POET: Preference Optimization for Enhanced Text-to-Image Generation

Ruibo Chen[1,2,⋆] , Jiacheng Pan[1,⋆] , Heng Huang[2,†] , Zhenheng Yang[1,†]

[1]TikTok, [2]University of Maryland, College Park
{ruibo.chen,panjiacheng.666,yangzhenheng}@bytedance.com,
heng@umd.edu

Abstract. Recent advances in text-to-image (T2I) generation have achieved impressive results, yet existing models often struggle with simple or underspecified user prompts due to a distributional gap with their descriptive training captions. This frequently leads to suboptimal image-text alignment, aesthetics, and overall visual quality. To bridge this gap, we propose POET (**P**reference **O**ptimization for **E**nhanced **T**ext-to-Image generation), an automated prompt rewriting framework that leverages large language models (LLMs) to refine user inputs before feeding them into frozen T2I backbones. POET introduces a carefully designed composite reward system and an iterative Direct Preference Optimization (DPO) training pipeline, enabling the rewriter to learn model-preferred prompt structures directly from multimodal feedback without requiring costly high-quality supervised fine-tuning (SFT) data. Extensive evaluations across diverse T2I models and benchmarks show that our prompt rewriter consistently improves image-text alignment, visual quality, and aesthetics, outperforming strong baselines. Furthermore, we demonstrate strong transferability by showing that a rewriter trained on one T2I backbone generalizes effectively to others without needing to be retrained. These findings highlight that POET is an effective, robust, and practical model-agnostic strategy for improving T2I systems.

Keywords: Text-to-Image Generation · Prompt Rewriting · Reinforcement Learning

1 Introduction

Text-to-image (T2I) generation has progressed rapidly along several fronts, including diffusion models [6, 13, 28], autoregressive models [3, 26], and emerging paradigms such as next-scale prediction [23], all of which demonstrate strong performance in terms of image quality and fidelity. However, one important direction remains relatively underexplored: a simple, *model-agnostic*, and *training-free* pipeline that systematically addresses common shortcomings, such as weak

⋆ Equal Contribution. Work done while Ruibo Chen was interning at TikTok.
† Corresponding Authors.

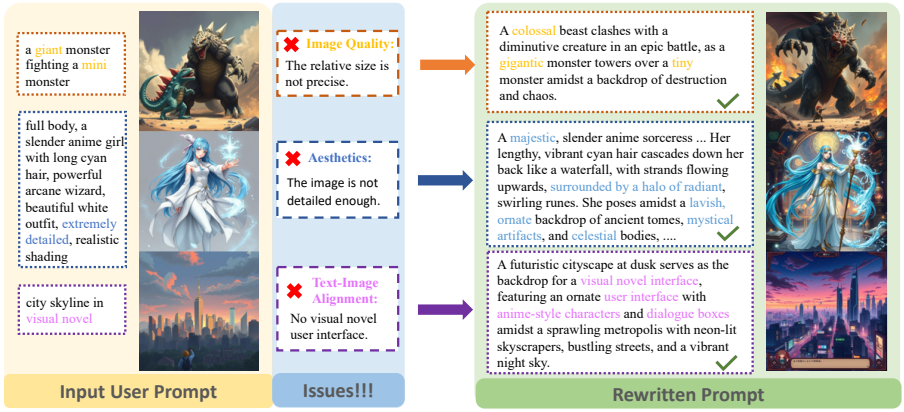


Fig. 1: Overview of the POET framework for automated prompt rewriting. Input user prompts often lead to suboptimal generation results due to issues in image quality, aesthetics, or text-image alignment. POET rewrites these prompts to enhance detail and semantic clarity, resulting in significantly improved visual outputs across various categories.

text–image alignment and inconsistent aesthetics, by modifying only the input text.

In this work, we study an input-side strategy that treats the T2I model as a black box and refines the user prompt, leaving the model parameters unchanged. This direction carries a *two-fold significance*: (i) from a practical perspective, real-world prompts are often short, vague, or underspecified. This creates a significant distributional gap between user inputs and the highly descriptive captions T2I models natively prefer. Rewriting these queries into clearer, more complete instructions mitigates alignment errors and stabilizes stylistic consistency; (ii) from a methodological perspective, it is of great interest to systematically study the achievable gain from optimizing the inputs alone.

Prior work refines prompts by using the In-Context Learning ability of LLMs to expand or restructure the input text [11, 28], and by employing human-in-the-loop systems [2, 7] that iteratively adjust prompts with feedback from complex engineering pipelines and human evaluators. These methods are not fully automated and depend on substantial human effort. Other approaches [1, 5] train dedicated rewriters with supervised fine-tuning (SFT) on curated pairs of short and refined prompts. However, because different T2I models encode different preferences for "good" prompts, collecting high-quality, model-specific annotations is costly, may not transfer well across backbones, and ties improvements to a particular target model and dataset.

To bridge this distributional gap without the prohibitive costs of SFT, we propose POET (Preference Optimization for Enhanced Text-to-Image generation), an automated, reinforcement learning-driven prompt rewriting framework. Rather than manually engineering prompts or retraining the T2I backbone, POET optimizes the prompt that feeds a *frozen* T2I model. Concretely, we train a prompt

rewriter using iterative Direct Preference Optimization (DPO) [19]. Starting from a short user-provided prompt, the rewriter generates a set of candidate refined prompts. Each candidate is then used to synthesize images with a frozen T2I model, which are subsequently evaluated by multimodal LLM judges according to a composite reward function that integrates (i) image quality, (ii) aesthetics, and (iii) text–image alignment. Pairwise preferences inferred from these reward scores are employed to update the rewriter iteratively via DPO. Empirical evaluations on real-world user prompts demonstrate that our approach attains state-of-the-art performance, while avoiding the substantial costs associated with SFT data collection and curation. We further summarize our contributions below:

- We introduce POET, an **RL-driven prompt rewriting framework** to enhance T2I model performance, which is both **model-agnostic** and **training-free** for the T2I generation models. Our policy optimizes only the input text, improving both diffusion and autoregressive T2I backbones *without* SFT pairs, parameter updates, or architectural access.
- Through extensive experiments across multiple LLMs, T2I models, and benchmarks, we show that POET consistently improves T2I performance in terms of image quality, aesthetics, and text–image alignment, achieving state-of-the-art results.
- We conduct a systematic study of POET's transferability. Specifically, our RL-trained rewriter *transfers* across diverse T2I backbones, quantifying cross-model robustness without per-model adaptation.

2 Related Work

Prompts play a pivotal role in determining the quality and fidelity of outputs in downstream tasks. Consequently, a growing body of research [12, 16, 27] has investigated techniques for prompt refinement in natural language processing. In the context of text-to-image (T2I) generation, prompt refinement methods can be broadly categorized as follows.

2.1 In-Context Learning

In-Context Learning (ICL) has demonstrated strong effectiveness across a wide range of textual tasks. Extending this concept to text-to-image generation, Zeng et al. [32] formally define T2I ICL and introduce a benchmark for evaluating the in-context capabilities of T2I models. Several recent models, including Image-Qwen [28], Emu2 [22], and ELLA [11], report performance gains attributable to ICL. Furthermore, Lee et al. [15] apply in-context few-shot learning strategies to further enhance T2I generation quality. While effective, ICL relies heavily on prompt sensitivity and context window limits, motivating more systematic optimization approaches.

2.2 Chain-of-Thought Reasoning

Chain-of-thought (CoT) reasoning has proven highly effective for complex text-based tasks. Guo et al. [9] provide the first systematic study on adapting CoT techniques to autoregressive image generation, introducing specialized reward models that iteratively verify and refine outputs during the generation process. Complementarily, ImageGen-CoT [17] proposes an automated pipeline for constructing high-quality datasets to fine-tune unified multimodal large language models, thereby enhancing their contextual reasoning abilities and improving image generation quality. Recent frameworks like UniGen [24] further explore bridging language and vision via unified generation pathways. In contrast to these methods, which often intervene directly in the multi-step generation or reasoning process, `POET` isolates the optimization entirely to the input prompt, treating the T2I backbone as a frozen environment.

2.3 Prompt Refinement

Learning-free Methods. A number of works [2, 7] explore interactive refinement approaches that leverage human input to produce more detailed and effective prompts for image synthesis. Manas et al. [18] propose a backpropagation-free optimization strategy that automatically rewrites prompts to improve prompt–image consistency. Similarly, Chen et al. [4] present a method that exploits historical user interactions to rewrite prompts, thereby better aligning generated images with individual user preferences. However, these methods are not fully automated and often depend on substantial human effort or historical data.

Supervised Fine-Tuning. A straightforward strategy for automated prompt refinement is supervised fine-tuning (SFT). For instance, DALL-E 3 [1] employs an auxiliary image captioner trained to generate high-quality captions, which are subsequently used to recaption images and enhance T2I training. Datta et al. [5] introduce a Prompt Expansion framework that automatically generates multiple aesthetically rich prompt variants from a single user query, thereby improving generation quality. Despite their success, SFT approaches require the expensive curation of high-quality, model-specific annotations, which may not transfer well across different T2I backbones.

Reinforcement Learning. Reinforcement learning (RL) has also been applied to prompt refinement. An early study by Hao et al. [10] utilized PPO [20] to fine-tune a GPT-2 model for prompt rewriting. Building on this direction, Wu et al. [30] combined SFT and PPO to mitigate prompt toxicity and generate safer images. Lee et al. [14] proposed a multi-reward RL framework that jointly fine-tunes a T2I diffusion model and a prompt-expansion network. A concurrent work, RePrompt [29], explores RL-based refinement for T2I prompts. Our framework, `POET`, distinguishes itself by bypassing SFT entirely and employing Direct Preference Optimization (DPO) solely on the rewriter, avoiding the instability of PPO and the need to jointly update the T2I model. A detailed comparison between our method and RePrompt is provided in Appendix Section 8.

3 Method

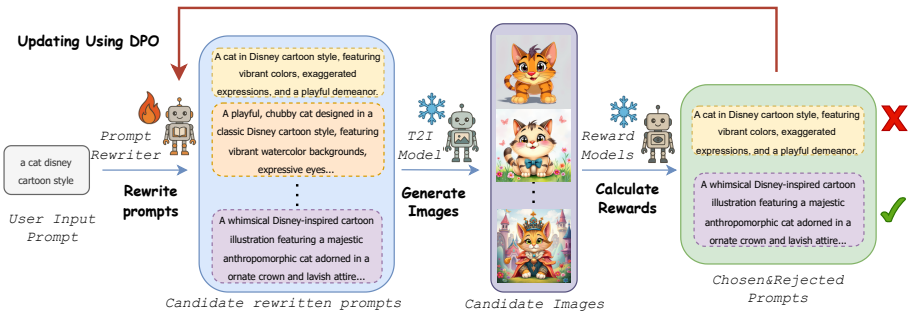


Fig. 2: POET Training Pipeline. Given a user input prompt, the rewriter generates multiple candidate refinements, which are used by a frozen T2I model to synthesize images. Reward models evaluate these images to produce pairwise preferences. The chosen and rejected prompt pairs are then used to update the rewriter via Direct Preference Optimization, and the process is repeated iteratively for multiple rounds. The T2I model is kept frozen.

The core of POET is an iterative Direct Preference Optimization (DPO) pipeline that trains an LLM-based prompt rewriter without relying on Supervised Fine-Tuning (SFT). Our results demonstrate that the rewriter can successfully learn the complex task of prompt expansion and alignment purely through reinforcement learning.

For each short input prompt, the policy generates n candidate rewrites, which are then passed to a frozen T2I model to synthesize images. A multimodal LLM (MLLM) judge evaluates these images and provides pairwise preferences across multiple criteria, acting as the reward signal. The aggregated preferences yield a single winner–loser pair per prompt, which is used to optimize the DPO loss. This procedure is repeated over multiple rounds to progressively refine the rewriter's understanding of the T2I model's implicit preferences. The overall POET pipeline is illustrated in Figure 2.

3.1 Iterative DPO without SFT

SFT typically requires model-specific data curation, which is both costly and prone to human biases tied to a particular backbone, thereby hindering transferability. Moreover, our preliminary experiments revealed that SFT diminishes the policy's exploratory capacity, resulting in overfitting to the training prompts rather than exhibiting robust generalization at test time. To overcome these limitations, POET bypasses SFT entirely and trains the rewriter exclusively through iterative DPO.

Concretely, at each training round, the rewriter proposes multiple candidate rewrites for a user prompt x. Images synthesized from these rewrites by a frozen

T2I model are scored to produce pairwise preferences. We then update the policy via DPO using the following objective:

$$\mathcal{L}_{\text{DPO}}(\pi_\theta; \pi_{\text{ref}}) = -\mathbb{E}_{(x,y_w,y_l)\sim\mathcal{D}}\left[\log\sigma\Big(\beta\log\frac{\pi_\theta(y_w \mid x)}{\pi_{\text{ref}}(y_w \mid x)} - \beta\log\frac{\pi_\theta(y_l \mid x)}{\pi_{\text{ref}}(y_l \mid x)}\Big)\right], \tag{1}$$

where x is the input prompt, y_w and y_l represent the chosen and rejected rewritten prompts respectively, and π_{ref} is the reference model distribution.

We intentionally do not adopt GRPO [8, 21] or its variants, as these methods are most effective when a reliable scalar reward is available. Instead, we rely on *pairwise* image comparisons to significantly reduce variance, which align natively with DPO. Furthermore, Tong et al. [25] report that in image generation settings involving reasoning, DPO yields stronger in-domain performance than GRPO, empirically validating our choice for the `POET` framework.

3.2 Reward Design

We adopt the *MLLM-as-a-judge* paradigm and employ Qwen2.5-VL-72B-Instruct as the evaluation model. During training, the judge is provided with an instruction alongside *two* candidate images, outputting a pairwise preference or a tie.

To comprehensively capture the primary aspects of user intent and visual fidelity, we define four distinct reward dimensions: (1) image quality, r_{Quality}; (2) general image–text alignment, $r_{\text{General-Alignment}}$; (3) physical image–text alignment, $r_{\text{Physical-Alignment}}$; and (4) aesthetics, $r_{\text{Aesthetics}}$.

Image Quality. The image quality reward evaluates whether the generated image is visually plausible across five dimensions: (1) correctness of human or animal body parts, (2) plausibility of geometric details (e.g., absence of unintended wavy lines, consistency in symmetrical objects), (3) adherence to basic physical laws such as reflections and shadows, (4) semantic and logical coherence of the scene, and (5) correctness of fine-grained details without noticeable noise or distortions.

General Image–Text Alignment. This reward evaluates the degree to which an image adheres to the input prompt at a holistic level. Because assessing a complex prompt directly can be challenging for the judge, we employ a two-step procedure:

1. *Decomposition:* Given the input prompt, the judge generates a concise list of key questions capturing its essential components. For instance, for the prompt "four apples on a table," the decomposition yields: "Are there four apples?", "Is there a table?", and "Are the four apples on the table rather than below it?"
2. *Comparison:* The decomposed questions, alongside the candidate image pair, are provided to the judge to produce a pairwise preference decision.

Physical Image–Text Alignment. This reward targets concrete physical relationships that text-to-image models frequently fail to capture. It evaluates: (i) the correctness of object counts, (ii) spatial relations (e.g., left/right, on/under, in front/behind), and (iii) attribute binding (i.e., whether the specified color, size, or style is accurately associated with the correct object).

Image Aesthetics. The aesthetics reward assesses the relative visual appeal and stylistic quality of two candidate images, taking into account composition, richness of detail, and overall artistic merit.

Based on these rewards, we train two distinct variants of POET: a *general rewriter*, which prioritizes semantic faithfulness and overall image quality, and an *aesthetics rewriter*, which emphasizes visual appeal while maintaining baseline alignment and quality. We train these two policies by combining the rewards as follows:

$$\begin{cases} r^{\text{General}} = r_{\text{Quality}} + r_{\text{General-Alignment}} + r_{\text{Physical-Alignment}} \\ r^{\text{Aesthetics}} = r_{\text{Quality}} + r_{\text{General-Alignment}} + r_{\text{Physical-Alignment}} + r_{\text{Aesthetics}}. \end{cases} \tag{2}$$

More details on the specific prompts designed for the MLLM-as-a-judge reward are provided in Appendix 12.

3.3 Selection of Chosen and Rejected Prompts

For each original prompt, we construct a single chosen-rejected pair for DPO training using pairwise image comparisons, which provide much more reliable supervision than pointwise scoring.

Given a prompt x, POET samples n rewritten candidates and renders n corresponding images. From these, we form all $n(n-1)$ *ordered* pairs. Each ordered pair is evaluated by all reward models associated with the current rewriter variant. For each comparison, the reward model issues a preference: the preferred image receives $+1$, the other receives -1, and a tie assigns 0 to both.

We aggregate these votes across all reward models and all ordered pairs to obtain a cumulative score for each rewritten candidate. The candidate with the highest score is selected as the chosen prompt y_w, while the one with the lowest score is designated as the rejected prompt y_l. The triplet (x, y_w, y_l) is subsequently used as a DPO training instance. The complete procedure for selecting the chosen and rejected prompts is summarized in Algorithm 1.

3.4 Iterative DPO Pipeline

At each iteration of the DPO training, we draw a batch of user prompts and employ the current rewriter policy to generate n candidate rewrites for each prompt. Subsequently, Algorithm 1 is applied to the pairwise image comparisons to identify a single preferred and a single rejected rewrite per prompt. The policy is then updated by minimizing the DPO loss defined in Equation 1. The overall iterative training process for POET is summarized in Algorithm 2.

Algorithm 1 Selection of chosen and rejected rewritten prompts

Require: Original prompt x, rewritten candidates $\{y_1, y_2, \ldots, y_n\}$, text-to-image model g, reward models $\{\mathrm{RM}_k\}_{k=1}^{K}$
  for $i = 1$ to n **do**
    Generate image $\mathrm{Img}_i \leftarrow g(y_i)$
    Initialize rewards $r_i^k \leftarrow 0$ for all $k = 1, \ldots, K$
  end for
  for $i = 1$ to n **do**
    for $j = 1$ to n **such that** $j \neq i$ **do**
      for $k = 1$ to K **do**
        if $\mathrm{Img}_i \succ \mathrm{Img}_j \mid \mathrm{RM}_k, x$ **then**
          $r_i^k \leftarrow r_i^k + 1, \quad r_j^k \leftarrow r_j^k - 1$
        else if $\mathrm{Img}_i \prec \mathrm{Img}_j \mid \mathrm{RM}_k, x$ **then**
          $r_i^k \leftarrow r_i^k - 1, \quad r_j^k \leftarrow r_j^k + 1$
        end if ▷ No update if $\mathrm{Img}_i \approx \mathrm{Img}_j \mid \mathrm{RM}_k, x$
      end for
    end for
    Aggregate rewards: $R_i \leftarrow \sum_{k=1}^{K} r_i^k$
  end for
  Select chosen prompt $y_w \leftarrow y_{i_w}$, $i_w \leftarrow \arg\max_i R_i$
  Select rejected prompt $y_l \leftarrow y_{i_l}$, $i_l \leftarrow \arg\min_i R_i$
  return (y_w, y_l)

4 Experiments

To evaluate the effectiveness of POET, we conduct comprehensive experiments across multiple text-to-image architectures and benchmarks. Detailed configurations, including the specific LLMs used for the prompt rewriter, the frozen T2I models, training datasets, baseline comparisons, and hyperparameters, are provided in Appendix 7.

4.1 Benchmark Results

We benchmark POET against recent state-of-the-art methods on GenEval. Under a standardized evaluation protocol (Table 1), POET achieves the highest overall GenEval score, outperforming all competing methods across both model families by a substantial margin.

We further test POET on more datasets with different T2I backbones. As shown in Tables 2 and 3, POET achieves strong performance on GenEval, T2I-CompBench++, and the TIFA-Benchmark in terms of text–image alignment, and attains lower FID scores on MS-COCO-30K for image quality. Across all benchmarks, our approach consistently improves alignment, for example increasing the FLUX.1-dev GenEval score from 0.70 to 0.79, while simultaneously enhancing image quality as reflected by reduced FID.

Algorithm 2 Iterative DPO for POET

Require: Base LLM f_θ, training prompt set $\mathcal{D}$, samples-per-round m, candidates-per-prompt n
  for each round **do**
    Update reference policy $f_{ref} = f_\theta$.
    Sample prompts for each round: $\{x_i\}_{i=1}^{m} \sim \mathcal{D}$.
    for $i = 1$ to m **do**
      Generate rewritten prompts: $\{y_i^j\}_{j=1}^{n} \sim f_\theta(\cdot \mid x_i)$.
      Get chosen and rejected rewritten prompts (y_i^w, y_i^l) with Alg. 1.
    end for
    for each step **do**
      Sample a batch B from $\{(x_i, y_i^w, y_i^l)\}_{i=1}^{m}$,
      Update f_θ with f_{ref} and B using Eq. 1.
    end for
  end for
  return f_θ

4.2 MLLM-as-a-Judge Evaluations

On the Pick-a-Pic v2 dataset, using GPT-4o as a pairwise judge, POET consistently outperforms the original test prompts. Crucially, these improvements are achieved entirely on the input side, requiring no additional training or parameter updates to the underlying T2I backbones. As reported in Table 4, the *general rewriter* variant attains the highest win rate for text-image alignment while simultaneously enhancing overall image quality and aesthetics. Conversely, the *aesthetics rewriter* is highly effective at maximizing visual appeal, achieving a striking aesthetics win rate of 0.818, while still maintaining competitive performance on both image quality and alignment.

These robust gains extend to the PartiPrompts benchmark (detailed in Table 5). GPT-4o evaluations demonstrate that POET delivers consistent improvements across all tested T2I backbones. Specifically, the win rates of our rewritten prompts compared to the original, underspecified user inputs exceed 0.5 across every evaluation dimension.

We further observe that simple in-context learning (ICL) prompts frequently improve generation quality. Contemporary T2I models often underperform when conditioned on short or underspecified prompts, which are prevalent in real-world usage, whereas straightforward prompt expansions provide consistent benefits. This phenomenon likely arises from a distributional gap, as training captions are generally longer and more descriptive than prompts provided by real users.

4.3 Human Evaluations

To verify that our automated MLLM-as-a-judge metrics accurately reflect true human preference, we conducted a rigorous human evaluation study. As shown in Table 6, human annotators consistently prefer images generated using POET refined prompts over the original baselines across all three dimensions. Notably, the

Table 1: GenEval results compared with state-of-the-art methods. Using the general rewriter with Llama-3-70B-Instruct as the LLM backbone, `POET` achieves the best overall GenEval score (higher is better). **Bold** is the best and underline the second-best. †Results reported by Wu et al. [29]. ‡Results reported by Guo et al. [9].

Method	T2I Backbone	Single object	Two object	Counting	Colors	Position	Color attribution	Overall
Baseline‡	Show-o	0.95	0.52	0.49	0.82	0.11	0.28	0.53
Baseline	FLUX.1-dev	1.00	0.87	0.76	0.84	0.22	0.49	0.70
PARM‡	Show-o	0.99	0.77	0.68	0.86	0.29	0.45	0.67
PARM++‡	Show-o	0.99	0.71	0.69	**0.95**	0.36	0.49	0.70
Idea2Img†	FLUX.1-dev	-	-	-	-	-	-	0.69
Zero-Shot	FLUX.1-dev	0.98	0.89	0.71	0.82	0.47	0.49	0.73
ICL	FLUX.1-dev	**1.00**	0.89	**0.80**	0.83	0.54	0.41	0.74
RePrompt†	FLUX.1-dev	0.98	0.87	0.77	0.85	**0.62**	0.49	0.76
`POET` (Ours)	FLUX.1-dev	**1.00**	**0.95**	0.78	0.88	0.58	**0.56**	**0.79**

human judgments closely mirror the GPT-4o evaluations. This strong correlation not only confirms the real-world effectiveness of our framework, but also validates the evaluation process.

4.4 Qualitative Analysis

We provide qualitative examples in Appendix 11 that trace the evolution of rewritten prompts throughout the iterative DPO training process. Additionally, we visualize the stylistic differences between the outputs of the general rewriter, which prioritizes faithfulness, and the aesthetics rewriter, which focuses on visual appeal. Detailed examples can be found in Appendix Table 17.

5 Analyses

5.1 Reward Type

Appendix Table 11 presents an ablation study of the reward terms. Removing the quality reward substantially decreases the image quality win rate, from 0.494 to 0.364. Excluding either the general or physical alignment rewards reduces alignment, from 0.561 to 0.521 and to 0.504 respectively, confirming that each reward effectively drives its intended objective. Incorporating the aesthetics reward yields a pronounced improvement in aesthetics, increasing from 0.476 to 0.818, but simultaneously reduces alignment from 0.561 to 0.424, highlighting a trade-off. Qualitative analysis in Appendix Table 17 suggests that the aesthetics reward often enriches scenes by introducing additional objects or ornamentation, which can dilute the main subject and thereby weaken alignment. To accommodate different objectives, we report two variants: a general rewriter optimized for faithfulness and overall quality, and an aesthetics rewriter tailored to visual appeal.

Table 2: GenEval results for the general rewriter across T2I backbones. POET consistently outperforms the original prompts and ICL baselines, and performance improves with LLM size; Llama-3-70B-Instruct achieves the best results.

Model	Single object	Two object	Counting	Colors	Position	Color attribution	Overall
FLUX.1-schnell	0.99	0.87	0.61	0.79	0.35	0.43	0.67
+ ICL	0.99	0.87	0.55	0.81	0.50	0.50	0.69
+ POET (Llama 3 3B)	**1.00**	0.89	0.63	0.81	0.47	**0.59**	0.73
+ POET (Llama 3 8B)	**1.00**	**0.93**	0.58	**0.86**	0.55	0.55	0.74
+ POET (Llama 3 70B)	0.99	0.91	**0.65**	0.77	**0.64**	0.57	**0.75**
FLUX.1-dev	**1.00**	0.87	0.76	0.84	0.22	0.49	0.70
+ ICL	**1.00**	0.89	**0.80**	0.83	0.54	0.41	0.74
+ POET (Llama 3 70B)	**1.00**	**0.95**	0.78	**0.88**	**0.58**	**0.56**	**0.79**
SD-3.5-medium	**1.00**	0.84	**0.68**	0.85	0.24	**0.61**	0.70
+ ICL	0.96	**0.92**	0.60	**0.86**	0.42	0.54	0.72
+ POET (Llama 3 70B)	**1.00**	**0.92**	**0.68**	0.85	**0.57**	0.56	**0.76**

5.2 Rewritten Prompt Length

Appendix Figure 6 illustrates a steady increase in the average token length of rewritten prompts over training in POET. The rewriter progressively incorporates additional specificity, such as attributes, relations, and constraints, and this increase in length correlates with higher win rates.

5.3 Full Finetune vs. LoRA

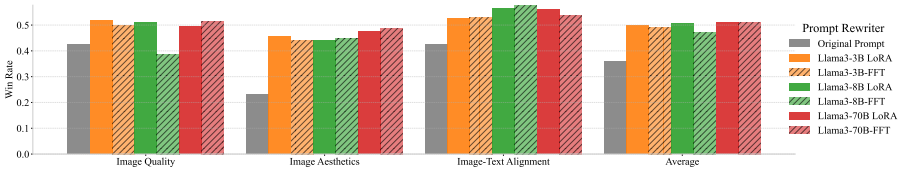


Fig. 3: Full finetune vs. LoRA. win rates against DALL-E 3. Evaluated on the Pick-a-pic v2 datasets with GPT-4o judge. We use Llama3-70B-Instruct as the general rewriter.

We conduct a controlled comparison between LoRA and full-model finetuning under identical experimental conditions, as shown in Figure 3. Full-model finetuning does not yield consistent improvements, whereas LoRA attains comparable performance while requiring substantially less memory and computation. These results suggest that LoRA constitutes a strong and practical choice for the training in POET.

Table 3: Results on T2I-CompBench++, TIFA-Benchmark, and FID on MS COCO 30K using the general rewriter with Llama-3-70B-Instruct as the LLM backbone. POET improves text–image alignment and image quality across all evaluated T2I backbones.

	T2I-CompBench++						TIFA	COCO
	Color↑	Shape↑	Texture↑	Spatial↑	Numeracy↑	Complex↑	Score↑	FID↓
FLUX.1-schnell	0.7492	0.5673	0.6911	0.2754	0.6062	0.3614	0.8803	20.57
+ POET	**0.7614**	**0.6010**	**0.7063**	**0.3216**	**0.6161**	**0.3713**	**0.8868**	**17.76**
FLUX.1-dev	0.7647	0.5053	0.6336	0.2763	0.6130	0.3586	0.8572	24.38
+ POET	**0.7978**	**0.5834**	**0.6929**	**0.3206**	**0.6343**	**0.3694**	**0.8809**	**19.57**
SD-3.5-medium	0.7988	0.5800	0.7206	0.2889	0.6033	0.3614	0.8782	17.81
+ POET	**0.8040**	**0.5855**	**0.7346**	**0.3322**	**0.6320**	**0.3707**	**0.8878**	**17.13**
JanusPro	0.5294	0.3247	0.4168	0.1579	0.4380	0.3778	0.8457	19.28
+ POET	**0.7861**	**0.5892**	**0.7220**	**0.2773**	**0.5983**	**0.3911**	**0.8845**	**16.71**

Table 4: Pick-a-Pic v2 results with Llama-3-70B-Instruct as the LLM backbone. Reported values are GPT-4o–judged win rates vs. DALL·E 3. The general rewriter achieves the highest text–image alignment while also improving quality and aesthetics; the aesthetics rewriter delivers the best aesthetics while maintaining quality and alignment. **Bold** is the best and underline the second-best.

	Image Quality	Image Aesthetics	Image-Text Alignment	Average	LAION
Dalle-3	0.500	0.500	0.500	0.500	6.19
FLUX.1-schnell	0.469	0.314	0.419	0.401	6.07
+ ICL	0.475	0.307	0.422	0.401	6.07
+ POET (General Rewriter)	0.494	0.476	**0.561**	0.510	6.08
+ POET (Aesthetics Rewriter)	**0.495**	**0.818**	0.424	**0.579**	**6.39**
FLUX.1-dev	0.513	0.350	0.360	0.408	6.22
+ ICL	0.531	0.363	0.457	0.450	6.17
+ POET (General Rewriter)	0.536	0.491	**0.575**	0.534	6.06
+ POET(Aesthetics Rewriter)	**0.576**	**0.800**	0.391	**0.589**	**6.41**
SD-3.5-medium	0.424	0.230	0.425	0.359	5.74
+ ICL	0.487	0.327	0.455	0.423	5.92
+ POET(General Rewriter)	**0.504**	0.456	**0.542**	0.501	5.95
+ POET(Aesthetics Rewriter)	0.481	**0.799**	0.401	**0.560**	**6.23**
JanusPro	0.193	0.181	0.392	0.255	5.89
+ ICL	**0.213**	0.180	0.335	0.243	5.97
+ POET(General Rewriter)	0.183	0.268	**0.421**	0.291	5.95
+ POET(Aesthetics Rewriter)	0.200	**0.612**	0.317	**0.377**	**6.23**

Table 5: PartiPrompts [31] results. Values are GPT-4o–judged win rates (higher is better) of our rewritten prompts vs. the original prompts. POET consistently improves image quality, alignment, and aesthetics.

	Image Quality	Image Aesthetics	Image-Text Alignment	Average
FLUX.1-schnell	0.524	0.583	0.592	0.566
FLUX.1-dev	0.510	0.584	0.625	0.573
SD-3.5-medium	0.599	0.743	0.542	0.628
JanusPro	0.616	0.674	0.687	0.659

Table 6: Human evaluation and MLLM-as-a-judge results using the FLUX.1-dev backbone with Llama-3-70B-Instruct on the Pick-a-Pic v2 dataset. Values represent the pairwise win rates of POET refined prompts against the original user prompts. The human annotator preferences closely align with the GPT-4o automated judge, demonstrating consistent improvements in image quality, aesthetics, and text-image alignment.

	Evaluator	Image Quality	Image Aesthetics	Image-Text Alignment	Average
FLUX.1-dev	-	0.500	0.500	0.500	0.500
+ POET	GPT-4o	0.523	0.561	0.658	0.581
+ POET	Human	0.552	0.561	0.618	0.577

5.4 LLM Backbones

To understand the impact of the rewriter's architectural family and reasoning capacity, we evaluated POET using three distinct LLM backbones. As detailed in Table 7, all POET variants consistently surpass the FLUX.1-schnell baseline across every metric.

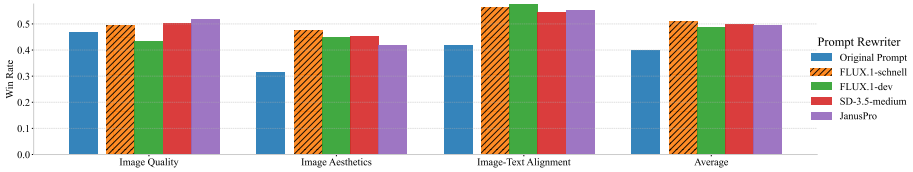
Llama-3-70B-Instruct achieves the highest overall average. The comparatively lower aesthetic performance of the DeepSeek model suggests that heavy reasoning-oriented tuning may occasionally conflict with the exploratory prompt enrichment required to maximize visual appeal. These results confirm that while POET is effective across various LLMs, the choice of backbone allows practitioners to trade off between text-image alignment and aesthetic enhancement.

5.5 Transferability

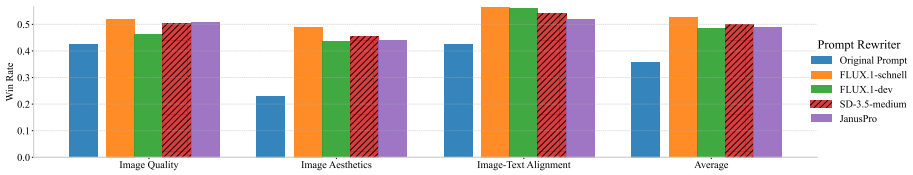
We evaluate transferability by training multiple rewriters, each paired with a different T2I backbone, and subsequently freezing the rewriter for testing on a fixed target backbone. We consider two target models, FLUX.1-schnell and Stable Diffusion 3.5. For each target, we report performance in three conditions: when the training and target T2I backbones are identical, when they differ, and when using the original unmodified prompt as a baseline. Results are presented in Figure 4 and Appendix Table 9.

Table 7: Performance comparison of POET across different LLM backbones using the FLUX.1-schnell T2I model. Values represent GPT-4o-judged pairwise win rates against DALL·E 3. While Llama 3 70B achieves the best overall text–image alignment and average score, Qwen 3 32B demonstrates highly efficient performance, securing the highest image quality score despite a smaller parameter count. **Bold** indicates the best score in each column.

	Image Quality	Image Aesthetics	Image–Text Alignment	Average
FLUX.1-schnell	0.469	0.314	0.419	0.401
+ POET (Qwen 3 32B)	**0.517**	0.424	0.534	0.492
+ POET (DeepSeek 70B)	0.502	0.380	0.506	0.463
+ POET (Llama 3 70B)	0.494	**0.476**	**0.561**	**0.510**



(a) Win rate against DALL-E 3 using FLUX.1-schnell as test-time T2I.



(b) Win rate against DALL-E 3 using SD-3.5-medium as test-time T2I.

Fig. 4: Win rates for rewriters trained on various T2I backbones but evaluated on a fixed target, either (a) FLUX.1-schnell or (b) SD-3.5-medium. Evaluation is on the Pick-a-pic v2 dataset using a GPT-4o judge. Results show strong transferability.

Across both targets, rewritten prompts consistently improve over the original prompts, irrespective of whether the training and testing backbones match. Performance differences under train–test mismatch are generally small and often comparable to the matched setting. These findings suggest that the rewriter acquires prompt refinements that generalize across T2I backbones, enabling reuse without per-model retraining. We hypothesize that this transferability arises because T2I models are commonly trained on image–caption pairs from similar distributions, rendering our rewrites broadly portable.

6 Conclusion

In this work, we introduce POET, an automated, RL-driven prompt rewriting framework designed to bridge the distributional gap between underspecified user prompts and the highly descriptive captions preferred by text-to-image models.

By leveraging iterative Direct Preference Optimization guided by a composite MLLM-as-a-judge reward system, our approach learns to optimize inputs without relying on costly supervised fine-tuning. Consequently, POET consistently improves image quality, text-image alignment, and aesthetics across diverse T2I backbones while leaving their underlying parameters entirely frozen. Extensive experiments further demonstrate the framework's strong cross-model transferability. POET establishes input-side optimization as a highly practical, model-agnostic strategy for advancing T2I systems.

Acknowledgements

Ruibo Chen and Heng Huang were partially supported by NSF IIS 2347592, 2348169, DBI 2405416, CCF 2348306, CNS 2347617, RISE 2536663.

References

1. Betker, J., Goh, G., Jing, L., Brooks, T., Wang, J., Li, L., Ouyang, L., Zhuang, J., Lee, J., Guo, Y., et al.: Improving image generation with better captions. Computer Science. https://cdn. openai. com/papers/dall-e-3. pdf **2**(3), 8 (2023)
2. Brade, S., Wang, B., Sousa, M., Oore, S., Grossman, T.: Promptify: Text-to-image generation through interactive prompt exploration with large language models. In: Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology. pp. 1–14 (2023)
3. Chen, X., Wu, Z., Liu, X., Pan, Z., Liu, W., Xie, Z., Yu, X., Ruan, C.: Janus-pro: Unified multimodal understanding and generation with data and model scaling. CoRR **abs/2501.17811** (2025). https://doi.org/10.48550/ARXIV.2501.17811, https://doi.org/10.48550/arXiv.2501.17811
4. Chen, Z., Zhang, L., Weng, F., Pan, L., Lan, Z.: Tailored visions: Enhancing text-to-image generation with personalized prompt rewriting. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7727–7736 (2024)
5. Datta, S., Ku, A., Ramachandran, D., Anderson, P.: Prompt expansion for adaptive text-to-image generation. In: Ku, L., Martins, A., Srikumar, V. (eds.) Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2024, Bangkok, Thailand, August 11-16, 2024. pp. 3449–3476. Association for Computational Linguistics (2024). https://doi.org/10.18653/V1/2024.ACL-LONG.189, https://doi.org/10.18653/v1/2024.acl-long.189
6. Esser, P., Kulal, S., Blattmann, A., Entezari, R., Müller, J., Saini, H., Levi, Y., Lorenz, D., Sauer, A., Boesel, F., Podell, D., Dockhorn, T., English, Z., Rombach, R.: Scaling rectified flow transformers for high-resolution image synthesis. In: Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024. OpenReview.net (2024), https://openreview.net/forum?id=FPnUhsQJ5B
7. Feng, Y., Wang, X., Wong, K.K., Wang, S., Lu, Y., Zhu, M., Wang, B., Chen, W.: Promptmagician: Interactive prompt engineering for text-to-image creation. IEEE Transactions on Visualization and Computer Graphics **30**(1), 295–305 (2023)

8. Guo, D., Yang, D., Zhang, H., Song, J., Zhang, R., Xu, R., Zhu, Q., Ma, S., Wang, P., Bi, X., et al.: Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. arXiv preprint arXiv:2501.12948 (2025)
9. Guo, Z., Zhang, R., Tong, C., Zhao, Z., Gao, P., Li, H., Heng, P.: Can we generate images with cot? let's verify and reinforce image generation step by step. CoRR **abs/2501.13926** (2025). https://doi.org/10.48550/ARXIV.2501.13926, https://doi.org/10.48550/arXiv.2501.13926
10. Hao, Y., Chi, Z., Dong, L., Wei, F.: Optimizing prompts for text-to-image generation. In: Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., Levine, S. (eds.) Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023 (2023), http://papers.nips.cc/paper_files/paper/2023/hash/d346d91999074dd8d6073d4c3b13733b-Abstract-Conference.html
11. Hu, X., Wang, R., Fang, Y., Fu, B., Cheng, P., Yu, G.: Ella: Equip diffusion models with llm for enhanced semantic alignment. arXiv preprint arXiv:2403.05135 (2024)
12. Kong, W., Hombaiah, S.A., Zhang, M., Mei, Q., Bendersky, M.: Prewrite: Prompt rewriting with reinforcement learning. In: Ku, L., Martins, A., Srikumar, V. (eds.) Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics, ACL 2024 - Short Papers, Bangkok, Thailand, August 11-16, 2024. pp. 594–601. Association for Computational Linguistics (2024). https://doi.org/10.18653/V1/2024.ACL-SHORT.54, https://doi.org/10.18653/v1/2024.acl-short.54
13. Labs, B.F.: Flux. https://github.com/black-forest-labs/flux (2024)
14. Lee, S.H., Li, Y., Ke, J., Yoo, I., Zhang, H., Yu, J., Wang, Q., Deng, F., Entis, G., He, J., et al.: Parrot: Pareto-optimal multi-reward reinforcement learning framework for text-to-image generation. In: European Conference on Computer Vision. pp. 462–478. Springer (2024)
15. Lee, S., Lee, J., Bae, C.H., Choi, M., Lee, R., Ahn, S.: Optimizing prompts using in-context few-shot learning for text-to-image generative models. IEEE Access **12**, 2660–2673 (2024). https://doi.org/10.1109/ACCESS.2023.3348778, https://doi.org/10.1109/ACCESS.2023.3348778
16. Li, C., Zhang, M., Mei, Q., Kong, W., Bendersky, M.: Learning to rewrite prompts for personalized text generation. In: Chua, T., Ngo, C., Kumar, R., Lauw, H.W., Lee, R.K. (eds.) Proceedings of the ACM on Web Conference 2024, WWW 2024, Singapore, May 13-17, 2024. pp. 3367–3378. ACM (2024). https://doi.org/10.1145/3589334.3645408, https://doi.org/10.1145/3589334.3645408
17. Liao, J., Yang, Z., Li, L., Li, D., Lin, K., Cheng, Y., Wang, L.: Imagegen-cot: Enhancing text-to-image in-context learning with chain-of-thought reasoning. CoRR **abs/2503.19312** (2025). https://doi.org/10.48550/ARXIV.2503.19312, https://doi.org/10.48550/arXiv.2503.19312
18. Mañas, O., Astolfi, P., Hall, M., Ross, C., Urbanek, J., Williams, A., Agrawal, A., Romero-Soriano, A., Drozdzal, M.: Improving text-to-image consistency via automatic prompt optimization. CoRR **abs/2403.17804** (2024). https://doi.org/10.48550/ARXIV.2403.17804, https://doi.org/10.48550/arXiv.2403.17804
19. Rafailov, R., Sharma, A., Mitchell, E., Manning, C.D., Ermon, S., Finn, C.: Direct preference optimization: Your language model is secretly a reward model. In: Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., Levine, S. (eds.) Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023 (2023), http://papers.nips.cc/paper_files/paper/2023/hash/a85b405ed65c6477a4fe8302b5e06ce7-Abstract-Conference.html

20. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O.: Proximal policy optimization algorithms. CoRR **abs/1707.06347** (2017), http://arxiv.org/abs/1707.06347
21. Shao, Z., Wang, P., Zhu, Q., Xu, R., Song, J., Bi, X., Zhang, H., Zhang, M., Li, Y., Wu, Y., et al.: Deepseekmath: Pushing the limits of mathematical reasoning in open language models. arXiv preprint arXiv:2402.03300 (2024)
22. Sun, Q., Cui, Y., Zhang, X., Zhang, F., Yu, Q., Wang, Y., Rao, Y., Liu, J., Huang, T., Wang, X.: Generative multimodal models are in-context learners. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14398–14409 (2024)
23. Tian, K., Jiang, Y., Yuan, Z., Peng, B., Wang, L.: Visual autoregressive modeling: Scalable image generation via next-scale prediction. In: Globersons, A., Mackey, L., Belgrave, D., Fan, A., Paquet, U., Tomczak, J.M., Zhang, C. (eds.) Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024 (2024), http://papers.nips.cc/paper_files/paper/2024/hash/9a24e284b187f662681440ba15c416fb-Abstract-Conference.html
24. Tian, R., Gao, M., Xu, M., Hu, J., Lu, J., Wu, Z., Yang, Y., Dehghan, A.: Unigen: Enhanced training & test-time strategies for unified multimodal understanding and generation. arXiv preprint arXiv:2505.14682 (2025)
25. Tong, C., Guo, Z., Zhang, R., Shan, W., Wei, X., Xing, Z., Li, H., Heng, P.A.: Delving into rl for image generation with cot: A study on dpo vs. grpo. arXiv preprint arXiv:2505.17017 (2025)
26. Wang, X., Zhang, X., Luo, Z., Sun, Q., Cui, Y., Wang, J., Zhang, F., Wang, Y., Li, Z., Yu, Q., Zhao, Y., Ao, Y., Min, X., Li, T., Wu, B., Zhao, B., Zhang, B., Wang, L., Liu, G., He, Z., Yang, X., Liu, J., Lin, Y., Huang, T., Wang, Z.: Emu3: Next-token prediction is all you need. CoRR **abs/2409.18869** (2024). https://doi.org/10.48550/ARXIV.2409.18869, https://doi.org/10.48550/arXiv.2409.18869
27. Wang, Y., Zhang, H., Pang, L., Guo, B., Zheng, H., Zheng, Z.: Maferw: Query rewriting with multi-aspect feedbacks for retrieval-augmented large language models. In: Walsh, T., Shah, J., Kolter, Z. (eds.) AAAI-25, Sponsored by the Association for the Advancement of Artificial Intelligence, February 25 - March 4, 2025, Philadelphia, PA, USA. pp. 25434–25442. AAAI Press (2025). https://doi.org/10.1609/AAAI.V39I24.34732, https://doi.org/10.1609/aaai.v39i24.34732
28. Wu, C., Li, J., Zhou, J., Lin, J., Gao, K., Yan, K., ming Yin, S., Bai, S., Xu, X., Chen, Y., Chen, Y., Tang, Z., Zhang, Z., Wang, Z., Yang, A., Yu, B., Cheng, C., Liu, D., Li, D., Zhang, H., Meng, H., Wei, H., Ni, J., Chen, K., Cao, K., Peng, L., Qu, L., Wu, M., Wang, P., Yu, S., Wen, T., Feng, W., Xu, X., Wang, Y., Zhang, Y., Zhu, Y., Wu, Y., Cai, Y., Liu, Z.: Qwen-image technical report (2025), https://arxiv.org/abs/2508.02324
29. Wu, M., Wang, L., Zhao, P., Yang, F., Zhang, J., Liu, J., Zhan, Y., Han, W., Sun, H., Ji, J., et al.: Reprompt: Reasoning-augmented reprompting for text-to-image generation via reinforcement learning. arXiv preprint arXiv:2505.17540 (2025)
30. Wu, Z., Gao, H., Wang, Y., Zhang, X., Wang, S.: Universal prompt optimizer for safe text-to-image generation. In: Duh, K., Gómez-Adorno, H., Bethard, S. (eds.) Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), NAACL 2024, Mexico City, Mexico, June 16-21, 2024. pp. 6340–6354. Association for Computational Linguistics (2024). https://doi.org/10.18653/V1/2024.NAACL-LONG.351

31. Yu, J., Xu, Y., Koh, J.Y., Luong, T., Baid, G., Wang, Z., Vasudevan, V., Ku, A., Yang, Y., Ayan, B.K., Hutchinson, B., Han, W., Parekh, Z., Li, X., Zhang, H., Baldridge, J., Wu, Y.: Scaling autoregressive models for content-rich text-to-image generation. Trans. Mach. Learn. Res. **2022** (2022), https://openreview.net/forum?id=AFDcYJKhND
32. Zeng, Y., Kang, W., Chen, Y., Koo, H.I., Lee, K.: Can mllms perform text-to-image in-context learning? arXiv preprint arXiv:2402.01293 (2024)