

Visual Artists' Impressions On Incorporating Large-Scale Text-to-Image Generation Models into Illustration Editors

지도교수 : 서진욱

이 논문을 공학학사 학위 논문으로 제출함.

2023년 6월 27일

서울대학교 자유전공
자유전공학부
김동훈

2023년 8월

Visual Artists' Impressions On Incorporating Large-Scale Text-to-Image Generation Models into Illustration Editors

지도교수 : 서진욱

이 논문을 공학학사 학위 논문으로 제출함.

2023년 6월 27일

서울대학교 자유전공
자유전공학부
김동훈

2023년 8월

Abstract

Visual Artists' Impressions On Incorporating Large-Scale Text-to-image Generation Models into Illustration Editors

Kim Dong Hun

College of Liberal Studies

Seoul National University

Large-scale Text-to-Image Generation Models (LTGMs) enable users to generate high-quality images with a prompt describing their desired output. These models have been extensively tested in various domains, demonstrating their potential as a tool to assist in image creation. However, limited research has been conducted on whether LTGMs enhance image creation for expert designers. Our objective was to investigate to which extent LTGMs benefit expert designers in creating desired images. For this purpose, we assembled a group of 12 visual designers and assigned them tasks involving the creation of illustrations, both with and without the use of a LTGM. We developed a pixel art illustration editor specifically for these tasks, and utilized a customized pixel style Stable Diffusion LTGM to generate visual assets. Following the completion of the tasks, we conducted interviews with the participants to gather their feedback and analyzed their impressions regarding the integration of LTGMs into their creative process. In conclusion, we propose that researchers should consider supporting multiple LTGMs tailored to specific purposes and making LTGMs customizable to individual users when incorporating them into a designer-centered illustration editor.

keywords : Large-scale text-to-image generation model; illustration editor; visual artists; interview study

Table of Contents

1. Introduction	1
2. Related Work	2
2.1. Human-AI Collaboration for Creation	2
2.2. LTGMs for Aiding Designer's Image Creation	3
3. Study Method	4
3.1. Participants	4
3.2. Task Design	4
4. Results	7
4.1. Positive Views on Creating with LTGMs	7
4.2. Negative Views on Creating with LTGMs	8
4.3. Further thoughts on how LTGMs should be serviced	10
5. Discussion	11
6. Limitations and Further Work	12
7. Conclusion	13
References	14

1. Introduction

Image creation is a significant challenge for many designers, as it requires both originality and alignment with the expectations of the intended audience. Large-scale Text-to-image Generation Models (LTGMs), such as DALL-E [6], have gained popularity in the hopes of alleviating the difficulties associated with this task. These models are trained on extensive datasets and allow users to generate images based on textual prompts that describe their desired visual concepts. Adobe recently announced Adobe Firefly [1], a tool that enables image generation and texture application to calligraphies using text prompts. Similarly, Microsoft recently announced Microsoft Designer [5], which assists in creating design templates based on text prompt inputs. Numerous studies [2, 3, 4, 9, 10, 11, 12] have demonstrated the efficacy of LTGMs in aiding users to produce high-quality images. However, there is a lack of research on whether these LTGMs truly facilitate the image creation process for professional designers who are already skilled in producing images independently.

Previous studies have primarily focused on how LTGMs can assist non-professionals, highlighting their ability to enable novice individuals to achieve professional-level quality with the aid of generative models [3, 10]. LTGMs empower users to attain a level of proficiency comparable to that of experts. Unfortunately, there is a scarcity of research papers that specifically explore the benefits of LTGMs for experts in their respective domains. Some studies [4, 9] that do examine whether LTGMs benefit experts in the visual domain predominantly concentrate on the ideation process, rather than the subsequent manual modification or customization process, where designers incorporate the generated image and refine it to complete a desired final outcome.

We developed a pixel-art editor for examining the impact of LTGMs on the image creation process of designers. Initially, we analyzed several popular illustration editor tools such as Photoshop, Illustrator, Krita, InkScape, Aseprite to determine the essential functionalities required in an illustration tool. Common functionalities found in these tools included editing, selecting, and a history feature for undo/redo operations. Following this established convention, we designed a pixel-art illustration editor that

incorporated all of the aforementioned functionalities. We opted for a pixel-based editor rather than a vector or raster-based approach to ensure fairness among participants, as favoring one approach over another could potentially bias the results based on individual participants' previous work experience.

Based on the findings from our interview study, we propose two guidelines for practitioners interested in integrating LTGMs into illustration editors. In order to enable designers to seamlessly incorporate LTGMs into their creative workflow we recommend that the editor 1) support multiple LTGMs for specific purposes, 2) make LTGMs customizable to the user.

2. Related Work

2.1. Human-AI Collaboration for Creation

Several research papers have examined how Human-AI Collaboration can enhance users' abilities to achieve high-quality outputs, with some studies specifically targeting young children. Chao Zhang et al. [11] developed Bio Sketchbook that enabled children to engage in observational learning with the assistance of generative models. Children would capture a picture of something they were interested in, and the model would generate a contour drawing for them to continue the observation by sketching. They also introduced StoryDrawer [12], a service where children could create stories with the help of generative models. In this particular service, the model not only generated images but also generated speech texts to support children's thought processes. Both applications highlight the potential of utilizing generative AI models as collaborative assistants to enhance individual learning experiences.

Other research papers on Human-AI Collaboration have primarily focused on non-expert users leveraging AI assistance across various domains. One such example is De-Stijl [7], which explores the use of AI in the visual domain. De-Stijl employs AI to recommend color palettes based on the images placed on the canvas. Users can select suitable colors for their scenes and apply them to texts or shapes they add to the canvas. Another study by Santiago III, Jose Ma, et al. [6] demonstrates how generative

AI models can aid in the text domain. By providing specific prompts tailored to the context of Dungeon and Dragons storytelling, the researchers outline design guidelines for developers interested in integrating generative AI for interactive storytelling purposes.

The aforementioned examples illustrate the effectiveness of Human-AI Collaboration in generating high-quality outputs across diverse domains. Technology enables individuals to surpass their limitations and expand their creative boundaries. However, it is worth noting that a majority of research papers in this area have focused on assisting non-experts rather than experts.

2.2. LTGMs for Aiding Designer's Image Creation

There is limited research on the benefits of LTGMs for expert designers. One notable study in this area is BrainFax [8], which explored how process-oriented generative AI could support collaboration among expert designers. The researchers employed generative AI models as a visual communication tool for sharing ideas among designers. Unlike result-oriented models that generate images randomly based on a prompt input, they utilized a process-oriented model that allowed users to modify specific parts of the generated images through detailed prompts. This approach facilitated effective communication and consensus-building among designer participants, enabling them to achieve the desired generated output. It is important to note, however, that this research primarily focused on the use of LTGMs for ideation and visual communication, rather than directly addressing the image editing process employed by designers.

Opal [4] investigated how LTGMs can assist designers in creating news illustrations. The study tested the ability of a GPT-3 based model to generate appropriate keywords for a news article and evaluated Opal's capability to generate usable news illustrations based on user prompts that referenced the extracted keywords. Users were able to generate images with improved prompts using the extracted keywords, and participants positively viewed Opal's capability to generate suitable news illustrations. However, this research did not directly involve LTGMs in the image editing process, as participants could only modify the overall tone or style of the image output only.

Ko, Hyung-Kwon, et al. [3] explored how expert designers envision the co-creation process with LTGMs. After introducing the functionalities of DALL-E to visual artists, a semi-structured interview was conducted to capture the impressions of the artists regarding LTGMs. The interviews were then analyzed to identify the potential benefits and limitations of incorporating LTGMs into the design process. While this research shed light on the experiences and considerations designers may encounter when using LTGMs, it was based on a speculative thought process rather than direct interactions.

Many studies on how LTGMs can assist expert designers have predominantly focused on using LTGMs as generators of the final image rather than as interactive companions that designers actively engage with and manually modify to achieve their desired output.

3. Study Method

3.1. Participants

In our study, we recruited a group of 12 visual designers, consisting of 7 males and 5 females. All participants had majored in visual design and had a minimum of two years of experience using illustration tools. The age range of the participants was 24-29. The study was conducted remotely using an online meeting platform. We conducted an open-ended interview to gather the participants' impressions of LTGMs. The participants were given tasks that involved creating illustration images both with and without the assistance of a LTGM, and were interviewed afterwards.

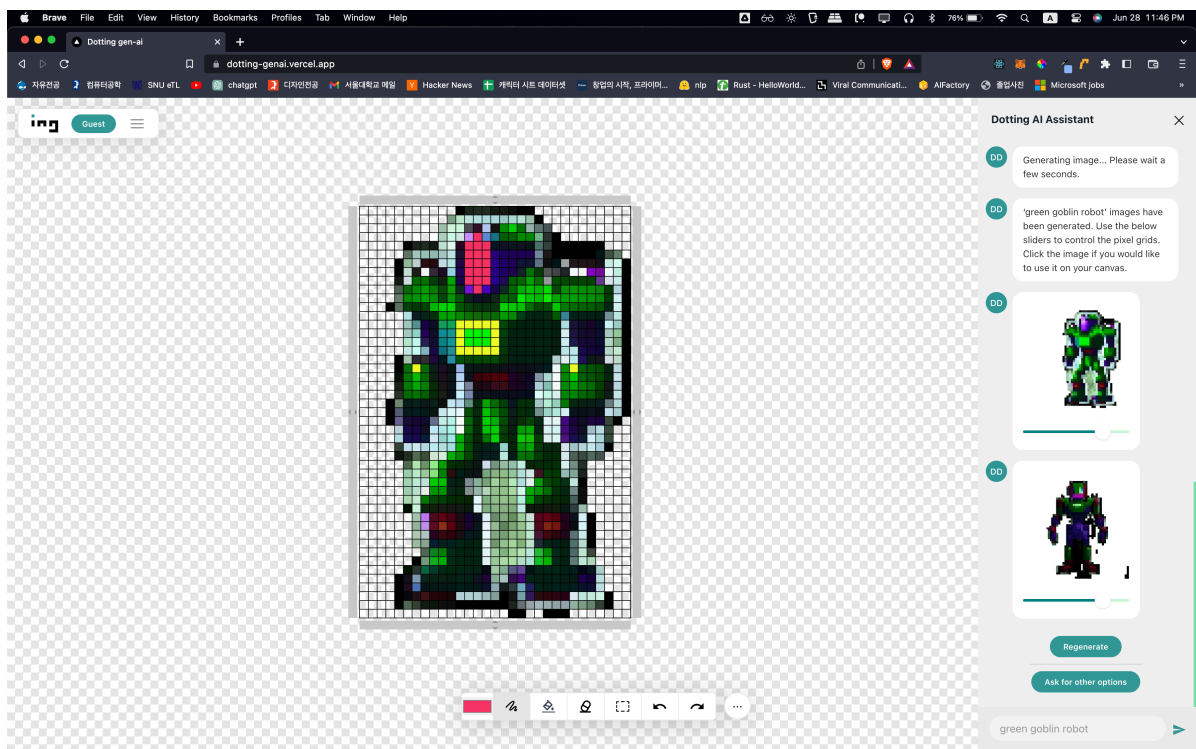
3.2. Task Design

For our research, we developed a customized pixel illustration editor specifically designed for the study. The goal was to ensure that participants' unfamiliarity with the

pixel art illustration editor would not hinder the creation process. To create the editor, we examined popular illustration editors such as Photoshop, Illustrator, Krita, InkScape, Aseprite. Through this examination, we identified three core features commonly found in illustration editors: edit, select, and undo/redo functionality. Following the convention, we designed the pixel art illustration editor to include these basic core features.

Among various visual domains, we chose pixel art for the tasks to avoid potential discrepancies in results that could arise from using vector-based or raster-based illustration editors, which may favor participants with specific familiarity in a particular approach. Additionally, a pixel art editor, where pixels are drawn based on mouse strokes, was deemed intuitive and user-friendly for first-time users, making it suitable for testing the impact of LTGMs on the image creation process.

We employed a Stable Diffusion Model trained using dreambooth for generating images in a pixel style as the LTGM for the study. To simplify the tasks, participants were required to create a character image rather than a complete scene. Furthermore, we implemented a pipeline to remove the background of the generated image from the Stable Diffusion model. This allowed users to receive character images with transparent backgrounds for their inputted prompts. Participants could edit the generated images on the canvas by dragging-and-dropping the generated images to the canvas.



(Figure 1) Pixel art editor with access to LTGM (Task 2)

Participants were given two tasks related to illustrating a “robot”. The first task required participants to create a robot illustration without the assistance of the LTGM. After a brief two-minute familiarization period with the pixel art tool, participants were instructed to create their own robot illustration.

For the second task, participants were asked to illustrate a robot again, but this time with the help of the LTGM. Following another two-minute familiarization period, this time with the LTGM, participants were guided to create a robot illustration by interacting with the LTGM.

There were no time restrictions for creating the images in each task, but participants were encouraged to spend a maximum of approximately 10 minutes on each task.

Participants had the opportunity to interact with the LTGM only during the second task. They could input their prompt in a designated bar located on the right side of the website. The bar served as a chat log where users could input their prompt and receive generated images as replies. If participants wanted to utilize the generated images in their creation process, they could drag one of the images from the right bar and place it on the editor canvas by dropping it on a desired location. Once placed in on the canvas, the pixels of the generated image were colored in the pixel editor, and participants could further edit the color of each individual pixel manually.

Following the completion of the two tasks, an open-ended interview was conducted to gather participants’ feedback and impressions regarding their experience using the LTGM for the image creation process. The interview aimed to explore their thoughts on using the LTGM, their overall impression of the collaborative experience with the LTGM, and any other insights they wished to share.

4. Results

4.1. Positive Views on Creating with LTGMs

4.1.1. Image Reference Tool

All participants expressed unanimous agreement that LTGMs can be beneficial as an image reference tool in the image creation process. They highlighted the difficulty of starting with a blank canvas and emphasized how the editor's LTGM helped alleviate this challenge by providing image suggestions based on their prompts. Participants often rely on images from external sources to inspire their designs, but with the assistance of LTGMs, they expressed the hope of obtaining image references without the need to search elsewhere. Some participants (P2, P9, P10) specifically mentioned that the experience of inputting a prompt and receiving suggestions made them feel like they were collaborating with a companion.

P2, *"It felt like as if I was communicating with another companion about 'What should I draw?'."*

Furthermore, a few participants (P2, P3) mentioned that they considered LTGMs to be potentially better image reference tools compared to the ones they currently use in their work. They explained that existing tools like Pinterest often provide image references based on their previous search results, leading to similar and predictable suggestions. In contrast, LTGMs offered unexpected and diverse image references that deviated from their preferences.

P3, *"I use Pinterest a lot when I find image references. The problem is that since Pinterest learns from my previous search results, I get image references that are similar to the ones I searched before. LTGMs, on the other hand, give me image references that I have not expected."*

LTGMs sometimes fall short in accurately capturing the precise intention of the input prompt, resulting in images that do not align with the user's intended direction.

However, some participants (P2, P7) expressed that this slight discrepancy in interpreting their intentions could actually enhance the value of LTGMs as an image reference tool. The unexpected image results that deviate from their initial expectations serve as a source of inspiration and fresh perspective for designers.

P7, *“When seeking insights for creating something, designers often draw inspiration from other topics, sometimes even from completely unrelated sources. LTGMs have the potential to offer designers something entirely unintended, sparking inspiration through these unexpected results.”*

4.1.2. Encourage Using Untried Styles

Participants found it beneficial that the editor's LTGM encouraged them to explore new styles and approaches. The generated pixel art images produced by the LTGM introduced participants to styles they had never tried before. This aspect of the LTGM's influence prompted participants (P5, P6, P9) to consider alternative viewpoints, use colors they rarely incorporated in their works, and even explore different dimensions in their illustrations.

P9, *“In the first task, without the LTGM, I approached drawing a robot from a 2D perspective. However, in the second task with the LTGM, it presented me with 3D versions of robots, which inspired me to attempt drawing a 3D robot.”*

4.2. Negative Views on Creating with LTGMs

4.2.1. Fear of Public Opinion Against Generated Images

Participants expressed concerns about copyright issues and the potential for their work to resemble existing creations when collaborating with LTGMs. They emphasized the importance of maintaining their own style as a creator, as it serves as a testament to the authenticity of their work. Unintentionally infringing upon someone

else's work is a significant fear for designers, as it can result in public condemnation and a loss of credibility.

P12 , *"As illustrators, our greatest fear is unknowingly infringing upon someone else's work. When a work is found to be an infringement, it loses its authenticity, and 'authenticity' is precisely what the public expects from illustrators."*

Some participants (P5, P8, P12) highlighted incidents where the public expressed discontent with the use of generative AI in creative industries. Examples were given of entertainment companies in Korea, such as Melon and Naver Webtoon, attempting to incorporate generative AI into album covers or webtoon covers. The public reacted strongly, criticizing the lack of consistency and authenticity in the styles produced by the AI. Companies had to issue apologies to the public for their attempts to use generative AI in artworks.

4.2.2. Lack of Personalization

Participants expressed their desire for a more customized generative model that learns their unique style and return images that align with their personal artistic preferences. They emphasized that while the editor's LTGM helped find image references and explore untried styles, incorporating the generated images into their work was met with hesitation. Participants wanted the AI model to be trained specifically on their style, reducing repetitive work and providing practical assistance.

P6 , *"While untried styles suggested by the editor's LTGM were helpful, having an AI model that learns my style and generates images in my own style would be more practical since it would alleviate repetitive work."*

In addition to learning user's style, participants (P3, P6, P7, P9, P10) expressed the desire for more detailed features that would allow the LTGMs to better understand the context of their prompts. Some participants (P7, P9) mentioned that there were specific generated images they found interesting but wanted the ability to tweak certain features to better suit their needs. However, since the editor's LTGM only returned new images based on the input prompt, participants felt limited in their ability to modify and use the generated images they were interested in.

4.2.3. Inefficiency

Some participants (P5, P11, P12) expressed a preference for manual creation over using generated images and retouching them. They argued that they can create images faster manually once they have a clear vision of what they want to draw. They felt that using AI-generated images to depict subjects that they have a concrete visualization of in their minds would slow down the creations process, as the LTGMs may misinterpret their prompts. One participant (P11) even mentioned a desire for the model to generate ready-for-production results that require no further modification.

P11, *“Personally, I see working with LTGMs for creating a work as outsourcing rather than collaboration. It’s difficult to collaborate on an already created work when there’s a specific image in my head.”*

4.3. Further Thoughts on How LTGMs Should Be Serviced

4.3.1. Multiple LTGMs for different purposes

Many participants expressed a desire for multiple LTGMs, each specializing in distinct features or stages of the design process. They saw value in having separate LTGMs for finding image references and for assisting in the actual creation process. P2 even proposed further subdivisions within the image reference LTGM, with one model providing exact image references based on the inputted prompt and another model offering random image references based on the context of the prompt.

Moreover participants (P1, P2, P6, P9, P12) emphasized the importance of a personalized LTGM that could learn their individual style and assist them in reducing repetitive work. They envisioned a two-stage model, referred to as “ideation stage” and the “development stage.” The ideation stage would involve exploring multiple ideas and concepts with the assistance of the LTGM, while the development stage would focus on leveraging the model’s knowledge of the user’s personal style to help create a final work in a manner consistent with the user’s established approach.

P9, *“It would be great if we had two stages of the model. One is for the ideation*

stage, where one could explore many ideas with the model. Next is the development stage, where another model could help create my work based on my style. This model would have knowledge of my work and assist me in creating a work in a way that aligns with my established approach.”

4.3.2. Transparent LTGMs

Some participants (P5, P12) expressed concerns about copyright infringement and suggested that LTGMs should provide more transparency in the image generation process. They proposed that if LTGMs could show the original images used as references to generate an output, designers would have a better understanding of the style and source of those images. This transparency would enable designers to avoid inadvertently infringing the copyright of another person’s work.

P5, “If the final output is a finished work, there should be a clear explanation of who created it and who owns it. Thus, it would be great if LTGMs could tell me how the image was created and what images they chose to reference for creating the final output. This can help me avoid any possible copyright issues.”

5. Discussion

Designers value the role of LTGMs in two specific areas to facilitate their work. Firstly, they expect LTGMs to assist in the initial ideation stage by providing image references. Manual collection of image references can be time-consuming and biased, based on personal preferences. LTGMs can offer a convenient solution by generating original images from text input prompts. Designers appreciate unexpected and diverse range of images that can inspire their creative process.

Secondly, designers seek to reduce their manual workload by utilizing LTGMs. They consider their personal style to be a crucial aspect of their work and are

reluctant to use generated images that do not reflect their unique artistic style. Designers desire LGTMs that can recognize and learn their individual styles, allowing for customized image generation that aligns with their artistic preferences. The primary motivation for using generative models in their work is to alleviate repetitive tasks, and personalized LTGMs would significantly contribute to this goal.

While designers appreciate the potential benefits of LTGMs, concerns regarding copyright infringement persist. They fear unintentionally infringing upon the work of others and believe that transparency in the image generation process is essential. Some designers propose that LTGMs should provide information about the original images used as references, enabling them to avoid potential copyright issues. By having insight into the sources and styles of generated images, designers can make informed decisions and avoid unintentional copyright violations.

6. Limitations And Future Work

Our study contains some limitations. Firstly, the use of a custom tool for the pixel art illustration may limit the generalizability of the findings. Further studies could consider using a range of existing tools to assess the impact of LTGMs across various illustration platforms. Secondly, the choice of a robot as the subject for the illustration tasks may have influenced the participants' familiarity and level of interest in the topic. Participants who were more fond of drawing robots might have had different experiences compared to those who were less familiar or less interested in the subject. Thirdly, the participants were relatively young. As such, they may be more open to new technologies like AI and have different perspectives and preferences compared to senior designers with extensive experience in the field. Including designers with diverse backgrounds and experience levels would provide a more comprehensive understanding of the impact of LGTMs in the image creation process. Lastly, the LTGM used for our study was a customized pixel-based Stable Diffusion model. The use of a specific LTGM could potentially lead to biased impressions and findings. Further research could explore a range of LTGMs to assess their impact on the design process.

7. Conclusion

In our paper, we present the perspective of designers regarding their collaboration with LTGMs in the creation of illustration images. Through interviews with 12 designers, we conducted a study where participants were asked to create pixel illustrations with and without the assistance of a fine-tuned pixel-style LTGM. Following the tasks, we conducted open-ended interviews to gather insights on designers' positive and negative views on working with LTGMs. Furthermore, we analyzed the features that designers considered necessary for integrating these models into their work processes. The findings indicated that designers viewed LTGMs positively as effective image reference tools, but identified challenges such as the lack of personalization, concerns about copyright infringement, and the need for models to learn and reflect the user's artistic styles. Designers expressed the desire for multiple LTGMs with different specialities to cater their specific needs.

References

- [1] Adobe. 2023. Adobe Firefly. <https://firefly.adobe.com/>
- [2] Hong, Yeong-Gi, et al. "Spot The Difference: AI, Please Make This for Me!." Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems. 2023.
- [3] Ko, Hyung-Kwon, et al. "Large-scale text-to-image generation models for visual artists' creative works." Proceedings of the 28th International Conference on Intelligent User Interfaces. 2023
- [4] Liu, Vivian, Han Qiao, and Lydia Chilton. "Opal: Multimodal Image Generation for News Illustration." Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology. 2022.
- [5] Microsoft. 2023. Microsoft Design. <https://designer.microsoft.com/>
- [6] OpenAI. 2023. DALL-E. <https://openai.com/dall-e-2>
- [7] Santiago III, Jose Ma, et al. "Rolling the Dice: Imagining Generative AI as a Dungeons & Dragons Storytelling Companion." arXiv preprint arXiv:2304.01860 (2023).
- [8] Shi, Xinyu, et al. "De-Stijl: Facilitating Graphics Design with Interactive 2D Color Palette Recommendation
- [9] Verheijden, Mathias Peter, and Mathias Funk. "Collaborative Diffusion: Boosting Designerly Co-Creation with Generative AI." Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems. 2023.
- [10] Wang, Fengjie, et al. "Slide4N: Creating Presentation Slides from Computational Notebooks with Human-AI Collaboration." Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. 2023.
- [11] Zhang, Chao, et al. "Bio sketchbook: an ai-assisted sketching partner for children's biodiversity observational learning." Interaction Design and Children. 2021.
." Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. 2023.
- [12] Zhang, Chao, et al. "StoryDrawer: A Child-AI Collaborative Drawing System to Support Children's Creative Visual Storytelling." Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. 2022.

국문초록

초거대 자연어 기반 이미지 생성 모델들은 (LTGM) 사용자가 텍스트 프롬프트로 높은 퀄리티의 이미지를 받게끔 해준다. LTGM을 사용해서 얻을 수 있는 이점은 다양한 분야에서 검증되었고, 사용자들이 LTGM을 이미지 생성하는 하나의 도구로 사용될 수 있다는 것이 명실상부해졌다. 하지만, LTGM이 전문 디자이너들이 이미지를 제작하는 과정에 기여하는지에 대한 연구는 부족하다. 이 논문은 LTGM이 실제 디자인 전문가들의 이미지 제작 과정을 도와주는지 알아본다. 연구를 진행하기 위해 12명의 디자인 전문가를 대상으로 LTGM이 없는 상태로 한번 이미지를 그리게끔 하고, LTGM이 있는 상태로 이미지를 그리게끔 해보았다. 이후 인터뷰를 진행하여 참여자들이 LTGM과 함께 작업하는 것에 대해서 어떤 인상을 받았는지 물어보았다. 연구 결과, 전문 디자이너들이 LTGM을 에디터에서 사용할 수 있게끔 하려면 제각각 다른 목적을 위한 LTGM들이 준비되어 있어야 하며, LTGM이 유저의 스타일을 학습하여 개인화할 수 있는 방향으로 제작되어야 한다는 것을 발견했다.

주요어 : 초거대 자연어 기반 이미지 생성 모델; 일러스트 에디터; 디자이너; 인터뷰 조사