

# The Creativity of Text-to-Image Generation

Jonas Oppenlaender

joppenlu@jyu.fi

University of Jyväskylä

Jyväskylä, Finland



Figure 1: Digital image generated from the prompt “art on the wall” with Midjourney [32].

## ABSTRACT

Text-guided synthesis of images has made a giant leap towards becoming a mainstream phenomenon. With text-to-image generation systems, anybody can create digital images and artworks. This provokes the question of whether text-to-image generation is creative. This paper expounds on the nature of human creativity involved in text-to-image art (so-called “AI art”) with a specific focus on the practice of *prompt engineering*. The paper argues that the current product-centered view of creativity falls short in the context of text-to-image generation. A case exemplifying this shortcoming is provided and the importance of online communities for the creative ecosystem of text-to-image art is highlighted. The paper provides a high-level summary of this online ecosystem drawing on Rhodes’ conceptual four P model of creativity. Challenges for evaluating the creativity of text-to-image generation and opportunities for research on text-to-image generation in the field of Human-Computer Interaction (HCI) are discussed.

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

*Academic Mindtrek 2022, November 16–18, 2022, Tampere, Finland*

© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-9955-5/22/11...\$15.00

<https://doi.org/10.1145/3569219.3569352>

## CCS CONCEPTS

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

## KEYWORDS

prompt engineering, text-to-image generation, text-guided image synthesis, creativity, Midjourney, AI art, generative art

## ACM Reference Format:

Jonas Oppenlaender. 2022. The Creativity of Text-to-Image Generation. In *25th International Academic Mindtrek conference (Academic Mindtrek 2022), November 16–18, 2022, Tampere, Finland*. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3569219.3569352>

## 1 INTRODUCTION

Text-to-image generation systems based on deep generative models have become popular means for creating digital images and artworks [8, 53]. Given an input prompt in natural language, these generative systems are able to synthesize digital images of high aesthetic quality. An ecosystem of tools and resources around text-to-image generation has emerged online and a wide variety of text-to-image generation systems are now available as open source in executable notebooks. Google’s Colaboratory<sup>1</sup> (Colab), in particular, is an online service that was instrumental to the growth and popularity of text-to-image generation due to Colab allowing execution of Python-based code free of charge. For this reason,

---

<sup>1</sup><https://colab.research.google.com>

Colab has become one of the main drivers of innovation in the enthusiastic text-to-image art community. This community consists of *practitioners* (i.e., novices, amateurs, semi-professionals, but also skilled artists and professionals) with or without technical expertise. Many practitioners share their digital artworks on social media and in dedicated communities, such as Midjourney [32]. Given the growing capabilities and ease-of-use of text-to-image generation systems, digital art synthesized with deep generative models is on the verge of becoming a mainstream phenomenon.

Common to text-to-image generation systems is that practitioners can create digital images and artworks with little to no understanding of the underlying technologies, simply by writing prompts in natural language (a practice called prompt engineering [20], prompt programming [48], prompt design [61], or prompting for short). Given the ease of use and emerging ubiquity of text-to-image generation, the question arises: **If anybody can produce digital images that resemble masterful pieces of art by simply feeding textual prompts into an opaque system, is text-to-image art creative? What is the nature of the human creativity involved in generating images with text-to-image synthesis?**

In this paper, I aim to answer these questions.

In the scholarly literature, creativity is often split into two distinct parts. According to this two-part definition of creativity, an artifact is creative if it is both original (sometimes labeled as novel, unusual, or unique) and effective (or useful, fitting, or appropriate) [54]. This definition relates to Csikszentmihalyi's sociocultural model of creativity in which only an idea or artifact that is deemed appropriate to the field is allowed to enter the domain [9]. The two-part definition of creativity has seen vast adoption in the scholarly literature because it enables researchers to operationalize the concept of creativity into two measurable parts. Given that measurement requires an observable product, the two-part definition is taking a product-centered view [11]. It is the outcome of the creative process (i.e., the observable artifact) that is used to determine if something is creative.

I argue that this product-centered paradigm of creativity is not enough to assess the creativity of the artifact in the case of text-to-image generation. This paper exemplifies a case in which the two-part operationalization of creativity fails to measure the full extent of 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' conceptual model of the "4 P" of creativity [50] provides a fitting framework for this investigation. The framework looks at creativity from four perspectives – product, person, process, and press (i.e., the environment). In the context of text-to-image generation, all four perspectives are necessary to capture the full extent of human creativity (with special emphasis on the latter three). Online communities, in particular, form an increasingly important social constituent of the online creative ecosystem around text-to-image generation.

This paper expounds on the nature of human creativity prevalent in the sub-culture of text-to-image art (sometimes referred to as "AI art" or "generative art" [1, 17] within the online community). When creating art from text, practitioners cede control (in part) to artificial intelligence (AI) [17]. Consequently, the human creativity in text-to-image synthesis 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-to-image generation. These novel creative practices are shaped and informed by a growing ecosystem of community-driven resources and tools. In particular, online communities and online resources are factors with growing importance influencing the human creativity involved in text-to-image generation.

This paper makes the following contributions:

- A case in which the popular product-based operationalization of creativity fails to measure the human creativity involved in text-to-image generation is exemplified.
- Rhodes' four P framework [50] is used to expound on the nature of human creativity involved in text-to-image generation. Special focus is placed on the iterative and interactive practice of "prompt engineering" and the online community of practitioners of this novel creative practice.
- Image-level and portfolio-level curation are highlighted as two important creative practices involved in the creative process of text-to-image generation.

Additionally, attention is pointed to the growing importance of communities in the emerging ecosystem of text-to-image generation as a catalyst for creativity and learning. Online communities and resources are important factors affecting the human creativity of text-to-image art. An outline of five different roles taken by members in the AI art community is provided. Finally, the practical challenges of evaluating the creativity of images synthesized with text-to-image generation systems are discussed and an outline of opportunities for future research is provided.

The paper is structured as follows. After first discussing related work (Section 2) and the methodology (Section 3), I 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-to-image synthesis lies in the interaction of practitioners with the deep learning based system and the creative practices arising from this interaction, such as prompt engineering and curation. I expand on the nature of human creativity involved in text-to-image generation in Section 5. I then discuss opportunities for future research on text-to-image generation 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 Text-to-Image Generation

The text-guided synthesis of images using deep learning has made significant advances since the inception of Generative Adversarial Networks (GANs) [18] in 2014 and Google's Deep Dream [34] in 2015. It was the release of OpenAI's CLIP [46] in January 2021 that spurred immense technical progress in text-to-image generation. CLIP is a contrastive language-vision model conceived for the task of classifying images by providing the names of the visual categories to be recognized. CLIP was trained on a large corpus of images and text from the World Wide Web. Due to the Web-scale size of its training data, the CLIP model has learned a wide variety of visual concepts from natural language supervision and can associate

them with their names. Given an image and a set of labels (e.g., “a photo of a dog”), CLIP can predict the label that is most likely to be paired with the image. It is this ability to visually associate natural language with images that resulted in CLIP being used in generative systems.

In the context of text-to-image generation, CLIP found its first significant application in GAN-based image generation systems. When used as a discriminator component in a generative deep learning architecture, CLIP can guide the generator component 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 generating digital art. Ryan Murdoch created “Big Sleep” [5, 36], 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 was very popular in 2021 and became one of the standard techniques of generating artworks until the method was superseded by diffusion-based systems [10]. One can argue that VQGAN-CLIP was instrumental to advancing the emerging field of text-to-image art [8]. The source code of VQGAN-CLIP was 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-to-image art. The current state of the art are diffusion models, such as CLIP guided diffusion [7] and latent diffusion [52].

A growing number of generative systems are available as open source in Jupyter notebooks on Google’s Colaboratory<sup>2</sup> (Colab), on Huggingface<sup>3</sup>, and other places online where they can be downloaded or executed free of charge and with relatively high ease of use. Due to this low barrier of entry, Colab notebooks contribute to a democratization of digital art production – anybody can create digital images and artworks with text-to-image generation systems. This raises the question about the level and nature of human creativity involved in text-based generative art (or “AI art” as it is often called within the community of practitioners). This paper draws on Rhodes’ model of creativity as an analytical framework to answer this question, as outlined in the following section.

## 2.2 The Four P of Creativity

Creativity is a complex multi-faceted phenomenon that is difficult to define [40]. In the field of Human-Computer Interaction (HCI), a multitude of definitions of creativity are in use which has contributed to a “fuzziness” around the concept of creativity in the scholarly literature [14].

One important model of creativity is Csikszentmihalyi’s “systems model” of creativity [9]. The model consists of three parts: the individual person having an idea or creating an artifact, the domain (embedded in the broader context of culture), and the field which consists of individuals (gatekeepers) assessing the creativity of the idea or artifact before it enters the domain. According to this sociocultural model of creativity, “social confirmation is necessary for something to be called creative” because “creativity does not happen inside people’s heads, but in the interaction between

a person’s thoughts and a sociocultural context” [9]. Contrary to the above, one can argue that creativity can also take place on a personal level when an individual creates something personally valuable and meaningful (so-called small-c creativity [26] or everyday creativity [51]). This is opposed to Big-C (or eminent) creativity which requires a large contribution to society for something to be considered creative (e.g., a Nobel-prize winning discovery) [26, 51].

The above model and definition of creativity are coarse and do not capture the multitude of activities that lead to the creation of a creative artifact. One useful conceptual model for reflecting on the entirety of the concept of creativity is Rhodes’ model of the “four P” of creativity [50]. The model aims to explain what it means to be creative by looking at four components. *Person* pertains to a human beings’ personality, intellect, habits, attitudes, and other factors that affect a person’s creativity. *Process* pertains to the process of creating ideas and artifacts (e.g., thinking, motivation, learning, etc.). *Press* refers to the human being’s environment which can influence the person and the mental processes. Last, *product* pertains to “artifacts of thought” – ideas that have been communicated to other people in observable form (e.g., as a poem, painting, or sculpture) [50].

As in the above sociocultural model and the two-part definition of creativity, all four aspects in Rhodes’ conceptual model are linked to an observable product [11]. 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 observable 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 [45, 54]. Following this definition, an artifact needs to be both novel (original, unique, etc.) and effective (appropriate, valuable, useful, 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. Following the Big-C perspective of creativity, this someone is a larger group of people or society as a whole (the gatekeepers in the field), and from a small-c perspective, it is the creator of the creative artifact [26].

## 3 METHOD

As self-disclosure to contextualize this paper, I first provide details about my research on text-to-image generation and my background before I describe the research method.

### 3.1 Researcher’s background

I am a Computer Scientist with a background in Human-Computer Interaction (HCI) and Human-Centered Computing (HCC). My past research interest lies in **human creativity**, not computational creativity [6]. What interests me in text-to-image generation is the interaction of humans with artificial intelligence (AI) and the lived and internalized practices that may arise from repeated interactions with AI. While I have created several thousands of images with text-to-image generation systems (as explained in the remainder of this section), I do not consider myself an artist and I pursue text-to-image generation merely for intrinsic pleasure.

<sup>2</sup><https://colab.research.google.com>

<sup>3</sup><https://huggingface.co>

### 3.2 Online Ethnography

This work is grounded in an online ethnography [29] of the text-to-image art community conducted between October 2021 and March 2022. I discovered and learned about text-to-image art in mid-2021 in the postings from Katherine Crowson (@RiversHaveWings) and others on Twitter. Since October 2021, I have been experimenting with text-to-image generation systems. I initially started to experiment with VQGAN-CLIP due to its performance, ease of use, and popularity (at the time). 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 practitioners on Twitter, one is left none-the-wiser how the digital artworks were created. The launch of the Midjourney<sup>4</sup> community proved to be a eureka moment in this regard. Members of the Midjourney community generate images by typing prompts into Discourse-based chat rooms. The images are displayed shortly after in the chat. 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 social learning resource, unlocking the social creativity of what previously was a lone-creator practice. Communities like Midjourney are a game-changer in advancing the practice of prompt engineering for text-to-image generation. I was invited into the Midjourney beta program in mid-March 2022, and have since practiced and studied how prompt engineering for text-to-image art is conducted in practice.

Drawing on Rhodes' multiperspective model of creativity, my research seeks to develop a broad overview of the emerging creative ecosystem of text-to-image synthesis and text-based generative art (AI art) to find answers to the research questions mentioned in the introduction. This ecosystem consists of online resources and online communities with many participants. There are thus many factors affecting the creativity of practitioners in these communities. Given this complexity, this work takes a pluralistic research approach, combining different qualitative methods. Following a participant-observational approach [29], online communities (Twitter and Midjourney) were studied. Part of this work is based on the prompt writing practices that I observed, first on Twitter and later in the Midjourney community. Additionally, resources emerging from the online community (e.g., guides and other helpful resources created by practitioners) were studied and reviewed. The review also involved both scholarly literature, particular on prompt engineering and text-to-image generation, and gray literature (e.g., blog posts and news articles). Data was gathered in field notes which were iteratively revisited and extended.

## 4 FEEDING RANDOM SNIPPETS OF TEXT TO THE MACHINE

In the context of text-to-image synthesis, the product-centered definition 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 and social creativity involved in text-to-image generation. In 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-to-image generation.

Text-to-image generation systems have become exceedingly good at potentially turning any given textual input into high-fidelity images, 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 words to a text-to-image generation system.
- (2) A person listens to music and types verbatim lyrics into a text-to-image generation system.

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-to-image generation system, *is this creative?*

These two scenarios serve to illustrate that little to no human creativity may be involved in producing art with text-to-image synthesis. The second scenario is comparable to Searle's thought experiment of the “Chinese Room” [55]. In this philosophical argument, a person receives a task while being locked in a room with no access to information from the outside world and, by following detailed instructions, produces sophisticated outputs without having any specific skills or knowledge. Similarly in the case of text-to-image generation, little or no creativity may be involved besides some basic literacy skills.

I have extensively experimented with entering music lyrics into text-to-image generation systems (VQGAN-CLIP, CLIP Guided Diffusion, and Midjourney). From this personal experience and observations in the Midjourney community, I can tell that surprising, interesting, and high-quality artworks can be produced from lyrics and random textual inputs. 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 practitioners do not necessarily need to stretch their imagination and exercise their creativity to produce high-quality results with state-of-the-art text-to-image generation systems. In this case, the interaction with the system is both skill-free (besides some basic literacy) and not creative. The end product – i.e., the generated digital image – is an imperfect proxy for the person's creativity. Much of the recent criticism of text-to-image synthesis voiced by artists (e.g., [59]) can be predicated on this point. Text-to-image synthesis allows hobbyists to create digital artworks in seconds as compared to dozens or even hundreds of hours using traditional methods. The case presented in this section serves as motivation to expand upon the human creativity involved in text-to-image generation in the following section.

<sup>4</sup><https://www.midjourney.com>



*What if everything around you isn't 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)



*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)



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



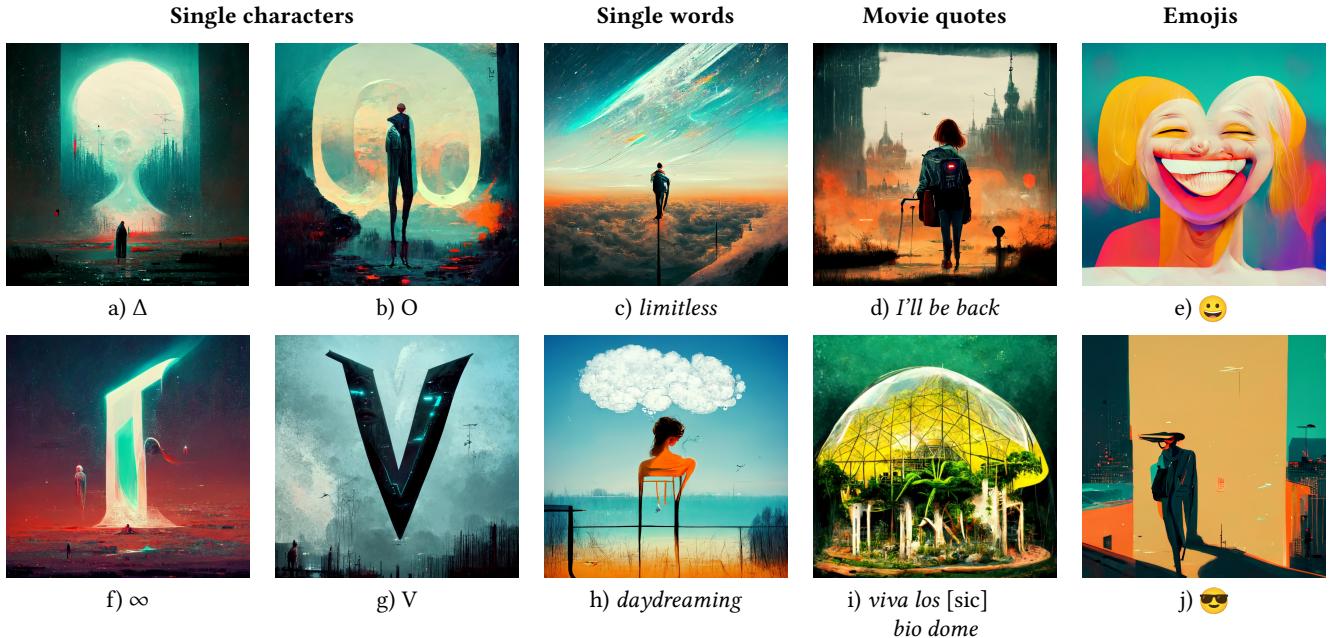
*Soaked in soul, he swims in my eyes by the bed*  
(lyrics from the song 'Wake Up Alone' by Amy Winehouse)

Figure 2: Examples of images generated from parts of music lyrics with Midjourney's CLIP guided diffusion [32]. Figure a) demonstrates the use of additional prompt modifiers. Figures b) and c) were created without prompt modifiers.

## 5 THE HUMAN CREATIVITY OF TEXT-TO-IMAGE GENERATION

I argue that the creativity of text-to-image generation arises from the text-based interaction of human users with text-to-image generation systems and the human-computer co-creativity that is determined by the constellation of user and AI [25, 49]. Increasingly,

online communities are also becoming determining factors in how users find inspiration and learn the craft of prompt engineering. It is this increasingly social practice of prompt engineering that makes text-to-image generation creative. In the following, Rhodes' 4 P model serves as a framework to expand upon the human creativity involved in text-to-image generation.



**Figure 3:** Examples of images generated with Midjourney CLIP guided diffusion without style modifiers [32].

### 5.1 Product: The Digital Image

Text-to-image generation with the aim of synthesizing digital art has made significant advances in 2021. The state-of-the-art of text-to-image generation systems are able to synthesize images of high aesthetic quality, especially if prompt modifiers are being used to alter the style or boost the quality of the image (c.f. Figure 2 and [31, 41]). 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 human creativity. Instead, the images are a result of computational creativity [6].

Today's artworks created with text-to-image synthesis are still 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-to-image artworks apart from creations made by humans [37]. Therefore, text-to-image artworks (i.e., the “product” of text-to-image generation systems) are an imperfect measure for human creativity. Instead of looking at the image as creative artifact, we need to turn to the other three P in Rhodes’ model (person, process, and press) to assess the extent of human creativity involved in text-to-image generation.

### 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-to-image generation system with very long and specific prompts,

while others effortlessly produce beautiful images with rather minimalist 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 different prompt modifiers [41]. Together, this knowledge and the skills constitute the practice of “prompt engineering” [20, 31] – that is, the creative practice of writing effective textual input prompts for text-to-image generation systems. This skill also goes beyond mere knowledge of textual prompts. For instance, knowledge of which aspect ratio to choose for a specific subject and an understanding of the system’s training data and configuration parameters is key to produce high-fidelity 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 [41]. For instance, one popular prompt modifier used in the text-to-image art community is “trending on artstation.” The community of practitioners found that adding this modifier to a prompt will increase the quality of generated images. Many other prompt modifiers that modify the style and quality of generated images have since been discovered and are regular being applied in the community. The following section highlights the iterative nature of applying prompt modifiers to prompts.

### 5.3 Process: Iterative Prompt Engineering and Image Curation

The creative process is vast and may involve a multitude of different workflows, practices, and tools. The current view of image generation systems, as, for instance, held by Emad Mostaque (the founder of Stability.ai) is that text-to-image generation models are only the first part in a creative pipeline that extends over multiple AI-based

systems and applications [35]. A practitioner may, for instance, generate a first sketch with Stable Diffusion [52], run the output through a model that was trained on improving faces (such as GFP-GAN [60]), and then proceed to perform more advanced techniques (e.g., inpainting and outpainting) before finally editing the result in a graphics editor. In this section, I want to focus on two parts of the creative process that I deem to be native to text-to-image generation: iteration and curation.

**5.3.1 Iterative prompt engineering.** Text-to-image art is driven by exploration. Practitioners probe the model’s latent space with textual prompts to see what works (and what not) [57]. This process is iterative [1] and links ideas between subsequent prompts [21]. 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 the AI to steer in unanticipated directions. For instance, practitioners may adapt their prompts to serendipitously follow opportunities presented by the AI (thereby avoiding fixation [24]). Iteration is also inherently present in how images are created by AI, as described in the following section.

**5.3.2 Image-level and portfolio-level curation.** Images created with text-to-image synthesis do not materialize in their final form. Instead, text-to-image generation systems are iterative and take the previous output as input for the subsequent step. One can therefore visually observe how the images slowly emerge over time. In this regard, text-to-image synthesis can be compared with the process of developing a Polaroid photograph and the physical processing of film material in a film laboratory (see Figure 4). At any point in this process, the practitioner may decide that the result will not be satisfactory and abort the “development process.” This is 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]. Practitioners therefore often have to select a single image from the set of generated images, and the best image is not necessarily the one generated in the last generation step. The creativity of text-to-image art, therefore, is also a creativity of *image-level curation* within the wider process of text-to-image generation (see Figure 4).

Further, *portfolio curation* is an important constituent of the creativity of text-to-image art. Not every generated image is up to the standard of the person writing the prompt. Consequently, practitioners of text-to-image art carefully curate a portfolio of their best works, while discarding (or deciding not to publish) other images. The curative creative aspect of text-to-image 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 an online gallery<sup>5</sup>, only the images that the Midjourney member chose to upscale are shown by default, thereby reflecting the member’s curated works. Due to curation playing an important role in the image creation process, holistic metrics of curation [28] may be applicable to assess the creativity of the practitioner’s portfolio curation.

<sup>5</sup><https://www.midjourney.com/app/>

## 5.4 Press: The Emerging Text-to-Image Ecosystem

Besides the recent leap in the technical development of text-to-image generation systems, online communities dedicated to text-to-image art are one of the most significant developments. A lively online ecosystem around text-to-image art has emerged. This ecosystem is enabled by the technologies described in Section 2.1 and includes online communities on both general social media platforms (e.g., Twitter and Reddit) and in dedicated online communities. Further, the creative ecosystem includes learning resources, tools, and online services. Together, these online communities and resources contribute to and support the practitioners’ creativity [4, 56].

**5.4.1 Communities dedicated to text-to-image art.** Dedicated online communities have emerged around text-based generative art. Members in these communities share their prompts and artworks (or even generate images in the community itself). These online communities include, for instance, Reddit’s /r/MediaSynthesis, Twitter, Midjourney [32], and many other Discord-based communities. Discord, in particular, has proven useful to the formation of communities as it allows to quickly spin up a chat-based community around a certain topic. The key roles of members in the broader text-to-image online community can be described as follows:

- (1) The community is enabled by *innovators* who introduce novel techniques and technologies to the community. For instance, Ryan Murdock and Katherine Crowson generously shared their software code as open source which has significantly advanced the field of text-to-image generation, as mentioned in Section 2.1.
- (2) Not every innovator publishes their code in Colab notebooks. This is where *porters* step in and convert code from repositories on GitHub to executable notebooks on Colab. Since the field moves quickly, porters often refer to their notebooks as “quick,” “rushed,” or “dirty” implementations 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 and tools provided by innovators and the systems provided by service providers to generate digital images and artworks.

The above roles in the community are fluid. For instance, an innovator 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-to-image generation systems creatively. One such skill is the creative practice of “prompt engineering” [20, 31, 41]. As mentioned in sections 5.2 and 5.3, prompt engineering is a skill learned with experimentation and by engaging with prior work [27]. Because enthusiastic creators share their prompts and practices, online communities are a fertile ground for novice practitioners to learn from other members of the community in order to overcome the learning curve associated with prompt engineering. But while online communities act as catalysts



**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, unwanted 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.**

for creativity, the distributed nature and sequential presentation of messages in the communities make it difficult to pursue specific learning goals (such as learning about prompt modifiers). Consequently, practitioners and online communities increasingly strive to create dedicated resources that better support their members, as described in the following section.

**5.4.2 Dedicated learning resources.** The Midjourney community is on its way to becoming a cultural hub due to its attention to features that benefit the community as a whole. Midjourney has, for instance, introduced member profiles, exploratory maps, and information on style modifiers that help its members to explore and learn the craft of prompt engineering. Besides online communities, community members have created learning resources on prompt engineering. For instance, practitioners have created a number of resources for specific goals, such as style experimentation (e.g., [12, 15]) and teaching prompt engineering (e.g., the DALL-E prompt book [43]). Some of these resources follow an approach similar to the systematic experimentation by Liu and Chilton [31]. 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) [12] and Harmeet Gabha’s list of Disco Diffusion modifiers [16]. 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-to-image generation discoverable. Hub pages act as indexes that provide collections of links

to Colab notebooks. Examples of hub page are Lj Miranda’s list of VQGAN-CLIP implementations [33], pharmapsychotic’s “Tools and Resources for AI Art” [44], and Reddit user Wiskkey’s “List of Stable Diffusion systems” [47].

**5.4.3 Tools and services.** A growing number of services and tools are emerging on the Web to support practitioners in creating text-to-image art. Besides the aforementioned Midjourney community, which has amassed a membership approaching 1 million community members [23], such services include Artbreeder<sup>6</sup> (a service that allows its users to fork other users’ AI-generated artworks), NightCafe<sup>7</sup>, Pollinations<sup>8</sup>, Visions of Chaos<sup>9</sup> (a desktop software for experimenting with different deep learning models), WOMBO Dream<sup>10</sup> and Starry AI<sup>11</sup> (mobile applications for generating text-to-image art), and multimodal.art<sup>12</sup> (a user interface built on top of Google’s Colab). Tools include, for instance, CLIP retrieval<sup>13</sup> that helps to find out how the CLIP model understands a text prompt, and CLIP prefix captioning<sup>14</sup>, a model that applies CLIP for captioning images with the aim of helping to produce better prompts from this initial image. The growing ecosystem of services and tools assists in making text-to-image generation more accessible to

<sup>6</sup><https://www.artbreeder.com>

<sup>7</sup><https://creator.nightcafe.studio>

<sup>8</sup><https://pollinations.ai>

<sup>9</sup><https://softology.pro/voc.htm>

<sup>10</sup><https://www.wombo.art>

<sup>11</sup><https://www.starryai.com>

<sup>12</sup><https://multimodal.art>

<sup>13</sup><https://rom1504.github.io/clip-retrieval>

<sup>14</sup>[https://github.com/rmokady/CLIP\\_prefix\\_caption](https://github.com/rmokady/CLIP_prefix_caption)

non-technically minded practitioners. 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-to-image art community to thrive.

## 6 DISCUSSION

Art is subjective and, first and foremost, meaningful to its creator (according to the small-c paradigm of creativity). We humans enjoy pretty things and intrinsic motivation is one key driver to engage in the production of text-to-image art. However, some practitioners also pursue other interests – the money, fame, and glory of successful artists (Big-C perspective). Artworks synthesized with text-to-image generation systems (or edits thereof) are being sold as NFT (non-fungible cryptographic tokens built on blockchain-based technology) in online marketplaces dedicated to digital art [30]. In combination with NFT technology, text-to-image art has, therefore, contributed to boosting the digital creative economy.

With text-to-image generation systems, anybody can create art. The easy access to this technology makes it possible for anybody to sell artworks, for instance, as NFTs. This raises questions about the level and nature of the human creativity involved in text-to-image synthesis. Melvin Rhodes wrote in 1961 that each “strand” of the four-part model of creativity “has unique identity academically, but only in unity do the four strands operate functionally” [50]. 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 by the “standard” product-based definition of creativity [54]. 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-to-image art.

### 6.1 Challenges for Evaluating the Creativity of Text-to-Image Art

Text-to-image art is the result of an opaque process. When evaluating the creativity of text-to-image art, we run into three challenges (system, prompts, and process) related to information asymmetries between the creator and the viewer of the artwork.

**6.1.1 System.** Based on the digital image alone, we can infer little about the generative system used to synthesize the image. Some text-to-image generation systems have dozens of configuration parameters that can individually 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-to-image art from beginner’s use with default settings.

**6.1.2 Prompts.** 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 practitioners 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. In other cases, the creator may have shared the prompt (or parts thereof), but not the full set of prompt modifiers. Prompt modifiers are keywords that determine the style

and quality of the generated image [41]. Knowing the full set of prompt modifiers is therefore crucial for understanding how the image was generated.

Further complicating the assessment of creativity is the fact that some systems can accept images as input prompts in addition – or as substitute – to textual prompts. Text-to-image generation systems are inherently iterative by design. Images can therefore be used as initial 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-to-image generation systems accept one or several target images as 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 practitioner came up with the prompt. A text-to-image artwork could, as demonstrated in Section 4, be the result of a near skill-free interaction with the text-to-image generation system, the practitioner could have gotten lucky on the first try, or the practitioner could have just copied someone else’s prompt or prompt modifiers. On the other hand, the artwork can also be the result of an arduous iterative process – a focused conversation with the AI system – that constitutes the process of prompt engineering for text-to-image generation.

For assessing the full extent of human creativity involved in text-to-image art, we need information on all three of the above aspects.

### 6.2 Opportunities for Future Research on Text-to-Image Generation

From the perspective of Human-Centered Computing, the phenomena described in this paper are a “blend of technology, humans, and community” [19]. This section describes future research opportunities structured around these three aspects.

**6.2.1 Technology – better understanding the user’s intent.** Practitioners interact with text-to-image generation systems via textual prompts. Today, the systems’ 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 [57]. Natural language understanding (NLU) is a part of the field of natural language processing (NLP) concerned with making machines understand human language. In the context of text-to-image generation, 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 text-to-image generation system.

**6.2.2 Community – understanding the practice of prompt engineering and the design of co-creative systems.** Text-guided generative art is among the first areas in which the practice of prompt engineering is being exercised “in the wild.” Since text-to-image generation involves natural language, practices and lessons learned from large-scale language models (LLMs) – and in particular GPT-3 [3, 48] –

could be applied to prompt engineering for text-to-image synthesis. For instance, OpenAI maintains a collection of input templates for use with GPT-3 [38] and similar resources and tools are emerging in the text-to-image art community. Some first design guidelines for prompt engineering in the context of text-to-image generation have been published [31], but there is still much to learn about the creative practices emerging in the communities around text-to-image art and how systems can be designed to support practitioners. For instance, the interaction with text-to-image generative systems takes place through web-based interfaces. Research on these user interfaces could inform the design of novel co-creative systems [49] and creativity support tools [4, 39, 42, 56]. For instance, dedicated user interfaces have been developed aiming at better supporting practitioners in creating text-to-image art (e.g., the MindsEye interface provided by multimodal.art<sup>15</sup>).

**6.2.3 Humans – co-creation and its effect on society.** We are witness to an emergence of a new type of creative 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 [22]. Back then, artists saw their livelihood being threatened by photography and – in fear of photography superseding the art of painting – ridiculed photography as not being art [58]. In hindsight, photography unlocked a vast amount of creativity that was traditionally corralled and coveted by painters. While some artists were negatively affected (especially those who specialized in producing photo-realistic portraits), the invention of photography opened opportunities to many more. It is likely that generative AI will lead to a similar singularity event in our human history. This change could deeply affect the role of the creator and our relation to images and creative work could change as a result, with implications for society as a whole.

AI-based systems, as an extension of our human cognitive workbench, have the potential to augment and support our creativity [13, 56]. 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 as well.

## 7 CONCLUSION

The product-centered view of creativity is not enough to fully assess the creativity of art synthesized with text-to-image generation systems. This work 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-to-image generation. The paper argued that in order to evaluate the full extent of human creativity involved in text-to-image generation,

we need to look not only at the digital images as outcome, but also the creator's process and environment. Drawing on Rhodes' conceptual "four P" model, this paper expanded on the nature of human creativity involved in text-to-image generation. Image-level and portfolio-level curation, for instance, are ways in which creators express their creativity. The human creativity of text-to-image art lies in the interaction with text-to-image generation systems and in online communities. Increasingly, online communities act as a catalyst for learning the craft of "prompt engineering." This craft, practiced by members of a growing online community of text-to-image practitioners, offers an exciting area of research in HCI.

## 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/10.1080/146260902867915>
- [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, Dorotya 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, Tatsumi Hashimoto, Peter Henderson, John Hewitt, Daniel E. Ho, Jenny Hong, Kyle Hsu, Jing Huang, Thomas Icard, Saahil Jain, Dan Jurafsky, Pratyusha Kalluri, Siddharth Karamchetti, 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 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 Murdoch, and Michael Cook. 2021. Generative Search Engines: Initial Experiments. In *Proceedings of the 12th International Conference on Computational Creativity (ICCC '21)*. Association for Computational Creativity, 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](https://colab.research.google.com/drive/12a_Wrfi2_gwwAuN3VvMTwVMz9TfqctNj)
- [8] Katherine Crowson, Stella Biderman, Daniel Kornis, Dashell 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] Mihaly Csikszentmihalyi. 1996. *Creativity: Flow and the Psychology of Discovery and Invention*. HarperCollins Publishers, New York, NY, USA.
- [10] Prafulla Dhariwal and Alexander Nichol. 2021. Diffusion Models Beat GANs on Image Synthesis. In *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (Eds.), Vol. 34. Curran Associates, Inc., 8780–8794. <https://proceedings.neurips.cc/paper/2021/file/49ad23d1ec9fa4bd8d7d02681df5cfa-Paper.pdf>
- [11] 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](https://doi.org/10.1007/978-1-4939-3261-0_2)
- [12] Remi Durant. 2021. Artist Studies by @remi\_durant. <https://remidurant.com/artists/>
- [13] Douglas C. Engelbart. 1962. Augmenting Human Intellect: A Conceptual Framework. Air Force Office of Scientific Research, AFOSR-3233.
- [14] 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*

<sup>15</sup><https://multimodal.art>

- (Hong Kong, China) (*DIS '18*). Association for Computing Machinery, New York, NY, USA, 1235–1257. <https://doi.org/10.1145/3196709.3196732>
- [15] Harmeet Gabha. 2022. Disco Diffusion 70+ Artist Studies. <https://weirdwonderfulai.art/resources/disco-diffusion-70-plus-artist-studies/>
- [16] Harmeet Gabha. 2022. Disco Diffusion Modifiers. <https://weirdwonderfulai.art/resources/disco-diffusion-modifiers/>
- [17] 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>
- [18] 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.
- [19] Mark Guzdial. 2013. Human-Centered Computing: A New Degree for Licklider's World. *Commun. ACM* 56, 5 (2013), 32–34. <https://doi.org/10.1145/2447976.2447987>
- [20] Gwern. 2020. GPT-3 Creative Fiction. <https://www.gwern.net/GPT-3>
- [21] Lydia Paine Hagtvædt, 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>
- [22] Aaron Hertzmann. 2018. Can Computers Create Art? *Arts* 7, 2 (2018), 25 pages. <https://doi.org/10.3390/arts7020018>
- [23] David Holz. 2022. Hey everyone, this server is getting BIG! Announcement on Midjourney. <https://discord.com/channels/662267976984297473/952771221915840552/1003404579897102387>
- [24] 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](https://doi.org/10.1016/0142-694X(91)90003-F)
- [25] Anna Kantosalo, Prashanth Thattai Ravikumar, Kazjon Grace, and Tapiola 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.
- [26] 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>
- [27] Andriud 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>
- [28] Andriud 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 (2014), 48 pages. <https://doi.org/10.1145/2591677>
- [29] Robert V. Kozinets. 2010. *Netnography: Doing Ethnographic Research Online*. Sage, London, UK.
- [30] Logan Kugler. 2021. Non-Fungible Tokens and the Future of Art. *Commun. ACM* 64, 9 (2021), 19–20. <https://doi.org/10.1145/3474355>
- [31] Vivian Liu and Lydia B Chilton. 2022. Design Guidelines for Prompt Engineering Text-to-Image Generative Models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 384, 23 pages. <https://doi.org/10.1145/3491102.3501825>
- [32] Midjourney. 2022. Midjourney.com. <https://www.midjourney.com>
- [33] Lj Miranda. 2021. List of VQGAN+CLIP Implementations. <https://lvmiranda921.github.io/notebook/2021/08/11/vqgan-list/>
- [34] 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>
- [35] Emad Mostaque. 2022. “A sneak peek at some of the futures to come for #DreamStudio...”. <https://twitter.com/EMostaque/status/1561778925906395140>
- [36] Ryan Murdock and Phil Wang. 2021. Big Sleep. <https://github.com/lucidrains/big-sleep>
- [37] 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>
- [38] OpenAI. n.d. Examples – OpenAI API. <https://beta.openai.com/examples>
- [39] Jonas Oppenlaender. 2020. Crowd-Powered Creativity Support Systems. In *Companion Proceedings of the 12th ACM SIGCHI Symposium on Engineering Interactive Computing Systems (Sophia Antipolis, France) (EICS '20 Companion)*. ACM, New York, NY, USA, Article 15, 4 pages. <https://doi.org/10.1145/3393672.3398646>
- [40] Jonas Oppenlaender. 2021. *Crowdsourcing Creative Work*. Ph.D. Dissertation. University of Oulu, Oulu, Finland. <https://doi.org/10.48550/ARXIV.2203.16495>
- [41] Jonas Oppenlaender. 2022. A Taxonomy of Prompt Modifiers for Text-to-Image Generation. <https://doi.org/10.48550/ARXIV.2204.13988>
- [42] Jonas Oppenlaender, Maximilian Mackeprang, Abderrahmane Khiat, Maja Vuković, Jorge Goncalves, and Simo Hosio. 2019. DC2S2: Designing Crowd-powered Creativity Support Systems. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (Glasgow, Scotland UK) (CHI EA '19)*. ACM, New York, NY, USA, Article W06, 8 pages. <https://doi.org/10.1145/3290607.3299007>
- [43] Guy Parsons. 2022. The DALL-E 2 Prompt Book. <https://dallery.gallery/the-dalle-2-prompt-book/>
- [44] pharmapsychotic. 2022. Tools and Resources for AI Art. <https://pharmapsychotic.com/tools.html>
- [45] 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](https://doi.org/10.1207/s15326985ep3902_1)
- [46] 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.
- [47] Reddit user Wiskkey. 2022. List of Stable Diffusion systems. [https://www.reddit.com/r/StableDiffusion/comments/wqaizj/list\\_of\\_stable\\_diffusion\\_systems/](https://www.reddit.com/r/StableDiffusion/comments/wqaizj/list_of_stable_diffusion_systems/)
- [48] Laria Reynolds and Kyle McDonnell. 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.
- [49] 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.* (2022), 27 pages. <https://doi.org/10.1145/3519026>
- [50] Mel Rhodes. 1961. An Analysis of Creativity. *The Phi Delta Kappan* 42, 7 (1961), 305–310.
- [51] Ruth Richards. 1990. Everyday creativity, eminent creativity, and health: “After-view”; for CRJ Issues on creativity and health. *Creativity Research Journal* 3, 4 (1990), 300–326. <https://doi.org/10.1080/10400419009534363>
- [52] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2021. High-Resolution Image Synthesis with Latent Diffusion Models. [arXiv:2112.10752 \[cs.CV\]](https://arxiv.org/abs/2112.10752)
- [53] Robin Rombach, Andreas Blattmann, and Björn Ommer. 2022. Text-Guided Synthesis of Artistic Images with Retrieval-Augmented Diffusion Models. <https://doi.org/10.48550/ARXIV.2207.13038>
- [54] 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>
- [55] John R. Searle. 1980. Minds, brains, and programs. *Behavioral and Brain Sciences* 3, 3 (1980), 417–424. <https://doi.org/10.1017/S0140525X00005756>
- [56] Ben Shneiderman, Gerhard Fischer, Mary Czerwinski, Mitch Resnick, Brad Myers, Linda Candy, Ernest Edmonds, Mike Eisenberg, Elisa Giaccardi, Tom Hewitt, Pamela Jennings, Bill Kules, Kumiko 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](https://doi.org/10.1207/s15327590ijhc2002_1)
- [57] Charlie Snell. 2021. Alien Dreams: An Emerging Art Scene. <https://ml.berkeley.edu/blog/posts/clip-art/>
- [58] Jordan G. Teicher. 2016. When Photography Wasn't Art. <https://daily.jstor.org/when-photography-was-not-art/>
- [59] Chris Vallance. 2022. “Art is dead Dude” – the rise of the AI artists stirs debate. <https://www.bbc.com/news/technology-62788725>
- [60] Xintao Wang, Yu Li, Honglun Zhang, and Ying Shan. 2021. Towards Real-World Blind Face Restoration with Generative Facial Prior. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR '21)*.
- [61] Ningyu Zhang, Luoqiu Li, Xiang Chen, Shumin Deng, Zhen Bi, Chuanqi Tan, Fei Huang, and Huajun Chen. 2021. Differentiable Prompt Makes Pre-trained Language Models Better Few-shot Learners. <https://doi.org/10.48550/ARXIV.2108.13161>