arXiv:2206.02904v1 [cs.HC] 13 May 2022

The Creativity of Text-based Generative Art

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

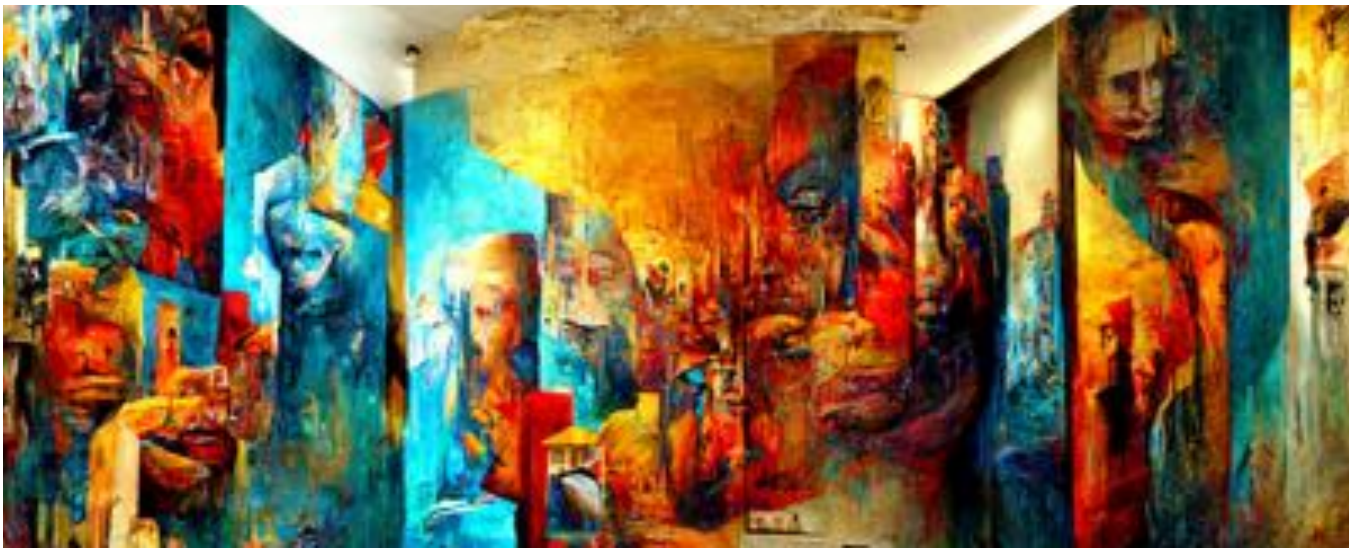


Figure 1: Digital image generated from the prompt "art on the wall" using Midjourney's CLIP guided diffusion [28].

ABSTRACT

Text-based generation of digital images has made a giant leap towards becoming a mainstream phenomenon. With text-based generative systems, anybody can create digital images and artworks. This provokes the question of whether text-based generative art is creative. This paper expounds on the nature of human creativity involved in text-based generative art with a specific focus on the practice of prompt engineering, drawing on Rhodes's conceptual model of creativity. The paper critiques the current product-centered view of creativity which may fall short in the context of text-based generative art. An case exemplifying this shortcoming is provided and future opportunities for research on text-based generative art are outlined.

CCS CONCEPTS

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

KEYWORDS

generative art, creativity, text-to-image synthesis, co-creation

1 INTRODUCTION

Text-based generative systems have become popular means for creating digital artworks. Given a textual input prompt, these systems are able to generate digital images with high aesthetic quality and subjective attractiveness. With certain textual inputs, the images can imitate the style of artworks by past and present artists [26]. An ecosystem of tools and resources around text-based generative art is emerging online. Many systems are available as open source in notebooks on Google's Colaboratory[1] (Colab). Colab is an online service that allows execution of Python-based code free of charge. For this reason, Colab has become one of the main drivers of innovation in the enthusiastic text-based generative art community. This community consists of *practitioners* (i.e., novice users, but also skilled artists) who share their digital artworks on social media. Given its growing capabilities and ease-of-use, text-based generative art is on the verge of becoming a mainstream phenomenon.

Common to most text-based generative art systems is that users can create digital images and artworks with little to no understanding of the underlying technologies, simply by writing textual prompts in natural language (a practice called *prompt engineering* [17] or *prompt programming* [37]). Given the ease of use and emerging ubiquity of text-based generative systems, the question arises: **If anybody can produce digital images that resemble masterful pieces of art simply by feeding textual prompts into a "black box" system, is text-based generative art creative? What is the nature of the human creativity involved in generating images with text-based generative systems?** In this paper, I aim to answer this question.

The paper critiques the reigning product-centered paradigm of creativity in which an artifact is considered creative if it is both novel and appropriate [40]. In the context of text-based generative

[1] https://colab.research.google.com

art, this paper exemplifies a case in which this two-part operationalization of creativity fails to measure the human creativity. Instead of only looking at the end product of the creative process, we need to expand our view to include the whole creative process. Rhodes's conceptual model of the "four P" of creativity [39] provides a fitting framework for this investigation. The "four P" look at creativity from four perspectives – the product, person, process, and press (i.e., the creative environment). In the context of text-based generative art, all four perspectives are necessary to capture the full extent of human creativity (with special emphasis on the latter three).

This paper expounds on the nature of human creativity prevalent in the sub-culture of text-based generative art. When creating text-based generative art, users cede control (in part) to artificial intelligence (AI) [14]. Consequently, the human creativity in text-based generative art lies not in the end product (i.e., the digital image), but arises from the interaction of humans with the AI and the resulting practices that evolve from this interaction (e.g., prompt engineering). For instance, image-level and portfolio-level curation are important practices relating to the process of text-based generative art. The practices are shaped and informed by a growing ecosystem of community-driven resources and tools. Online communities are an important factor influencing the human creativity of text-based generative art.

In summary, the paper makes the following contributions:

- We sketch a case (with two exemplary scenarios) in which the popular product-based operationalization of creativity fails to measure the human creativity involved in text-based generative art.
- Using Rhodes's four P framework [39], we expound on the nature of human creativity involved in text-based generative art. The elaboration places special focus on the iterative and interactive practice of prompt engineering.
- We highlight image-level and portfolio-level curation as two important factors involved in the creative process of text-based generative art.
- Online communities and resources are another important factor affecting the human creativity of text-based generative art. We provide an outline of five different roles taken by members in the text-based generative art community.
- Last, we discuss the practical challenges of evaluating the creativity of text-based generative art and outline opportunities for future research.

The paper is structured as follows. After first discussing related work in Section 2, I provide details about my own background and experience as a form of self-disclosure in Section 3. I then discuss and illustrate a case in which a person creates a digital artwork in a non-creative way (Section 4). Motivated by this case, I argue that the human creativity of text-based generative art lies in the interaction of users with the AI-based system and the practices arising from this interaction, such as prompt engineering and curation. I expand on the nature of human creativity involved in text-based generative art in Section 5. I then discuss opportunities for future research on text-based generative art in the field of Human-Computer Interaction (HCI) and the broader implications of text-based co-creation with AI-based systems in Section 6. The paper concludes in Section 7.

2 RELATED WORK

2.1 Generative Art

Generative art is an "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" [14]. Generative art is a broad term that may encompass many different types of art [1]. One perspective, for instance, may define generative art as art created with computer code (e.g., with the language Processing) [1, 14]. This, of course, raises the question of "what is art?" Following Galanter [14], I do not seek an answer to this question in this paper and leave the definition of art as an open question surrounded by many timely issues of debate – such as whether human authorship and agency is required for a work to be protected under copyright [34].

This paper is particularly focused on *text-based* generative art. In this type of generative art, digital images are generated with artificial intelligence (AI) based on textual prompts in natural language. The paper leaves any advanced techniques, practices, and post-processing steps to future work. For instance, many text-based generative systems accept not only textual inputs, but additionally initial and target images to guide the generative system to produce images in a certain style or with a given scene composition. Further, practitioners of text-based generative art may use the resulting images only as first step in their creative process and, for instance, continue to manually edit images in Adobe Photoshop or employ GAN-based imagine inpainting techniques. Such advanced practices and techniques are not subject of this paper and left to future work. The subsequent section provides a technical overview of how text-based generative art is created with AI-based systems.

2.2 Text-based Generative Art

Generative art created with artificial intelligence (AI) has made significant advances since the inception of Generative Adversarial Networks (GANs) [15] in 2014 and Google's Deep Dream [30] in 2015. It was the release of OpenAI's CLIP [36] in January 2021 that spurred technical progress in text-based generative art. CLIP (Contrastive Language-Image Pretraining) is a machine learning method and model conceived for the task of visual classification. To this end, the CLIP neural network was trained on a large corpus of images and text from the World Wide Web. Due to the size of its training data, the CLIP model has developed a robust understanding of many concepts present in its training data. For instance, CLIP knows what a 'skyscraper' looks like because the model has encountered many buildings in different shapes and sizes (and the associated user-generated text labels) in the training data. CLIP has encoded this understanding in the latent space of its neural network. Given an image and a set of labels, CLIP can identify which labels best match the image. It is this ability to visually classify images and 'understand' natural language that resulted in CLIP being used in combination with generative adversarial networks (GANs). When used as discriminator in a generative system, CLIP guides the generator to produce digital images that best match a given textual prompt.

Shortly after OpenAI released CLIP in January 2021, AI enthusiasts created GAN+CLIP based systems for the specific purpose of text-based generative art. Ryan Murdoch created "The Big Sleep"

[5, 31], a combination of a GAN called BigGAN and CLIP. This inspired Katherine Crowson to connect an even more powerful neural network (VQGAN) with CLIP [8]. The combination of VQGAN and CLIP has become one of the popular techniques of text-based generative art. One can argue that VQGAN-CLIP was instrumental to advancing the emerging field of text-based generative art [8]. The source code of VQGAN-CLIP is available online, and many generative systems and methods have since been developed based on the pioneering work of Murdoch and Crowson. Today, practitioners can choose from a growing variety of systems for generating text-based generative art, such as CLIP guided diffusion [7], pixray [44], and pytti [27].

A growing number of systems are available as open source in notebooks on Google's Colaboratory where they can be executed free of charge and with high ease of use. Due to this low barrier of entry, the notebooks contribute to a democratization of digital art production – anybody can create digital images and artworks with text-based generative systems. This raises the question about the level and nature of human creativity involved in text-based generative art. This paper draws on Rhodes's conceptual model of creativity, as outlined in the following section.

2.3 The Four P of Creativity

Creativity is a complex multi-faceted phenomenon that is difficult to define [12]. One useful conceptual model for reflecting on the entirety of the concept of creativity is Rhodes's model of the "four P" of creativity: *person* (personality, intellect, habits, attitudes, etc.), *process* (thinking, motivation, learning, etc.), *press* (the human being's environment), and *product* ("artifacts of thought") [39].

All four aspects in Rhodes's conceptual model, however, are linked to an observable product [9]. The underlying assumption is that the embodiment of an idea – i.e., the outcome of a process, such as a painting, poem, sculpture, or an idea scribbled on a napkin – is a proxy for the creativity of a person. Following this perspective, the product can serve as a measure for human creativity. One popular operationalization of this measure in the scholarly literature is the two-part "standard" definition of creativity [35, 40]. Following this definition, an artifact needs to be both novel (original, unique, etc.) and useful (appropriate, effective, etc.) to be considered creative. According to this definition, it is therefore not enough for a product to be novel, it also needs to be appropriate, useful, or valuable for someone. In Kaufman and Beghetto's small-c paradigm [22], this someone is the creator of the creative artifact and from a Big-C perspective, it is a larger group of people or society as a whole.

3 SELF-DISCLOSURE AND ETHNOGRAPHIC RESEARCH

As a form of self-disclosure to contextualize this paper, I provide details about my own background and my research on text-based generative art in this section.

3.0.1 Researcher's background. I am a Computer Scientist and my background is in Human-Computer Interaction (HCI). My past research interest lies in **human creativity**, not computational creativity [6]. What interests me in text-based generative art is the interaction of humans with AI-based systems and the lived and internalized practices that may arise from repeated interactions with AI. While I have created digital images with text-based systems (as explained in the remainder of this section), I do not consider myself an artist and I pursue text-based generative art merely for intrinsic pleasure.

3.0.2 Ethnographic research. This work is grounded in my ethnographic research of the text-based generative art community. I discovered and learned about text-based generative art on Twitter in mid-2021 in the postings from Katherine Crowson (@Rivers HaveWings) and other users on Twitter. Since October 2021, I have been experimenting with text-based generative systems. I initially started to experiment with VQGAN-CLIP due to its performance, ease of use, and popularity (at the time) in the emerging community of text-based generative art. Later, I also experimented with CLIP guided diffusion and other systems available on Google Colab. For learning about writing effective prompts, I turned to Twitter. However, there is a natural limit to the learning curve on Twitter: since the full prompt is often not shared by artists and practitioners, one is left none-the-wiser how the digital artworks were created. The launch of the Midjourney[2] community proved to be a eureka moment in this regard. I was fortunate to be invited into the Midjourney beta program in mid-March 2022, and have since practiced and studied how prompt engineering for text-based generative art is conducted in practice.

Members of the Midjourney community generate images by typing prompts into Discourse-based chat rooms. The results are displayed shortly after in the chat. All members can see the images created by other members together with the respective prompts. Therefore, the prompts are shared in the community which makes the community a vast learning resource. Communities like Midjourney are a game-changer in advancing the practice of prompt engineering for text-based generative art. The following is based on the prompt writing practices that I observed, first on Twitter and later in the Midjourney community, and my own opinions learned from this experience.

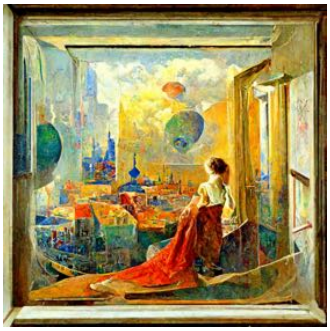
4 FEEDING RANDOM SNIPPETS OF TEXT TO THE MACHINE

In the context of text-based generative art, the product-centered view of creativity is not enough to fully assess the human creativity involved in creating a digital artwork. The whole "four P" of creativity are needed to capture the extent of the human creativity involved in text-based generative art. In the this section, I motivate and illustrate this point with a case in which the product-centered definition of creativity fails to capture the full extent of the creativity involved in text-based generative art.

Text-based generative systems have become exceedingly good at potentially turning any given textual input into a digital image with high aesthetic qualities, no matter what input is given. For instance, consider the following two scenarios:

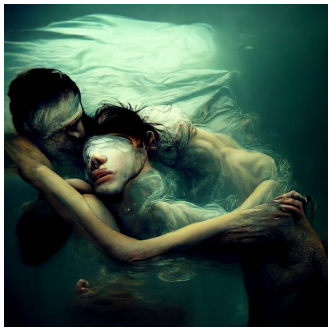
(1) A person takes a random snippet of text, such as words from an encyclopedia, and feeds the text to a text-based generative system.
(2) A person listens to music and types verbatim lyrics into a text-based generative system.

[2] https://www.midjourney.com

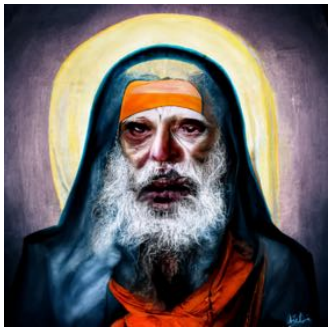


a) *What if everything around you isnt quite as it seems What if all the world you used to know Is an elaborate dream. Painting by Haralampi Oroschakoff, trending on wikiart* (lyrics from the song 'Right Where It Belongs' by Nine Inch Nails)

b) *When I see the man in the mirror I see an animal clearer* (lyrics from the song 'Hot' by CunninLynguists)



c) *Soaked in soul, he swims in my eyes by the bed* (lyrics from the song 'Wake Up Alone' by Amy Winehouse)



d) *How long will this take, baba? And how long have we been sleeping? Do you see me hanging on to every word you say? How soon will I be holy? How much will this cost, guru? How much longer till you completely absolve me?* (lyrics from the song 'Baba' by Alanis Morissette)

Figure 2: Examples of images generated from parts of music lyrics. All images were generated with Midjourney CLIP guided diffusion [28]. Figure a) demonstrates the use of additional prompt modifiers. Figures b) and c) were created without prompt modifiers.

In the latter case, the information passes through the person's perceptual system where it is, of course, subject to interpretation, misinterpretation, bias, and other conscious and unconscious processes. But assuming that the person correctly understands the music lyrics and merely transmits them to the text-based generative system, *is this creative?*

These two scenarios serve to illustrate that little to no human creativity may be involved in producing text-based generative art. The second scenario is comparable to Searle's "Chinese Room" [41]. In this philosophical argument and thought experiment, a person is locked in a room and, by following detailed instructions, produces sophisticated outputs without any specific skills or knowledge. Similarly in the case of text-based generative art, no creativity may be involved, besides some basic literacy skills and knowledge.

I have experimented with feeding music lyrics to text-based generative systems (VQGAN-CLIP, CLIP Guided Diffusion, and Midjourney). From this personal experience and observations made in the Midjourney community, I can tell that surprising, interesting, and high-quality text-based generative art can be produced from lyrics and random texts. Figures 2 and 3 illustrate this point with hand-picked examples of textual input prompts and their respective images. The images in Figure 2 were created from music lyrics. The images in Figure 3 were created with single characters (figures a, b, f, and g), single words (c and h), short quotes from movies (d and i) and emojis (e and j).

The figures exemplify that users do not (necessarily) need to stretch their imagination and exercise their creativity to produce results with high visual aesthetics and quality with state-of-the-art text-based generative systems. In this case, the interaction with the system is both skill-free (besides some basic literacy) and not creative. The product – i.e., the resulting digital image – is an imperfect proxy for the person's creativity. The case presented in this section serves as motivation to expand upon the human creativity of text-based generative art in the following section.

5 THE HUMAN CREATIVITY OF TEXT-BASED GENERATIVE ART

I argue that the creativity of text-based generative art is deeply rooted in the text-based *interaction* of human users with text-based generative systems and the human-computer co-creativity that arises from this interaction [21, 38]. It is the iterative practice of prompt engineering [17, 26, 37] that makes text-based generative art creative. In the following, Rhodes's 4-P model serves as a framework to expand upon the human creativity involved in text-based generative art.

5.1 Product: The Digital Image

Text-based generative art has made significant advances in 2021. State-of-the-art text-based generative systems are able to generate images of high aesthetic quality, especially if prompt modifiers are being used to boost the quality of the image (c.f. Figure 2 and [26]). From the product-based perspective of creativity, the resulting images are, of course, creative in themselves. But, as exemplified in Section 4, the images may not be the fruit of the human's creativity. Instead, the images are a result of computational creativity [6].

Today's text-based artworks are often not perfect, especially if human anatomy and faces are depicted. Additional post-processing steps are often necessary in these cases (e.g., to correct the eyes of human subjects). But it can be expected that we will, in the near future, no longer be able to tell text-based generative artworks apart from creations made by humans (e.g., see [32]). Therefore, text-based generative artworks (i.e., the "product" of AI-based systems) are an imperfect measure for the human creativity in this case. Instead of looking at the image as creative artifact, we need to turn

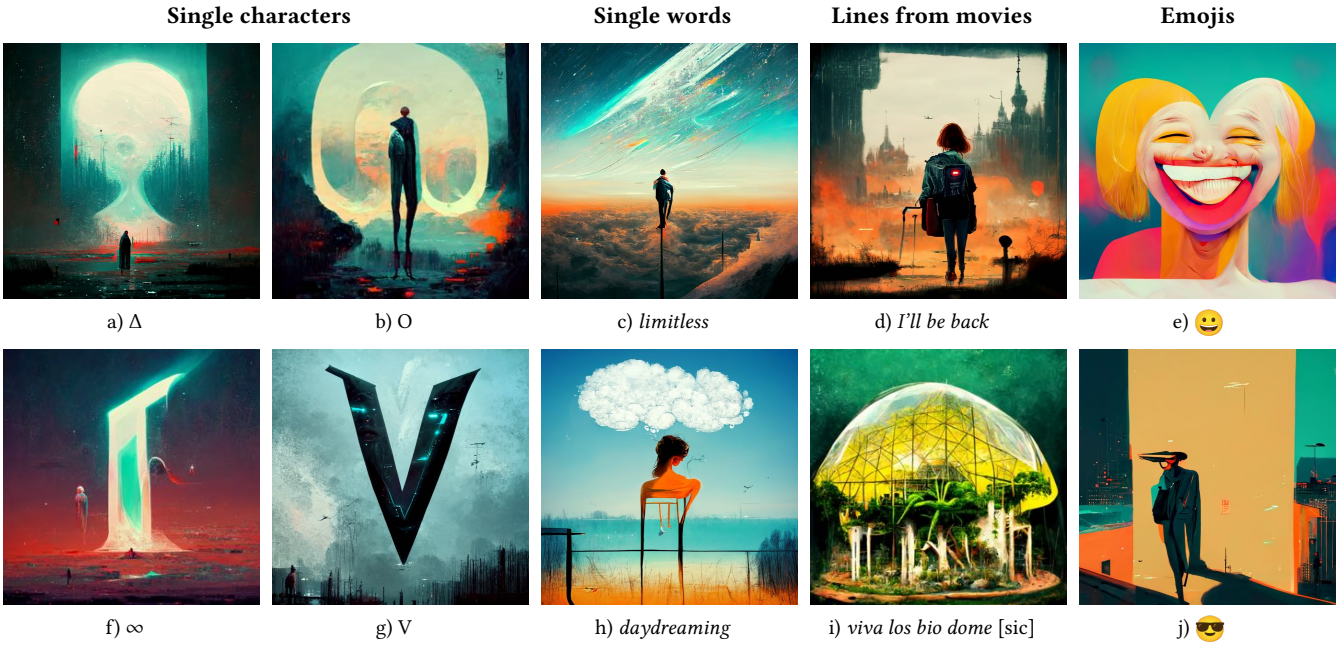


Figure 3: Examples of images generated with Midjourney CLIP guided diffusion without style modifiers [28].

to the other three P in Rhodes's model (person, process, and press) to assess the extent of human creativity involved in text-based generative art.

5.2 Person: The Practitioner

In my study of and participation in the Midjourney community, I observed some practitioners produce an abundance of creative, innovative, and exotic prompts, while others seem to struggle or feel inhibited. For instance, it can be observed that some novice practitioners attempt (and often fail) to coax images out of the text-based generative system with long and specific prompts, while others effortlessly produce beautiful images with rather minimalistic prompts. Writing effective prompts is a skill linked to a person's knowledge of the training set and the neural networks' latent space, but also the person's knowledge of and experience with prompt modifiers.

Together, this knowledge and the skills constitute the practice of prompt programming [37] or prompt engineering [17] – that is, the creative practice of writing textual input prompts in a certain format. This skill also goes beyond mere knowledge of text-based prompts. For instance, knowing which aspect ratio to choose for a specific prompt and knowledge of the system's configuration parameters is key to making the system produce coherent images. Prompt Engineering is a learned skill, because it is not immediately apparent how to write effective prompts and which keywords make good prompt modifiers. For instance, one popular prompt modifier used in the text-based generative art community is "trending on artstation." The community found that adding this modifier to a prompt will increase the quality of generated images. Many other prompt modifiers that modifiy the style and quality of generated images have since been discovered. The following section highlights the iterative nature of applying prompt modifiers to prompts.

5.3 Process: Iterative Prompt Engineering and Image Curation

5.3.1 Iterative prompt engineering. Text-based generative art is driven by exploration. Practitioners probe the model's latent space with textual prompts to see what works (and what not) [43]. This process is iterative [1] and links ideas between subsequent prompts [18]. Several iterations are often needed to arrive at a subjectively satisfactory result. The process is also affected by chance and the practitioner's willingness to "go with the flow" and allow the conversation with AI to steer in unanticipated directions. For instance, practitioners may adapt their prompts to serendipitously follow opportunities presented by the AI (thereby avoiding fixation [20]). Iteration is also inherently present in how images are created by the text-based generative system, as described in the following section.

5.3.2 Image-level and portfolio-level curation. Images created with text-based generative systems do not materialize in their final form. The systems are iterative and take the previous output as input for subsequent steps. Users can therefore visually observe how the images slowly emerge over time. In this regard, text-based generative art can be compared with the process of developing a Polaroid photograph or the physical processing of film material in a film laboratory (see Figure 4). At any point in this process, the user may decide that the result will not be satisfactory and abort the "development process." This is especially the case with GAN-based systems such as VQGAN-CLIP which can degenerate to sub-optimal solutions at some point in the image generation process [8]. Users therefore often have to choose an image from the set of generated

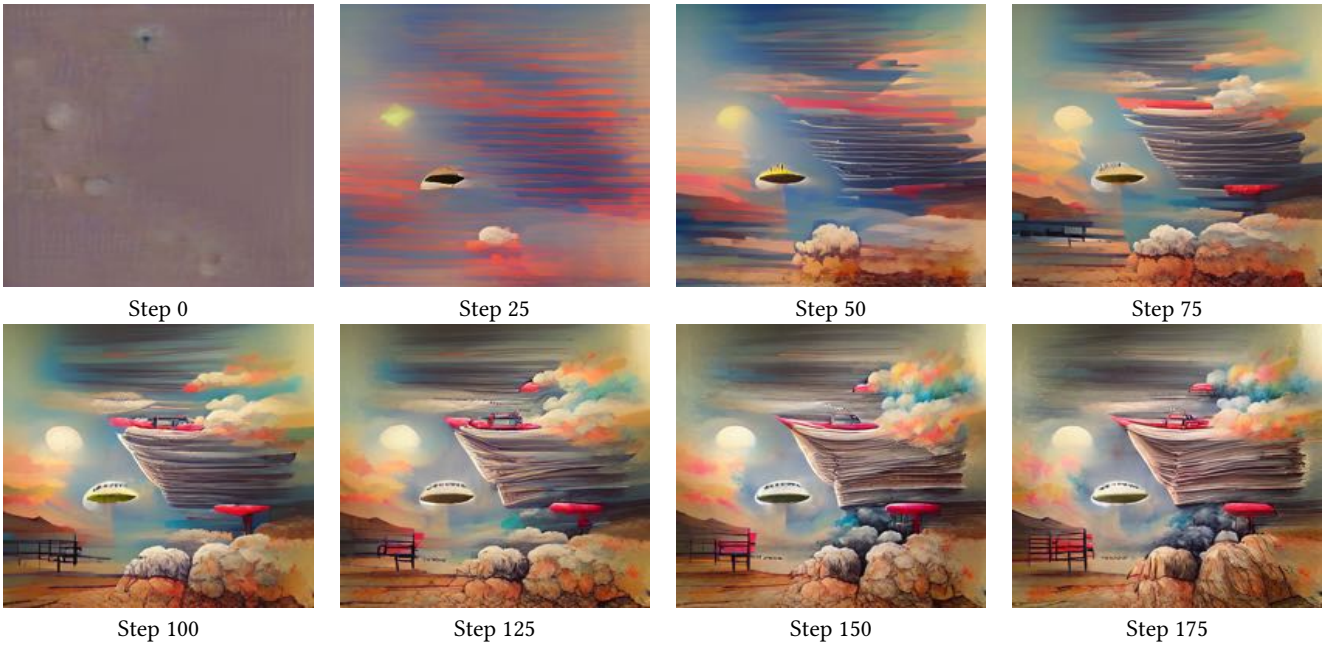


Figure 4: An example of image-level curation. The image was created with VQGAN-CLIP [8] with the prompt "ufo landing, fantastic, on Vellum, trending on /r/art." The image slowly emerged from the initial random noise used to seed the GAN (step 0). After 100 steps, a different style and details started to emerge and the author aborted the process after 175 steps. The author then selected step 50 as "best" (i.e., most interesting and artistic) image in the batch.

images, and the best image is not necessarily the one generated in the last generation step. The creativity of text-based generative art, therefore, is also a creativity of *image-level curation* within the wider process of image generation (see Figure 4).

Further, *portfolio curation* is an important constituent of the creativity of text-based generative art. Not every image is up to the standard of the person writing the prompt. Consequently, practitioners of text-based generative art carefully curate a portfolio of their best works, while discarding (or deciding not to publish) other images. The curative creative aspect of text-based generative art is similar to a photographer's process of weeding out and curating images from a larger collection of images. Midjourney's community has this curation process built-in. In a gallery[3], only the images that the Midjourney user chose to upscale are shown, thereby reflecting the user's curated works. Due to curation playing an important role in the image creation process, holistic metrics of curation [24] may be applicable to assess the creativity of the practitioner's portfolio curation.

5.4 Press: The Emerging Text-based Generative Art Ecosystem

An online ecosystem around text-based generative art is emerging. This ecosystem is enabled by the technologies described in Section 2.2 and includes online communities, learning resources, and tools and services. Together, these online resources contribute to and support the practitioners's creativity [4, 42].

[3]https://gallery.midjourney.com

5.4.1 Communities around text-based generative art. Dedicated online communities have emerged around text-based generative art. Members in these communities share their artworks (and sometimes the prompts). Online communities around text-based art include, for instance, Midjourney [28], Reddit's /r/MediaSynthesis, and Twitter.

The key roles of members in the emerging text-based generative art community can be described as follows:

(1) The community is enabled by *innovators and technology leaders*. For instance, Ryan Murdock and Katherine Crowson generously shared their software code as open source which has advanced the field of text-based generative art, as mentioned in Section 2.2.
(2) Not every innovator publishes their code on Google Colab. This is where *porters* step in and convert code from repositories on GitHub to executable notebooks on Google Colab. Since the field moves quickly, porters often refer to their notebooks as "rushed" or "dirty" implementations. This is also done as a way to not diminish the effort of the software code's original creator.
(3) *Conservators* seek to preserve the code created by innovators in, for instance, GitHub repositories.
(4) *Service and resource providers* create resources and applications using the open source code provided by innovators.
(5) *Practitioners* (enthusiasts, hobbyists, but also artists) use the notebooks provided by innovators and the systems provided by service providers to generate digital images and artworks.

The above roles in the community are fluid. An innovator, for instance, may briefly take the role of a porter to make a new technology available in a Colab notebook and advance the field.

Online communities act as a fertile ground for learning the skills necessary to use text-based generative systems creatively. One such skill is writing effective prompts or the practice of prompt engineering [17, 26] or prompt programming [37]. As mentioned in sections 5.2 and 5.3, prompt engineering is a skill learned with experimentation and by engaging with prior work [23]. Because enthusiastic creators share their prompts and practices, online communities are a fertile ground for learning from other users and for novice users to overcome the learning curve associated with prompt engineering for text-based generative art. But while online communities are useful resources, the distributed nature of messages in the communities make it difficult to pursue specific learning goals (such as learning about prompt modifiers). Consequently, practitioners have created resources for specific goals, as described in the following section.

5.4.2 Dedicated learning resources. Besides online communities, community members have created learning resources on prompt engineering. Some of these resources follow an approach similar to the systematic experimentation by Liu and Chilton [26]. For instance, practitioners have shared their experiments on artist names that may be used as style modifiers in input prompts. Such resources include, for instance, Remi Durant's artist studies (a list of artist names for generating images in a certain artistic style) [10] and Harmeet Gabha's list of Disco Diffusion modifiers [13].

Hub pages are another type of resource. Since not all Colab notebooks are indexed in search engines, hub pages were created in an effort to make the quickly growing field of text-based generative art discoverable. Hub pages act as indexes that provide collections of links to Colab notebooks. An example of a hub page is Lj Miranda's list of VQGAN-CLIP implementations [29].

5.4.3 Tools and services. A growing number of services and tools are emerging on the Web to support practitioners in creating text-based generative art. To name but a few, such services include Artbreeder[4] (a service that allows its users to fork other users's AI-generated artworks), Midjourney[5], NightCafe[6], Pollinations[7], Visions of Chaos[8] (a desktop software for experimenting with different deep learning models), WOMBO Dream[9] and Starry AI[10] (mobile applications for generating text-based artworks), conjure.art[11], and multimodal.art[12] (a user interface built on top of Google's Colab). This growing ecosystem of services and tools assists in making text-based generative art more accessible to non-technically minded users. The ecosystem also acts as a catalyst that draws in new practitioners and advances the field as a whole. It is the specific combination of people, technology, services, tools, and resources that formed a healthy ecosystem for the text-based generative art community to thrive.

[4]https://www.artbreeder.com
[5]https://www.midjourney.com
[6]https://creator.nightcafe.studio
[7]https://pollinations.ai
[8]https://softology.pro/voc.htm
[9]https://www.wombo.art
[10]https://www.starryai.com
[11]https://www.conjure.art
[12]https://multimodal.art

6 DISCUSSION

Art is subjective and, first and foremost, meaningful to its creator (small-c creativity). We humans enjoy pretty things and intrinsic motivation is one key driver to engage in the production of text-based generative art. However, some practitioners also pursue other interests – the money, fame, and glory of successful artists (Big-C perspective). Text-based generative art (or edits thereof) is being sold as NFT (non-fungible cryptographic tokens built on blockchain-based technology) in online marketplaces dedicated to digital art [25]. In combination with NFT technology, text-based generative art has, therefore, contributed to boosting the digital creative economy.

With text-based image synthesis systems, anybody can create art. This raises questions about the level and nature of the human creativity involved in text-based generative art. Melvin Rhodes wrote in 1961 that each "strand" of the four-part definition of creativity "has unique identity academically, but only in unity do the four strands operate functionally" [39]. By focusing on the product-centered operationalization of the concept of creativity, science seems to have lost track of this big picture. We obsess so much over p-values that we seem to have forgotten that human creativity can, as demonstrated in this paper, be deeply rooted in factors that are not covered with the "standard" product-based definition of creativity [40]. But not only is science's one-sided view a subject for concern, but there are also concrete challenges when assessing the creativity of text-based generative art.

6.1 Challenges for Evaluating the Creativity of Text-based Generative Art

Text-based generative art is the result of an opaque process. When evaluating the creativity of generative art, we run into three challenges (system, prompt, process) related to information asymmetries between the creator and the viewer of the artwork. However, for assessing the full extent of human creativity involved in text-based generative art, we need information on all three aspects.

6.1.1 System. Based on the digital image alone, we can infer little about the system used to generate the image and the system's configuration. Some text-based generative systems have dozens of parameters that can be adjusted to optimize the results. Clearly, knowing and adjusting the configuration parameters plays a crucial role in distinguishing skillful and purposeful mastery of text-based generative art from beginner's use with default settings.

6.1.2 Prompt(s). Many practitioners of text-based generative art do not share their prompts. Among these practitioners, prompts may be considered as a trade secret – especially among those creators who pursue commercial interests. Therefore, when we encounter an artwork posted on social media, we often do not know the prompt used to generate the image – or the creator shares the prompt (or parts thereof), but not the full set of prompt modifiers.

Further complicating the assessment of creativity is the fact that some systems can accept images as input prompts in addition, but not as substitute, to textual prompts. Text-based generative systems are inherently iterative by design. Images can therefore be used as

input for the system to seed the first iteration. Such initial images can, for instance, be used to direct the image's scene composition or to manipulate and distort existing images [8]. Further, some text-based generative systems accept one or several target images as input prompts. These visual prompts act like text prompts in that they provide a (visual) guide and a target for the system to optimize its losses. It is difficult and often impossible to tell what kind of inputs were used (i.e., text, images, or combinations thereof) to generate an image.

6.1.3 Process. From the digital image alone, we know little about the process used to create the image. For instance, we do not know how the creator came up with prompts. A text-based generative artwork could, as demonstrated in Section 4, be the result of a near skill-free interaction with the system, the creator could have gotten lucky on the first try, or the creator could have just copied someone else's prompt or prompt modifiers. However, the artwork can also be the result of an arduous iterative process – a focused conversation with the system – that constitutes the process of prompt engineering for text-based generative art.

6.2 Opportunities for Future Research on Text-based Generative Art

From the perspective of human-centered computing, the phenomena described in this paper are a "blend of technology, humans, and community" [16]. This section describes future research opportunities structured around these three aspects.

6.2.1 Technology – better understanding the user's intent. Practitioners of text-based generative art interact with generative systems via text-based prompts in natural language. Today, the systems's understanding of textual input prompts is far from perfect and the concepts learned by the neural network may not correspond to concepts in our visual world [43]. Natural language understanding is a part of the field of natural language processing (NLP) concerned with making machines understand human language. In the context of text-based generative art, advances in natural language understanding are needed to minimize the gap between the user's implicit intent, the explicitly stated textual prompt, and the visual output of the generative system.

6.2.2 Community – understanding the practice of prompt engineering and designing co-creative systems. Text-based generative art is among the first areas in which the practice of prompt engineering is being exercised "in the wild." Since text-based generative art involves natural language, lessons and practices from autoregressive language models, in particular GPT-3 [3, 37], could be applied to prompt engineering for text-based generative art. For instance, OpenAI maintains a collection of input templates for use with GPT-3 [33], and similar resources and tools are emerging in the text-based art communities. Some first design guidelines for prompt engineering in the context of generative models have been published [26], but there is still much to learn about the creative practices emerging in the communities around text-based generative art.

The interaction with text-based generative systems takes place through Web-based interfaces. Research on these user interfaces could therefore inform the design of novel co-creative systems [38] and creativity support tools [4, 42]. For instance, dedicated user interfaces are already emerging aiming at better supporting practitioners in creating text-based art (e.g., multimodal.art[13]).

6.2.3 Humans – co-creation and its effect on society. We are witness to an emergence of a new type of creator economy that requires no prior experience or skill – a democratization of art and creative production. Some say that this paradigm shift is of a magnitude even greater than the shift when photography was first introduced [19]. The role of the creator and our relation to images and creative work will change as a result, with implications for society as a whole.

AI-based systems, as an extension of our cognitive workbench, have the potential to augment and support our creativity [11, 42]. But our increasing interaction with and involvement in co-creative systems may also have an effect on our behavior, language, knowledge, and skills. Similar to how the World Wide Web changes society (as much as society changes the Web), interaction with AI in natural language will shape the future digital society, the way we interact with computers, and the way we work. We may, in the future, even see a bidirectional effect of prompt engineering on society. As we become more accustomed to phrasing our inputs for foundation-scale machine learning models [2], the way we use language and communicate with other humans may consequently change.

7 CONCLUSION

The product-centered view of creativity is not enough to fully assess the creativity of text-based generative art. The paper demonstrated this point with a scenario in which the product-centered definition of creativity fails to capture the full extent of the human creativity involved in text-based generative art. The paper argued that, in order to evaluate the full extent of human creativity involved in text-based generative art, we need to look not only at the digital images as outcome, but also the creator's process and environment. Drawing on Rhodes's conceptual "four P" model, this paper expanded on the nature of human creativity involved in the creation of text-based generative art. Image-level and portfolio-level curation, for instance, are ways in which creators express their creativity. The human creativity of text-based generative art lies in the interaction of users with text-to-image synthesis systems and the prompt engineering practices arising from the interactions.

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

[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 S. Chatterji, Annie S. Chen, Kathleen Creel, Jared Quincy Davis, Dorottya 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 D. 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 S. Krass, Ranjay Krishna, Rohith Kuditipudi, and et al. 2021. On the Opportunities and Risks of Foundation Models. *CoRR* abs/2108.07258 (2021), 212 pages. arXiv:2108.07258

[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

[13] https://multimodal.art

Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. https://doi.org/10.48550/ARXIV.2005.14165

[4] John Joon Young Chung, Shiqing He, and Eytan Adar. 2021. The Intersection of Users, Roles, Interactions, and Technologies in Creativity Support Tools. In *Designing Interactive Systems Conference 2021 (DIS '21)*. Association for Computing Machinery, New York, NY, USA, 1817–1833. https://doi.org/10.1145/3461778.3462050

[5] Simon Colton, Amy Smith, Sebastian Berns, Ryan Murdock, and Michael Cook. 2021. Generative Search Engines: Initial Experiments. In *Proceedings of the 12th International Conference on Computational Creativity (ICCC '21)*. 237–246.

[6] Simon Colton and Geraint A. Wiggins. 2012. Computational Creativity: The Final Frontier?. In *Proceedings of the 20th European Conference on Artificial Intelligence* (Montpellier, France) *(ECAI'12)*. IOS Press, NLD, 21–26.

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

[8] 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

[9] Igor N. Dubina and Suzanna J. Ramos. 2016. Creativity Through a Cultural Lens: The Dichotomy of "The West" and "The East". In *Creativity, Innovation, and Entrepreneurship Across Cultures: Theory and Practices*, Igor N. Dubina and Elias G. Carayannis (Eds.). Springer New York, New York, NY, 29–34. https://doi.org/10.1007/978-1-4939-3261-0_2

[10] Remi Durant. 2021. Artist Studies by @remi_durant. https://remidurant.com/artists/

[11] Douglas C. Engelbart. 1962. Augmenting Human Intellect: A Conceptual Framework. Air Force Office of Scientific Research, AFOSR-3233.

[12] Jonas Frich, Michael Mose Biskjaer, and Peter Dalsgaard. 2018. Twenty Years of Creativity Research in Human-Computer Interaction: Current State and Future Directions. In *Proceedings of the 2018 Designing Interactive Systems Conference* (Hong Kong, China) *(DIS '18)*. Association for Computing Machinery, New York, NY, USA, 1235–1257. https://doi.org/10.1145/3196709.3196732

[13] Harmeet Gabha. 2022. Disco Diffusion Modifiers. https://weirdwonderfulai.art/resources/disco-diffusion-modifiers/

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

[15] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative Adversarial Nets. In *Advances in Neural Information Processing Systems*, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger (Eds.), Vol. 27. Curran Associates, Inc.

[16] 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

[17] Gwern. 2020. GPT-3 Creative Fiction. https://www.gwern.net/GPT-3

[18] Lydia Paine Hagtvedt, Karyn Dossinger, Spencer H. Harrison, and Li Huang. 2019. Curiosity made the cat more creative: Specific curiosity as a driver of creativity. *Organizational Behavior and Human Decision Processes* 150 (2019), 1–13. https://doi.org/10.1016/j.obhdp.2018.10.007

[19] Aaron Hertzmann. 2018. Can Computers Create Art? *Arts* 7, 2 (2018), 25 pages. https://doi.org/10.3390/arts7020018

[20] David G. Jansson and Steven M. Smith. 1991. Design fixation. *Design Studies* 12, 1 (1991), 3–11. https://doi.org/10.1016/0142-694X(91)90003-F

[21] Anna Kantosalo, Prashanth Thattai Ravikumar, Kazjon Grace, and Tapio Takala. 2020. Modalities, Styles and Strategies: An Interaction Framework for Human-Computer Co-Creativity. In *Proceedings of the Eleventh International Conference on Computational Creativity (ICCC 2020)*. Association for Computational Creativity, 57–64.

[22] James C. Kaufman and Ronald A. Beghetto. 2009. Beyond Big and Little: The Four C Model of Creativity. *Review of General Psychology* 13, 1 (2009), 1–12. https://doi.org/10.1037/a0013688

[23] Andruid Kerne, Nic Lupfer, Rhema Linder, Yin Qu, Alyssa Valdez, Ajit Jain, Kade Keith, Matthew Carrasco, Jorge Vanegas, and Andrew Billingsley. 2017. Strategies of Free-Form Web Curation: Processes of Creative Engagement with Prior Work. In *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition* (Singapore, Singapore) *(C&C '17)*. Association for Computing Machinery, New York, NY, USA, 380–392. https://doi.org/10.1145/3059454.3059471

[24] Andruid Kerne, Andrew M. Webb, Steven M. Smith, Rhema Linder, Nic Lupfer, Yin Qu, Jon Moeller, and Sashikanth Damaraju. 2014. Using Metrics of Curation to Evaluate Information-Based Ideation. *ACM Trans. Comput.-Hum. Interact.* 21, 3, Article 14 (jun 2014), 48 pages. https://doi.org/10.1145/2591677

[25] Logan Kugler. 2021. Non-Fungible Tokens and the Future of Art. *Commun. ACM* 64, 9 (aug 2021), 19–20. https://doi.org/10.1145/3474355

[26] 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

[27] David Marx. 2021. PyTTI-Tools. https://github.com/pytti-tools

[28] Midjourney. 2022. Midjourney.com. https://www.midjourney.com

[29] Lj Miranda. 2021. List of VQGAN+CLIP Implementations. https://ljvmiranda921.github.io/notebook/2021/08/11/vqgan-list/

[30] Alexander Mordvintsev, Christopher Olah, and Mike Tyka. 2015. Inceptionism: Going Deeper into Neural Networks. https://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html

[31] Ryan Murdock and Phil Wang. 2021. Big Sleep. https://github.com/lucidrains/big-sleep

[32] Sophie J. Nightingale and Hany Farid. 2022. AI-synthesized faces are indistinguishable from real faces and more trustworthy. *Proceedings of the National Academy of Sciences* 119, 8 (2022), e2120481119. https://doi.org/10.1073/pnas.2120481119

[33] OpenAI. n.d.. Examples – OpenAI API. https://beta.openai.com/examples

[34] Shira Perlmutter, Suzanne Wilson, and Kimberley Isbell. 2022. Second Request for Reconsideration for Refusal to Register A Recent Entrance to Paradise (Correspondence ID 1-3ZPC6C3; SR # 1-7100387071).

[35] Jonathan A. Plucker, Ronald A. Beghetto, and Gayle T. Dow. 2004. Why Isn't Creativity More Important to Educational Psychologists? Potentials, Pitfalls, and Future Directions in Creativity Research. *Educational Psychologist* 39, 2 (2004), 83–96. https://doi.org/10.1207/s15326985ep3902_1

[36] 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. In *Proceedings of the 38th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 139)*, Marina Meila and Tong Zhang (Eds.). PMLR, 8748–8763.

[37] 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*. Association for Computing Machinery, New York, NY, USA, Article 314, 7 pages.

[38] Jeba Rezwana and Mary Lou Maher. 2022. Designing Creative AI Partners with COFI: A Framework for Modeling Interaction in Human-AI Co-Creative Systems. *ACM Trans. Comput.-Hum. Interact.* (feb 2022), 27 pages. https://doi.org/10.1145/3519026

[39] Mel Rhodes. 1961. An Analysis of Creativity. *The Phi Delta Kappan* 42, 7 (1961), 305–310.

[40] Mark A. Runco and Garrett J. Jaeger. 2012. The Standard Definition of Creativity. *Creativity Research Journal* 24, 1 (2012), 92–96. https://doi.org/10.1080/10400419.2012.650092

[41] John R. Searle. 1980. Minds, brains, and programs. *Behavioral and Brain Sciences* 3, 3 (1980), 417–424. https://doi.org/10.1017/S0140525X00005756

[42] Ben Shneiderman, Gerhard Fischer, Mary Czerwinski, Mitch Resnick, Brad Myers, Linda Candy, Ernest Edmonds, Mike Eisenberg, Elisa Giaccardi, Tom Hewett, Pamela Jennings, Bill Kules, Kumiyo Nakakoji, Jay Nunamaker, Randy Pausch, Ted Selker, Elisabeth Sylvan, and Michael Terry. 2006. Creativity Support Tools: Report From a U.S. National Science Foundation Sponsored Workshop. *International Journal of Human–Computer Interaction* 20, 2 (2006), 61–77. https://doi.org/10.1207/s15327590ijhc2002_1

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

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