

Title	インタラクティブなプロンプトの視覚的探索を通じたテキストから画像生成
Author(s)	黄, 柏飛
Citation	
Issue Date	2025-03
Type	Thesis or Dissertation
Text version	author
URL	<a href="http://hdl.handle.net/10119/19719">http://hdl.handle.net/10119/19719</a>
Rights	
Description	Supervisor: 謝 浩然, 先端科学技術研究科, 修士 (知識科学)

In recent years, generative Artificial intelligence (AI) has demonstrated significant progress in text-to-image generation, merging advancements in computer vision and natural language processing. As above, cutting-edge models such as Stable Diffusion and DALL·E synthesize high-quality images with stunning aesthetics, offering unprecedented opportunities for creativity and innovation. However, crafting prompts that accurately reflect user intent and align with the model’s interpretation remains a significant challenge, particularly for novice users lacking prompt engineering expertise. This challenge is further compounded when users aim to explore their creativity and generate ideas iteratively, as the current workflow often requires repeated modifications and refinements, resulting in a time-intensive trial-and-error process. These limitations highlight the need for tools to bridge the gap between user intent and generative model outputs, enabling experts and non-experts to harness the potential of generative AI for creative tasks effectively.

To address these issues, PromptNavi is proposed as an interactive system designed to assist users in enhancing and refining prompts through visual exploration and iterative optimization within text-to-image generative models. PromptNavi introduces a novel approach to prompt refinement, leveraging an attribute interpolation system powered by large language models. This system analyzes initial user inputs and suggests enhancements to align prompts with desired outcomes. A node is a fundamental unit representing the system’s image or prompt component. Each node encapsulates visual attributes and semantic information, allowing users to manipulate and refine elements interactively. Furthermore, PromptNavi provides a dynamic, node-based visual interface that transforms the traditionally repetitive and opaque cycle of prompt engineering into an intuitive and interactive experience. Users can adjust prompt attributes, transfer them to generate other images, and iteratively refine their inputs based on visual feedback. The strength of connections between nodes visually represents how modifications to prompts influence the generated outputs, allowing users to understand better and control the relationship between prompts and images than traditional text-to-image generation.

Attributes refer to an image’s key visual and semantic properties, such as color, style, and composition. PromptNavi empowers users to discover and integrate these attributes effectively. The system significantly reduces the cognitive and temporal demands of prompt engineering by enabling real-time

visual feedback and intuitive attribute manipulation. Additionally, PromptNavi promotes deeper insights into text-to-image generative models, ensuring accessibility for novice users while providing advanced customization options—such as fine-tuned attribute weighting, multi-node attribute blending, and hierarchical prompt adjustments—for experienced users seeking greater control over image generation. This approach aims to democratize access to generative AI technologies, enabling a broader audience to realize their creative potential fully.

In conclusion, this thesis makes the following contributions:

- This thesis proposes PromptNavi, a node-based visual interface that clarifies the relationships among various prompt elements. It also supports interactive prompt exploration and refinement, making it easier for novice users to generate images from text.
- By leveraging LLM-based attribute interpolation, the system efficiently refines users’ prompts while helping them better understand the connection between images and prompts.
- Through text-image similarity, the system provides bidirectional alignment to ensure consistency and offers users a richer selection of attributes.
- By supporting multiple backends (e.g., Stable Diffusion, DALL-E, pre-trained models), addressing the diverse creative needs of different users.
- A user study involving 16 participants demonstrates the effectiveness of PromptNavi and its advantages over commonly used baseline tools.