A programmable interface for creative exploration

Gerard Serra Batou Research Barcelona gerard@batou.xyz Oriol Domingo Batou Research Barcelona oriol@batou.xyz Pol Baladas Batou Research Barcelona pol@batou.xyz

Abstract

Current advances in generative models, specially Large Language Models (LLM), has allowed the emergence of new AI-powered tools to support creativity. These models generate both text and image given a human prompt containing their ideas using natural language. However, LLM often require expertise to fine-tune prompting in order to the responses better match user needs, which can be a barrier for non-expert users.

In this paper, we present a spatial canvas with a set of AI-powered tools with predefined prompts to support creative exploration for non-expert users. Our results show how non-expert users can explore a creative space by combining their own input with the generated responses.

1 Introduction

In recent years, the use of LLM-based applications has steadily risen. Generative text models such as GPT-3[1] or LaMDA [2] or image models such DALL \cdot E 2 [3], Imagen [4] or Stable Diffusion [5] allowed the emergence of new tools to interact with these models and integrate them in creative workflows.

Despite the fact that these models can empower and augment human creative process, they rely on human capabilities to express their creative intention via a prompt that is shared with the model. This approach often involves a fine-tuning process that consists of humans redefining their prompt based on the generated responses from the model. An iterative process for finding out an accurate way to prompt the model to get the expected generated responses is needed. This could be a challenge specially for non-expert creators or in early explorative phases where there's only a vague definition of the creative intention rather than a well structured idea of what we want.

We believe that generative models are great for concept exploration helping creators to better define their intentions, transforming a vague definition to a well-defined output. To do that, new interaction models that allow to overcome fixation and creative blocks allowing creators to move forward in their process are needed. In addition, despite the fact that there is a wide range of interfaces to interact with these models without the possibility of interconnecting them in a single space.

In our work, we present Fermat, a programmable spatial canvas that leverages the benefits of using an spatial interface for exploration with the power of different LLM models to allow the creative community to experiment and generate multimodal content.

2 Our approach

We present a scenario for image generation ¹ that illustrate how humans and AI-powered tools can share initiative [6] in the creative process within Fermat. This template allows non-expert creators to

¹An online demo is available in https://app.fermat.ws/permalink/?perma=632b129979cb7e67296c96be

³⁶th Conference on Neural Information Processing Systems (NeurIPS 2022).

a cat with a hat	Generate Prompt	A cat with a hat / monochrome / ink drawing / / black and white / super detailed / pet / cute / HD / domestic	Cenerate Image	
	Senerate Prompt		Senerate Image	

Figure 1: Generating content from a user selection

overcome creative block by generating text [1] and images [5] based on user-defined content. To do that, users must select content within the canvas and use any of the provided tools with predefined prompts to generate content. By predefining prompts and embedding them into buttons, users can directly choose which action to perform without having to define their own prompts to interact with a LLM. As an example, in Figure 2, user starts with a basic description and the system provides with a more detailed prompt suitable for an image generation model such as Stable Diffusion [5]. Hence, the end-user is able to chain different AI models within the same canvas during their workflow.

This template provides users with four AI-powered buttons to generate content:

- Generate prompt: generate a more detailed description from a simple one (Figure 2)
- **Text variations**: given a text, generate variation content As an example, given Van Gogh it provides other artists such Pablo Picasso, Salvador Dalí or Leonardo da Vinci.
- Get keywords: given an image or text, get keywords and split them in different blocks to allow recombination.
- Generate Image: given a prompt description, generate an image

Users can decide to select one or multiple blocks within the canvas as input to the generative model. By combining the generated content from AI-powered buttons with their own content, users can assume a role of directing creative exploration. They can decide when to get inspiration from the generative model or provide their own expertise when needed as shown in Figure 2.



Figure 2: Fermat toolkit for Image Generation.

3 Conclusion

In this paper, we propose combining multiple LLM based tools within a spatial canvas to support nonexpert users to generate images from simple descriptions. We leverage models generative capabilities with direct manipulation of content within this spatial canvas to encourage creative exploration. As shown in Figure 2, having an spatial canvas allows users to both curate and generate content by performing simple actions in the space. In the future, we are planning to explore new interaction models that use other spatial affordances to extend our current work.

4 Ethical implications

Despite the fact that using these models can provide powerful ways to explore creative spaces they are trained in large datasets which are not often available for general public. Then, users from these new generative tools may not be aware from previous artistic work which influenced this generated content. In order to support discoverability of original artistic work, we can provide references to these original datasets. In addition, once a new image is generated, we can perform reverse image search to the original dataset to encourage users to explore artists' original work and also get inspiration from them.

References

[1] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... Amodei, D. (2020). Language models are few-shot learners. Advances in neural information processing systems, 33, 1877-1901.

[2] Thoppilan, R., De Freitas, D., Hall, J., Shazeer, N., Kulshreshtha, A., Cheng, H. T., ... Le, Q. (2022). Lamda: Language models for dialog applications. arXiv preprint arXiv:2201.08239.

[3] Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., Chen, M. (2022). Hierarchical text-conditional image generation with clip latents. arXiv preprint arXiv:2204.06125.

[4] Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E., ... Norouzi, M. (2022). Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. arXiv preprint arXiv:2205.11487.

[5] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 10684-10695).

[6] Deterding, S., Hook, J., Fiebrink, R., Gillies, M., Gow, J., Akten, M., ... Compton, K. (2017, May). Mixed-initiative creative interfaces. In Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (pp. 628-635).