Anyone can Design! With a little help from Generative AI

Ajay
zwz3wu@virginia.edu

Anushruti
rba7cb@virginia.edu

Changhong
hmw4yz@virginia.edu

Mohamad
anz8av@virginia.edu

Uday
rfn3ua@virginia.edu

Yash
ys4yh@virginia.edu

ABSTRACT

With a recent surge in generative AI systems, new collaboration possibilities have emerged between humans and AI for creative content. The role of AI has changed from decision-maker to human supporter. However, there are very few studies that have explored these collaborations. Further, generating creative content with AI requires a good understanding of the algorithms and their features. In this work, we studied human-computer co-creativity for poster design tasks. We recruited two participant groups and assigned them the task of designing two different types of posters. One of the groups was provided with a taxonomy designed by us to guide the design process. Later, we compared the completion time, design process, and prompt used between the taxonomy and non-taxonomy group. Further, we conducted a qualitative survey to gauge the participant's experience.

Author Keywords

HCI; Generative AI; Machine Learning; Stable Diffusion

INTRODUCTION

The rapid expansion of generative AI algorithms has facilitated the development of a new research area of generative art. AI is able to generate high-quality images, sounds, videos, and text that is difficult to differentiate from human-generated content. This development has enabled AI to collaborate with humans for design tasks. It has shifted AI's role from a decision-maker to a creative collaborator. Almost all leading HCI and AI conferences, such as CHI and NeurIPS, are organizing workshops or talks focused on Generative AI and HCI areas [1, 4].

Generative AI is defined as an AI system that uses existing media to create new, plausible media. With generative AI, algorithms learn the underlying pattern present in the data and can produce similar-looking novel content. Over the last decade, we have seen a shift in methodology from expert systems based on patterns and heavy curating towards stochastic and generative models such as Generative Adversarial Networks

ISBN 978-1-4503-2138-9.
DOI: 10.1145/1235

Figure 1: World's first magazine cover designed by Artificial Intelligence.

and Diffusion models. The development of these algorithms is further fueled by the excitement of prominent research academic and corporate research labs working on developing billion-parameter models trained on billions of images and text data. DALL-E by OpenAI, Imagen by Google, and Stable Diffusion by StabilityAI are some recently proposed popular algorithms [10, 11, 12].

However, how AI comprehends the data is different from how humans perceive the data. Due to this, every new surreal looking generated art receives wide excitement from the broader research and artistic community. Hence, to facilitate the wide adoption of these algorithms and their collaboration with humans, it is essential to study the working of these algorithms and their results for different contexts. As part of this work, we focus on the text-to-image generative model and study it for generating posters. Initially, we envisioned studying poster generation for a breadth of problems, such as news articles, magazines, and movies. However, our interim experiments introduced us to the amount of variability involved in the poster generation process for different contexts. Hence, we constrained our study to the New Yorker's cover story poster design. We chose New Yorker for our experiments because of its unique style and aesthetics. New Yorker is a widely popular magazine across the globe for its journalism, commentary, cartoons, and poetry. The cover designs and cartoons of the New Yorker are highly-popular among the readers and are famous for their impact.

We used an open-sourced deep-learning diffusion model trained on large magnitudes of data, allowing us the freedom to experiment with a large number of visual concepts. The stable diffusion model, since its release, has garnered wide reception and has already been adopted for multiple applications. In addition, model extensions are available for design applications such as Canva and Photoshop, allowing users to generate images for their tasks. These text-to-image machine learning models are trained to generate aesthetic real-looking images. A huge corpus of images and their associated captions are collected and used for training. After training, the model can generate an image based on the user's prompt. However, while we may compose any number of visual concepts, it is not guaranteed that any prompt passed through a text-to-image AI will produce a quality outcome. Hence, it requires hit-and-trial iterations to understand the combination of keywords, subjects, and styles so the model can generate high-quality content. In our study, we tried to develop a recipe in the form of taxonomy for guiding this generation process and for assisting designers with some inspiration and ideas for effectively using this model. Further, we studied and compared the design differences and required efforts between the taxonomy and non-taxonomy participant group.

RELATED WORKS

There have been few publications in this area, though there is active ongoing research being done by machine learning and human-computer interaction researchers. Vivian et al. recently developed guidelines for prompt engineering for these text-to-image-based models [6]. They compared different permutations of phrasings, model parameters, and image styles. They found no significant differences between different prompt phrasing and recommended focusing on subject and style keywords. They concluded that the model performed better on certain styles than others. In the follow-up to this study, they attempted to develop Opal, a tool for news illustrations [7]. Their developed tool takes in headlines and article content and, based on context, recommends keywords, styles, and tones for the image generation process.

Pavlichenko et al. investigated different methods for identifying the best prompts [9]. Finally, Oppenlaender et al. attempted to design a taxonomy of prompt modifiers for studying different modifiers. They conducted an ethnographic study of the community's prompt engineering practices on Twitter

Figure 2: New Yorker cover stories used for poster generation task.

[8]. As a result, they derived six different types of modifiers that significantly impact the generated art: subject term indicating the desired subject, style modifier capturing certain style, reference image for guiding the model, quality boosters for increasing details, repetition for strengthening the associations, and a magic term for introducing stochasticity in the design process. Building over these works, we explore prompt engineering for a poster generation task.

In their June edition, Cosmopolitan recently used the first-ever AI-generated poster for their magazine cover (Fig. 1). The cover was designed using OpenAI's text-to-image model - DALL-E 2. This design paved the way for future design covers and how designers will interact with these algorithms. The cover claimed to be designed in 20 seconds which is the time taken by the algorithm to generate the image given a prompt. However, a detailed video by the designers and AI practitioners later clarified that the task of coming up with a desirable prompt was a long iterative process. And herein lies the motivation for our project. In the current literature on prompt engineering, there have not been many formal studies proposing a framework for the design of prompts. Instead, independent practitioners have written many informal suggestions, tweets, and blogs detailing learning from their experiments. As part of this study and during our taxonomy design, we tried to develop a recipe or a guideline for assisting practitioners in generating better AI art. Prompt engineering is a brute-force process since the language used is free and open to interpretation. Our goal was to study the prompts design and build a taxonomy to assist designers.

STUDY DESIGN

In our study, we recruited 20 participants for the design task. The participants were randomly divided into two groups - the taxonomy group and the non-taxonomy group. Both groups

Figure 3: Image generated with base prompt - "a girl standing in raining dollar bills"

Figure 4: Image generated using taxonomy prompt - "A portrait of a girl standing in raining dollar bills, charcoal"

were assigned the same design task. The task comprised designing two New Yorker-style cover story posters using the Stable Diffusion text-to-image model [2]. Two different styles of posters were selected to study the impact of context in generative art design. One of the covers was a portrait-style poster (Sun Dappled), while the other was a landscape-style poster (Summer Walk), Fig. 2. We provided all the participants with the article title, context, and original poster. The participants were encouraged to develop a new design that they believed best suited the article. The original poster was only shared to inspire them. Our task pdf is present here.

Each participant was given access to the Stable Diffusion text-to-image model via the Huggingface online portal. Further, the taxonomy group was introduced to our taxonomy, and they were provided with guidelines to generate good-looking AI art. In contrast, the non-taxonomy group was neither introduced nor aware of this taxonomy during their design task. All the participants were required to complete the tasks within a maximum of 10 minutes per poster. The threshold for task completion was user satisfaction; as soon as the participant was satisfied with the resulting art, they would stop.

Upon task completion, all participants had to submit their final text prompt, design, and completion time. After the design task, they completed a qualitative survey. Survey collected demographic information, task difficulty (on a scale of 1 to 7), strategies adopted to come up with prompts, and the most valuable words and modifiers for design generation. Further, participants with taxonomy were required to share their thoughts on taxonomy, its usefulness and provide suggestions.

TAXONOMY DESIGN

The first step of developing the taxonomy involved conducting a detailed literature survey. Oppenlaender et al.'s work on taxonomy was helpful in breaking down the prompt into different parts, such as modifiers, and style words, among others, and iterating on each of them individually [8]. Inspired by Vivian et al.'s design tool Opal, we spent substantial time exploring the possibility of a text generation model, such as GPT3, assisting designers in writing prompts [7]. However, we didn't pursue this for our final proposal since GPT-3 style models tend to recommend words that improve the prompt's explanation, not necessarily the quality of images. Moreover, it will be required to finetune these models on a large corpus of prompts to make them a better prompt generators. A broad survey of informal resources introduced us to a breadth of artists and style modifiers [5]. Hinging on the taxonomy structure proposed in [8], we developed a collection of artist and style modifiers that are helpful for the poster generation process. The specific context of designing for the New Yorker poster helped us ground our study and search for these enhancers for a particular task. Furthermore, many websites with prompts and their generated art have come up for browsing through old-developed art. We used one of these websites to study the aesthetic impact of enhancers on a range of images [3]. Also, it allowed us to search the prominent keywords used by practitioners while generating a picture for the New Yorker context.

Based on our learnings from our survey and experiment, we converged on a prompt structure - "[Prefix] [Subject], [Enhancers]." The subject captures the primary purpose of the image, with prefixes and enhancers improving the quality. The subject description should be detailed and descriptive to guide the generative process. The inclusion of adjectives, action words, and time of the day, such as sunrise, can direct the model in a targeted direction. Prefix helps us in changing the type of image. For instance, adding "a portrait of" to a prompt

Effect	*Df*	*Sum Sq*	*Mean Sq*	*F*	*Pr (>F)*
Image Type (IT)	1	34.21	34.21	4.64	0.038*
Taxonomy (T)	1	10.87	10.87	1.48	0.232
IT:T Interaction	1	1.00	1.00	0.14	0.715
Residuals	36	36	265.09	7.36	

Table 1: Summary of ANOVA results. *: significant at 0.05 level

can substantially improve the quality of the generated face. Enhancers in the form of artist names and style modifiers help in dictating the mood of the image. For instance, the inclusion of artist names such as "Picasso" can change the image to look as if Pablo Picasso painted it. Style modifiers, such as "artstation" or "unreal engine," can make the picture similar to a game environment picture.

So overall, in our design guidelines, we start with a sample base prompt and the image generated for it. For example, the sample base prompt could be "A girl standing in raining dollar bills," and the image generated is shown in Fig 3. This base prompt captures the message of the picture. Then iteratively, we enhance this prompt using a list of prefixes and enhancers. In our design guidelines, we provided taxonomy participants with some popular prefixes, such as "A portrait of" or "An illustration of," and showed them how these prefixes change the base image. Similarly, we provide examples for enhancers such as "Artist Enhancers," which generate images in the style the artist would have drawn them, and "Modifier Enhancers," such as "oil painting" or "4K Resolution". In addition, we provided a link that summarized a broad list of possible artist and modifier enhancers that could be used to produce a better-looking poster. Fig 4 illustrates how the taxonomy has enhanced the poster. The prompt provided was "A portrait of a girl standing in raining dollar bills, charcoal." Taxonomy shared with participants can be viewed here.

RESULTS

In this section, we summarize our quantitative and qualitative findings from our user study. From our 20 participants, we obtained 40 prompts, generated images, and completion times. Furthermore, all the participants completed the survey.

Quantitative Analysis

We studied the effect of image type and taxonomy on completion time. In the image type variable, we included if the participant was designing for a portrait cover or a landscape cover. The taxonomy variable indicated whether the participants had access to our designed taxonomy. Time taken by the participant for each cover was the dependent variable in our study. Only image type significantly affected the completion time ($F_{1,35} = 4.64, p = 0.038$). There were no other significant effects observed. The results are summarized in Table 1. These findings indicated that the image context plays a major role in the design process. Also, these findings pointed us to consider more factors in our future study, such as there needs to be an objective evaluation of the poster cover by experts for comparing completion time, taxonomy inclusion and generated image quality.

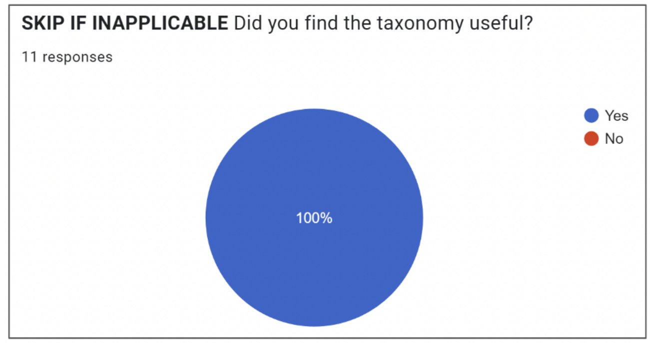

Figure 5: Qualitative Analysis on usefulness of Taxonomy

Qualitative Analysis

We analyzed the post-study survey filled out by all the participants. A major takeaway from this survey was that all the participants with access to our taxonomy found the taxonomy useful in the design process (Fig. 5). Moreover, on average, taxonomy candidates found the design task easier than non-taxonomy candidates. The task difficulty visualization can be seen in Figure 6. There was a mixed reaction, with a few taxonomy candidates finding the task very easy. Overall, there was a consensus among the taxonomy candidates in their design process. They followed the guidelines shared by our taxonomy, first describing the subject in great detail. And afterward, they iterated over the base image with different modifiers and enhancers. Interestingly, there was no consensus on which artist names or modifiers to use. This also indicated that different participants tend to approach these creative tasks differently. These findings helped us understand different factors of human-AI interaction in better detail.

DISCUSSION

The quantitative results show two important findings: First, completion time is context-dependent since image type significantly affected completion time. For instance, a task to create a portrait image is contextually different from a task to create a natural landscape image, which impacts completion time. This finding is a strong indicator that in future Human-AI interaction studies, a breadth of tasks should be considered and not restricted to a specific type. Second, completion time with self-satisfaction as the threshold is not the best metric to assess the effect of taxonomy. A person's design satisfaction can vary by many unknown factors. It could be driven by their mood, time of the day, or artistic experience. For example, a more artistic and experienced user may take longer to be content with a result than a user who is generally uninterested in art. Moreover, a plausible explanation for taxonomy candidates taking a long time is that the taxonomy causes the participants to be more meticulous and encourages them to iterate to reach a more satisfactory image. On the contrary, a non-taxonomy

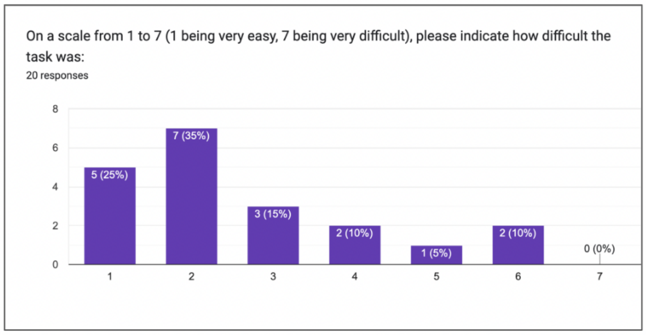

Figure 6: Qualitative Analysis on task difficulty

Figure 7: Posters generated by participants - 4 of these posters were generated by taxonomy candidates, and the other four were generated by non-taxonomy candidates.

candidate can take a long time because of their learning iterations in deriving a guideline for design. Hence, there needs to be an objective evaluation of the designed art instead of subjective satisfaction. Thus, a good follow-up study would be to evaluate the results produced by the users. The evaluators could be a committee of professional artists.

In addition to the limitations alluded to above, the study was limited by a small sample size of 20 participants (10 participants per taxonomy condition). A larger sample size may help reveal more significant results. Furthermore, an increase in the number of design tasks by each user can also help understand the learning happening during the design iterations. A taxonomy guides the process while the user designs his own intuitive design guidelines. This transitory learning happening during the human-AI interaction from a novice interactor to an experienced interactor can help HCI and AI community better design future tools. It may be beneficial to make the same participant do a non-taxonomy study first, followed by a taxonomy study, to realize the strengths of taxonomy. It will also lead to suggestions for future taxonomy design. Moreover, a sequential text analysis of prompt changes in a user design study can lead to a valuable prompt engineering study, capturing a great detail of human-AI interaction.

Our post-study survey captured many important factors in the user design process. However, it was brief and could have asked for more details from the user. In future studies, we could have questions on self-satisfaction, workload, fatigue, and whether the taxonomy impacted those measures.

CONCLUSION

In this study, we just scratched the tip of the iceberg. There are many unexplored questions and ideas around human-AI interaction that needs to be studied in the future. Furthermore, generative AI algorithms are witnessing unprecedented growth, gaining massive popularity in diverse domains. These algorithms are being used to generate different modalities of data - sound, text, video, and images. As part of our study, we only explored the generative AI algorithm (Stable Diffusion text-to-image model) to generate a New Yorker poster. Our experimental results indicated that taxonomy assists practitioners in designing better posters and can improve the human-AI interaction experience. The taxonomy acted as a source of inspiration for designers to come up with new innovative ideas and subsequently enhance the quality of the image. We found the type of image plays a significant role in the design process. Different design tasks and topics require variable effort. Hence, they need to be studied in detail to improve the interaction experience. We strongly believe that human-AI interaction will have a strong presence in HCI studies in the next decade. And AI as a technology will revolutionize many design spaces, changing how humans interact with these tools.

ACKNOWLEDGMENTS

We would like to thank all the participants who took out time and participated in our study and shared their feedback with us. We would also like to thank Professors and TAs for their valuable feedback, which kept us motivated to reach our final goal. Finally, we would also like to show our sincere gratitude towards our classmates for their valuable peer reviews and comments.

REFERENCES

[1] 2022. Generative AI and HCI Workshop. `https://generativeaiandhci.github.io/`. (2022).

[2] 2022. Hugging Face Stable Diffusion Model. `https://huggingface.co/spaces/stabilityai/stable-diffusion-1`. (2022).

[3] 2022. LexicaArt Website. `https://lexica.art/`. (2022).

[4] 2022. NeurIPS Creativity Workshop. `https://neuripscreativityworkshop.github.io/2022/`. (2022).

[5] 2022. ParrotZone Website. `https://proximacentaurib.notion.site/proximacentaurib/parrot-zone-74a5c04d4feb4f12b52a41fc8750b205`. (2022).

[6] Vivian Liu and Lydia B Chilton. 2022. Design Guidelines for Prompt Engineering Text-to-Image Generative Models. In *CHI Conference on Human Factors in Computing Systems*. 1–23.

[7] Vivian Liu, Han Qiao, and Lydia Chilton. 2022. Opal: Multimodal Image Generation for News Illustration. *arXiv preprint arXiv:2204.09007* (2022).

[8] Jonas Oppenlaender. 2022. The Creativity of Text-based Generative Art. *arXiv preprint arXiv:2206.02904* (2022).

[9] Nikita Pavlichenko and Dmitry Ustalov. 2022. Best Prompts for Text-to-Image Models and How to Find Them. *arXiv preprint arXiv:2209.11711* (2022).

[10] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. 2022. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125* (2022).

[11] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10684–10695.

[12] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily Denton, Seyed Kamyar Seyed Ghasemipour, Burcu Karagol Ayan, S Sara Mahdavi, Rapha Gontijo Lopes, and others. 2022. Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. *arXiv preprint arXiv:2205.11487* (2022).