Trash to Treasure:
Using text-to-image models to inform the design of physical artefacts

Anonymous submission

Abstract

Text-to-image generative models have recently exploded in popularity and accessibility. Yet so far, use of these models in creative tasks that bridge the 2D digital world and the creation of physical artefacts has been understudied. We conduct a pilot study to investigate if and how text-to-image models can be used to assist in upstream tasks within the creative process, such as ideation and visualization, prior to a sculpture-making activity. Thirty participants selected sculpture-making materials and generated three images using the Stable Diffusion text-to-image generator, each with text prompts of their choice, with the aim of informing and then creating a physical sculpture. The majority of participants (23/30) reported that the generated images informed their sculptures, and 28/30 reported interest in using text-to-image models to help them in a creative task in the future. We identify several prompt engineering strategies and find that a participant's prompting strategy relates to their stage in the creative process. We discuss how our findings can inform support for users at different stages of the design process and for using text-to-image models for physical artefact design.

Background and Motivations

Text-to-image deep learning models are generative AI techniques that synthesize images from text inputs. Implementations of this technology such as Midjourney, DALLE-2, and Stable Diffusion (Rombach et al. 2022) have exploded in popularity in the last year. The company Stability AI has released the models, weights, and code for Stable Diffusion, allowing it to be used publicly, and Midjourney's Discord interface reached over 1 million users. Increased access to this technology has accelerated the development of open-source tools and resources for creativity and design. As a result, we have seen artists using AI-generated imagery as part of their visual design process. A few examples include a *Cosmopolitan* magazine cover designed in tandem with DALLE-2, (Liu 2022), and an artist using Midjourney to win an art competition (Harwell 2022). This democratization has raised questions regarding how such models can be used in a wide variety of tasks, such as idea visualization (Epstein et al. 2020; Epstein, Schroeder, and Newman 2022; Rafner et al. 2021; Smith and Colton 2022).

These tools offer new possibilities for navigating the creative process. Design research suggests a need for flexible pathways for creative computational assistance *early* on in the design process (Mothersill and Bove 2017). Hwang (2022) also found that many AI-empowered co-creative tools focus on support for the later stages of design, when participants exploit, refine, or implement ideas, but are less involved in the earliest stages of co-creation, when participants are in the "explore" phase.

In contrast to using the generative model for directly creating digital media content, a growing body of work explores the use of generative AI for upstream tasks in the creative process, such as ideation and visualization, that have a range of downstream tangible outcomes. These can include tattoos, (TattoosAI 2022) fashion, (Levine 2022) community values (Epstein, Schroeder, and Newman 2022) and visualizations of the future (Rafner et al. 2021). A key challenge for using generative AI in upstream creative tasks, such as ideation and idea visualization, is the fact that the affordances of physical materials could differ from those of generated imagery, and users may be frustrated bridging that distance to bring their idea into reality. Work by Dang et al. (2022) shows that lack of support in the trial and error process of prompt engineering (tactics for refining prompts to synthesize desired outputs) can be frustrating for users, motivating further investigation into users' prompting strategies. Connecting a user's needs, based on their prompting strategy, to their design stage could help provide individualized support for a user's design goals.

Study Design

We conduct a pilot study to examine the impact of introducing AI-generated images into the early stages of a design task in a physical medium. The experiment was advertised at a community event at a research university as an activity to "turn trash into treasure" by making artistic sculptures out of discarded materials. Once participants had joined the activity, they were were instructed by the facilitator to choose 3-5 pieces of material from a box with diverse objects such as test tubes, pipe cleaners, foam pieces, and wires.

Once participants had chosen their sculpture materials, we explained that they would give us three text prompts which would generate three images, and we asked them to consider how these images could inform their sculpture design. We then asked participants for their first prompting phrase, which the facilitator used as input for Stable Diffusion. The facilitator then asked the following while the image was be-

ing generated: 1) "Why did you choose that prompt?" and 2) "What are you expecting to see?" Once the facilitator had written down the responses, the generated image was revealed to the participant. The facilitator repeated this process of prompting, reflection, and image reveal up to another two times. After the visualization stage, the participant was given 3 minutes to build a sculpture using their existing materials, as well as adhesives like tape and hot glue. Once the sculpture had been completed, the facilitator asked 1) "Was your sculpture informed by your generated images?" and 2) "Would you use a text-to-image model like Stable Diffusion for a creative task again?"

We set out to measure each participant's level of conceptual exploration through their prompting journey. The concept of semantic distance is a popular one in the evaluation of the creative processes (Kenett 2019) and is one we believe extends to search through semantic space. To operationalize this, we transformed user prompts into sentence embeddings,[1] then measured the cosine distance between a user's first, second, and third prompts, taking the average distance over these to characterize a user's level of conceptual exploration over semantic distance. We also qualitatively analyzed the notes from each participant's interaction as well as the sculptures they produced.

Results

Generated images informed final designs

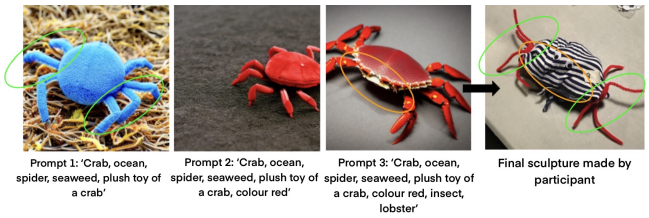
Of the 30 participants, 27 produced at least two prompt and image pairs. Of those 27, 24 produced three images (the remaining three produced one image). 23/30 participants self-reported that their sculptures were informed by the images they saw, and 28/30 participants reported that they would use text-to-image models again for a creative task.

Figure 1 shows an example of strong image influence on sculpture design for someone who did not have a sculpture idea going into the process. The rightmost image is a photo of the sculpture they built, which is striking in its visual similarity to the third generated image in particular. Some congruences, including color of materials which were not changeable, are highlighted in colored boxes on the image.

Figure 1: Visual elements inform sculpture design of a building

Figure 2 shows the journey of someone who knew from the onset they wanted to make a sculpture of a crab. The three images are outputs from the prompts: "crab, ocean, spider, seaweed, plush toy of a crab" →"crab, ocean, spider, seaweed, plush toy of a crab, colour red" →"crab, ocean, spider, seaweed, plush toy of a crab, colour red, insect, lobster." Despite already knowing what they wanted to make, the participant said the images gave reminders of a crab's features so they could create it in real life. The participant tweaked their second prompt to include "colour red" because their physical materials were also red. Common entities like crabs are easy to generate thanks to their high visual determinacy, but some users with more abstract ideas reported frustration visualizing what they had in mind, even if they knew what they wanted to see.

Figure 2: Visual elements inform sculpture design of a crab



Of the participants who reported that the images were not informative to their sculpture (n = 7), three explained that this was because images did not sufficiently relate to the materials given and they could not build what they saw, some expressing frustration translating what they wanted to see in an image into language that would produce the image they expected. Only 36.2% of images contained elements that the participants expected to see, indicating many visual elements were unexpected. 40% of participants (12/30) directly referenced their chosen materials, like "box" or "sponge." [2] This suggests a need to better support users in translating between physical materials and prompt wording when ideating for physical artefacts.

The reasons participants reported that images were informative were diverse. Some found that the lateral concepts introduced by the images gave them new conceptual ideas, perhaps indicating they were in an early design stage where exploration was particularly useful. Others honed existing ideas by getting implementation inspiration from the images, perhaps indicating a later design stage focused on execution. These findings confirm those in Epstein, Schroeder, and Newman (2022), which showed that AI-generated images are helpful in visualizing ideas for two main reasons: they give lateral insight as well as implementation ideas.

Two participants who did not find the images helpful built their sculpture immediately after their first image, and explained that the strength of their pre-visualization ideas meant that they believed no images would further inform

[1] We used the sentence transformer *'all-MiniLM-L6-v2'* from HuggingFace due to its performance capturing semantic meaning.

[2] We do not find statistically significant differences in the distribution of average cosine distances of user prompts and whether or not the prompt contained mentions of physical materials ($t = 1.01$, $p = 0.32$), material qualities ($t = -1.25$, $p = 0.22$), or colors ($t = 0.12$, $p = 0.91$).

them. This might suggest that seeing images is less helpful for users with an already concrete vision, but we do not find strong evidence to support that claim. Of the individuals with a design idea before, 3/8 found seeing the images helpful and 5 did not. Of the individuals with no design idea before, 15/23 found seeing the images helpful and 8 did not, a 44% difference. This provides suggestive evidence ($t = -1.365$, $p = 0.183$) that participants who did not have an idea may find the generated images more helpful.

Distinct prompting styles emerge

Participants' wording choices for prompts ranged in their degree of conceptual exploration. We observed and named some patterns we saw in users' prompting choices. The least amount of conceptual exploration that we observed was a pattern we call the "refiner" style. In these instances, the participant started with a prompt and made minor edits to it, and it was the most common pattern we saw (n = 15). For example: **Prompt 1:** 'geometric ominous dystopian creature metallic'. **Prompt 2:** 'geometric ominous dystopian creature *neon heart beating*'. **Prompt 3:** 'geometric ominous creature neon heart beating *rainbow*'.

The next pattern observed was the "rephraser" prompting style (n = 5). Here, conceptual subject matter remained the same, but changes to the wording or order of the prompt changed substantially. Example: **Prompt 1:** 'idyllic cyber vista with a bottle robot'. **Prompt 2:** 'dasani water bottle robot in front of the Windows XP wallpaper'. **Prompt 3:** 'robot that looks like a water bottle in front of a grassy hill'.

Some participants did even more conceptual exploration. The "explorer" prompting style (n = 2) we describe consisted of three conceptually unrelated prompts. Example: **Prompt 1:** 'my printed map required too much plastic'. **Prompt 2:** 'I'm exhausted but I'm still having fun'. **Prompt 3:** 'my hovercraft is full of eels'. We observe these described styles as an exploration gradient where "explorers" have the most semantic distance traveled, "rephrasers" have a main idea but still explore around it, and "refiners" focus on exploitation of a single string of words.

We observed that the remaining 7 participants did not fit in to these prompting categories. For example, the user who built the sculpture in Figure 1 prompted as follows: **Prompt 1:** "Hamburg is the most beautiful city in the world", **Prompt 2:** "sun and shiny water, colourful buildings made of trash and ropes", **Prompt 3:** "green red and white buildings in front of a shiny river with a white fence on the side". The first prompt appears to be exploratory, whereas the second two describe a more concrete scene in two similar but distinct ways, reminiscent of behavior we see in those using a "rephraser" style. The final three participants to account for only gave a single prompt, so their prompting style is harder to characterize.

Due to the fact that not all participants fit neatly into the prompting styles we described, we use a computational measure of semantic distance to objectively characterize user journeys in the rest of our analysis. We calculate the average conceptual distance a participant traveled through their prompting journey by taking the cosine distance between the text embeddings of first and second prompts, then the second and third prompts, then averaging the values. For participants with two prompts, we treat the single distance between the two prompts as the average distance. For participants with a single prompt, the distance was 0.

Figure 3: Histogram of participants' conceptual exploration across their prompting session, measured by average cosine distance. The prompting styles appear to occur at different average cosine distances.

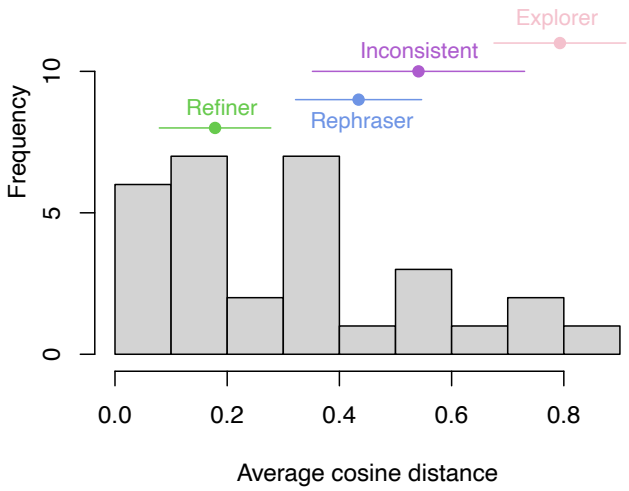


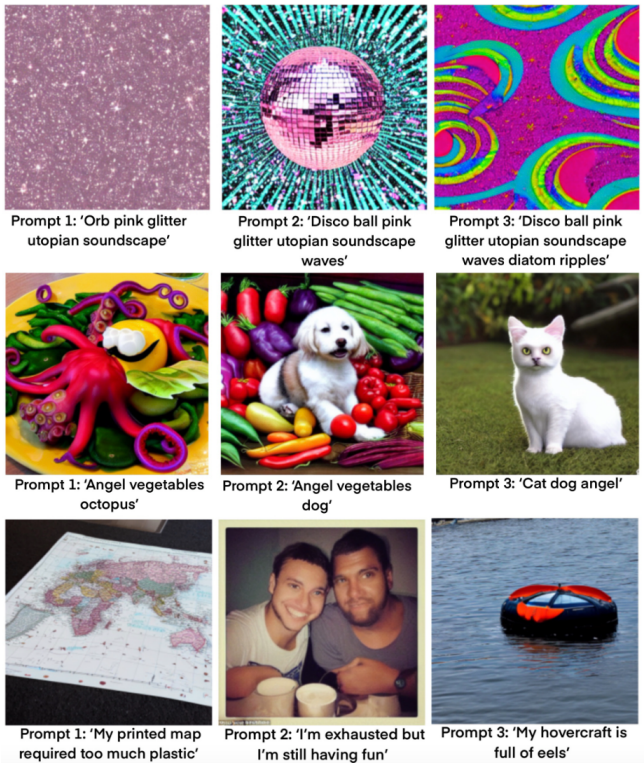
Figure 3 shows a histogram of average cosine distances across the prompting journeys of our 27 participants who gave at least 2 prompts, along with the average and standard deviation of the subgroup we qualitatively grouped into "refiner," "rephraser," "inconsistent," and "explorer." The participants we describe as "refiners" have the lowest average cosine distance of 0.178, while the participants we described as "explorers" had the highest average cosine distance of 0.793. The participants we describe as "rephraser" and "inconsistent" fall in the middle range.

We visualize three participant journeys through semantic space in Figure 5. We use TSNE to reduce dimensionality of the shared embedding space to two dimensions, and draw arrows from points representing a participant's first to second to third prompts. The path between green dots in Figure 5 show more extreme conceptual exploration using an "explorer" prompting style, blue dots show a less exploratory "rephraser" style, and purple dots show the "refiner" prompting style. The prompts and corresponding images for these three participants can be seen in Figure 4.

Prompting Style differs by Design Stage

We next investigate the relationship between the design stage a user reported and the average semantic distance they traveled. Because we gave participants the design task of creating a sculpture, participants skipped what Hwang (2022) calls the "Q&A" stage of defining the task. By constraining the category of materials and type of creative output (sculpture), we further define the design problem's "artefact type" and "media type," per the Mothersill and Bove

Figure 4: Prompts and image pairs for the different prompting styles illustrated in Figure 3. From top to bottom: 'refiner', 'rephraser' and 'explorer'



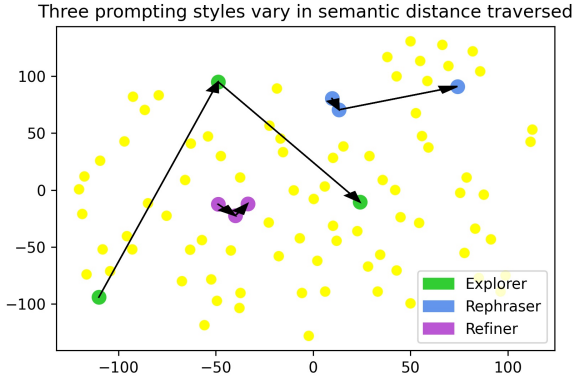
(2017) conceptualization of the design process. Those who came into the visualization stage of the exercise with no idea what to make started in the "wandering stage," wherein creatives explore possible strategies and incubate ideas. Some who started visualization with a sculpture idea skipped this stage and were already in the "hands-on stage" of developing solutions, or were already moving to the "camera-ready stage," which focuses on selecting and implementing ideas, to use Hwang's taxonomy. Using these ideas, we compare the average semantic distance traveled for the group of participants who had ideas before visualization to those who did not.

We found that the amount of conceptual exploration a participant did was lower for participants who said they had a sculpture idea at the visualization stage than those who did not ($t = -2.94$, $p = 0.006$). We also found that participants who stopped early, generating fewer than 3 images, employed less conceptual exploration in their images than those who did not ($t = 4.31$, $p < 0.001$). All participants who stopped after one ($n = 3$) or two images ($n = 4$) had an idea of what to make before starting their visualization.

Future Work and Limitations

It remains unclear how the use of generated imagery may impact tools and materials used for downstream physical creation. This could be studied in a similar experiment that

Figure 5: TSNE plot of embedded representations of three user prompting sessions where users range in degree of conceptual exploration while prompting



allows participants to choose their materials after visualization, instead of before. We also gave participants a maximum of three tries with Stable Diffusion in this experiment, but more complex behavior may emerge for users across more interactions with the technology. Anecdotally, we also saw that some users were distracted by the technology itself, so future work could directly observe the effect of user familiarity with the tool on the user journey.

The embedding representations we chose for prompts could also use the CLIP or BERT tokenizers that were used in Stable Diffusion for greater conceptual similarity to the tokenization process of text to image models, and a sensitivity analysis of individual tokens on these embeddings could be done to deepen conclusions.

Conclusion

Most participants found the images informative to their final output. The average semantic distance a participant traveled in their prompting journey was lower if they had a sculpture idea to start the activity than if they did not. This shows that prompting decisions relate to the design stage a participant is in. Participants who started with ideas used image generation as an opportunity to "exploit" or refine ideas, traveling less average semantic distance than those who were unsure what to build and used the images to explore. To better support creators, text-to-image tools could identify a user's semantic distance traveled over a prompting session to suggest hints that are useful to their current design stage.

The vast majority of participants in our study were interested in using a generative creative tool like this, and they had many ideas for ways they would do so—- from designing a new kind of battery to generating artwork for their bedrooms. With thoughtful prompting guidance, we hope that a wide of variety of creators can use text-to-image models to inspire visionary ideas and bring those ideas from our screens into reality.

References

Dang, H.; Mecke, L.; Lehmann, F.; Goller, S.; and Buschek, D. 2022. How to Prompt? Opportunities and Challenges of Zero-and Few-Shot Learning for Human-AI Interaction in Creative Applications of Generative Models. *arXiv preprint arXiv:2209.01390*.

Epstein, Z.; Levine, S.; Rand, D. G.; and Rahwan, I. 2020. Who Gets Credit for AI-Generated Art? *iScience*, 23(9): 101515.

Epstein, Z.; Schroeder, H.; and Newman, D. 2022. When happy accidents spark creativity: Bringing collaborative speculation to life with generative AI. *arXiv preprint arXiv:2206.00533*.

Harwell, D. 2022. He used AI to win a fine-arts competition. Was it cheating? *The Washington Post*.

Hwang, A. H.-C. 2022. Too Late to Be Creative? AI-Empowered Tools in Creative Processes. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI EA '22. New York, NY, USA: Association for Computing Machinery. ISBN 9781450391566.

Kenett, Y. N. 2019. What can quantitative measures of semantic distance tell us about creativity? *Current Opinion in Behavioral Sciences*, 27: 11–16. Creativity.

Levine, G. 2022. Futuristic Fashion Designs Generated by DALL-E's AI. https://80.lv/articles/futuristic-fashion-designs-generated-by-dall-e-s-ai/.

Liu, G. 2022. The World's Smartest Artificial Intelligence Just Made Its First Magazine Cover. *Cosmopolitan*.

Mothersill, P.; and Bove, V. M. 2017. Humans, Machines and the Design Process. Exploring the Role of Computation in the Early Phases of Creation. *The Design Journal*, 20(sup1): S3899–S3913.

Rafner, J.; Langsford, S.; Hjorth, A.; Gajdacz, M.; Philipsen, L.; Risi, S.; Simon, J.; and Sherson, J. 2021. Utopian or Dystopian?: Using a ML-Assisted Image Generation Game to Empower the General Public to Envision the Future. In *Creativity and Cognition*, Camp;C '21. New York, NY, USA: Association for Computing Machinery. ISBN 9781450383769.

Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; and Ommer, B. 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10684–10695.

Smith, A.; and Colton, S. 2022. The@ artbhot text-to-image twitter bot. In *Proceedings of the International Conference on Computational Creativity*.

TattoosAI. 2022. AI-powered tattoo artist. https://www.tattoosai.com/.

.