

PromptCrafter: Crafting Text-to-Image Prompt through Mixed-Initiative Dialogue with LLM

Seungho Baek, Hyerin Im, Jiseung Ryu, Juhyeong Park and Takyeon Lee

KAIST (Korea Advanced Institute of Science and Technology)



Motivation

Recent works have proposed a variety of interaction methods that help users understand the capabilities of models and utilize them. However, how to support users to efficiently explore the model's capability and to create effective prompts are still open-ended research questions. In this paper, we present *PromptCrafter*, a novel mixed-initiative system that allows step-by-step crafting of text-to-image prompt through Mixed-Initiative Dialogue with LLM. Through the iterative process, users can efficiently explore the model's capability, and clarify their intent.

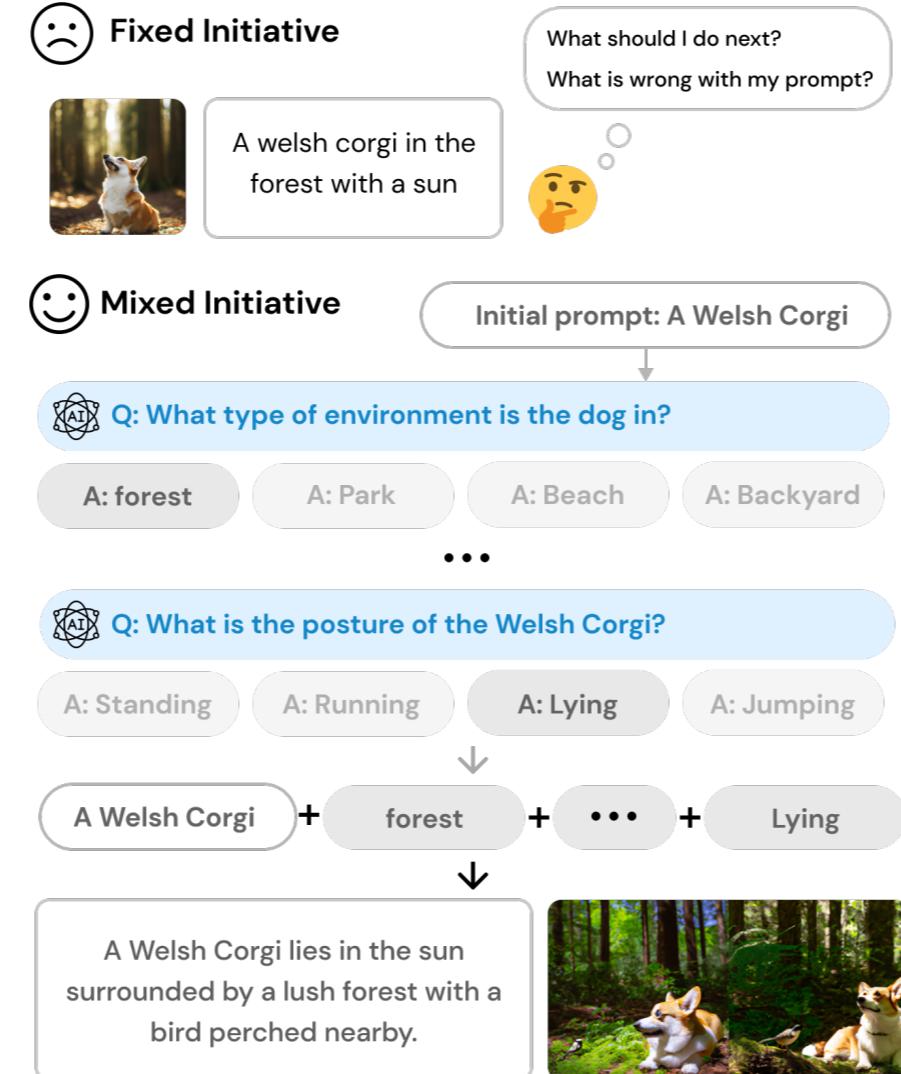
Problem

- P1:** Hard to identify a part of a prompt that causes undesired results, and to make necessary modifications.
- P2:** Compromise on incomplete outcomes rather than adding content that cannot guarantee better results and have difficulty in prompt engineering.
- P3:** Difficult to utilize LTGM's capability in the workflow, as only certain prompts can be used due to the issue with P2.

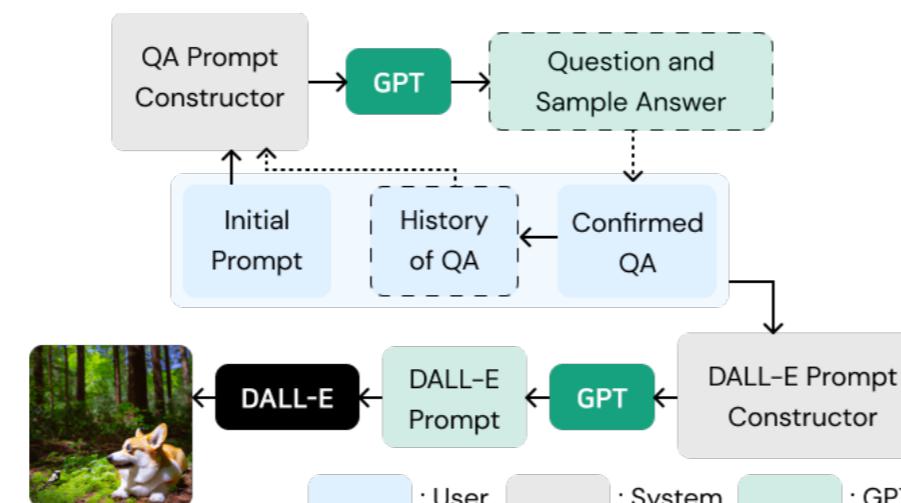
Design Goal

- G1:** Decomposing prompt completion into smaller steps that deal with single idea to make problem identification and correction easier.
- G2:** Providing examples and multiple results to be compared simultaneously in order to explore various image generation directions.
- G3:** Visualizing the workflow enables the easy understanding of the LTGM working process and supports for the generation of various results.

Mixed-Initiative QA

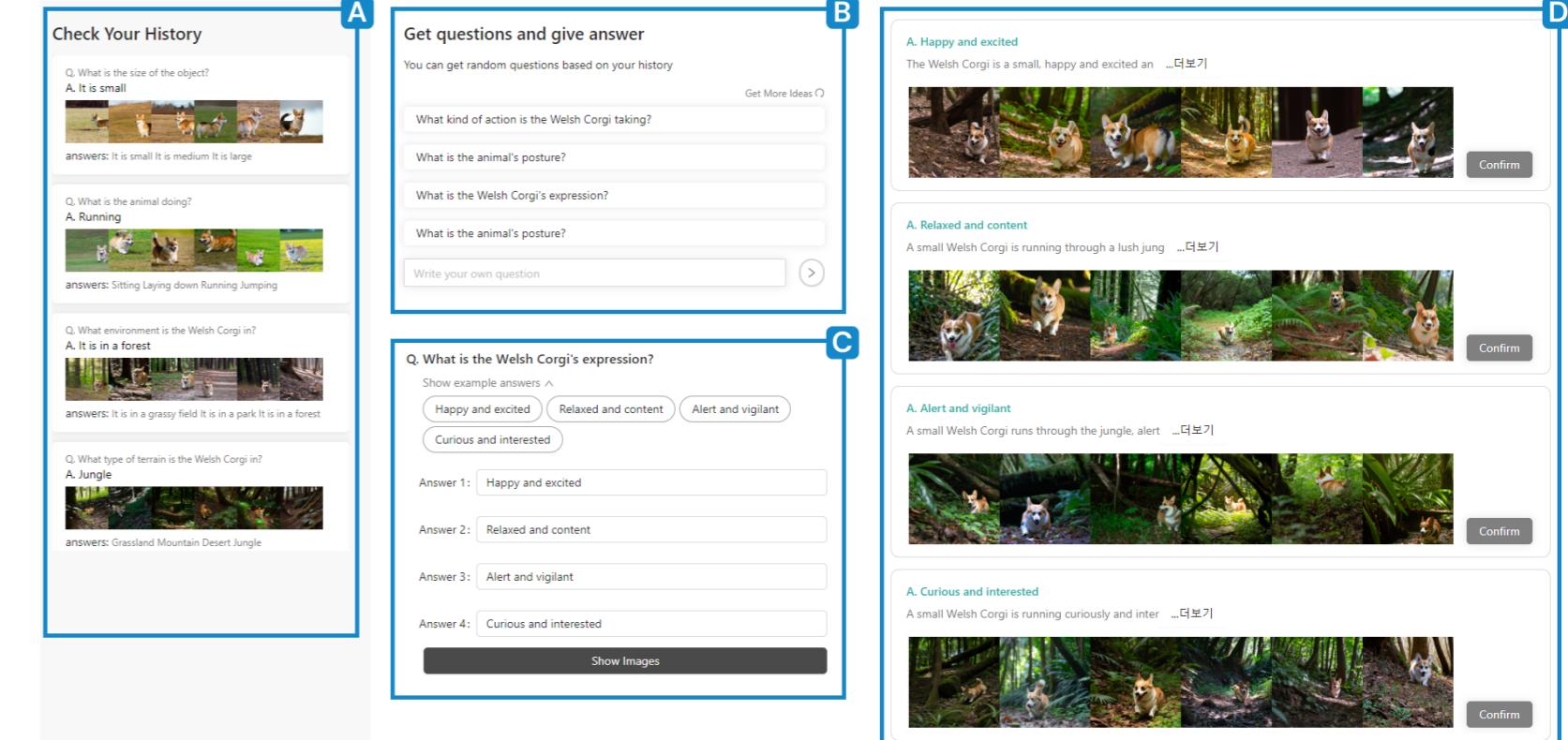


Implementation



PromptCrafter is a web application that was developed using React framework for its interface, and communicates with a back-end server that operates an AI model through API.

PromptCrafter



Question-Answer Panel (B and C)

Question panel (B) proposes four (or even more on request) LLM-generated clarifying questions. By selecting the most interesting question, users decide what information to be added at the current step – i.e., how to expand / clarify their intent. When a question is selected, LLM proposes four sample answers in the answer panel (C).

Confirmation Panel (D)

generates text-to-image prompt and six images for each answer based on previous QA histories and current responses. Users can compare the images, and press "Confirm" button of the most desired set of images.

History Panel (A)

shows the list of previous steps, and allows users to return to any step, in order to create images of different topics. And if desired, user can also revert back to that step to explore other answers, compare them, and create a different image than the initial work

Future Work

We plan to evaluate a user study to verify if the use of *PromptCrafter* is effective. The research goals for the user study are: 1) effectively support the LTGM prompt engineering process?, 2) support the process of clarifying ideas?, and 3) assists in understanding and utilizing LTGM?