Designing for Visual Thinkers: Overcoming Text-Centric Limitations in GenAI Tools

HYERIM PARK, VISUS, University of Stuttgart, Germany and BMW Group, Germany

MALIN EIBAND, BMW Group, Germany

This workshop paper examines the challenges of integrating generative AI (GenAI) into the design process, focusing on designers who are predominantly visual thinkers. Drawing from previous interviews with designers, we highlight challenges related to text-based inputs, difficulties in controlling prompts, and the lack of proper integration into design practices. Our paper explores ways to improve designers' interaction with GenAI tools, enabling them to achieve desired outcomes with less effort. We propose two approaches: one that enhances the use of text-based inputs through improved prompt generation with multimodal Large Language Models (LLMs) and another that introduces more intuitive, visually-driven methods like sketch and doodle inputs, reducing reliance on text prompts. This paper aims to foster discussions on how GenAI tools can better align with the needs of design professionals. Our work thus focuses on addressing the limitations of text-centric GenAI interfaces, making these tools more accessible and effective for visual thinkers, and ultimately improving their ability to leverage GenAI in the creative process.

Additional Key Words and Phrases: generative AI, creativity support tools, human-AI interaction

1 INTRODUCTION

Generative AI (GenAI) technologies are becoming increasingly popular and are significantly impacting fields like design. GenAI is a subset of AI technologies that can create diverse forms of content, such as images, text, and music, using inputs like text and images. Unlike traditional AI or machine learning (ML) models, which are typically employed for tasks such as predictive analytics, classification, and natural language processing (e.g., translation), GenAI specializes in producing *new* and original content comparable to human-created outputs [2, 8]. This capability holds considerable potential, particularly in creative fields, as it can speed up design processes, support concept development, and enable quick experimentation with multiple design alternatives [1, 22].

Designers commonly use GenAI tools for both image generation and natural language tasks. Well-known GenAI image generation tools include Dall-E [7], Midjourney [17], and Stable Diffusion [10], while Large Language Model (LLM) tools like ChatGPT (GPT-4) [9], LLaMa [16] and Claude [6] handle mainly text-based tasks such as text generation, translation, and summarization. Although these tools have expanded to support inputs like images and sketches, they still rely heavily on text input. Moreover, tools like Midjourney often require developer-like commands, such as shorthand terms for controlling parameters (e.g., `ar` for aspect ratio, `iw` for image weight). Designers, therefore, need to acquire a sophisticated skill set in a different area before they can effectively work with the system, delaying their ability to focus on their actual design tasks.

NordiCHI 2024 Workshop: Designing with AI-Based Tools, October 13–16, 2024, Uppsala, Sweden
DOI: https://doi.org/10.5281/zenodo.14186390

Authors' addresses: Hyerim Park, VISUS, University of Stuttgart, Stuttgart, Germany and BMW Group, Munich, Germany, hyerim.park@bmw.de; Malin Eiband, BMW Group, Munich, Germany, malin.eiband@bmw.de.

These text-oriented interfaces present challenges, particularly for designers who think and work visually. Previous research on the challenges designers face with GenAI tools highlights the difficulty many professionals encounter when trying to control prompts to generate the desired visual outputs [3, 15, 18, 22]. Our research [19], which included interviews with 16 professional designers, further revealed that many struggle to create effective text prompts that lead to satisfactory results, often resulting in increased time and effort spent on trial-and-error.

This paper argues that current GenAI tools are not fully inclusive of visual thinkers, leading to gaps in usability and adoption. We suggest shifting or extending the design of GenAI tools to focus more on visual-centric interfaces and reduce the complexity of generating and controlling prompts. Our discussion focuses on two key approaches:

- **Supporting text prompt generation using multimodal LLMs:** We suggest exploring how LLMs can assist in generating and refining text prompts, potentially reducing the effort and trial-and-error designers currently face.
- **Designing visual-input-centric GenAI interfaces:** We propose discussing new approaches for interface designs that allow designers to control GenAI tools more naturally through visual inputs like sketches, scribbles, and images. These approaches could create a more intuitive experience aligned with visual thinkers' cognitive processes.

This work aims to explore how these interface changes can bridge the gap between text-based GenAI systems and the needs of visual thinkers, offering new strategies to improve usability and better support designers in their creative workflows.

2 RELATED WORK

2.1 Challenges with GenAI Use in Design

Research has explored the experiences of various groups of designers, including visual designers, architects, UI/UX designers, and 2D/3D creators, in their use of GenAI tools, particularly those involving GenAI image models [11, 13, 22, 25]. Given that designers typically work with visual concepts, several challenges have been identified. For example, Zhang et al. [25] identified difficulties among 11 architectural design students using GenAI tools, noting that a frequent mismatch between intended designs and the generated outputs often leads to frustration.

Despite their potential, these tools, therefore, often demand significant time and effort from designers, who must refine their inputs repeatedly to reach the desired outcomes [4, 15, 18]. For instance, Oppenlaender [18] found that practitioners using text-to-image frequently refine GenAI-generated images by adding specific modifiers to their text prompts to align the results with their vision better. Similarly, Liu and Chilton [15] analyzed over 5,000 text-to-image generations and proposed design guidelines for effective prompt engineering, highlighting the trial-and-error process necessary to achieve coherent and satisfactory results.

These challenges reflect findings from our own qualitative research [19], where many designers expressed difficulties with text-based interfaces and controlling GenAI prompts. Visual thinkers, in particular, struggle to translate their ideas into effective text prompts, often resulting in time-consuming iterations and frustration when the generated outputs fail to match their vision. Additionally, controlling the prompts to achieve desired outcomes often requires extensive trial and error, adding to the effort and complexity of the design process.

2.2 Visual-Input Centered Interfaces in GenAI Tools

Research about the design process in general suggests that incorporating other modalities than text could improve the usability of GenAI tools for designers. For instance, Suwa et al. [21] observed that during this process, design sketches not only act as external memory or providers of visual cues for associating non-visual information but also create a physical environment in which design thoughts developed in real-time.

Several studies in the area of interaction with GenAI have therefore explored the use of inputs beyond just text. For example, Lee et al. [12] compared sketch-based inputs to text prompts in generating 3D designs, concluding that while text prompts were effective in sparking initial ideation, sketches played a crucial role in physically representing and refining design concepts. Similarly, Qiao et al. [20] showed that combining image prompts with text significantly enhanced the representation of specific subjects in GenAI-generated images. These findings highlight the potential benefits of incorporating various visual inputs into GenAI tools to support designers' creative processes better. Interestingly, in the context of story generation using LLMs, Chung et al. [5] used line sketch inputs to intuitively control story development with GPT, showing how visual input can enhance certain aspects of text generation beyond just image creation.

2.3 Problem Statements

We found that current text-input-centered GenAI tools are not well suited for designers, as they often struggle with effectively controlling and refining text prompts. Recently suggested sketch-based tools, such as those introduced by researchers [14] or commercial GenAI image tools like Vizcom [23], primarily focus on translating user sketches into varying levels of fidelity, including high-fidelity renderings or different artistic styles, from simple line sketches to detailed digital illustrations. However, these tools emphasize the conversion of sketches rather than using them as an intuitive form of prompt input to control and guide GenAI tools throughout the creative process.

To address these challenges, we propose exploring design approaches for GenAI interfaces that cater to visual thinkers, with a focus on aligning with their cognitive processes and creative workflows. In this workshop, we aim to discuss new design approaches for interfaces and input methods while also considering broader interface and control solutions within the GenAI landscape, particularly in relation to the HCI community. Rather than requiring designers to adapt to GenAI, we propose adapting GenAI tools to support designers' workflows.

3 DESIGNING GENAI INTERFACES FOR VISUAL THINKERS

To help designers overcome challenges with GenAI tools, we propose two key approaches to improve GenAI interfaces. The first approach focuses on supporting designers in generating effective text prompts, enabling them to better understand and adapt to current text-input-oriented tools. This would reduce trial and error and save time in the design workflow. The second approach involves creating interfaces that allow designers to control GenAI tools using visual inputs, such as sketches and doodles, offering a more intuitive and direct interaction for visual thinkers.

3.1 Approach 1: Supporting Effective Text Prompt Generation Using LLMs

In our preliminary interviews [19], designers shared strategies to overcome challenges with text prompting. A common approach involves leveraging multimodal LLMs for prompt engineering or generating descriptive text to guide image generation. Tools like ChatGPT or Midjourney's `describe` function[1] on Discord allow designers to upload reference images and generate tailored prompts. This reduces trial and error by producing prompts more closely aligned with the intended visual outcomes (see Figure 1).

With the rise of multimodal LLMs, there is potential to integrate sketch and doodle inputs to further support text prompt generation for visual thinkers. For example, Wang et al. [24] proposed multimodal prompting that refines text

[1]https://docs.midjourney.com/docs/describe

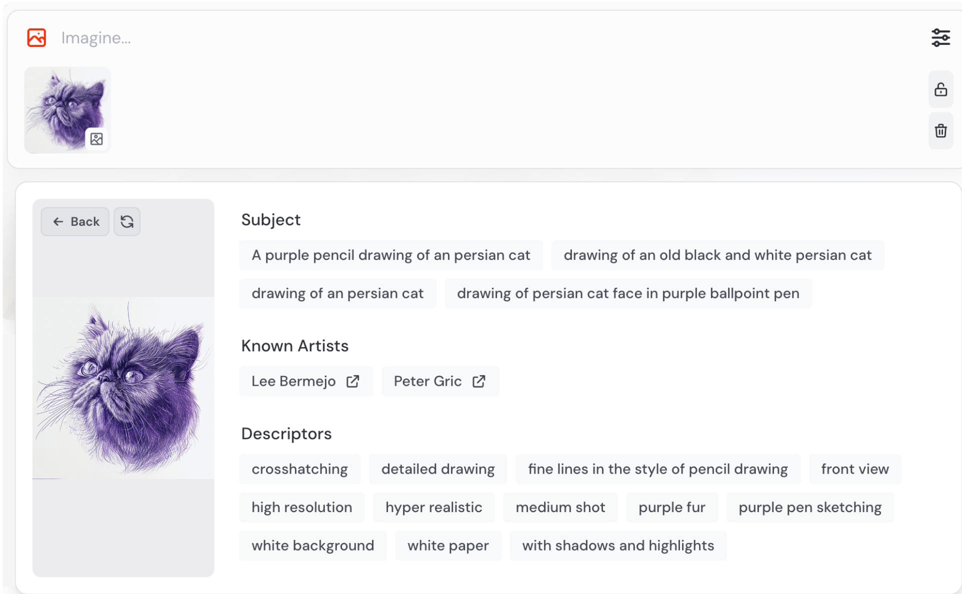

Fig. 1. The `describe` function in Midjourney's web version generates words and phrases based on an uploaded reference image. Designers can add these to prompts, refining text to better align with desired visual outcomes, reducing trial and error.

inputs and offers image variations. Using LLMs to refine or generate prompts based on visual cues, designers can create more precise outputs. When LLMs interpret sketches, they can extract key elements and translate them into refined text prompts for GenAI models. This reduces the ambiguity that often arises with purely text-based inputs and leads to more accurate reflections of the designer's vision.

Additionally, several designers in our preliminary interviews expressed confidence in improving their ability to work with text-based prompts [19]. Some were actively learning from communities or engaging in specialized training, showing a willingness to enhance their text-input skills. This suggests that, despite challenges, refining text-based interactions remains a valuable approach to improving GenAI workflows.

However, this approach also raises important questions about consistency. How can LLMs ensure reliable and consistent output from the same sketch or visual input? For instance, small variations in prompts can produce unexpectedly different results [18], making prompt control challenging. This is an important area for further exploration.

3.2 Approach 2: Integrating Visual Input into GenAI Interfaces

Incorporating visual inputs like sketches into GenAI tools can make the design process more intuitive and aligned with how many designers naturally work. For visual thinkers, relying solely on text-based prompts can be limiting. By integrating sketches alongside or, in some cases, without text inputs (an area that could benefit from further research), designers can better express and refine their ideas in more intuitive ways.

From a technical perspective, LLMs could be integrated into GenAI image generation tools to interpret visual inputs, such as rough sketches or doodles. The system would process these visual cues and automatically generate or refine text descriptions, which are typically required by GenAI tools. Over time, the system could learn from the designer's inputs, adapting to their style and preferences. This approach reduces the need for designers to manually translate their

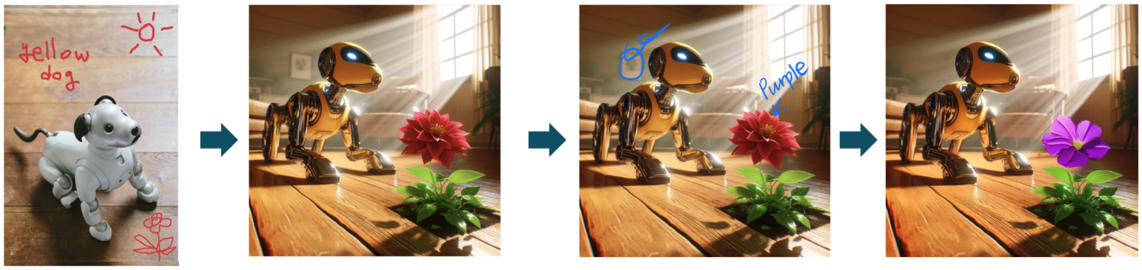

Fig. 2. Our concept for a GenAI image tool prototype to be developed in future research. The tool allows users to control image generation through scribbles or sketches, enabling them to generate or refine images in a more intuitive, visual-driven design process with less reliance on text prompts (visualized with the help of DALL-E).

visual ideas into detailed text prompts, as the LLMs interpret the concepts directly. Instead of constantly adjusting text prompts, designers can sketch their ideas and allow GenAI tools to interpret and build upon them. This approach helps designers focus directly on the actual visual task rather than the text prompts as an intermediary step.

Thus, visual inputs like sketches or doodles do not merely serve as preliminary sketches to be translated into high-fidelity renderings. Instead, they act as visual prompts that allow designers to control GenAI tools directly in a way that more closely mirrors their creative thought process. By using sketches as control mechanisms, designers could interact with GenAI more naturally, reducing their reliance on text-based inputs and enhancing their creative workflows (see Figure 2).

To the best of our knowledge, there is limited research exploring the use of visual prompts based on visual inputs in GenAI interfaces. We suggest further exploration to better understand the benefits and challenges of using visual inputs to guide GenAI tools. Future research could investigate how these inputs might enable GenAI systems to interpret complex visual cues with less reliance on text prompts, allowing for more intuitive, sketch-based interactions that align with designers' natural workflows.

4 CONCLUSION & FUTURE WORK

In this position paper, we highlighted the limitations of text-centric GenAI tools for visual thinkers and proposed two approaches to address these issues. We suggested leveraging multimodal LLMs to enhance text prompt generation and designing interfaces that incorporate visual inputs, such as sketches, for more intuitive interaction. Our work advocates for expanding input modalities beyond text, aligning GenAI tools more closely with the needs of visual designers. For future work, we plan to prototype and test visual-centric GenAI interfaces and conduct user studies to assess their impact on design workflows. These efforts aim to contribute to more inclusive and flexible GenAI tools within the HCI community.

REFERENCES

[1] Volker Bilgram and Felix Laarmann. 2023. Accelerating Innovation With Generative AI: AI-Augmented Digital Prototyping and Innovation Methods. *IEEE Engineering Management Review* 51, 2 (June 2023), 18–25. https://doi.org/10.1109/EMR.2023.3272799

[2] Rina Caballar and ibm. 2024. Generative AI vs. predictive AI: What's the difference? https://www.ibm.com/blog/generative-ai-vs-predictive-ai-whats-the-difference/www.ibm.com/blog/generative-ai-vs-predictive-ai-whats-the-difference

[3] Minsuk Chang, Stefania Druga, Alexander J. Fiannaca, Pedro Vergani, Chinmay Kulkarni, Carrie J Cai, and Michael Terry. 2023. The Prompt Artists. In *Proceedings of the 15th Conference on Creativity and Cognition (C&C '23)*. Association for Computing Machinery, New York, NY, USA, 75–87. https://doi.org/10.1145/3591196.3593515

[4] John Joon Young Chung and Eytan Adar. 2023. PromptPaint: Steering Text-to-Image Generation Through Paint Medium-like Interactions. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. ACM, San Francisco CA USA, 1–17. https://doi.org/10.1145/3586183.3606777

[5] John Joon Young Chung, Wooseok Kim, Kang Min Yoo, Hwaran Lee, Eytan Adar, and Minsuk Chang. 2022. TaleBrush: Sketching Stories with Generative Pretrained Language Models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–19. https://doi.org/10.1145/3491102.3501819

[6] claude. 2024. Meet Claude. https://www.anthropic.com/claude

[7] DALL·E. 2024. DALL·E 2. https://openai.com/dall-e-2

[8] Stefan Feuerriegel, Jochen Hartmann, Christian Janiesch, and Patrick Zschech. 2024. Generative AI. *Business & Information Systems Engineering* 66, 1 (Feb. 2024), 111–126. https://doi.org/10.1007/s12599-023-00834-7

[9] gpt 4. 2024. Hello GPT-4o. https://openai.com/index/hello-gpt-4o/

[10] Huggingface. 2024. The Stable Diffusion Guide. https://huggingface.co/docs/diffusers/v0.14.0/en/stable_diffusion

[11] Nanna Inie, Jeanette Falk, and Steve Tanimoto. 2023. Designing Participatory AI: Creative Professionals' Worries and Expectations about Generative AI. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–8. https://doi.org/10.1145/3544549.3585657

[12] Seung Won Lee, Tae Hee Jo, Semin Jin, Jiin Choi, Kyungwon Yun, Sergio Bromberg, Seonghoon Ban, and Kyung Hoon Hyun. 2024. The Impact of Sketch-guided vs. Prompt-guided 3D Generative AIs on the Design Exploration Process. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–18. https://doi.org/10.1145/3613904.3642218

[13] Jie Li, Hancheng Cao, Laura Lin, Youyang Hou, Ruihao Zhu, and Abdallah El Ali. 2024. User Experience Design Professionals' Perceptions of Generative Artificial Intelligence. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–18. https://doi.org/10.1145/3613904.3642114

[14] David Chuan-En Lin, Hyeonsu B Kang, Nikolas Martelaro, Aniket Kittur, Yan-Ying Chen, and Matthew K Hong. 2024. Inkspire: Supporting Designers to Prototype Product Designs through Sketching. (2024).

[15] Vivian Liu and Lydia B Chilton. 2022. Design Guidelines for Prompt Engineering Text-to-Image Generative Models. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–23. https://doi.org/10.1145/3491102.3501825

[16] meta. 2024. Llama 3.1. https://www.llama.com/

[17] Midjourney. 2024. Midjourney. https://www.midjourney.com/home?callbackUrl=%2Fexplore

[18] Jonas Oppenlaender. 2023. A Taxonomy of Prompt Modifiers for Text-To-Image Generation. *Behaviour & Information Technology* (Nov. 2023), 1–14. https://doi.org/10.1080/0144929X.2023.2286532 arXiv:2204.13988 [cs].

[19] Hyerim Park, Joscha Eirich, Andre Luckow, and Michael Sedlmair. 2024. "We Are Visual Thinkers, Not Verbal Thinkers!": A Thematic Analysis of How Professional Designers Use Generative AI Image Generation Tools. (2024).

[20] Han Qiao, Vivian Liu, and Lydia Chilton. 2022. Initial Images: Using Image Prompts to Improve Subject Representation in Multimodal AI Generated Art. In *Creativity and Cognition*. ACM, Venice Italy, 15–28. https://doi.org/10.1145/3527927.3532792

[21] Masaki Suwa, Terry Purcell, and John Gero. 1998. Macroscopic analysis of design processes based on a scheme for coding designers' cognitive actions. *Design studies* 19, 4 (1998), 455–483. Publisher: Elsevier.

[22] Macy Takaffoli, Sijia Li, and Ville Mäkelä. 2024. Generative AI in User Experience Design and Research: How Do UX Practitioners, Teams, and Companies Use GenAI in Industry?. In *Designing Interactive Systems Conference*. ACM, IT University of Copenhagen Denmark, 1579–1593. https://doi.org/10.1145/3643834.3660720

[23] vizcom. 2024. Vizcom. https://www.vizcom.ai/

[24] Zhijie Wang, Yuheng Huang, Da Song, Lei Ma, and Tianyi Zhang. 2024. PromptCharm: Text-to-Image Generation through Multi-modal Prompting and Refinement. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–21. https://doi.org/10.1145/3613904.3642803

[25] Chengzhi Zhang, Weijie Wang, Paul Pangaro, Nikolas Martelaro, and Daragh Byrne. 2023. Generative Image AI Using Design Sketches as input: Opportunities and Challenges. In *Creativity and Cognition*. ACM, Virtual Event USA, 254–261. https://doi.org/10.1145/3591196.3596820