Prompt Engineering for Text-Based Generative Art

Jonas Oppenlaender
joppenlu@jyu.fi
University of Jyväskylä
Jyväskylä, Finland

Figure 1: Examples of images generated with VQGAN+CLIP [9].

ABSTRACT

Text-based generative art has seen an explosion of interest in 2021. Online communities around text-based generative art as a novel digital medium have quickly emerged. This short paper identifies five types of prompt modifiers used by practitioners in the community of text-based generative art based on a 3-month ethnographic study on Twitter. The novel taxonomy of prompt modifiers provides researchers a conceptual starting point for investigating the practices of text-based generative art, but also may help practitioners of text-based generative art improve their images. The paper concludes with a discussion of research opportunities in the space of text-based generative art and the broader implications of prompt engineering from the perspective of human-AI interaction in future applications beyond the use case of text-based generative art.

CCS CONCEPTS

• **Human-centered computing** → *Natural language interfaces*; *Human computer interaction (HCI)*; • **Applied computing** → *Arts and humanities.*

KEYWORDS

prompt engineering, text-to-image synthesis, AI-generated art, language models, generative adversarial networks, GAN art

1 INTRODUCTION

Text-based generative art refers to the use of generative adversarial networks (GANs) to generate digital images from textual inputs. This textual interaction with the GAN-based system is similar to the interaction with autoregressive language models, such as OpenAI's GPT-3 [3]. In the case of text-based generative art, users describe a desired output (e.g., "`an oil painting of a beautiful landscape at dawn`") and the GAN-based system generates one (or several) images that match this input prompt via an adversarial learning approach. Examples of three images created from textual prompts are depicted in Figure 1.

To be effective, the textual input prompts need to be given in a certain format[1]. *Prompt engineering* [19, 30] is an emerging practice and research area on how to formulate effective input prompts for deep learning models. However, there is little formal guidance available on how to effectively formulate input prompts for text-based generative systems and artists often do not share their complete input prompts for their AI-generated artworks. Therefore, the learning curve can be steep and prompt engineering is a skill that is learned from other people's prompts and with experimentation.

With a specific focus on digital art generated with text-based generative systems, this paper contributes a taxonomy of input prompts used by practitioners in the text-based generative art community, based on a three-month ethnographic study of the community's prompt engineering practices on Twitter.

This short paper is structured as follows. We first provide a brief introduction into text-based generative art and prompt engineering in Section 2. After describing the methodological approach in Section 3, we present a taxonomy of five different types of prompt modifiers used by practitioners in the text-based generative art community (Section 5). The paper concludes with a discussion of opportunities for future research on text-based gneerative art and the broader implications beyond AI-generated art (Section 6).

2 BACKGROUND

2.1 Text-based Generative Art

AI art [22, 26] and generative art [1, 15] are broad concepts. Digital art and generative art are also often used synonymously [15]. In this paper, we follow Galanter's definition of generative art [15]:

[1] For instance, see https://beta.openai.com/examples/

> *Generative art refers to any art practice in which the artist cedes control to a system with functional autonomy that contributes to, or results in, a completed work of art. Systems may include natural language instructions, biological or chemical processes, computer programs, machines, self-organizing materials, mathematical operations, and other procedural inventions.*

From the lens of Human-Computer Interaction, this paper focuses on the *text-based* practice of generative art. Text-based generative art has experienced a strong technological push in early 2021 with the release of OpenAI's CLIP [28]. CLIP was conceived as an image classifier trained on a large amount of images and text from the World Wide Web. Due to the size and nature of the training set (connecting images and text), the CLIP neural network has learned to recognize concepts in images. and discriminative models has proven to be effective in generating digital images and artworks with high perceptual quality from textual input prompts. Because of this ability to recognize concepts in images, CLIP proved to be an effective discriminative component in GAN-based systems. One such system that has become enormously popular in 2021 is Katherine Crowson's VQGAN+CLIP [9]. In a VQGAN+CLIP system, VQGAN [14] generates images while CLIP acts as discriminator that judges how well the GAN-generated images match the given textual input prompt. Therefore, CLIP iteratively guides the GAN in activating latents in the neural network that produce the best-matching results for the given textual input prompt. This interaction between the two models in the text-to-image synthesis system is iterated until the discriminator can no longer tell the GAN-generated images apart from real images in the training set.

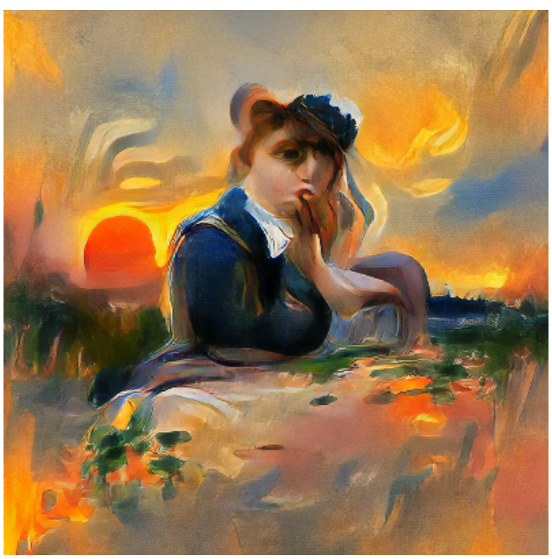

Figure 2: Artwork generated with VQGAN+CLIP [9] with the prompt "`pensive young woman at sunset in the style of Pierre Auguste Renoir, painting by Greg Rutkowski, trending on artstation.`"

2.2 Input Prompts for Generative Art

Inputs for text-to-image synthesis systems, such as VQGAN+CLIP [9], CLIP guided diffusion [7], PixelDraw [34], and ruDALL-E [31], are provided as textual input prompts. *Prompt engineering* [19] or *prompt programming* [30] refers to the emerging practice of writing input prompts in natural language to steer models into a desired direction. Figure 2 provides an example of a typical textual input prompt and the resulting AI-generated image.

Because the GAN operates iteratively on prior images, text-based generative systems also may accept initial images that seed the initial state of the GAN model. This can, for instance, be used to transfer a certain style to an image or direct the scene composition in the image. Further, target images can be given to the model besides the textual inputs to indicate a desired style to the model. For the sake of reducing complexity and avoiding confusion, this paper focuses only on *textual prompts* and leaves image-based prompts to future work.

3 METHOD

The research followed a two-fold approach. We combined auto-ethnographic experimentation (Section 3.1) and ethnographic research (Section 3.2), as described in this section.

3.1 Auto-ethnographic Research on Prompt Engineering for Text-based Generative Art

The first author conducted auto-ethnographic study [10–12] between October 2021 and December 2021. This personal ethnography [6] allowed the first author to get a "practitioner's perspective" [11] on text-based generative art by "learning from self-use" [21]. In the case of text-based generative art, auto-ethnography is an appropriate method because prompt engineering is an acquired skill that is associated with a learning curve. Prompt engineering is learned through iterative experimentation (akin to "brute-force trial and error" [19]) and by learning from other practitioners' input prompts (e.g., by applying input prompts found on social media).

The first author experimented with text-to-image synthesis and created digital images with a text-based generative system hosted on Google's Colaboratory (Colab). The first author started on average at least one Colab session every work day between October 4 and December 31, 2021. The free tier of Google Colab was used in all sessions. This limited the overall working time to between 2 and 3 hours per day, depending on the computational power of the assigned resources and whether penalties were incurred the previous day. On average, the time and computational resources were sufficient to generate between 12 and 24 images per day.

The auto-ethnographic research was informed by learning from the community on social media. To this end, we complemented the auto-ethnographic research with an online ethnographic research of the text-based generative art community on Twitter.

3.1.1 Self-Disclosure. While the first author has experimented with text-based generative systems and produced digital artworks with these systems, the first author is not an artist. The researcher's background is in Computer Science with focus on Human-Computer Interaction (HCI) and Social Computing. The research was conducted from a human-centered lens [18]. Our specific interest is in the text-based interactions of users with text-based generative systems.

3.2 Ethnographic Study of the Text-based Generative Art Community on Twitter

The auto-ethnographic research of prompt engineering was complemented with ethnographic research on the emerging text-based generative art community on Twitter. The aim of this social media ethnography [25, 27] was to learn more about the textual prompts used in the text-based generative art community. Insights derived from this community were used in the auto-ethnographic experimentation with a text-based generative system.

The ethnographic research was complemented with a review of the literature on the emerging practice of text-based generative art. While text-to-image synthesis is not a new area of research, text-based generative art is a very recent phenomenon. The online community around text-based generative only emerged in early 2021. With the exception of Liu and Chilton's design guidelines for prompt engineering [19], there is not much scholarly literature on the practice of text-based generative art as of yet. Therefore, our literature review focused on sources in the grey literature, such as posts on webblogs and other articles on the Web.

3.3 Research Material

3.3.1 Text-based Generative System. VQGAN+CLIP was selected for experimenting with text-based generative art. VQGAN+CLIP was created shortly after the release of CLIP in early 2021 and has since become very popular in the text-based generative art community. VQGAN+CLIP can be considered instrumental to the growth of the text-based generative art community. The system is available on Google's Colaboratory (Colab) where it can be executed free of charge. The notebook on Colab is very accessible and easy to use. VQGAN+CLIP also requires less memory than later systems, and it is therefore less likely that image generation will fail on Colab due to insufficient memory.

Many variations of the VQGAN+CLIP have been created. The research in this paper was conducted with a notebook titled "VQGAN and CLIP (z + quantize method with augmentations)" [8]. This particular notebook was originally created by Katherine Crowson, with "modifications by Eleiber # 8347" and a "friendly interface" by "Abulafia # 3734" and further modifications by Justin John.

3.3.2 Twitter Community. During the 3-month period of research, the first author followed postings on Twitter to learn about different prompts used in the text-based generative art community. To this end, the first author followed trending hashtags, such as #vqganclip, #VQGAN, #clipguideddiffusion, #digitalart, #AIArt, #generativeart, and #GANArt.

Not every practitioner of text-based shares their prompts, as mentioned in the Introduction. Especially if commercial interests are involved (e.g., selling the text-based art as non-fungible token (NFTs)), some practitioners keep their prompts a secret. Others are more liberal in sharing their prompts. It is the posts from this group of Twitter users that informed this research (e.g., the posts by Katherine Crowson (@RiversHaveWings), Hannah Johnston (@hannahjdotca), @nshepperd1, and John David Pressman (@jd_pressman), to name but a few).

Due to the relative scarcity of this material, the data was analysed as follows. A list of potential candidates for prompt modifiers was iteratively and inductively compiled and thematically grouped. Whenever a candidate for a novel type of prompt modifier was encountered in a post on Twitter, the first author revisited the list of potential candidates. Therefore, the list of prompt modifiers was iteratively and inductively expanded when new types of prompt modifiers were encountered. After some weeks of collecting data this way, the list of prompt modifiers did no longer grow, even if outstanding instances of outstanding prompts were encountered.

The author further engaged with the text-based generative art community, participated in discussions, and posted digital images created with a text-based generative system on Twitter. The AI-generated images (including intermediate steps) were collected together with their respective configuration settings (e.g., textual input prompts, GAN model, seed, etc.) used in each run of the VQGAN+CLIP system.

3.4 Documentation

The findings were documented in a PowerPoint presentation [24] with text and images to produce an evocative and aesthetic description of the ethnographic research. The documentation was iteratively extended and corrected, if needed. The first author's creation of and engagement with the presentation slides acted as a daily conversation with the research material. This allowed the first author to concurrently and iteratively develop and articulate an understanding of the subject matter both visually and textually. At the end of the research period, the first author engaged in a summative analysis of the field notes [11].

4 TAXONOMY OF PROMPT MODIFIERS

The text-based generative art community quickly found that the aesthetic qualities and subjective attractiveness of images can be improved by adding modifiers to the textual input prompts. These **prompt modifiers** aim to direct the GAN-based system in certain directions.

Our research points towards there being at least five different types of prompt modifiers (subject terms, style modifiers, quality boosters, repetitions, and magic terms) used by practitioners in the text-based generative art community.

Subject terms indicate the desired subject to the model (e.g., "`an old car in a meadow`" or "`a landscape at dawn`"). While it is possible to generate images without subject terms, the subject is essential for controlling the image generation process. On the other hand, since CLIP was trained on images in context of their respective descriptive text, subject terms can, in some cases, have less control over the outcome. One such case is the artist Zdzisław Beksiński who developed a unique and recognizable style but never provided titles for his works. For this reason, specific subjects can be less reliably reproduced in images that are styled to resemble Beksiński's artworks.

Style modifiers can be added to a prompt to produce images in a certain style This type of modifier will consistently reproduce a characteristic style. For instance, the modifier "`by Francisco Goya`" will generate digital images in the recognizable style of

a) b) c) d)

Images generated with VQGAN+CLIP [9], 175 iterations, CLIP model ViT-B/32, VQGAN model wikiart_16384, seed 6087304447281500163, text prompts:

a) "ufo landing"

b) "ufo landing, daguerreotype"

c) "ufo landing, daguerreotype, trending on /r/art"

d) "ufo landing, daguerreotype, by greg rutkowski, trending on /r/art"

Figure 3: Example of iterative prompt engineering for generating an image.

the late Spanish painter. Examples of this type of modifier include, for instance, "surrealistic painting", "oil painting on canvas", "#pixelart", "hyperrealistic", "ultra photorealistic", "ultra photorealistic", "in the style of a cartoon", "matte painting", "by Claude Lorrain", or "in the style of Hudson River School". Especially the two modifiers "by Greg Rutkowski" and "by James Gurney" have become very popular in the community of text-based generative art as a means to incentivize the generative system to produce images in a certain style and quality.

Quality boosters can be added to a prompt to increase aesthetic qualities and the level of detail. Examples of this type of modifier are the terms "trending on artstation" and "rendered in UnrealEngine", but also "AWESOME", "#wow", "wondrous", "eclectic, "detailed, "fantastic", "beautiful", "rendering", "rendered in Unreal Engine", or "trending on artstation".

This type of modifier can come in the form of "extra fluff" and verbosity in the prompt which may boost the amount of details and overall quality of the generated image. For instance, the prompt "painting of an exploding heart" could potentially be improved by adding the modifiers "eclectic, detailed, fiery, vfx, rendered in octane, postprocessing, 8k."

Repetition can potentially strengthen the associations formed by the generative system. For instance, the prompt "space whale. a whale in space"[2] by @nshepperd1 will likely produce subjectively better results than either of the two subject terms alone.

Magic terms introduce randomness to the image that can lead to surprising results. For instance, Twitter user @jd_pressman added the magic term "control the soul" to the prompt "orchestra conductor leading a chorus of sound wave audio waveforms swirling around him on the orchestral stage, control the soul, trending on artstation"[3]. The term was added to – in his words – produce "more magic, more wizard-ish imagery"[4].

As seen in the examples provided in this section, prompt modifiers can take a variety of forms. They can, for instance, be added as hash tags (e.g., #wow), simple attribution phrases (e.g., by [artist]), and more complex composite statements (e.g., in the style of [artist]). Further, not every part of a prompt has the same importance and there are specific affordances of the generative systems that are being used in prompt engineering in practice, as described in the following section.

5 THE PRACTICE OF PROMPT ENGINEERING

Prompt engineering is the iterative practice of applying prompt modifiers to input prompts. Prompt engineering resembles a conversation with the text-based generative system. Prompt modifiers are applied iteratively or based on best practices learned with experience (see Figure 3). Knowing what prompt modifiers work best for a given subject term is often the result of iterative experimentation, research in online communities, and the use of online tools and resources created for supporting the practice of prompt engineering.

This section provides an overview of how the different types of prompt modifiers are applied in practice with specific focus on the generation of static images with VQGAN+CLIP. We acknowledge that alternative means of text-to-image synthesis exist, for instance, diffusion-based systems and systems for creating keyframed animations (such as zooming). We leave these alternatives to future work. We further focus only on text-based generation of images, and leave any kind of post-processing steps (e.g., editing with graphics editors or more complex work flows consisting, for instance, of several chained GAN-based models) to future work.

While images can be generated from random text or even single characters and emojis, the **subject term** is elemental to the controlled generation of digital images. Consequently, a prompt

[2] https://twitter.com/nshepperd1/status/1456584388037148678

[3] https://twitter.com/jd_pressman/status/1457171648293924867

[4] https://twitter.com/jd_pressman/status/1457445367125921793

typically contains at least one subject term. Any other parts of the prompt are optional. It is, for instance, possible to generate artworks with the prompt "car." In practice, however, practitioners of text-based generative art use modifiers to improve the resulting images and to exercise control over the image creation process.

Modifiers are typically added with the intention to either boost the quality or modify the style of the generated image (as described in Section 4). Style modifiers and quality boosters do not form a disparate set. Rather, the two types of modifiers can have overlapping effects and the difference between the two types of prompt modifiers is sometimes not fully apparent. The modifier "`Greg Rutkowski`", for instance, exhibits this property. Greg Rutkowski[5] is a contemporary illustrator and concept artist who has been embraced by the text-based generative art community in their practice of prompt engineering. Images generated with the modifiers "`by greg rutkowski`" or "`in the style of greg rutkowski`" are of high quality, texture-rich and contain a high amount of details. As such, this style modifier is often used as a quality booster in the community.

Once a style modifier has been added, the style can be reinforced and "solidified" without losing expressivity. **Solidifiers**, such as repetitive terms, can be applied to any of the other types of modifiers (subjects, style modifiers, and quality boosters), although they are most commonly applied to subject terms.

Last, **magic terms** may be optionally added to increase the chance of surprising results. The use of magic terms will result in more variation in the output, while maintaining the overall style. An example of this type of modifier can be found in the prompt by Twitter user @DMTFL_AI:

> "`a retrofuturistic world next to a bleak dystopian world, divided down the middle in the vivid vaporwave style of gustave dore`"

In the words of @DMTFL_AI, "when given this many words, the consistency holds strong yet the input image variety gets extra interpretive"[6].

Each of the five types of prompt modifiers can be assigned **weights**. Weighted terms can be negative to exclude subjects and styles from being generated. For instance, VQGAN+CLIP tends to generate heart-shaped objects with red colour when the prompt contains the word "`love.`" By adding a negative weight to the prompt (e.g., "`heart:-1`"), the system can be instructed to not activate the corresponding latents in its neural network. The resulting images are thus free from heart-shaped objects.

Weighted terms can also be used to seamlessly mix styles. For instance, Twitter user @c0y0te6 mixed the styles of two artists in the prompt "`a painting of a high prestess [sic] summoning a demon by Ralph McQuarrie:75 | by Zdzislaw Beksinski:25`"[7]. The style of Ralph McQuarrie was, in this case, given precedence over the style of Zdzislaw Beksinski (with a ratio of 3:1).

Table 1 summarizes the iterative nature of prompt engineering (c.f. Figure 3). Subject terms are most important for the controlled generation of images and usually written as first step. Modifiers and solidifers are then added to the prompt, either iteratively (image after image) or from learned experience. Last, weights can be applied to exclude or mix subjects and styles.

Table 1: The iterative practice of prompt engineering.

Step	Purpose	Prompt modifier	Importance
1	Subject	subject term	required
2	Modifier	style modifier, quality booster	optional
3	Solidifier	repetition	optional
4	Variation	magic terms	optional
5	Weights	exclusion and mixing	optional

6 DISCUSSION

Text-based generative art has made strong progress in the year 2021. The availability and accessibility of text-based generative systems as a new computational medium, paired with a specific bundle of technologies that support the ecosystem of this "emerging art scene" [33], have resulted in an explosion of AI-generated artworks being shared online by hobbyists and artists. Dedicated online communities have rapidly formed around generating and sharing text-based generative art.

Using only natural language, it is now possible for anyone to generate images that resemble the artworks of past and contemporary artists without technical expertise and skills. The application of style modifiers is key to this emerging creative practice.

Prompt engineering is an emerging field. With the exception of the design guidelines by Liu and Chilton [19], there is little scholarly literature on prompt engineering for text-based generative art.

This paper provides an investigation of prompt engineering practices and contributes a taxonomy of five different types of prompt modifiers. The taxonomy of prompt modifiers acts as a conceptual starting point for future structured investigations into prompt engineering for text-based generative art. And indeed, the emerging community around text-based generative art as well as the practice of prompt engineering present numerous opportunities for future research, as outlined in the following section.

6.1 Opportunities for Future Research

The art community around text-based generative art makes an attractive study subject. Future research questions could, for instance, include:

- *The text-based generative art community and its prompt engineering practices:*
  - What are the values of the emerging AI-based art community, especially in regard to timely topics such as attribution of agency to the AI, copyright, and ethical behavior?
  - What are the prompt engineering work flows and strategies adopted by practitioners in the text-based generative art community?
  - What are the factors that make an input prompt successful in generating a 'beautiful' AI-generated artwork?
  - Do 'creative' prompts produce images and artworks that are more creative?
- *Broader ethical, ontological, and epistemological questions:*

[5] https://www.artstation.com/rutkowski
[6] https://twitter.com/DMTFL_AI/status/1474229217009319936
[7] https://twitter.com/c0y0te6/status/1481780797858275329

– Is AI-generated art creative? Given that anybody can generate art from textual prompts, what is the nature of human creativity involved in text-based generative art?
– Does the AI have agency in the process of creating art? How do the practitioners in the community see the AI in the process of creating text-based generative art?

Future work could further explore the text-based generative art community and its prompt engineering practices in more detail, using the taxonomy presented in this paper as a conceptual starting point or framework. For instance, practitioners may develop complex practices and work flows for creating their artworks (e.g., generating initial images with ruDall-E and finalizing them with CLIP-guided diffusion). Some GAN-based systems also allow image inpainting [36] which offers a level of interactivity beyond mere textual input prompts. Further, practitioners may make certain idiosyncratic choices when they create text-based generative art (e.g., selecting certain values as seed for the model or adapting the canvas size to certain subject terms). Some of these choices may fall into the realm of folk theories [13, 17] – that is, causal attributions that may or may not be true –, while other choices may be based on the practitioner's experimentation and experience with prompt engineering.

The emerging community around text-based generative art also offers an opportunity for HCI researchers to make technical contributions [35] in the form of creativity support tools, user interfaces, and interactive experiences to support the creation of text-based generative art, to teach novices the practice of prompt engineering, and to advance the emerging AI-generated art ecosystem. Research in this space could make a timely contribution to a novel computational medium and an emerging digital art form.

There are also broader interdisciplinary implications of ethical and societal relevance that go beyond prompt engineering and interaction with AI via textual prompts, as outlined in the following section.

6.2 Broader Implications for Human-AI Interaction

Research on prompt engineering has broader implications and is not only relevant to the field of AI-generated art, but also to the interaction of humans with deep learning models and artificial intelligence in general.

6.2.1 AI and the future of creative work. In the future, we may see neural networks with generative capabilities that transcend what we can imagine today. Such powerful AI-based systems will have implications for how we interact with computers, and therefore the future of creative work. Machine learning will not only transform the way we perform work online, but also the content of our work and the human agency in the work.

A recent example of an application that has such transformative potential is OpenAI's Codex [4, 23]. Codex is a large language model that interprets commands in natural language and generates programming code. In the future, instead of typing code, we will be able to describe a software and its expected outputs in natural language. Deep learning models, such as OpenAI's Codex, or even larger foundation models [2], will then generate executable software code based on the human's spoken or written input prompts. This change in the agency of humans and computers could be transformative to creative work, such as software development. Similar technology has already found application in GitHub's CoPilot[8], an "AI pair programmer" that assists its users in auto-completing programming code.

There is much potential in deep learning to disrupt and transform entire sectors of the creative economy. For instance, generative systems could create entire interactive story-driven worlds and games from short text prompts. Low-code and no-code tools for creating online products and experiences will become increasingly common in the future, and declarative machine learning systems may – as a next wave of machine learning – bring machine learning to non-coders [20]. Gartner estimated that by 2024, 80% of technology products and services will be built by people who are not technology professionals [16]. Interaction with opaque machine learning models will become increasingly more common in future use cases and applications of artificial intelligence. Engineering effective prompts will, therefore, play an increasingly important role in the future.

6.2.2 Other use cases. The use case of text-based generative art is but one of many use cases, with implications for the future of creative work and human-AI interaction in general. The latter can be viewed from many different perspectives, such as human-centered AI [32], human-AI partnerships [29], and human-AI cooperation [5]. Irrespective of the term used to describe our relationship with AI, we will increasingly interact with opaque models through natural language in the future.

Research on how input prompts can be effectively formulated is therefore timely and important. Research in this space will advance our understanding of how people can effectively interact with machine learning models. The research will inform the design of future applications and systems powered by Artificial Intelligence (AI) and advance the understanding on the needs of different stakeholders using these AI-driven systems. A better understanding of users could not only advance research into interpretability and explainability of AI-driven systems, but also provide valuable insights into how web-based systems can be designed for fairness, accountability, and transparency. These three constituents could be a step stone to support the formation of trusting social relationships with artificial intelligence.

7 CONCLUSION

This paper presented a taxonomy of input prompt modifiers used in the community of text-based generative art on Twitter. We discussed how these modifiers are being applied in practice. The paper discussed several opportunities for future research on the fast-moving text-based generative art community and touched on the broader implications for creative work and interaction of humans with computers beyond the use case of text-based generative art.

REFERENCES

[1] Margaret A. Boden and Ernest A. Edmonds. 2009. What is generative art? *Digital Creativity* 20, 1-2 (2009), 21–46. https://doi.org/10.1080/14626260902867915

[8]copilot.github.com

[2] Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, Erik Brynjolfsson, Shyamal Buch, Dallas Card, Rodrigo Castellon, Niladri Chatterji, Annie Chen, Kathleen Creel, Jared Quincy Davis, Dora Demszky, Chris Donahue, Moussa Doumbouya, Esin Durmus, Stefano Ermon, John Etchemendy, Kawin Ethayarajh, Li Fei-Fei, Chelsea Finn, Trevor Gale, Lauren Gillespie, Karan Goel, Noah Goodman, Shelby Grossman, Neel Guha, Tatsunori Hashimoto, Peter Henderson, John Hewitt, Daniel E. Ho, Jenny Hong, Kyle Hsu, Jing Huang, Thomas Icard, Saahil Jain, Dan Jurafsky, Pratyusha Kalluri, Siddharth Karamcheti, Geoff Keeling, Fereshte Khani, Omar Khattab, Pang Wei Koh, Mark Krass, Ranjay Krishna, Rohith Kuditipudi, Ananya Kumar, Faisal Ladhak, Mina Lee, Tony Lee, Jure Leskovec, Isabelle Levent, Xiang Lisa Li, Xuechen Li, Tengyu Ma, Ali Malik, Christopher D. Manning, Suvir Mirchandani, Eric Mitchell, Zanele Munyikwa, Suraj Nair, Avanika Narayan, Deepak Narayanan, Ben Newman, Allen Nie, Juan Carlos Niebles, Hamed Nilforoshan, Julian Nyarko, Giray Ogut, Laurel Orr, Isabel Papadimitriou, Joon Sung Park, Chris Piech, Eva Portelance, Christopher Potts, Aditi Raghunathan, Rob Reich, Hongyu Ren, Frieda Rong, Yusuf Roohani, Camilo Ruiz, Jack Ryan, Christopher Ré, Dorsa Sadigh, Shiori Sagawa, Keshav Santhanam, Andy Shih, Krishnan Srinivasan, Alex Tamkin, Rohan Taori, Armin W. Thomas, Florian Tramèr, Rose E. Wang, William Wang, Bohan Wu, Jiajun Wu, Yuhuai Wu, Sang Michael Xie, Michihiro Yasunaga, Jiaxuan You, Matei Zaharia, Michael Zhang, Tianyi Zhang, Xikun Zhang, Yuhui Zhang, Lucia Zheng, Kaitlyn Zhou, and Percy Liang. 2021. On the Opportunities and Risks of Foundation Models. arXiv:2108.07258 [cs.LG]

[3] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. arXiv:2005.14165 [cs.CL]

[4] Mark Chen, Jerry Tworek, Heewoo Jun, Qiming Yuan, Henrique Ponde de Oliveira Pinto, Jared Kaplan, Harri Edwards, Yuri Burda, Nicholas Joseph, Greg Brockman, Alex Ray, Raul Puri, Gretchen Krueger, Michael Petrov, Heidy Khlaaf, Girish Sastry, Pamela Mishkin, Brooke Chan, Scott Gray, Nick Ryder, Mikhail Pavlov, Alethea Power, Lukasz Kaiser, Mohammad Bavarian, Clemens Winter, Philippe Tillet, Felipe Petroski Such, Dave Cummings, Matthias Plappert, Fotios Chantzis, Elizabeth Barnes, Ariel Herbert-Voss, William Hebgen Guss, Alex Nichol, Alex Paino, Nikolas Tezak, Jie Tang, Igor Babuschkin, Suchir Balaji, Shantanu Jain, William Saunders, Christopher Hesse, Andrew N. Carr, Jan Leike, Josh Achiam, Vedant Misra, Evan Morikawa, Alec Radford, Matthew Knight, Miles Brundage, Mira Murati, Katie Mayer, Peter Welinder, Bob McGrew, Dario Amodei, Sam McCandlish, Ilya Sutskever, and Wojciech Zaremba. 2021. Evaluating Large Language Models Trained on Code. arXiv:2107.03374 [cs.LG]

[5] Jacob W. Crandall, Mayada Oudah, Tennom, Fatimah Ishowo-Oloko, Sherief Abdallah, Jean-François Bonnefon, Manuel Cebrian, Azim Shariff, Michael A. Goodrich, and Iyad Rahwan. 2018. Cooperating with machines. *Nature Communications* 9, 1 (2018), 12 pages. https://doi.org/10.1038/s41467-017-02597-8

[6] Lyall Crawford. 1996. Personal ethnography. *Communication Monographs* 63, 2 (1996), 158–170. https://doi.org/10.1080/03637759609376384

[7] Katherine Crowson. 2021. CLIP Guided Diffusion HQ 256x256. https://colab.research.google.com/drive/12a_Wrfi2_gwwAuN3VvMTwVMz9TfqctNj

[8] Katherine Crowson. 2021. VQGAN+CLIP(Updated).ipynb. https://colab.research.google.com/github/justinjohn0306/VQGAN-CLIP/blob/main/VQGAN%2BCLIP(Updated).ipynb

[9] Katherine Crowson, Stella Biderman, Daniel Kornis, Dashiell Stander, Eric Hallahan, Louis Castricato, and Edward Raff. 2022. VQGAN-CLIP: Open Domain Image Generation and Editing with Natural Language Guidance. https://doi.org/10.48550/ARXIV.2204.08583

[10] Norman K. Denzin and Yvonna S. Lincoln. 2017. *The SAGE Handbook of Qualitative Research* (5th ed.). SAGE, Thousand Oaks, CA.

[11] Margot Duncan. 2004. Autoethnography: Critical Appreciation of an Emerging Art. *International Journal of Qualitative Methods* 3, 4 (2004), 28–39. https://doi.org/10.1177/160940690400300403

[12] Carolyn Ellis, Tony E. Adams, and Arthur P. Bochner. 2011. Autoethnography: An Overview. *Historical Social Research / Historische Sozialforschung* 36, 4 (138) (2011), 273–290. http://www.jstor.org/stable/23032294

[13] Motahhare Eslami, Karrie Karahalios, Christian Sandvig, Kristen Vaccaro, Aimee Rickman, Kevin Hamilton, and Alex Kirlik. 2016. First I "like" It, Then I Hide It: Folk Theories of Social Feeds. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 2371–2382. https://doi.org/10.1145/2858036.2858494

[14] Patrick Esser, Robin Rombach, and Björn Ommer. 2021. Taming Transformers for High-Resolution Image Synthesis. arXiv:2012.09841 [cs.CV]

[15] Philip Galanter. 2016. *Generative Art Theory*. John Wiley & Sons, Ltd, Chapter 5, 146–180. https://doi.org/10.1002/9781118475249.ch5

[16] Gartner. 2021. Gartner Says the Majority of Technology Products and Services Will Be Built by Professionals Outside of IT by 2024. Press release. https://www.gartner.com/en/newsroom/press-releases/2021-06-10-gartner-says-the-majority-of-technology-products-and-services-will-be-built-by-professionals-outside-of-it-by-2024

[17] Susan A. Gelman and Cristine H. Legare. 2011. Concepts and folk theories. *Annual Review of Anthropology* 40 (2011), 379–398. https://doi.org/10.1146/annurev-anthro-081309-145822

[18] Mark Guzdial. 2013. Human-Centered Computing: A New Degree for Licklider's World. *Commun. ACM* 56, 5 (may 2013), 32–34. https://doi.org/10.1145/2447976.2447987

[19] Vivian Liu and Lydia B. Chilton. 2021. Design Guidelines for Prompt Engineering Text-to-Image Generative Models. https://doi.org/10.48550/ARXIV.2109.06977

[20] Piero Molino and Christopher Ré. 2021. Declarative Machine Learning Systems. *Commun. ACM* 65, 1 (dec 2021), 42–49. https://doi.org/10.1145/3475167

[21] Carman Neustaedter and Phoebe Sengers. 2012. Autobiographical Design in HCI Research: Designing and Learning through Use-It-Yourself. In *Proceedings of the Designing Interactive Systems Conference* (Newcastle Upon Tyne, United Kingdom) *(DIS '12)*. Association for Computing Machinery, New York, NY, USA, 514–523. https://doi.org/10.1145/2317956.2318034

[22] Anna Notaro. 2020. State-of-the-art: AI through the (artificial) Artist's Eye. In *Proceedings of EVA London 2020*. 322–328. https://doi.org/10.14236/ewic/EVA2020.58

[23] OpenAI. 2021. Codex. https://openai.com/blog/openai-codex/

[24] Jonas Oppenlaender. 2022. Prompt Engineering for Text-to-image Synthesis. https://doi.org/10.6084/m9.figshare.18899801

[25] Sarah Pink, Heather Horst, John Postill, Larissa Hjorth, Tania Lewis, and Jo Tacchi. 2016. *Digital Ethnography: Principles and Practice*. SAGE, London, UK.

[26] Fabrizio Augusto Poltronieri and Max Hänska. 2019. *Technical Images and Visual Art in the Era of Artificial Intelligence: From GOFAI to GANs*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3359852.3359865

[27] John Postill and Sarah Pink. 2012. Social Media Ethnography: The Digital Researcher in a Messy Web. *Media International Australia* 145, 1 (2012), 123–134. https://doi.org/10.1177/1329878X1214500114

[28] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning Transferable Visual Models From Natural Language Supervision. arXiv:2103.00020 [cs.CV]

[29] Sarvapali D. Ramchurn, Sebastian Stein, and Nicholas R. Jennings. 2021. Trustworthy human-AI partnerships. *iScience* 24, 8 (2021). https://doi.org/10.1016/j.isci.2021.102891

[30] 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* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 314, 7 pages. https://doi.org/10.1145/3411763.3451760

[31] Sber AI. 2021. Russian DALL-E. https://rudalle.ru/en/

[32] Ben Shneiderman. 2020. Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. *International Journal of Human–Computer Interaction* 36, 6 (2020), 495–504. https://doi.org/10.1080/10447318.2020.1741118

[33] Charlie Snell. 2021. Alien Dreams: An Emerging Art Scene. https://ml.berkeley.edu/blog/posts/clip-art/

[34] Tom White. 2021. pixray GitHub repository. https://github.com/pixray/pixray

[35] Jacob O. Wobbrock and Julie A. Kientz. 2016. Research Contributions in Human-Computer Interaction. *Interactions* 23, 3 (2016), 38–44. https://doi.org/10.1145/2907069

[36] Lisai Zhang, Qingcai Chen, Baotian Hu, and Shuoran Jiang. 2020. *Text-Guided Neural Image Inpainting*. Association for Computing Machinery, New York, NY, USA, 1302–1310. https://doi.org/10.1145/3394171.3414017