 2021.

[3] A. Wasif, Muhammad Hamid, and Asad Abbas. Human-ai collaboration in creative arts and content generation. 11 2023.

[4] Ben Mann, N Ryder, M Subbiah, J Kaplan, P Dhariwal, A Neelakantan, P Shyam, G Sastry, A Askell, S Agarwal, et al. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*, 2020.

[5] Laria Reynolds and Kyle McDonell. Prompt programming for large language models: Beyond the few-shot paradigm. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–7, 2021.

[6] Vivian Liu and Lydia B Chilton. Design guidelines for prompt engineering text-to-image generative models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA, 2022. Association for Computing Machinery. URL: `https://doi.org/10.1145/3491102.3501825`, `doi:10.1145/3491102.3501825`.

[7] Jules White, Quchen Fu, Sam Hays, Michael Sandborn, Carlos Olea, Henry Gilbert, Ashraf Elnashar, Jesse Spencer-Smith, and Douglas C. Schmidt. A prompt pattern catalog to enhance prompt engineering with chatgpt, 2023. `arXiv:2302.11382`.

[8] Triyono Bramantyo. Digital art and the future of traditional arts. *Journal of Music Scholarship*, 42(1):96–110, 2021.

[9] Christopher Andrews. An embodied approach to ai art collaboration. In *Proceedings of the 2019 Conference on Creativity and Cognition*, C&C '19, page 156–162, New York, NY, USA, 2019. Association for Computing Machinery. URL: `https://doi.org/10.1145/3325480.3325506`, `doi:10.1145/3325480.3325506`.

[10] Tantri Ardianti, Budi Purwoko, and Umi Anugerah Izzati. The effects of playing playdough on the ability to recognize geometric shapes in kindergarten. *Studies in Philosophy of Science and Education*, 4(1):17–21, 2023.

[11] Chengbin Du, Yanxi Li, Zhongwei Qiu, and Chang Xu. Stable diffusion is unstable. *Advances in Neural Information Processing Systems*, 36, 2024.

[12] Ian Apperly. *Mindreaders: the cognitive basis of" theory of mind"*. Psychology Press, 2010.

[13] Lindsey Byom and Bilge Mutlu. Theory of mind: mechanisms, methods, and new directions. *Frontiers in Human Neuroscience*, 7, 2013. URL: `https://www.frontiersin.org/articles/10.3389/fnhum.2013.00413`, `doi:10.3389/fnhum.2013.00413`.

[14] Chris Frith and Uta Frith. Theory of mind. *Current biology*, 15(17):R644–R645, 2005.

[15] Eleonora Lima. Ai art and public literacy: the miseducation of ai-da the robot. *AI and Ethics*, pages 1–14, 2024.

[16] Miloš Todorović. Ai and heritage: A discussion on rethinking heritage in a digital world. *Uluslararası K¨ult¨urel ve Sosyal Araştırmalar Dergisi*, 10(1):1–11.

[17] Rebecca Chamberlain, Caitlin Mullin, Bram Scheerlinck, and Johan Wagemans. Putting the art in artificial: Aesthetic responses to computer-generated art. *Psychology of Aesthetics, Creativity, and the Arts*, 12(2):177, 2018.

[18] Paul Cohen. Harold cohen and aaron. *Ai Magazine*, 37(4):63–66, 2016.

[19] Louise Sundararajan. Harold cohen and aaron: Collaborations in the last six years (2010–2016) of a creative life. *Leonardo*, 54(4):412–417, 2021.

[20] Aikaterini Petridou. Ai-da: The world's first humanoid ai artist. will ai become art's next medium? 2020.

[21] Sveinn Steinar Benediktsson. *Human creativity in the augmented age: Interrogating the works of Maurice Conti and Sougwen Chung*. PhD thesis, 2019.

[22] Campbell R Eubanks. What does it mean to be a human in the age of artificial intelligence? 2024.

[23] Kathryn Wojtkiewicz. How do you solve a problem like dall-e 2? *The Journal of Aesthetics and Art Criticism*, 81(4):454–467, 2023.

[24] Ar Mohesh Radhakrishnan. Is midjourney-ai the new anti-hero of architectural imagery & creativity? *GSJ*, 11(1):94–104, 2023.

[25] James Betker, Gabriel Goh, Li Jing, Tim Brooks, Jianfeng Wang, Linjie Li, Long Ouyang, Juntang Zhuang, Joyce Lee, Yufei Guo, et al. Improving image generation with better captions. *Computer Science. https://cdn. openai. com/papers/dall-e-3. pdf*, 2(3):8, 2023.

[26] Charles Anderson, Thomas Taylor, and Christopher Moore. Advancing ai image generation: Unveiling the mechanisms and innovations of stable diffusion technology.

[27] Yang Song, Jascha Narain Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *ArXiv*, abs/2011.13456, 2020. URL: `https://api.semanticscholar.org/CorpusID:227209335`.

[28] Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas M¨uller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, et al. Scaling rectified flow transformers for high-resolution image synthesis. *arXiv preprint arXiv:2403.03206*, 2024.

[29] Katja Thoring, Sebastian Huettemann, and Roland M Mueller. The augmented designer: a research agenda for generative ai-enabled design. *Proceedings of the Design Society*, 3:3345–3354, 2023.

[30] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Gaugan: semantic image synthesis with spatially adaptive normalization. In *ACM SIGGRAPH 2019 Real-Time Live!*, pages 1–1. 2019.

[31] Seung Won Lee, Tae Hee Jo, Semin Jin, Jiin Choi, Kyungwon Yun, Sergio Bromberg, Seonghoon Ban, and Kyung Hoon Hyun. The impact of sketch-guided vs. prompt-guided 3d generative ais on the design exploration process. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA, 2024. Association for Computing Machinery. URL: `https://doi.org/10.1145/3613904.3642218`, `doi:10.1145/3613904.3642218`.

[32] Vivian Liu, Jo Vermeulen, George Fitzmaurice, and Justin Matejka. 3dall-e: Integrating text-to-image ai in 3d design workflows. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, DIS '23, page 1955–1977, New York, NY, USA, 2023. Association for Computing Machinery. URL: `https://doi.org/10.1145/3563657.3596098`, `doi:10.1145/3563657.3596098`.

[33] Maliheh Ghajargar and Jeffrey Bardzell. Making ai understandable by making it tangible: Exploring the design space with ten concept cards. In *Proceedings of the 34th Australian Conference on Human-Computer Interaction*, pages 74–80, 2022.

[34] Maddalena Torricelli, Mauro Martino, Andrea Baronchelli, and Luca Maria Aiello. The role of interface design on prompt-mediated creativity in generative ai. In *Proceedings of the 16th ACM Web Science Conference*, pages 235–240, 2024.

[35] Dakuo Wang, Elizabeth Churchill, Pattie Maes, Xiangmin Fan, Ben Shneiderman, Yuanchun Shi, and Qianying Wang. From human-human collaboration to human-ai collaboration: Designing ai systems that can work together with people. In *Extended Abstracts of the 2020 CHI Conference*

on Human Factors in Computing Systems, CHI EA '20, page 1–6, New York, NY, USA, 2020. Association for Computing Machinery. URL: `https://doi.org/10.1145/3334480.3381069`, `doi:10.1145/3334480.3381069`.

[36] Maurice Jakesch, Megan French, Xiao Ma, Jeffrey T Hancock, and Mor Naaman. Ai-mediated communication: How the perception that profile text was written by ai affects trustworthiness. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2019.

[37] Francis Schmidt, Jim Jagger, Jim McCambell, and Craig Slagel. 3d animation: difficult or impossible to teach and learn? In *ACM SIGGRAPH 2004 Panels*, page 3. 2004.

[38] Thomas Suselo, Burkhard Claus W¨unsche, and Andrew Luxton-Reilly. Why are 3d transformations in computer graphics difficult? an analysis of a decade of exam questions. In *Proceedings of the 24th Australasian Computing Education Conference*, pages 181–190, 2022.

[39] Jun Gao, Tianchang Shen, Zian Wang, Wenzheng Chen, Kangxue Yin, Daiqing Li, Or Litany, Zan Gojcic, and Sanja Fidler. Get3d: A generative model of high quality 3d textured shapes learned from images. *Advances In Neural Information Processing Systems*, 35:31841–31854, 2022.

[40] Akio Kodaira, Chenfeng Xu, Toshiki Hazama, Takanori Yoshimoto, Kohei Ohno, Shogo Mitsuhori, Soichi Sugano, Hanying Cho, Zhijian Liu, and Kurt Keutzer. Streamdiffusion: A pipeline-level solution for real-time interactive generation. *arXiv preprint arXiv:2312.12491*, 2023.

[41] Sabah Al-Fedaghi. A conceptual foundation for the shannon-weaver model of communication. *International journal of soft computing*, 7(1):12–19, 2012.

[42] ELHMAIDI Khadija and TIJANI Omar. Literature review on knowledge transfer models: From shannon-weaver (1949) to kostova (1999). *African Scientific Journal*, 3(21):548–548, 2023.

[43] Kaledio Egon, J Russell, and R Julia. Ai in art and creativity: Exploring the boundaries of human-machine collaboration. *OSF Preprints*, 20, 2023.

Appendices

A. Touchdesigner script

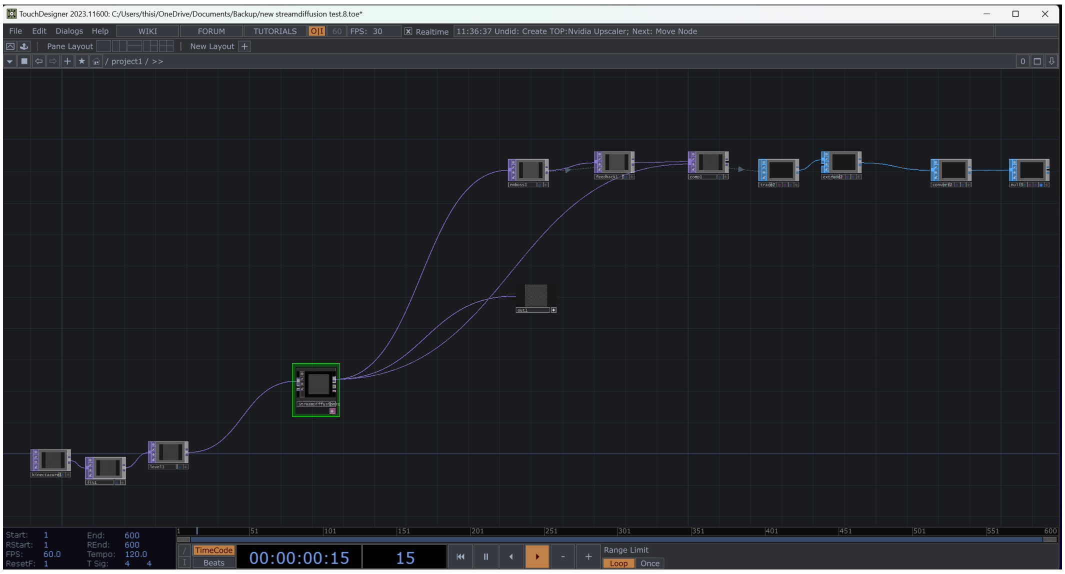


Figure A.1 Touchdesigner Development

B. Questionnaires

Questionnaires for preliminary study, Sandai Prototype and Sandai Prototype Update.

https://drive.google.com/drive/folders/1i-VSzH6RouVyTDdK$_{k}ZqIWvSAo4y2pYY?usp = drive_{l}ink$