Text-to-Image Generation: Perceptions and Realities

JONAS OPPENLAENDER, University of Jyväskylä, Finland
AKU VISURI, University of Oulu, Finland
VILLE PAANANEN, University of Oulu, Finland
RHEMA LINDER, University of Tennessee, United States
JOHANNA SILVENNOINEN, University of Jyväskylä, Finland

Generative AI is an emerging technology that will have a profound impact on society and individuals. Only a decade ago, it was thought that creative work would be among the last to be automated – yet today, we see AI encroaching on creative domains. In this paper, we present the key findings of a survey study on people's perceptions of text-to-image generation. We touch on participants' technical understanding of the emerging technology, their ideas for potential application areas, as well as concerns, risks, and dangers of text-to-image generation to society and the individual. The study found that participants were aware of the risks and dangers associated with the technology, but only few participants considered the technology to be a risk to themselves. Additionally, those who had tried the technology rated its future importance lower than those who had not.

Additional Key Words and Phrases: text-to-image, generative AI, generative deep learning, AI co-creation, computational creativity

1 INTRODUCTION

Recent advancements in generative artificial intelligence (GenAI) have yielded significant progress in various domains. This has the potential to greatly impact a wide range of industries, particularly in the creative domain. Only a decade ago, the general consensus was that knowledge and creative work would be among the last to be automated [1, 6]. However, current developments in GenAI have turned this prediction on its head [5]. Progress in generative AI has exploded in recent years and we increasingly see GenAI being applied in creative domains, such as arts and research. One particularly popular domain is text-to-image generation, as evident in generative systems that can synthesize images from text prompts, such as Midjourney [11], Stable Diffusion [16], and DALL-E 2 [15]. Outputs from state-of-the-art diffusion models are often indistinguishable from those generated by humans [12, 13]. Some call this development "AI's Jurassic Park moment" [8] – an adapt-or-die moment that could potentially result in massive job loss across many sectors. However, many people are still oblivious to the generative powers of state-of-the-art systems.

In October 2022, we conducted a survey study on the perception of the text-to-image generation technology among different groups of individuals, including artists as well as people with no prior experience and those who self-reported having experience with the technology. The survey focused on people's understanding of the emerging technology, its potential uses, and the dangers of the technology for the individual and society. We present the key findings of the survey in this paper.

2 RELATED WORK

Related research has explored various aspects of artificial intelligence and its use in visual art. These studies have focused on understanding the perception and attitudes towards art generated by AI [7], authorship, agency, and intention in AI-generated art [9], and the potential bias towards such art [14]. Additionally, Rostamzadeh et al. discussed the ethical

Manuscript submitted to ACM

implications of AI in the creative application of computer vision [17] and Epstein et al. discussed the question of who should be credited for AI-generated works [3]. Our paper provides a novel empirical perspective on this related work.

3 METHOD

In autumn of 2022, we invited visitors to complete an online survey at the Researchers Night, a local annual event in which researchers present their research to the public. The questionnaire consisted of 26 questions, including three open-ended items. Participation was incentivized with a raffle for three Amazon vouchers, each worth 30 EUR.

We qualitatively analyzed the responses to the three open-ended survey items using in vivo coding [2]. The first author read and then iteratively coded all responses. Multiple codes were assigned, if needed, and iteratively improved by visualizing the codes in histogram charts. Due to the manageable amount of data and straight-forward answers, the coding did not require multiple raters and an analysis of inter-rater reliability [10].

4 PARTICIPANTS

35 participants (P1–P35, aged 19 to 50, $M = 33.7$ years, $SD = 9.3$ years) completed the online survey. Participants had diverse educational backgrounds, the most common being computer science, literature, and information systems. Fourteen participants held a Bachelor's degree, 10 held a Master of Science, 4 held a Master of Arts, 3 held a doctoral degree, and one held no academic degree. Twenty-four participants were students. A third (34.3%) of the participants had used text-to-image generation before. The most popular system used was DALL-E Mini/Craiyon (used by 7 participants), followed by DALL-E 2 (5 participants), Dream/Wombo (3 participants), and Stable Diffusion (2 participants). Participants estimated they had written an average of 20 prompts ($Max = 80$, $SD = 22$). Participants were, therefore, inexperienced with the emerging technology. Participants who had tried text-to-image generation were younger than those who had not tried the technology ($p < 0.05$). Ten participants considered themselves artists and had created paintings, digital art, drawings, writing, and other forms of art.

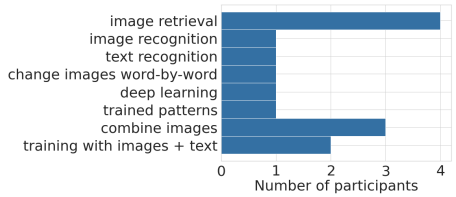
5 KEY FINDINGS

5.1 Understanding of Text-to-image Generation and Application Areas of the Emerging Technology

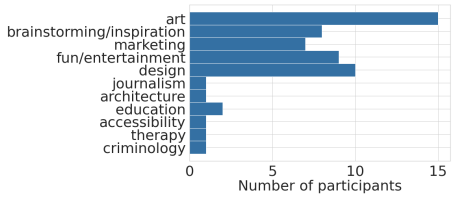
Most participants did not have a strong understanding of how text-to-image generation works. In their description of the technology, all but four participants did not distinguish between training and inference (i.e., image generation). Participants most often related the technology to image retrieval, followed by combining or mixing existing images (see Figure 1a). Creative areas dominated when it comes to participants' thoughts about potential application areas for text-to-image generation (see Figure 1b). Participants thought the technology was suited for creating artworks, illustrations, and other visual media. These could be applied in brainstorming or product development, but also marketing and design. The entertainment industry was also seen as an application area, for instance to make animations and games. The technology would also make a fun pass-time, according to participants. Less common, but still interesting, application areas included therapy, education, journalism, and criminology *"to reconstruct crime scenes"* (P33).

5.2 Criticisms and Concerns about Text-to-image Generation

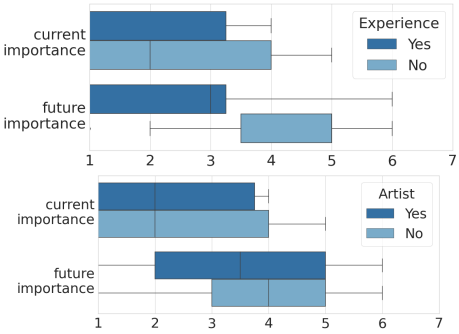
5.2.1 Professional importance of text-to-image generation. Most participants responded that text-to-image generation is currently not important for their profession, but will play an increasingly important role in the future (see Figure 1c). Interestingly, those who had tried image generation before found text-to-image generation not as important for their



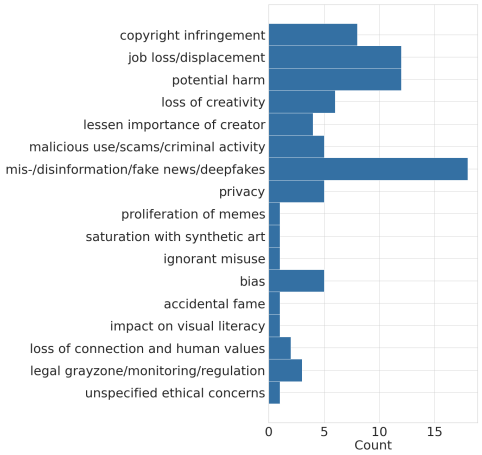
(a) Participants' understanding of text-to-image generation.



(b) Participants' thoughts on potential application areas.



(c) Current and future professional importance of image generation for participants who did and did not try text-to-image generation before (top) and self-identified artists and non-artists (bottom), from 1 – Not At All Important to 7 – Extremely Important.



(d) Participants' concerns about text-to-image generation.

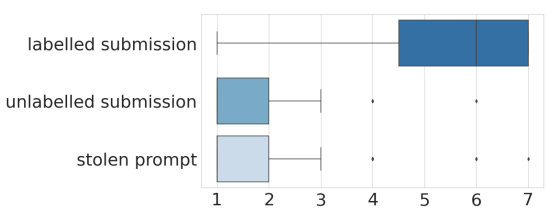
Fig. 1. Key results of the study on participants' perceptions of text-to-image generation, including their understanding of the technology, potential applications, perceived importance in their profession, and concerns.

professional future, as opposed to those who had not tried it before. This difference was significant ($p < 0.05$) and not found among self-declared artists.

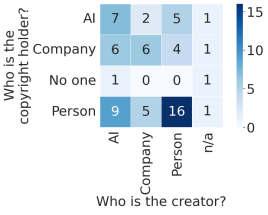
5.2.2 Concerns about the emerging technology. The majority of participants did not think that text-to-image poses a personal danger to themselves. But participants voiced many concerns about the effect of this emerging technology on society (see Figure 1d). The use of AI-generated imagery for opinion manipulation, fake news, and "deep fakes" was leading cause for concern. Many participants warned that synthetic images could be spread naively (misinformation) or for malicious purposes (disinformation). Another concern was unemployment due to increases in productivity. Participants mentioned that generative AI is cheaper and faster, and this could lead companies to not commission works from humans. Artists and designers were said to be particularly affected. Related to the potential loss of jobs, many participants noted that text-to-image generation operates in a legal gray zone with copyright infringement being one major concern. P15, for instance, mentioned that some *"people have directly used the artists name in the prompt to get an image to resemble the artists work as much as possible [without consent]."* As potential long-term effect, some participants mentioned there could be a loss of appreciation for artists and their work. The synthetic images could *"lessen the importance of the creator and the creative act"* (P4). Artists, *"who already suffer from poor income and low appreciation"* (P22), would be particularly vulnerable.

Some participants mentioned the potential effects of text-to-image generation on individuals and culture. For the individual, the effects could include harm, such as depression and other illnesses related to mental health. The AI

could be used to produce offensive, abusive, and inappropriate images that are *"not sensitive to people beliefs"* (P3). The harm could be accidental, such as the negative effects of accidental fame and leakage of private information, but also intentionally abusive, such as cyberbulllying. As for society, several participants thought there could be a *"decline in human creativity"* (P2, P23, P27). The technology *"could curtain artist imagination, when an AI can create art better than humans"* (P3). However, the AI was thought to be *"ultimately limited in its aesthetics"* (P20). This low diversity in synthetic imagery could contribute to *"narrow the viewpoint of the world"* if *"a lot of images start to look the same"* (P15). Synthetic images could lead to a *"biased and one-sided visual culture"* (P23). As P20 noted, generative AI *"has a great danger of enforcing certain values"* by showing "mainly white European bodies with certain aesthetics." (P15). The low diversity in synthetic imagery was seen to have a potentially negative impact on visual literacy if *"school book visualizations are [made] in the future with AI"* (P20). As P26 put it, text-to-image generation is *"a movement away from the things that make us human, e.g, human emotions being reflected in human-made art. The knock-on effects of unemployment, depression, cause by this lack of connection with human values and needs to create and be creative"* (P26).



(a) Rating of the ethicality of submitting a text-to-image artwork to an art fair on a Likert scale from 1 – Not Ethical At All to 7 – Highly Ethical.



(b) Participants responses to the multiple-choice questions of "who is the copyright holder and the creator?"

Fig. 2. Participants thoughts on the ethicality of labeling AI-created images (left) and copyright and creatorship (right).

5.2.3 Ethicality of disclosing AI generation. About half of the participants (n = 19; 54.3%) were of the opinion that it should be disclosed when something was created with AI. Ten participants had no strong opinion about this, and six participants (16.7%) thought that AI-generated images do not need to be labeled as such. However, when presented with the scenario of a person submitting a digital artwork to an art contest, participants thought that it was unethical to submit without disclosing that the image was created with AI (see Figure 2a). Not labeling a submission as created by AI was seen as equally unethical as submitting an artwork created from somebody else's prompt.

6 DISCUSSION AND CONCLUSION

Our survey study aimed to understand people's perceptions of text-to-image generation technology. While participants did not see immediate harm for themselves, they had varied opinions on the implications of the technology for society. It seems that when it comes to pinpointing the risks and dangers of text-to-image generation as an emerging technology, it was easier for participants to enumerate the potential problems of other people as compared to self-reflectively analyzing the impact of the technology on their own life. Interestingly, participants who had tried the technology rated its future importance lower compared to those who had not tried it. This observation conforms with the general hype cycle of technology in which expectations of an emerging technology undergo a trough of disillusionment before the technology's potential is realized [4]. Our findings suggest that while there is some awareness of the emerging technology, more education and awareness is needed to help people understand the capabilities and potential implications of text-to-image generation and generative AI.

REFERENCES

[1] M. Arntz, T. Gregory, and U. Zierahn. 2016. The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis. OECD Social, Employment and Migration Working Papers, No. 189. , 34 pages. https://doi.org/10.1787/5jlz9h56dvq7-en

[2] Kathy Charmaz. 2006. *Constructing Grounded Theory.* SAGE Publications Ltd.

[3] Ziv Epstein, Sydney Levine, David G. Rand, and Iyad Rahwan. 2020. Who Gets Credit for AI-Generated Art? *iScience* 23 (2020). https://doi.org/10.1016/j.isci.2020.101515

[4] Jackie Fenn and Alexander Linden. 2003. Understanding Gartner's Hype Cycles. Strategic Analysis Report R-20-1971. https://www.gartner.com/en/documents/396330/understanding-gartner-s-hype-cycles

[5] Organisation for Economic Co-operation and Development (OECD). 2021. Artificial intelligence and employment: New evidence from occupations most exposed to AI. Policy Brief on the Future of Work.

[6] Carl Benedikt Frey and Michael Osborne. 2013. The Future of Employment: How susceptible are jobs to computerisation?

[7] Joo-Wha Hong and Nathaniel Ming Curran. 2019. Artificial Intelligence, Artists, and Art: Attitudes Toward Artwork Produced by Humans vs. Artificial Intelligence. *ACM Trans. Multimedia Comput. Commun. Appl.* 15, 2s, Article 58 (jul 2019), 16 pages. https://doi.org/10.1145/3326337

[8] Gary Marcus. 2022. AI's Jurassic Park moment. https://garymarcus.substack.com/p/ais-jurassic-park-moment

[9] Jon McCormack, Toby Gifford, and Patrick Hutchings. 2019. Autonomy, Authenticity, Authorship and Intention in Computer Generated Art. In *Computational Intelligence in Music, Sound, Art and Design*, Anikó Ekárt, Antonios Liapis, and María Luz Castro Pena (Eds.). Springer International Publishing, Cham, 35–50.

[10] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and Inter-Rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 72 (nov 2019), 23 pages. https://doi.org/10.1145/3359174

[11] Midjourney. 2022. Midjourney.com.

[12] Jonas Oppenlaender. 2022. The Creativity of Text-to-Image Generation. In *25th International Academic Mindtrek Conference (Academic Mindtrek 2022)*. Association for Computing Machinery, New York, NY, USA, 192–202. https://doi.org/10.1145/3569219.3569352

[13] Jonas Oppenlaender. 2022. A Taxonomy of Prompt Modifiers for Text-To-Image Generation. , 15 pages. https://doi.org/10.48550/ARXIV.2204.13988

[14] Martin Ragot, Nicolas Martin, and Salomé Cojean. 2020. AI-Generated vs. Human Artworks. A Perception Bias Towards Artificial Intelligence?. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3334480.3382892

[15] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. 2022. Hierarchical Text-Conditional Image Generation with CLIP Latents. https://doi.org/10.48550/ARXIV.2204.06125

[16] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2021. High-Resolution Image Synthesis with Latent Diffusion Models. arXiv:2112.10752

[17] Negar Rostamzadeh, Emily Denton, and Linda Petrini. 2021. Ethics and Creativity in Computer Vision. (2021). https://doi.org/10.48550/ARXIV.2112.03111