



Improving Text-to-Image Consistency via Automatic Prompt Optimization

Oscar Mañas*, Pietro Astolfi*, Melissa Hall, Candace Ross, Jack Urbanek, Adina Williams, Aishwarya Agrawal, Adriana Romero-Soriano, Michal Drozdal

Motivation

- SOTA text-to-image (T2I) models excel at generating high-quality, aesthetically appealing images
- However, T2I models still struggle to **generate images consistent with the prompt**, or they need many “trials” (random seeds).

“a white cat and a black dog”



Stable Diffusion XL v1.0

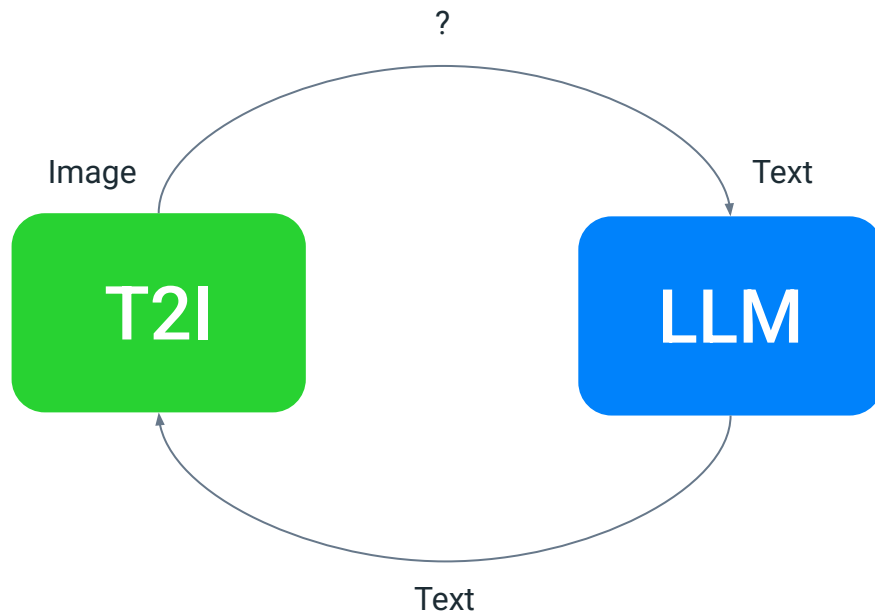
“a horse riding an astronaut”



Stable Diffusion 3.5 Large

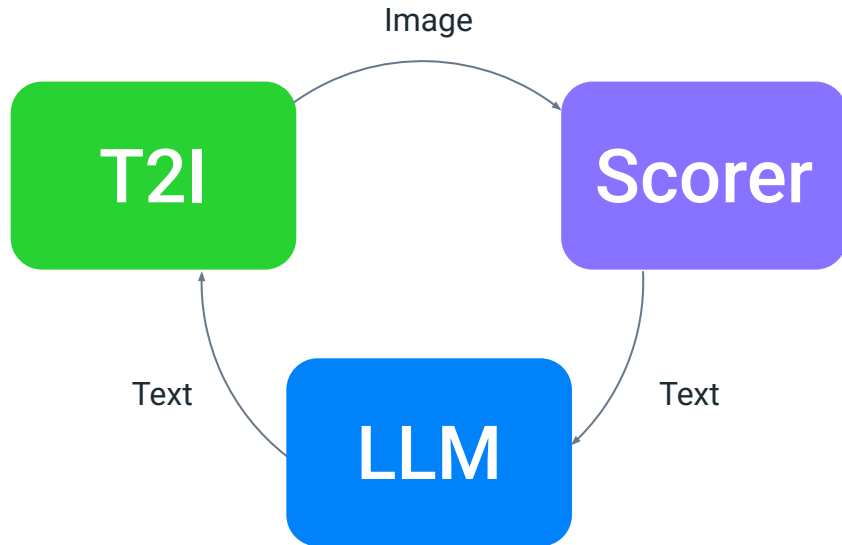
Approach

- T2I + LLM:
 - Optimize the user prompt with an LLM to generate more consistent images

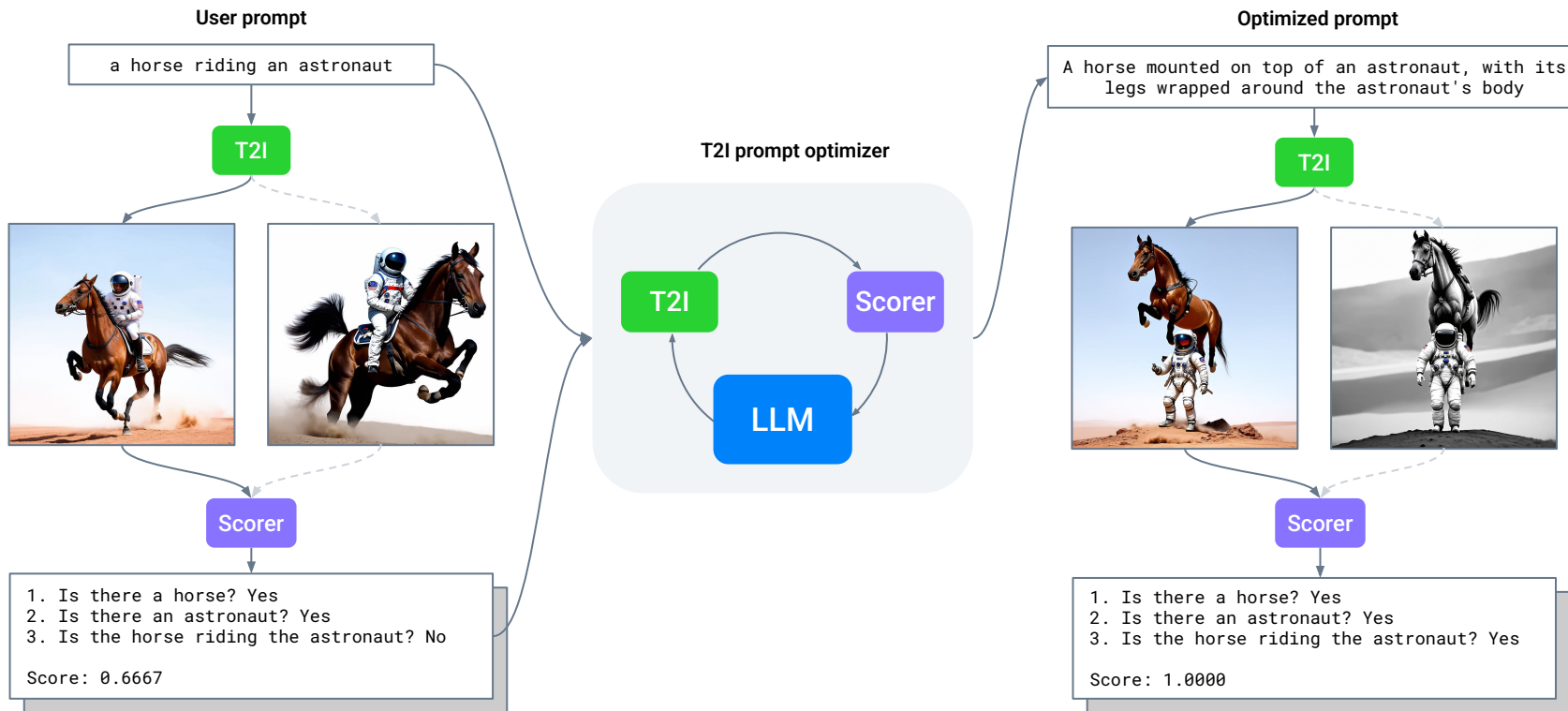


Approach

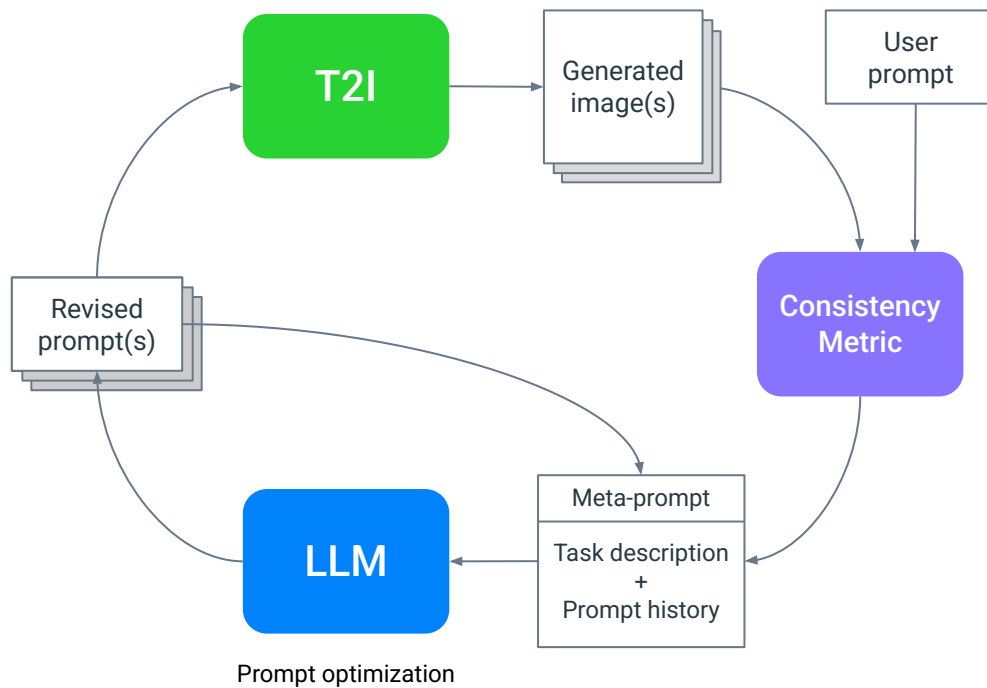
- T2I + LLM:
 - Optimize the user prompt with an LLM to generate more consistent images



Method

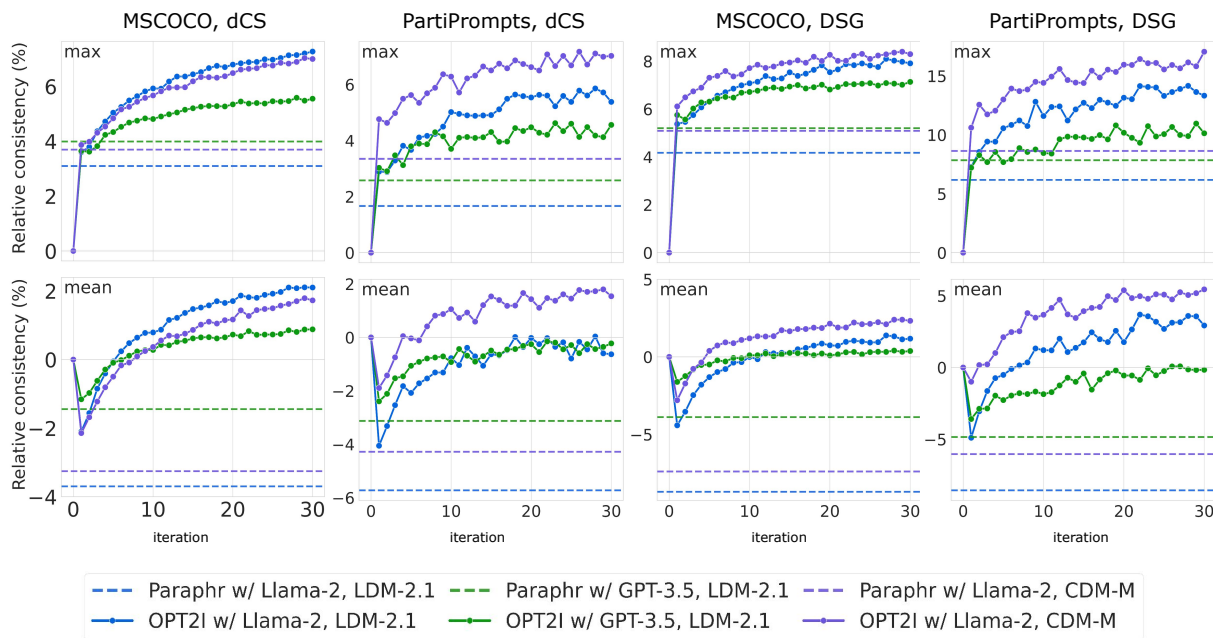


Method



Results

- Optimization curves tracking max and mean **relative improvement in T2I consistency** across revised prompts per iteration
- **Overall upward trend** confirms that the LLM in OPT2I is capable of **optimizing T2I prompts**



Thank you!

