Maximizing Effective Usage of AI based Image Generation through Prompt Engineering Analysis

SAINATH GANESH, LAVANYA RAMKUMAR, ELAINA WITTMER, SRISTI INGLESHWAR, SAM YUAN, and DEEPAK SUBRAMANIAN, University of Illinois Urbana-Champaign, USA

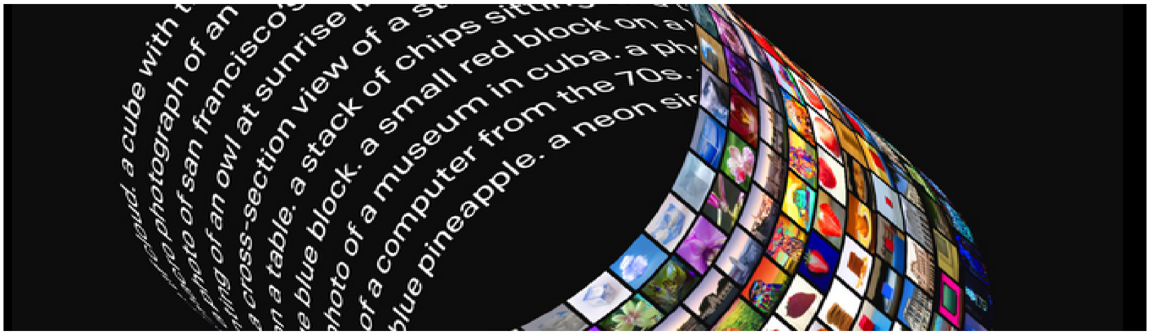

Fig. 1. Representation of the generation of images from text, taken from the DALL-E page on OpenAI.com.

1 ABSTRACT

Prompt engineering has become a popular subject of research as the popularity of LLMs such as ChatGPT grow, and their use becomes more widespread. This is, at least in part, because LLMs tend to operate as black boxes and "high-quality" prompts must be arrived at through trial-and-error. We investigate what consists of a high-quality prompt in the realm of image generation by indirectly observing users in this trial-and-error process to determine whether users are able to utilize AI to meet a specific image generation goal, how they approach creating a prompt to do so, and whether the user feels a sense of ownership over the image created. Results of our study show that users who were able to create highly-successful prompts did so by using relatively short prompts and starting with a simple idea then adding more description with each iteration. We also find that users who had generated more images and invested more time into their prompts, feel a sense of ownership over the image to a higher degree than those who spent lesser time. From our empirical analysis, we have created certain design guidelines for future user interfaces to consider to maximize the outcomes of using AI based Image generators.

Additional Key Words and Phrases: HCI, Artificial Intelligence, AI Image Generator, Open AI, Dall E, Prompt Engineering

2 INTRODUCTION

AI image generators have the capacity to produce high-quality images in a matter of seconds through the combination of elements from images that already exist all over the web. Currently, the power of these models is primarily accessed through text prompts; a user describes the image they would like to produce in plain text and the model produces an image output using the words provided in the description. On the surface, this task may seem similar to crafting a search engine query, a task that many people can perform proficiently. However, crafting a high-quality prompt for an LLM can be much more difficult as minor changes to the input can greatly affect the generated result and it's currently unclear where the limits of an LLM's understanding lie.

Authors' address: Sainath Ganesh, sg75@illinois.edu; Lavanya Ramkumar, lr15@illinois.edu; Elaina Wittmer, enw3@illinois.edu; Sristi Ingleshwar, sristii2@illinois.edu; Sam Yuan, jintaoy3@illinois.edu; Deepak Subramanian, deepaks3@illinois.edu, University of Illinois Urbana-Champaign, USA.

This work investigates how users approach crafting a prompt for an AI image generator by providing the user with a task that requires both prompt engineering and image evaluation. This experiment is guided by the following research questions:

This work is guided by the following research questions:

- **RQ1**: Are people able to generate images using AI that satisfy their creative goals?
- **RQ2**: How do people approach crafting a prompt for image generation?
- **RQ3**: Do users feel a sense of ownership over the AI-generated images?

The answers to these questions will help inform whether users are able to effectively utilize AI image generation for general purposes and what changes might be made to user interfaces to better help the user craft prompts they will be satisfied with. A negative result for RQ1 will show that either AI image generators still have limitations that make them inaccessible to the average user, the average user requires assistance from the UI to guide them in utilizing the AI effectively, or both. RQ2 will inform which approaches don't work when it comes to image generation, and which approaches do work. This will inform future design choices that can help guide the user towards making choices that lead to more satisfying results, rather than relying on the current "sandbox" approach. RQ3 will help inform whether users think of the AI as a tool or a collaborator. If the user thinks of the AI as a collaborative partner, they may expect back and forth communication and some level of compromise, as previous research suggests [13].

3 RELATED WORK

Previous work in the HCI-AI intersection has examined topics such as text generation [18], code-to-natural language translation [15, 16], song-writing [3], and even dance [8]. In the text and image domains specifically, there has been heavy emphasis on prompt engineering, or the idea that prompts fed to an LLM can be "engineered" to increase the chances the model will produce a desirable output. LLMs are not yet well-equipped to handle long or complex tasks, making crafting high-quality prompts a challenge for many end users. [18]

Prompts follow a few different forms. Zero shot prompts are most similar to using a search engine – the user provides the AI with a description of what they would like it to do and the model produces a result. Few-shot prompts, on the other hand, give the AI a pattern and ask it to learn from this pattern and apply it to a new scenario. One example of this is to give the model an English sentence and a French translation before giving it a different English sentence. In response to this, the model understands that it should provide a French translation of the second sentence. Both of these strategies can be applied to relatively simple patterns and requests. To address AI's struggles to adequately address complex or very long tasks, recent research has developed methods that chain simple zero shot prompts together to solve a much larger or complex problem [12, 18].

Recent work has focused on the relationship between AI and end users as AI becomes more well-known but perhaps not well-understood [2, 4, 7, 9–11, 14, 19]. One problem is that AI technologies are often black boxes, the inner workings of which are unknown to even ML experts [2, 19]. Recent work has focused on the idea of "AI literacy," meaning that a user just needs to know how to *use* the model, not necessary the intricacies of how it works [7]. This is becoming more and more important as AI models such as ChatGPT grow in popularity and become more ubiquitous. As novice users become acquainted with these systems, transparency is not enough. For example, the average Google user does not have to understand the PageRank algorithm to be able to effectively create a Google search query to obtain information. However, unlike search engine results, AI generations are more opaque and their inner workings even more mysterious, making user interfaces highly important for guiding the user towards crafting more successful prompts to give to the AI.

Previous work has examined the effect of different prompts for the task of image generation; Liu & Chilton [6] investigate the effectiveness of prompts using what we will refer to as a "keyword" style, meaning that prompts are given in sentence fragments such as "love abstract art." In Liu & Chilton's work specifically, they investigate different permutations of MEDIUM, STYLE, SUBJECT to determine which combinations of keywords lead to better results. The results of this paper indicate that prompts that focus on keywords and style have the greatest impact on the quality of the image, where quality is determined as a function of accuracy to a particular artistic style. In this way, the work is an evaluation of the capabilities of AI as well as specific features of prompts that lead to more effective prompts.

In contrast, we are leaving the task of image generation as an open sandbox with which the user can explore the limitations of image generation AI and prompting strategies on their own. In this work, we focus more on the user's approach and understanding more than the AI's. Whether the AI matches a specific style exactly is inconsequential unless the user has a highly-specific request. We make no assumptions about the artistic goals of the user. As such, we're interested in i) is the user able to produce a result they're happy with and ii) if so, how did they do so; we do not aim to evaluate the artistic quality of their results.

We are also interested in examining the sense of ownership people feel about the images they generated. With AI-generated art specifically, there is a tension in ownership between the user, the model, and the artists who produced the art which the model is trained on. This issue has led to recent lawsuits on behalf of the artists who unwittingly contributed to these datasets [1, 17]. Much emphasis is placed on the tension between the artists and the companies behind these AI models, but little attention has been paid to whether the user feels a sense of ownership towards these images as well. As such, we explore this question by asking the user to reflect on their level of contribution towards the generated image.

4 METHODOLOGY

We narrowed down our focus of this project to understanding how we can generate the most satisfactory image through an AI Image generator (DALL-E) by engineering a prompt which is provided as the input to the Generative Model. To this effect, we built an easy to use web platform that housed this Generative AI Image Model along with a system to collect feedback and automatically aggregate them across multiple image generations performed by a user. Building this web application provided us the flexibility to easily track user progress as well as provide interactive features such as examples to help out users in their generations. This also allowed us to create a timeline view to showcase their image generations and their corresponding prompt which gave the user quick access to glance through their work while providing feedback.

The users were given the following scenario "Imagine you are teaching an intro AI course. Think of what kind of visual representation of AI you would like to put on the first slide of your presentation. You want students to feel inspired or excited about the course and the future of AI, so you want an image that will capture this feeling". We asked them to keep trying different inputs by engineering the prompt and were instructed that they could end the generation process if the following three factors were in agreement:

- The vision of the image in their head
- The scenario that was provided to them
- The actual image generated by the AI

After this, each user was asked to fill out feedback on the satisfaction of their generations and the perceived ownership, i.e., how much they believed they contributed to the generation vs the AI. The images and prompts the user generated during the course of the experiment were displayed to the side of the survey so the participant could reflect on these as they answered the survey, as shown in Figure 2. We collected quantitative feedback using a Likert scale of 1 to 7 as well as subjective feedback to gain further insights. We also collected any other feedback the user wanted to provide other than satisfaction and ownership. Screenshots of the entire platform cannot be displayed here due to space concerns, but the page can be visited here: https://crowd-design.web.app/.

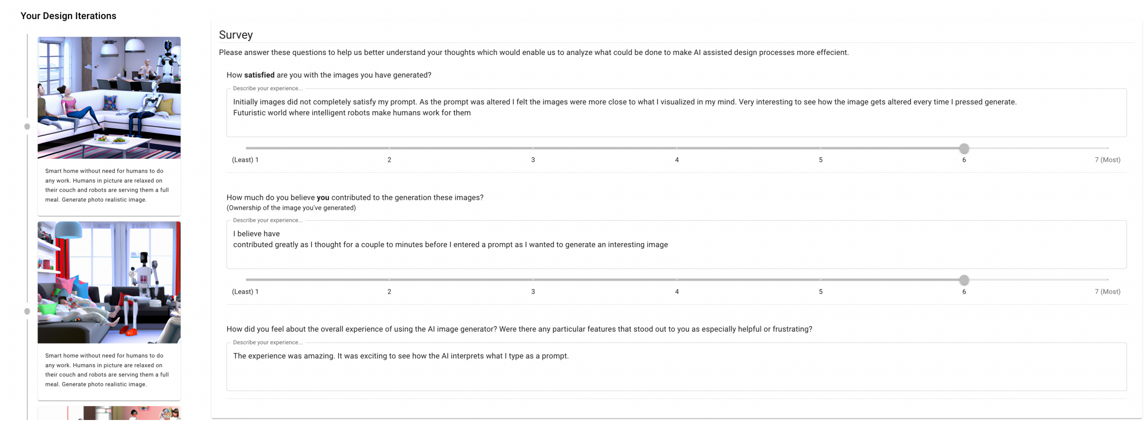

Fig. 2. The timeline view for a participant. The images and prompts are displayed to the left, which the user can scroll through. The feedback survey is displayed to the right.

Whilst the feedback provided information on the thought process of the user, we also collected other metrics in the background such as time spent on each generation, number of generations produced and number of words, keywords used to correlate with their feedback and derive insights. The team qualitatively coded the survey data to identify common and divergent themes between users. We also qualitatively coded the prompts themselves to observe the types of changes users made between iterations, such as revising the prompt, adding more information, or re-running the same prompt entirely. We also noted features of interest as the user developed their prompts from beginning to end, including use of abstract vs. concrete elements, the number of desired elements, and whether the prompt was constructed using a string of keywords or full sentences.

The experiment was conducted physically with 25 participants from the authors' social networks and their responses were collected over the course of two weeks. Although the population sampled for this study is of moderate size, we made sure to invite people from various backgrounds to ensure diversity and therefore more generalizable findings.

5 RESULTS

5.1 Quantitative Results

Taking a look at the central tendencies of the results (Figure 4a), the mean satisfaction score was at 4.89 with a standard deviation of 1.48 with the median value at 5. Similarly, ownership score was also at mean value of 4.36 with a standard deviation of 1.38 with the median being 4. So most people were moderately satisfied and held a sense of ownership over their generations. It was also interesting to note, no participant provided a score of 7 to ownership, i.e., no one believed they contributed completely toward the generation of the image.

Fig. 3. Top design generations from users who were highly satisfied with their creations

The average number of designs generated by a person was 6.60 with a median of 6. The average time taken for a user to generate a single image was 2.55 minutes and the average time to complete the whole experiment was 18.24 minutes. The mean number of keywords used was 11.24 with the median at 10.

5.1.1 Correlations. The heatmap in Figure 4 b shows the correlation between the various metrics we used to quantitatively analyze the results of the experiment. We see a decent correlation of 0.29 between ownership and satisfaction and a very good correlation between ownership and time spent. This highlights the fact that people perceived their contribution to be higher when they invested time into designing the prompts for the generations.

5.1.2 Ownership. Figure 5a clearly highlights and showcases the correlation between ownership and satisfaction among the participants. Participants who were satisfied with their generations also tended to have higher perceived ownership over their generations. This is also the case when participants spent more time on their generations (Figure 5b); as result they believe they have contributed more towards the generation of the image.

We see a similar situation when we take a look at the number of designs a user generated (Figure 5c). Ownership tends to follow an upward trajectory directly proportional the number of designs they generated. As it takes more time to produce more designs, this is to be expected. From these three graphs we can infer that ownership becomes a function of outcome and the amount of effort (time spent, number of designs) they put in. In other words, users believed they contributed a lot more when they invested more into their generations and got an image that they liked.

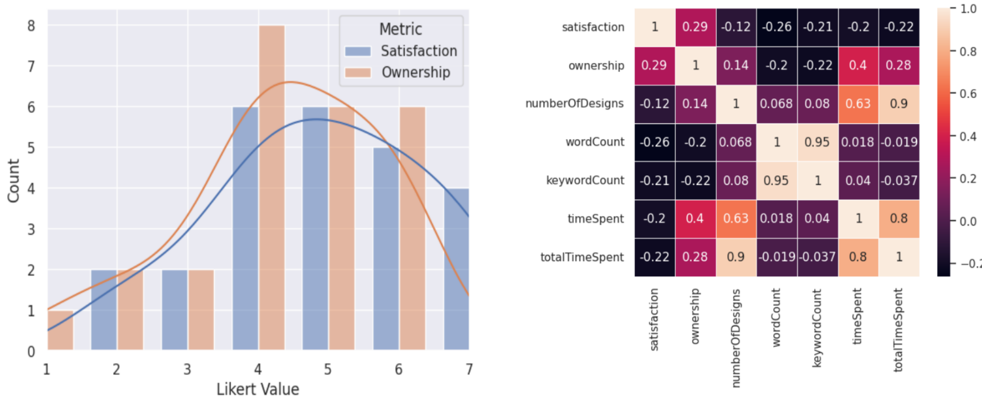

Fig. 4. Figure a (left): Bar chart of satisfaction and ownership counts with trend lines superimposed. Both metrics show normal distribution though both trend slightly right, indicating higher satisfaction and ownership scores. Figure b (right): A heatmap of variables observed including reported satisfaction, reported ownership, the number of designs a person created, the average word count for a prompt, the average keyword count, the average time spent on each prompt, and the total time spent on the task.

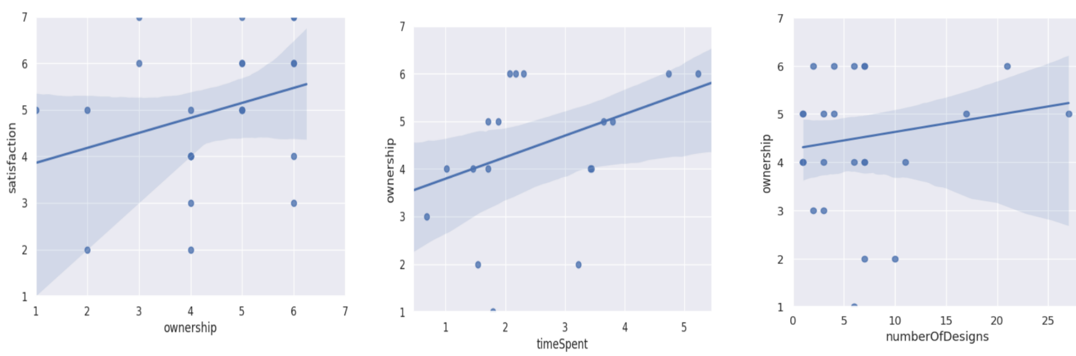

Fig. 5. Figure a (left) shows the relationship between satisfaction and ownership. Figure b (middle) shows the correlation between amount of time spent on each iteration and sense of ownership. Figure c (right) shows the correlation between the number of designs created per person and the sense of ownership they reported.

5.1.3 Satisfaction. The satisfaction feedback we got from users was not as straightforward as ownership and showed more nuanced aspects of using an AI Image Generator.

We can see in Figure 6a, when the number of designs increases, we see the satisfaction scores reach a peak at around 4-6 designs and then drop. To explain this phenomenon, we investigated the number of keywords used (Figure 6b) for the generation. The AI Image generator uses a tokenizer that takes in the keywords (i.e., nouns, adjectives, etc.) and tries to understand the context and generate an image from its understanding of the prompt. Looking at these keywords, we see a similar peak at around 8-10 keywords followed by a steady decline as the number of keywords increase. We believe too many keywords causes an information overload on the AI which causes the image to be a hodgepodge of various elements from the prompt, causing the image to be unsatisfactory.

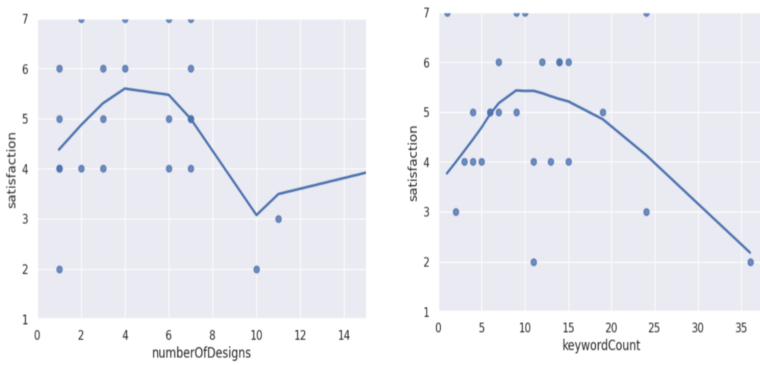

Fig. 6. Figure a (left) shows the correlation between the number of designs a person made and their level of satisfaction. Figure b (right) shows the average number of keywords per prompt and the correlation with satisfaction.

We observed that as the number of generations increases, people tend to keep appending scenarios and design guidelines to the prompt. This causes an increase in the number of keywords used in the prompt, explaining why there is a decline in the satisfaction scores reported by the users. We can infer that there is a sweet spot at around 8 - 12 keywords where the satisfaction reported by the users reaches a maxima. This can be used as an empirical guideline to optimize for while using Dall-E.

5.2 Prompt Analysis

We group the participants based on their reported satisfaction scores, 2-3 being unsatisfied, 4-5 being somewhat satisfied, and 6-7 being satisfied. Walking through each group, we can observe approaches that worked for the user, approaches which require improvement, and some which should be avoided.

5.2.1 Unsatisfied. The unsatisfied group (P1-P3) gave prompts which were either hyper-specific, had vague vision, or were focused on a particular weak point of the AI generators. For example, P1's final prompt was the following:

> P1: "Create an image that shows the capabilities of AI in solving complex problems. I want to see a futuristic world where AI guides humans through augmented intelligence, vast data, and super precise calculations. The image should have large displays, augmented reality, and show visualizations. I want the image to be an isometric top-down view of a large, futuristic office where human experts use the data and feedback to solve problems. I want the image to evoke a feeling of awe and hope."

This participant reported being disappointed because the AI did not depict the humans and machines interacting, even though this element is unclear in the prompt. The prompt is also rather lengthy and includes many references to abstract objects, such as augmented intelligence, precise calculations, and visualizations. Their vision is highly specific and the AI was not able to produce a result which met these expectations. On the other hand, P2 gave meandering prompts with no clear vision, adding elements to the prompt with each iteration without revising or deleting previous elements, which is evident in their final prompt:

> P2: "Images and examples of AI being used in the workplace. Images of people using AI in the workplace. Workplace, future, artificial intelligence. Artificial intelligence. Workplace AI. Exciting and colorful images of AI. Robots, AI, working. Images of robots."

P2 is focused on a vague idea of robots in a future workplace, but the structure of the prompt is confusing. This user reported they found the process of constructing a prompt confusing and could have used guidance on both structuring a prompt as well as assistance in creating a mental image in the first place. They also reported that the images were sometimes scary and/or full of nonsense words, which is a limitation of generative image models in general. P3 also alluded to similar limitations of these models stating "I don't think it understands emotions." P3 was disappointed the model could not generate an image that captured the feeling of "awe and hope," which are vague concepts.

5.2.2 Somewhat Satisfied. P4-P11 reported being somewhat satisfied, rating the process 4-5 on the Likert scale. These participants reported that the AI got it "almost correct," but was missing certain details described such as "facial features and body features of humans [which] were smeared and abstract" (P8). Similarly, P11 stated:

> "[The model] failed to include everything that is listed in the description (tigers). Also would appreciate it if more details were added to the image (Especially regarding the main subject - architecture)."

The aforementioned tigers were background elements and not the central focus of the image. For example, one of P11's prompts was:

> "Show ancient Indian architecture if it were made by the fictional character Howard Roark. Add as many details as possible. Should be able to see sunset, birds, people, two tigers, three elephants."

One of the limitations of the model is that unequal attention is paid to each element of the prompt. In this particular instance, the image generated contained architecture, a sunset, and two elephants while the rest of the elements were absent.

5.2.3 Highly Satisfied. For this section we look at P16-P18's prompts in detail as these participants reported being the most satisfied. Both P16 and P17 started with the highly generic prompts "head" and "virtual reality gear," respectively. From there, both participants added very few additional elements. P16 described a robot head that is laughing and smiling at jokes while P17 described a gold "andro humanoid robot" wearing virtual reality headphones. It seems that these participants explored the capabilities of the AI through adding elements, rather than removing them.

P18 admitted to using ChatGPT to create a starter prompt, which we ignore as it is unrelated to their second prompt, which was "Create a poster for the movie "ex machina" in the style of cyberpunk, focusing on the AI aspect of the movie." This prompt is simple and contains a specific reference to a movie, a style to follow, and a focus on AI. The only addition this user made to their prompt was "Make it very positive, and do not add any text to the picture." Unlike P3, P18 was satisfied with the model's interpretation of a positive emotion, and the request to remove text was a common trend in the data.

5.2.4 Overall Satisfaction. Based on the feedback provided, it seems that the AI-generated images were inconsistently meeting the expectations of the users. Some users were amazed by the images produced, while others found them scary or lacking in detail. However, as users provided more specific input and adjusted their prompts, they reported that the images became more accurate and aligned with their expectations. Overall, users had mixed experiences with some feeling that the images were amazing and specific to their requests, while others found them unrealistic or lacking in certain aspects.

5.2.5 Feedback. The feedback on the AI image generator is quite varied, with users expressing both positive and negative experiences. Some users were excited by the generator's ability to interpret prompts and generate unique images, finding it fascinating to observe how the AI weighed the importance of different words in the prompt to produce an output. They appreciated the AI's ability to capture designs very well and turn their descriptive points into something unique and visually appealing.

However, other users found the generator frustrating and difficult to use. They found it challenging to translate the images in their minds into prompts that the AI could understand and produce reasonable results for. The AI struggled with certain details, such as fingers and text, and often had trouble understanding how to create images that evoke human emotions or interesting settings. Some users felt that the AI's interpretation of prompts could be overly literal, leading to a lack of creativity and nuance in the generated images. They also found the AI to be slow in generating results, which could be frustrating when trying to experiment with different prompts.

Several users suggested improvements for the AI image generator, such as providing more description examples or keywords to help guide users in coming up with their prompts, and refining the AI's ability to understand and capture ideas rather than just designs. They believed that the AI had a long way to go in creating images based on vague or abstract prompts and that it struggled with generating more specific details for the images requested. Despite these limitations, many users expressed enthusiasm for the potential of generative AI to make our lives easier and open up new possibilities for creative expression. They were amazed by how far AI has come in recent years and were excited to see what developments would come next in this rapidly evolving field.

6 DISCUSSION

Our findings mirror the existing literature by emphasizing the need for AI literacy, specifically around the strengths and weaknesses of AI image generators. P2, for example, struggled with understanding how to use the model and what a prompt should look like. Prior research focuses on people such as P2 who lack understanding about what AI models are and asks how to best explain AI to new users. This can be achieved through moving away from a "sandbox" approach and giving the user options, rather than a plain text box. For new users especially, more structure which guides the user towards a specific prompt form such as Liu & Chilton's [6] MEDIUM, SUBJECT, STYLE prompt could be effective. Alternatively, recent work has suggested that prompts could consist of only one subject, entered via text, while the user can customize this subject through auto-suggested buttons relevant to the subject. [5]. Our findings suggest this approach would be helpful to novice users as this would reduce the need to create prompts from scratch. This approach could also be beneficial to all users if the number of buttons to use was fixed such that the user was unable to select more than 12 keywords.

Some of our participants, however, seemed to understand AI image generators on a basic level but were disappointed that the model could not address all of their expectations. This could be addressed by providing more information about the strengths and limitations about AI so that the user can adjust expectations to fit with what the model can achieve. Previous work has suggested that even users with some level of AI experience could benefit from guardrails which give the user a prompt template instead of an open sandbox [18], which in an image generation context could be used to restrict the number of subjects, styles, and colors used.

6.1 Suggested Improvements for Future Human-AI Interactions

Based on the results presented above, we offer some suggested improvements for Human-AI interactions in image generation which can be resolved through refinements to the user interface. We use data gathered from the most satisfied pool of participants to inform design decisions that could benefit all users, from those with little to no understanding of AI to those who have some experience with AI.

- **Specify the theme.** The survey results indicate that AI image generators frequently fail to create the image that the user has envisioned, especially highly generalized models such as Dall-E. It would be beneficial to allow users to specify the theme to use for its generation in order to vastly narrow down the search space it should use for its generations.
- **Suggest edits based on prompt length.** The quantitative results indicate that the most satisfactory prompts contain approximately 20-30 words in total and 8-12 keywords. We suggest the user receive feedback when their prompt reaches an excessive length, informing them that AI results are less effective with longer prompts. This could be done with an interface similar to programs such as Grammarly, which provides feedback as you type and offers suggestions.
- **Start small and add detail.** The qualitative results indicate that starting with a small idea and building on it leads to more satisfactory results. The user could be guided into starting small by being asked to select a visual focal point at the very beginning and then being guided towards adding more description, style words, colors, etc. This would help users who struggle with knowing where to begin as well as users who have too many ideas and should focus on one.
- **More communication about prompt options.** In forming a prompt, some users picked up on the fact that running the same prompt multiple times would lead to a different result, even if this was not explicitly communicated. It would be a benefit to more novice AI users to have buttons for running the prompt again or starting completely over, as these options were not obvious to some users.
- **More information about the limitations of AI.** The current user interface for ChatGPT includes example prompts, what the model is capable of, and limitations that the model faces. The user interface for DALL-E, however, simply states to "Start with a detailed description.". The limitations of DALL-E are addressed in following sections and we believe all of these limitations should be presented to the user so that they can adjust their expectations accordingly.
- **Guardrails against AI limitations.** Generative models struggle with depicting abstract concepts and feelings and are better at depicting tangible concepts. We propose guardrails on concepts such as style and mood so that the user can simply select from a set of options the model has been shown to be able to replicate. This would reduce the chance for disappointment by setting expectations.
- **Edit the image generated.** Once the AI has produced an image that meets their satisfaction, some users may wish to further refine it. It would be beneficial for the AI to enable users to modify the image by interpreting their new prompts and incorporating additional details such as altering the color or adding additional shadows.

6.2 Sense of Ownership

Our findings suggest that the more effort users put into the task, the greater sense of ownership they feel. This suggests that the user's sense of ownership could potentially be manipulated based on the amount of work the user is asked to do to create an image. If a UI is deployed which consists of a text box for a single subject along with some suggested options for customization, as mentioned in Liu (2023) [5], it is possible that this sense of ownership would decline. We leave further exploration of this effect to future work.

6.3 Limitations of Our Approach

Although our study provided valuable insights into the user experience of AI image generators, it is important to recognize the limitations of our methodology. Firstly, the relatively small sample size may hinder the generalizability of our findings. Secondly, our methodology did not control for potential confounding variables such as participants' prior experiences with AI image generators, which may have influenced our results by affecting the time spent and effectiveness of the prompts written by experienced participants. We also explored only one AI image generator, DALL-E, and our findings may not be applicable to other, more highly specified, image generators.

6.4 Strengths & Limitations of AI

When compared to other AI image generators, DALL-E has several distinctive features and advantages. One of the key strengths of DALL-E is its ability to generate high-quality and creative images from a wide variety of textual descriptions, which sets it apart from many other image generators that rely on highly-specific training datasets such as cartoons or paintings. DALL-E can generate a wide range of images, from abstract concepts to realistic scenes, and is not limited to specific domains or objects. DALL-E can also generate images with a high level of detail and specificity, which can be important in certain applications such as medicine or engineering. Additionally, DALL-E has gained widespread attention due to its impressive outputs and the level of novelty in its approach.

Although DALL-E represents a significant advancement in AI image generation, there are several limitations and challenges associated with this system. One of the main issues with DALL-E is its limited contextual understanding, which can lead to generated images that are unrealistic or do not accurately represent the intended concept. Additionally, DALL-E may struggle with generating highly realistic images for scenes it has not observed in training, complex textures, or reflections. DALL-E and other AI image generators also do not have full syntactic understanding, as requests to "not include X" often results in an image that contains "X." More known limitations of AI image generators are misrepresenting human hands as having 6 or more fingers and generating nonsensical text that is visually similar to the Latin alphabet, but is not meaningful in any natural language.

Furthermore, DALL-E and other AI image generators can be computationally expensive and require significant resources to run, which may limit their accessibility and practicality in certain applications. The generated images can also be inconsistent, with variations in quality and realism depending on the specific input. These are issues that a user interface cannot adequately address, but the user can be made aware of these issues through the interface itself and adjust expectations accordingly.

7 CONCLUSION

In this study, we have examined strategies users employ in interacting with a AI image generator. By examining iterative prompts with the generator and evaluating the users' senses of satisfaction and ownership, we found that from a prompting perspective, the best prompts are short and keyword-focused. From a process perspective, it is best to start with one concrete idea and add more detail to the image over multiple iterations. We suggest that to increase overall user satisfaction, UIs provide feedback when a prompt is too long or too detailed, communicating the limitations of the AI to the user. We believe the user interface could be more highly structured to reduce the cognitive load of the user, that other limitations of AI should be provided to the user before they begin creating prompts, and that the non-deterministic nature of these models should be more effectively shown. If successful, we believe these additions could increase user satisfaction and lead to increases in AI literacy. We leave evaluating these interventions for future work.

REFERENCES

[1] Matthew Butterick. 2023. *Stable Diffusion Litigation*. Retrieved May 4, 2023 from https://stablediffusionlitigation.com/

[2] Upol Ehsan, Philipp Wintersberger, Elizabeth A Watkins, Carina Manger, Gonzalo Ramos, Justin D. Weisz, Hal Daumé Iii, Andreas Riener, and Mark O Riedl. 2023. Human-Centered Explainable AI (HCXAI): Coming of Age. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, 1–7. https://doi.org/10.1145/3544549.3573832

[3] Cheng-Zhi Anna Huang, Hendrik Vincent Koops, and Ed Newton-Rex. 2020. AI SONG CONTEST: HUMAN-AI CO-CREATION IN SONGWRITING. (2020).

[4] Kori Inkpen, Stevie Chancellor, Munmun De Choudhury, Michael Veale, and Eric P. S. Baumer. 2019. Where is the Human? Bridging the Gap Between AI and HCI. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3290607.3299002

[5] Vivian Liu. 2023. Beyond Text-to-Image: Multimodal Prompts to Explore Generative AI. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. https://doi/10.1145/3544549.3577043

[6] Vivian Liu and Lydia B Chilton. 2022. 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)*. Association for Computing Machinery, New York, NY, USA, 1–23. https://doi.org/10.1145/3491102.3501825

[7] Duri Long, Takeria Blunt, and Brian Magerko. 2021. Co-Designing AI Literacy Exhibits for Informal Learning Spaces. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (Oct. 2021), 293:1–293:35. https://doi.org/10.1145/3476034

[8] Duri Long, Lucas Liu, Swar Gujrania, Cassandra Naomi, and Brian Magerko. 2020. Visualizing Improvisation in LuminAI, an AI Partner for Co-Creative Dance. In *Proceedings of the 7th International Conference on Movement and Computing (MOCO '20)*. Association for Computing Machinery, New York, NY, USA, 1–2. https://doi.org/10.1145/3401956.3404258

[9] Duri Long, Aadarsh Padiyath, Anthony Teachey, and Brian Magerko. 2021. The Role of Collaboration, Creativity, and Embodiment in AI Learning Experiences. In *Creativity and Cognition (C&C '21)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3450741.3465264

[10] Duri Long, Anthony Teachey, and Brian Magerko. 2022. Family Learning Talk in AI Literacy Learning Activities. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–20. https://doi.org/10.1145/3491102.3502091

[11] Michael Muller, Lydia B Chilton, Anna Kantosalo, Charles Patrick Martin, and Greg Walsh. 2022. GenAICHI: Generative AI and HCI. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, 1–7. https://doi.org/10.1145/3491101.3503719

[12] Laria Reynolds and Kyle McDonell. 2021. 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 (CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, 1–7. https://doi.org/10.1145/3411763.3451760

[13] Jeba Rezwana and Mary Lou Maher. 2022. Understanding User Perceptions, Collaborative Experience and User Engagement in Different Human-AI Interaction Designs for Co-Creative Systems. In *Creativity and Cognition* (Venice, Italy) *(Camp;C '22)*. Association for Computing Machinery, New York, NY, USA, 38–48. https://doi.org/10.1145/3527927.3532789

[14] Daniel Russell, Q. Vera Liao, Chinmay Kulkarni, Elena L. Glassman, and Nikolas Martelaro. 2023. Human-Computer Interaction and AI: What practitioners need to know to design and build effective AI system from a human perspective. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, 1–3. https://doi.org/10.1145/3544549.3574170

[15] Justin D. Weisz, Michael Muller, Stephanie Houde, John Richards, Steven I. Ross, Fernando Martinez, Mayank Agarwal, and Kartik Talamadupula. 2021. Perfection Not Required? Human-AI Partnerships in Code Translation. In *26th International Conference on Intelligent User Interfaces (IUI '21)*. Association for Computing Machinery, New York, NY, USA, 402–412. https://doi.org/10.1145/3397481.3450656

[16] Justin D. Weisz, Michael Muller, Steven I. Ross, Fernando Martinez, Stephanie Houde, Mayank Agarwal, Kartik Talamadupula, and John T. Richards. 2022. Better Together? An Evaluation of AI-Supported Code Translation. In *27th International Conference on Intelligent User Interfaces (IUI '22)*. Association for Computing Machinery, New York, NY, USA, 369–391. https://doi.org/10.1145/3490099.3511157

[17] Darian Woods, Adrian Ma, Noah Glick, Corey Bridges, and Kate Concannon. 2023. *Artists vs. AI*. Retrieved May 4, 2023 from https://www.npr.org/2023/01/30/1152653269/artists-vs-ai

[18] Tongshuang Wu, Michael Terry, and Carrie J Cai. 2022. AI Chains: Transparent and Controllable Human-AI Interaction by Chaining Large Language Model Prompts. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* 345 (April 2022). https://dl.acm.org/doi/abs/10.1145/3491102.3517582

[19] Wei Xu and Marvin Dainoff. 2023. Enabling Human-Centered AI: A New Junction and Shared Journey Between AI and HCI Communities. *Interactions* 30, 1 (Jan. 2023), 42–47. https://doi.org/10.1145/3571883