Taming Text-to-Image Synthesis for Novices: User-centric Prompt Generation via Multi-turn Guidance

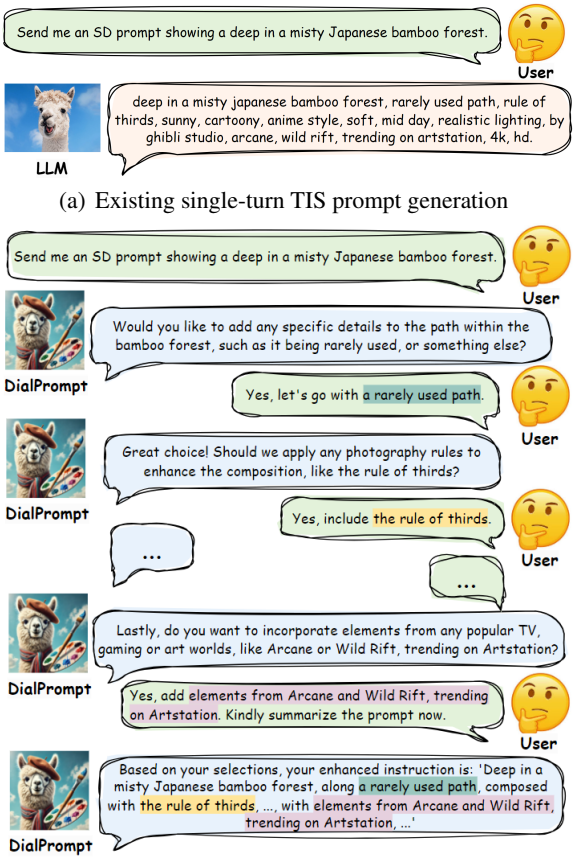
Yilun Liu[1]*, **Minggui He**[1,3]*, **Feiyu Yao**[1]†, **Yuhe Ji**[1], **Shimin Tao**[1], **Jingzhou Du**[1], **Duan Li**[1], **Jian Gao**[1], **Li Zhang**[1], **Hao Yang**[1], **Boxing Chen**[2], **Osamu Yoshie**[3]
[1] Huawei, China
[2] Huawei Canada, Canada
[3] Waseda University, Japan
liuyilun3@huawei.com, minggui_he@fuji.waseda.jp, frankyao.ece@gmail.com

Abstract

The emergence of text-to-image synthesis (TIS) models has significantly influenced digital image creation by producing high-quality visuals from written descriptions. Yet these models are sensitive on textual prompts, posing a challenge for novice users who may not be familiar with TIS prompt writing. Existing solutions relieve this via automatic prompt expansion or generation from a user query. However, this single-turn manner suffers from limited user-centricity in terms of result interpretability and user interactivity. Thus, we propose DialPrompt, a dialogue-based TIS prompt generation model that emphasizes user experience for novice users. DialPrompt is designed to follow a multi-turn workflow, where in each round of dialogue the model guides user to express their preferences on possible optimization dimensions before generating the final TIS prompt. To achieve this, we mined 15 essential dimensions for high-quality prompts from advanced users and curated a multi-turn dataset. Through training on this dataset, DialPrompt improves user-centricity by allowing users to perceive and control the creation process of TIS prompts. Experiments indicate that DialPrompt improves significantly in user-centricity score compared with existing approaches while maintaining a competitive quality of synthesized images. In our user evaluation, DialPrompt is highly rated by 19 human reviewers (especially novices).

1 Introduction

The advent of text-to-image synthesis (TIS) models like Stable Diffusion (SD) (Rombach et al., 2022) has revolutionized the creation of digital images, enabling the generation of high-fidelity visuals from textual descriptions. However, as highlighted by recent studies (Ko et al., 2023; Liu et al., 2022), these models rely heavily on the quality of textual prompts provided by users. The specificity and relevance of these prompts may throw a significant impact on the fidelity and aesthetics of the generated images. Sometimes even adding some magical phrases in the prompts are key to a highly desirable image, such as "soft", "by ghibli studio" and "arcane" shown in Fig. 1(a).



(a) Existing single-turn TIS prompt generation

(b) Our DialPrompt

Figure 1: Two user cases of TIS prompt generation with (a) single-turn style and (b) multi-turn guidance style.

Thus, crafting the perfect model-preferred prompt for TIS models such as SD can be a challenging and nontrivial task for novice users who are not familiar with relevant keywords and prompt writing. While there has been research

*Equal contribution.
†Corresponding author.

Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing, pages 8806–8823
November 4-9, 2025

on manual principles of designing prompts to improve image quality (Liu and Chilton, 2022; Pavlichenko and Ustalov, 2023), an emerging trend is to assist novice users with automatic creation of model-preferred prompts from user-inputted descriptions (Cao et al., 2023; Rosenman et al., 2024; Hei et al., 2024). These approaches typically leverage Large Language Models (LLMs) to interpret user inputs and responds with prompts that are more in line with the TIS model's preferences, thereby enhancing quality of the generated images.

However, we found existing single-turn-based approaches have limitations in user-centricity, especially for novice users:

Firstly, interpretability remains a challenge. Despite their ability to generate complex prompts, novice users often struggle to understand the significance of specific phrases within a prompt and how they correlate with the attributes of the generated image. For instance, as shown in Fig. 1(a), after obtaining the complex prompt with a single-turn query, users may still be confused about the effectiveness of the added keywords, such as "rule of thirds", which actually controls the photography rule, and "arcane", which means adding elements from a popular television series. Furthermore, existing studies highlighted the challenge that users could face understanding barriers in why the model did not produce expected outputs, which hindered users' trust with models (Zamfirescu-Pereira et al., 2023; Weisz et al., 2023).

Secondly, existing methods suffer from a lack of interactivity. Single-turn manners do not engage users in the prompt generation process, leading to outputs that may not align with the user's visual preferences. For example, in Fig. 1(a), the user may desire a realistically styled image, but were provided with a prompt of a comic-style image. Advanced users know where to modify in the final prompt to reflect it, which might be challenging for novice users. This is also observed in the study of Strobelt *et al.* (2022), where they found that a prompt engineering tool should provide the user with the human-in-the-loop ability with rich feedback and user controlability to iteratively improve their prompts.

To address these shortcomings and enhance the user-centricity, we introduce DialPrompt, a dialogue-based TIS prompt generation model. DialPrompt seeks to improve upon the areas of interpretability and interactivity by conducting multiple rounds of queries to the user and gathering ample user preferences before generating the final prompt. To ensure user-centric experience, we studied 70k TIS prompts written by advanced users and mined 15 essential dimensions for crafting high-quality TIS prompts. Based on this finding, we curated a dataset containing 500+ multi-turn dialogues and trained DialPrompt. As shown in Fig. 1(b), our multi-turn dialogue flow is designed to provide step-by-step guidance on possible directions of prompt optimization within the 15 dimensions, such as content, structure, art style, and atmosphere, thereby ensuring a better interpretability of the components in generated final prompt to novice users. Also, DialPrompt allows novice users to easily influence the outcome based on their specific visual preferences, using plain language in the dialogues. Our contributions are:

- We identified 15 essential dimensions for high-quality TIS prompts from advanced users, which can guide future research on prompt engineering for TIS and lead to better visual effects of images, as indicated by the competitive image quality of DialPrompt and the ablation study in Table 5.
- We proposed a novel user-centric paradigm for TIS prompt generation that significantly enhances user experiences (with the human ratings improved by 48.4% against the existing method) by allowing for more interpretable and personalized image creation processes.
- We open-sourced a high-quality dataset containing over 500 multi-turn dialogues for creating user-desired TIS prompts, facilitating future user-centric research.[1]

2 Related Work

2.1 Prompt Engineering in TIS

Despite various architectures of TIS models proposed by researchers (Ramesh et al., 2021; Sauer et al., 2023; Rombach et al., 2022), due to the relatively limited capacity of text encoders (such as the CLIP text encoder (Radford et al., 2021a) in SD), they are still sensitive to quality of input prompts. Prompt engineering methods in TIS aim to ease users' burden by generating prompts that achieve appealing visual effects of generated images. Despite different training paradigms, they can be categorized into two classes in term of user

[1] https://github.com/superboom/DialPrompt

experiences. The first is prefix-based, where user inputs a short prefix of their desired prompt and the model completes the prompt (Rosenman et al., 2024; Datta et al., 2023; Hao et al., 2024). The second is instruct-based, where user inputs an instruction conveying their core ideas of creation and the model responds with a optimized prompt (Mañas et al., 2024; Cao et al., 2023; Hei et al., 2024).

Our work differs from existing approaches mainly in the user-machine interaction logic. Through a multi-turn dialogue, even novice users can be guided through in the optimization of prompt and fully express their preferences.

2.2 User-centric AI

The aim of user-centric AI is to build explainable AI systems that users can understand, trust, and effectively manage (Wang et al., 2019). Various designing philosophies are proposed to achieve towards user-centric AI, including visual designing such as user interfaces (Kim et al., 2023; Feng et al., 2023), and procedure designing such as dialogue systems (Dong et al., 2024). Among them, the technique of reverse question answering (QA) is of particular interest (Yin et al., 2019; Yao et al., 2022). In stead of answering user's questions, reverse QA systems ask user a series of questions in order to collecting preferences, thereby making the AI decision-making process more explainable and customized. Our work can be seen as a pioneering attempt to apply reverse QA into TIS prompt generation to improve user-centricity.

In addition, we also noticed recent endeavors of combining multi-turn interactions with TIS. For example, DialogGen (Huang et al., 2025) and ChatEdit (Cui et al., 2023) explores multi-turn dialogue systems for TIS, featuring a multi-task LLM performing tasks through dialogues, such as image editing, style transferring and chatting. In contrast, DialPrompt focuses on the specific task of TIS prompt generation, using a more fine-grained dialogue flow to control the steps within the generation of prompts for each image.

3 Methodology

3.1 Advanced User Observation

The objective of DialPrompt is to enhance the experiences of novice users (*i.e.*, improving interpretability and interactivity of TIS prompt generation) by engaging them in a guided and step-by-step dialogue towards a high-quality TIS prompt. A critical prerequisite of this process involves identifying the key dimensions that define a high-quality TIS prompt. We achieved this goal by mining wisdom from advanced players of TIS models. Our initial dataset was sourced from lexica.art[2], a widely used platform for discovering SD images and prompts created and shared by experienced users. This platform can provide a comprehensive view of current best practices in TIS prompt engineering. We began with a publicly available dataset from Hugging Face[3], which includes 70k advanced SD prompts collected from lexica.art, along with corresponding user instructions generated by LLMs. To enhance the dataset's utility, we performed semantic clustering to remove redundant entries, ultimately refining it to a representative subset of around 5k pairs of user instructions and TIS prompts.

We then actively work with a group of language experts, which are from the language service center of a top-tier corporation, and conducted a manual study on the 5k TIS prompts. All experts are educated professionals in linguistics, offering services such as translation, editing and technical writing. We further selected those who have basic experiences on text-to-image prompt writing from them through interviews.

These prompts were evenly assigned to each language expert, who was asked to review the assigned subset of prompts and annotate key dimensions appeared in each prompt. This step is repeated until a high agreement on these dimensions is achieved. Specifically, after finishing each round, we discussed with experts to conclude on their sets of dimensions by merging similar terms and unifying the expressions (*e.g.*, unifying "specificity" and "specific elements" and further merging into the dimension of "Detail"). With the updated dimension set, the experts conduct next round of annotation, where the prompts are reassigned to avoid biases. The initial agreement rate (*i.e.*, average similarity between two experts' dimension sets extracted from their assigned prompt subsets) is only 20.13%, and ends up with 85.24% after several rounds of iterations. Then, a reviewing round was conducted for error proofing. Finally, we obtained 15 specific dimensions in 4 major categories that are essential for crafting high-quality TIS prompts. Details of the 4 categories and 15 dimensions are below:

[2] https://lexica.art/

[3] https://huggingface.co/datasets/MadVoyager/stable_diffusion_instructional_dataset

- **Artistic Elements and Techniques**: This category encompasses the core components and methods of creating art, including **Style** (the visual appearance and artistic influences), **Art** (the various forms and media used), **Detail** (intricate aspects that enhance realism), and **Composition** (arrangement of elements for visual balance).

- **Creative Expression**: This kind is focused on how artists convey ideas and emotions, including **Creativity** (innovation and uniqueness in art), **Theme** (the central subject guiding the narrative), and **Mood** (the emotional tone set by the artwork).

- **Visual Impact**: This group covers factors that influence the viewer's perception, such as **Lighting** (use of light to affect atmosphere), **Focus** (primary points of interest), **Realism** (accuracy and lifelikeness), and **Color** (use of hues for emotional expression).

- **Context and Quality**: Background and quality of the artwork, including **Setting** (temporal and spatial context), **Resolution** (clarity and detail level), **Elements** (basic visual components like shapes and textures), and the **Artist** (whose style and skill shape the work).

These dimensions represent the key aspects that a high-quality TIS prompt should effectively address, thereby can guide through our construction of training dialogue dataset of DialPrompt. Since multiple dimensions exist in one prompt, intuitively, each turn of the conversation can discuss about one dimension. To ensure dialogue length, we established a filtering policy whereby any prompt that demonstrates enhancements in at least 5 of these dimensions is preserved, leading to a final selection of 596 high-quality advanced TIS prompt along with user instructions. These data entries will serve as source for the construction of multi-turn dialogues.

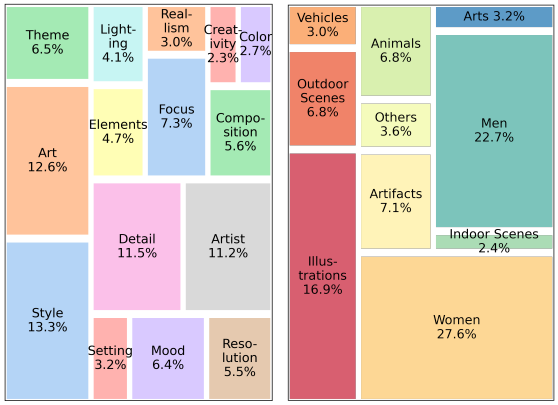
3.2 Construction of Dataset

Based on the curated 596 high-quality instruction-prompt pairs, we further build a multi-turn guidance prompt dataset (MTGPD). Each sample in the dataset is a representative dialogue between user and AI assistant, where the the AI assistant proactively asking users step-by-step questions to fulfill the initial user request and construct a final TIS prompt optimized in the 15 key dimensions as discussed above. The construction of MTGPD is comprised of two primary components: Dialogue Format Conversion and Human Calibration. These components work synergistically to ensure that each dialogue in MTGPD is of the highest quality, both in terms of structure and content.

Dialogue Format Conversion. The Dialogue Format Conversion process is designed to transform the 596 high-quality pairs of user instruction and advanced TIS prompt into a dialogue format. We use the assistance of advanced LLMs (Achiam et al., 2023) to automate this dialogue generation task (Appendix A). During the workflow, we utilize the previously extracted 15 key optimization dimensions as the foundation for the dialogue construction process. The dialogue construction process contains three steps: (1) For each pair of user instruction and advanced TIS prompt, the user instruction is input as the start of a conversation. (2) In each round, the user is presented with options corresponding to a specific optimization dimension within those annotated in the corresponding advanced TIS prompt. The user selects their desired options, gradually refining the TIS prompts. (3) After all dimensions existed in the advanced TIS prompts are discussed, the user terminates the conversation by "Please summarize the prompt for me" and the assistant outputs the final TIS prompt.

Human Calibration. To ensure the generated dialogue data meets high-quality standards, we implemented a human calibration process, which is critical for quality control. This process consists of three key steps: *(1) Format Control*, which ensures a one-query-one-answer structure by correcting or excluding dialogues with consecutive speaking turns; (2) *Relevance Control*, which filters out irrelevant content, such as mutual compliments or expressions of thanks, and retains only material related to TIS prompt optimization; and *(3) Summary Control*, ensuring each dialogue ends with a final prompt and adding a summary if not.

Analysis of MTGPD. Distribution of the 15 dimensions in the final curated 596 dialogues from the MTGPD dataset is displayed in Fig. 2(a), suggesting a mostly balanced coverage on the optimization dimensions within dialogues. As shown in Fig. 2(b), our MTGPD dataset covers a wide range of image topics, including human activities, natural objects and creative arts. Additional statis-



(a) Optimization Dimensions (b) Prompt Topic Category

Figure 2: Distributions of (a) optimization dimensions and (b) image topics in dialogues from MTGPD. Each dialogue has one topic and multiple turns. Each turn contains one dimension.

Average Number of Tokens in Dialogues	
per user message	9.91
per assistant message	28.26
Average Round of Dialogues	6.16
Average Number of Dimensions Per Dialogue	6.99

Table 1: Statistics of MTGPD

tics of MTGPD is in Table 1.

3.3 Multi-Turn Fine-Tuning of DialPrompt

Using this dataset, we developed a fine-tuning process upon open-source LLMs for building multi-turn dialogue ability of DialPrompt, since this dataset provides rich conversational examples that allows the model to learn how to offer step-by-step guidance to users in optimizing TIS prompts across the 15 optimization dimensions. The fine-tuning process incorporates a *multi-turn loss function* (Zheng et al., 2024b). During the training process, for a given dialogue sample from the split MTGPD training set, we apply a masking strategy to the user input. Dialprompt is then tasked with predicting only the assistant's responses. In the final loss computation, the total loss is calculated as the average of the cross-entropy losses for the predicted words in each assistant response throughout the conversation. This training strategy allows an efficient learning of assistant behaviors from the training sample by avoiding overly segmentation of multi-turn dialogues.

It should be noted that, according to the "superficial alignment hypothesis" proposed by LIMA (Zhou et al., 2023), this fine-tuning stage may only require a small amount of high-quality data to align the LLM behaviors with human expectations, given most of their knowledge learnt from the pre-training phase, which is why we didn't train DialPrompt with a very large dataset.

4 Experiments

4.1 Experimental Setting

Implementation Details. In our implementation of DialPrompt, the MTGPD dataset is randomly split into a training set and a test set by a ratio of 9:1. DialPrompt is then trained on the training set for 10 epochs, with a learning rate of 1×10^{-4}, and a batch size of 16. The model is initialized from LLaMA3-8B-Instruct (Dubey et al., 2024). Stable Diffusion-v3 (Esser et al., 2024) is the default TIS model.

User Preference Simulation. Given the multi-turn nature of DialPrompt, the generation of final TIS prompts require the other end of the dialogue, which is the participation of users. In mainstream multi-turn evaluation, the behavior of user end is fixed and irrelevant to AI responses, mostly asking pre-designed follow-up questions (Zheng et al., 2024a). However, in the evaluation of DialPrompt, user needs to express their preferences on the suggestions and choices proposed by DialPrompt in each round of dialogue, which is unpredictable. Thereby, in addition to human evaluation, we also utilize an advanced LLM (Achiam et al., 2023) as an agent to enable an efficient user preference simulation. The prompt used in simulation is listed in Appendix B. To ensure the convergence of dialogues and avoid possible biases, the behavior of the agent is strictly prompted as following: (1) start the dialogue by querying with a user input in the test set; (2) respond with a random preference during the dialogue and (3) end the dialogue by asking for summarizing the prompt after a maximum number N of turns (We use N=5, nearly the average dialogue length in Table 1).

Evaluation Dataset. In addition to the split test set from our MTGPD (which contains 60 samples and is denoted as MTGPD60), which is sourced from Lexica.art, another open-source TIS test set is also involved as an out-of-domain evaluation of DialPrompt. The out-of-domain test set, denoted as PP180, contains 180 prompts sampled from PartiPrompts (Yu et al., 2022), which is designed to

represent a wide range of topics, including different domains and features of language. For MTGPD60, we use the user instructions as the original user input to conduct TIS prompt generation. For PP180, we keep the prompts short by sampling only from the categories of Basic and Simple Detail, and directly use the short prompts as the user input prefix of prompt generation.

4.2 Evaluation on Synthesized Images

After obtaining the generated TIS prompts using DialPrompt or other methods, we input them into Stable Diffusion-v3 (Esser et al., 2024) to acquire the synthetic images. We seek to evaluate the quality of prompts via evaluating images, as a high-quality prompt should lead to visually-appealing and in-topic images, which is essential for a sound user experience of a prompt engineering tool.

Evaluation Metrics. In the evaluation, we consider two dimensions: *fidelity* and *aesthetic*. The dimension of *fidelity* measures the degree to which the the input prompt leads to an in-topic synthetic image. As naive or noised prompts (*e.g.*, containing unnecessary connection words) may disrupt the text encoder of the TIS model and lead to deviated or totally hallucinated output images (This is often used for adversarial attacks (Liu et al., 2023)), a high-quality prompt should produce relevant and in-topic images. Thus, we use CLIP Score (Radford et al., 2021b) as the metric of *fidelity*, which measures the semantic consistency between the textual prompt and the produced image. For *aesthetic*, we use Aesthetic Score (Murray et al., 2012), which is trained on ample human aesthetic feedbacks (a total of 52 million votes from both amateurs and professionals) to predict aesthetic score of images by average human aesthetic standards.

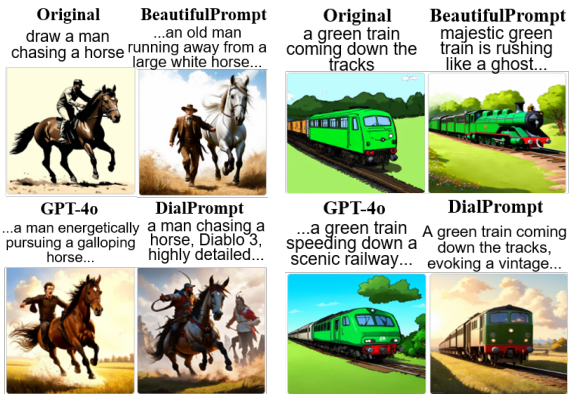
Baselines. We consider two groups of baselines: (1) *TIS Prompt Models*. We compare DialPrompt with three prefix-based approaches: PromptGen (AUTOMATIC1111, 2023), PromptExpansion (Datta et al., 2023) and MagicPrompt (Cao et al., 2023), plus BeautifulPrompt (Cao et al., 2023), which is a recent instruction-based model built upon LLMs. (2) *General-purpose LLMs*. Since most general-purpose proprietary LLMs nowadays are powerful in performing the task of prompt engineering (Liu et al., 2024a) and possess multimedia capabilities, we also include proprietary general-purpose LLMs in the evaluation, by directly instructing the LLMs via API to output an optimized prompt for Stable Diffusion.

Method	In-domain (MTGPD60)		Out-of-domain (PP180)	
	CS	AS	CS	AS
Original	0.264	5.913	0.281	5.562
General-purpose LLMs				
GPT-3.5-turbo	0.302	6.169	0.284	5.574
GPT-4-turbo	0.287	6.311	0.286	5.571
GPT-4o	**0.306**	6.236	**0.295**	5.731
Claude-3.5-Sonnet	0.280	6.157	0.288	5.595
DialPrompt (ours)	0.287	**6.578**	0.290	**6.188**
TIS Prompt Models				
PromptGen	0.265	5.925	0.270	5.351
PromptExpansion	0.267	5.932	0.276	5.568
MagicPrompt	0.255	6.000	0.278	5.528
BeautifulPrompt	0.263	6.528	0.255	6.185
DialPrompt (ours)	**0.287**	**6.578**	**0.290**	**6.188**

Table 2: Image evaluation on in-domain and out-of-domain test set. *CS* and *AS* stands for CLIP Score and Aesthetic Score. Best scores in each group are in *bold*. *Original* is images generated from original user inputs.

Result. As shown in Table 2, DialPrompt not only significantly improves the image quality of original user input, but also outperforms that of existing TIS prompt models in all test cases. DialPrompt's advantage indicates its competitive TIS prompt optimization ability, which can lead to stable and visually-appealing images. For comparison with general-purpose LLMs, despite being curated with far fewer data and training pipelines, DialPrompt still outperforms existing LLMs in Aesthetic Score and achieves a comparable performance in CLIP Score. As will be discussed in Section 4.4, the enhancement in Aesthetic Score can be attributed to the comprehensive prompt optimization dimensions in the training data of DialPrompt, while the advantage in CLIP Score is due to increased model stability through multiple turns.

Visualized Cases. 10 visualized cases are displayed in Appendix D, showing synthesized images and optimized prompts generated by different methods given original prompt. Fig. 3 offers just a brief preview due to space. The two cases show that images from DialPrompt exhibit a higher fidelity to the topic (*i.e.*, in Fig. 3(a), only DialPrompt correctly expressed the concept of "chasing") and a more sophisticated visual effect (*i.e.*, Fig. 3(b)).



(a) Man Chasing Horse (b) Green Train

Figure 3: Images produced by prompts from different approaches with the same input prompt. The original prompts are (a) draw a man chasing a horse and (b) a green train coming down the tracks.

Method	Clarity	Richness	Helpfulness	Overall
Reference Dialogue	8.67	8.48	8.67	8.65
PromptGen	2.58	2.09	2.50	2.50
PromptExpansion	3.50	2.78	3.47	3.35
MagicPrompt	3.45	2.79	3.46	3.30
BeautifulPrompt	4.10	3.15	3.86	3.79
Claude-3.5-Sonnet	6.05	4.08	5.13	5.06
GPT-3.5-turbo	6.23	4.25	5.26	5.25
GPT-4o	6.17	4.13	5.14	5.15
DialPrompt (ours)	**7.81**	**7.57**	**7.72**	**7.69**

Table 3: User-centricity score of different methods that generate TIS prompts for users. Scores for method X are from X *v.s.* reference, except that for reference dialogue is from DialPrompt *v.s.* reference.

4.3 User Experience Evaluation

In this section, we conduct empirical analysis on the user-centric experience of DialPrompt.

Evaluation Protocol Motivated by existing studies on user experiences (Kearsley and Shneiderman, 1998; Hildreth et al., 2000), we define the following dimensions to evaluate the extent of user-centricity an AI assistant demonstrates during interactions with users, regardless the styles (*i.e.*, single-turn or multi-turn): (1) *Clarity:* Language and layout clarity of AI's responses that allows users easily understanding generated content. (2) *Richness:* Richness of the AI recommended aesthetic elements during the interaction with users. (3) *Helpfulness:* Degree to which the AI can understand user's requirement and gives helpful guidance. Novice users are especially sensitive to the three dimensions since they are not familiar with TIS prompts, thereby relying on a clear and helpful interaction process with rich options provided by the system (Cockburn et al., 2014; Scheir, 2006). In the evaluation, each dimension receives a score on a scale of 1 to 10. Higher score means better performance.

Automatic Evaluation Following recent studies that utilize LLM-as-judges for evaluating LLM's capability (Liu et al., 2024b), we compose a prompt based on the above evaluation criterion to request evaluation results from advanced LLMs (Appendix C). In addition to scores from the three dimension, the judge is also requested to output an overall score and a reason (to mitigate hallucination). The evaluation is comparison-based. The judge is asked to compare user interaction processes from two AI assistants for every sample in the MTGPD60 test set. In the evaluation, we keep one of the assistant as the reference dialogues in MTGPD60 test set, which are human-calibrated, and the other assistant as the method to be tested. The dialogues to be evaluated are obtained by simulation (as described in Section 4.1) with the original user intentions in MTGPD60 test set as input. Single-turn methods are considered as "one-turn dialogues". To mitigate biases, the final rating is the average of two tests, with the input order of the two assistants swapped. For the baselines, in addition to existing TIS prompt generation approaches, we also include general-purpose LLMs, since they also possess multi-medial and dialogue abilities.

The results are shown in Table 3. Due to their superior language abilities, general-purpose LLMs receive higher user-centricity scores than existing TIS prompt generation approaches. Nevertheless, DialPrompt outperforms both general-purpose LLMs and other prompt generation models in terms of Clarity, Richness and Helpfulness, indicating an advantage in achieving interpretable and interactive user experiences. Moreover, the overall rating of DialPrompt (7.69) reaches 88.9% of the human-calibrated reference dialogues (8.65), which suggests DialPrompt's outstanding capabilities of user-centric TIS prompt generation.

Analysis of LLM-as-Judge's Justifications. In order to fully understand the high ratings DialPrompt received, we visualized the justifications from the judge when evaluating different methods, via the technique of word cloud diagram. As shown in Fig. 4, the word cloud diagrams consist of frequent adjectives extracted from justifications for each method, with the font size of the word in the

(a) TIS Prompt Model (b) General-purpose LLM

(c) DialPrompt

Figure 4: Word cloud diagrams showcasing frequent adjectives from LLM-as-judge's evaluation justifications of user experience for (a) the TIS prompt model (*e.g.*, PromptExpansion), (b) the general-purpose LLM (*e.g.*, Claude-3.5-Sonnet), and (c) DialPrompt.

diagram indicating its appearance frequency in the justifications. Fig. 4(a) and Fig. 4(b) displays the evaluation justifications for PromptExpansion and Claude-3.5-Sonnet, which represent TIS prompt models and general-purpose LLMs, respectively. The most frequent adjectives in their evaluations are "limited", "insufficient", "shallow", *etc.*, suggesting a limited user experience with insufficient *Helpfulness* and shallow *Richness*. In contrast, DialPrompt in Fig. 4(c) receives more positive feedbacks from the judge, such as "clear", "detailed", "guiding", indicating a supportive user experience.

Human User Evaluation Despite remarkable ratings given by LLMs, evaluations directly from human are irreplaceable. For this task, we recruited 19 well-educated volunteers. To avoid biases, both experienced users and novice users are included. The reviewers can be categorized into three groups based on their backgrounds. Group A contains seven professional visual designers from the design center of a top-tier corporation. They use TIS models such as SD to aid designing, and compose TIS prompts manually without tool assistance. Group B consists of six developers who are experienced users of TIS models. Around half of them have AI background and tried automatic prompt engineering. Group C are six novice users who do not regularly use TIS models and are hardly exposed to prompt engineering technologies. Each reviewer is asked to independently conduct at least 10 fully completed dialogues with DialPrompt and a baseline model (with the same input each time),

Reviewers	Clarity	Richness	Helpfulness
BeautifulPrompt			
Average	4.91	5.27	5.18
DialPrompt (ours)			
Group A (Designer)	8.29	6.43	7.43
Group B (Developer)	7.83	7.17	6.83
Group C (Novice)	8.58	8.33	7.50
Average	**8.23**	**7.31**	**7.25**

Table 4: Average scores from human reviewers after at least 10 completed dialogues with DialPrompt (ours) and BeautifulPrompt.

acquiring optimized TIS prompts for different images that they desire. BeautifulPrompt is utilized as the baseline model since it also aims at improving user-friendliness by allowing user instruction with natural language. Then, they rate on Clarity, Richness and Helpfulness after their experiences with the two models, according to the same criteria in Section 4.3. We did not require image generation and the TIS model to use if they desire images. There is no overlap between volunteers and authors.

Table 4 displays the result of human evaluation. Despite both supporting interaction with natural language, DialPrompt receives significantly higher average scores than BeautifulPrompt, indicating the advantage of multi-turn guidance. For DialPrompt, we further report score distributions of the three groups. visual designers from Group A give the lowest average scores in Richness and the highest scores in Helpfulness, which suggests that the dialogue-based guidance from DialPrompt is an encouraging paradigm to optimize their workflow of art designing, but the richness of aesthetic elements is still not so satisfying from the angle of professional designers. In contrast, developers from Group B rate the two dimensions reversely. Instead of focusing on aesthetic elements, their behaviors during the dialogues are more flexible, and are not limited to linear dialogue flows, leading to a lower Helpfulness. For Group C, the novice users give the highest average scores for the three dimensions among the three groups. **This aligns with the original intention of DialPrompt, which is improving the experience of novice users in TIS prompt composing**.

User Feedback In addition to ratings, we also encourage feedbacks from human reviewers, especially from the novice group. One of the reviewer commented: "Compared with the other tool that just throws out a complex prompt that is hard to

examine, the dialogue style of DialPrompt is obviously more helpful. Although it requires more time to discuss with the model, but it's worthwhile and inspiring. The dialogue gives me some new angles on improving my prompt, and after I walk through several rounds of conversations discussing about lighting, background and other details, I can fully understand the final complex prompt given by DialPrompt."

Another said: "DialPrompt is more controllable than existing prompt expansion models. The prompt expansion is often random in optimization directions, which causes troubles to revise or to rerun several times to get a satisfying one (This is particularly painful before I was familiar enough with writing prompts). DialPrompt allows me to reject options I dislike and choose the option I like. I am glad to see the final image easily influenced by my own desire."

And another commented: "I think this tool is especially more suitable for new users of SD than the prompt refining tools I tried. Compared with these tools where a baseline SD prompt should be firstly crafted by users as an input, for DialPrompt you only need to give a description of your intention in natural language, and the AI offers easy-to-read suggestions on building the SD prompt from multiple angles and guides you through the whole process."

We also receive suggestions from reviewers. Several reviewers commented that the dialogue flow is designed to be too linear and users should be allowed to interact more with the AI, such as asking for further details and conducting open-domain discussions. Another frequent comment is to visualize the suggested prompt in each round of dialogue for a better understanding of the optimization process. These feedbacks and suggestions shed light on future directions of our work, such as incorporating reinforcement learning and multi-media training.

4.4 Ablation Study

Method	CLIP Score	Aesthetic Score
Original	0.264	5.913
+Single-turn	0.250	6.522
+Multi-turn	**0.287**	**6.578**

Table 5: Image quality on MTGPD60 with **different training styles.**

Ablation on the Multi-turn Training Style Instead of utilizing the full multi-turn dialogues in MTGPD as training data, we keep only the initial user query and the final optimized prompt in the dataset to train a single-turn TIS prompt model. As shown in Table 5, this single-turn model still significantly improves the Aesthetic Score of images from original user inputs, which **suggests the effectiveness of the 15 mined prompt optimization dimensions** in the construction of MTGPD. Compared with single-turn, the multi-turn model improves largely in CLIP Score. Through multi-turn interaction with users, the prompt generation process forms a step-by-step chain-of-thought (Wei et al., 2022), thereby decreasing hallucinations and increasing fidelity.

TIS Model	Original		DialPrompt	
	CS	AS	CS	AS
LDM	0.278	6.122	**0.296**	**6.741**
SD-v1.5	0.267	5.213	**0.273**	**6.204**
SD-v2	0.275	6.134	**0.294**	**6.682**
SDXL	0.271	6.119	**0.295**	**6.700**
SD-v3	0.264	5.913	**0.287**	**6.578**

Table 6: Image quality of original user input and DialPrompt on MTGPD60 with **different TIS models**.

Ablation on Underlying TIS models We test the same prompt on a series of different TIS models: LDM (Rombach et al., 2022), SD-v1.5 (Rombach et al., 2022), SD-v2 (Rombach et al., 2022), SDXL (Podell et al., 2024) and SD-v3 (Esser et al., 2024). As shown in Table 6, images generated from DialPrompt's optimized prompts continuously outperforms that from original user inputs, indicating its strong compatibility to different TIS models.

5 Conclusion

In this paper, we seek to improve the user-centricity in TIS prompt engineering by proposing DiaPrompt, a novel dialogue-based TIS prompt generation model. DialPrompt not only shows advantages in improving the quality of synthetic images, but also provide a unique user experience through multi-turn guidance. Our user evaluation demonstrate that DialPrompt can not only assist novice users to easily optimize TIS prompt with their own ideas, but also aid professionals in their designing work through a comprehensive recommendation of aesthetic elements. Future work includes incorporating reinforcement and multi-modal training.

6 Limitations

We reveal the following limitations:

(1) Limited Freedom in Dialogues. Several human users commented that the dialogue flow is designed to be too linear for user to conduct open-ended discussions. However, the primary objective of the paper is to aid novice users with step-by-step guidance (*i.e.*, controlled dialogues), which is naturally conflict with free dialogues.

(2) Limited User Agent. The user simulation in this paper (Section 4.1) was implemented to be very simple (for technical feasibility) and may not reflect the complex user behaviors in the real world, reducing the soundness of the automatic evaluation. So, we included a human evaluation with different user backgrounds.

(3) Unaligned Individual Aesthetic Preference. The Aesthetic Score model we used is trained to reflect the average human aesthetic standards, which may not align with an individual user. However, it is challenging to directly measure and represent a wide range of individual aesthetic standards. The dimension of *Helpfulness* in our evaluation can reflect the degree to which an individual user can express his aesthetic preference.

7 Potential Risk & Ethical Consideration

While the development of DialPrompt offers significant improvements in the user experience of TIS models, it also raises several potential risks and ethical considerations that must be addressed.

Bias in Prompt Generation. As with many machine learning models, DialPrompt's performance is influenced by the data it is trained on. If the dataset of multi-turn prompts or user preferences used to train the model contains biased or skewed representations, the generated prompts could perpetuate stereotypes or reinforce harmful biases. For example, certain cultural or demographic representations may be underrepresented, leading to less accurate or diverse image generation. Careful curation of training data, including diverse and balanced inputs, is essential to reduce the risk of bias in the generated outputs.

Content Misuse and Harmful Applications. TIS models, including DialPrompt, can generate highly realistic images from textual prompts. While this has creative and practical benefits, it also opens the door for malicious use, such as creating misleading or harmful images that could spread misinformation or cause harm. DialPrompt's ability to generate realistic images should be accompanied by safeguards to prevent misuse, such as a content moderation system to flag harmful or inappropriate prompts and outputs.

References

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

AUTOMATIC1111. 2023. https://github.com/automatic1111/stable-diffusion-webui-promptgen.

Tingfeng Cao, Chengyu Wang, Bingyan Liu, Ziheng Wu, Jinhui Zhu, and Jun Huang. 2023. Beautifulprompt: Towards automatic prompt engineering for text-to-image synthesis. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: Industry Track*, pages 1–11.

Andy Cockburn, Carl Gutwin, Joey Scarr, and Sylvain Malacria. 2014. Supporting novice to expert transitions in user interfaces. *ACM Computing Surveys (CSUR)*, 47(2):1–36.

Xing Cui, Zekun Li, Pei Li, Yibo Hu, Hailin Shi, Chunshui Cao, and Zhaofeng He. 2023. Chatedit: Towards multi-turn interactive facial image editing via dialogue. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 14567–14583.

Siddhartha Datta, Alexander Ku, Deepak Ramachandran, and Peter Anderson. 2023. Prompt expansion for adaptive text-to-image generation. *arXiv preprint arXiv:2312.16720*.

Zhikang Dong, Xiulong Liu, Bin Chen, Pawel Polak, and Peng Zhang. 2024. Musechat: A conversational music recommendation system for videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12775–12785.

Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, et al. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.

Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, et al. 2024. Scaling rectified flow transformers for high-resolution image synthesis. In *Forty-first International Conference on Machine Learning*.

Yingchaojie Feng, Xingbo Wang, Kam Kwai Wong, Sijia Wang, Yuhong Lu, Minfeng Zhu, Baicheng Wang, and Wei Chen. 2023. Promptmagician: Interactive prompt engineering for text-to-image creation. *IEEE Transactions on Visualization and Computer Graphics*.

Yaru Hao, Zewen Chi, Li Dong, and Furu Wei. 2024. Optimizing prompts for text-to-image generation. *Advances in Neural Information Processing Systems*, 36.

Nailei Hei, Qianyu Guo, Zihao Wang, Yan Wang, Haofen Wang, and Wenqiang Zhang. 2024. A user-friendly framework for generating model-preferred prompts in text-to-image synthesis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 2139–2147.

Ellen C Hildreth, Jack MH Beusmans, Erwin R Boer, and Constance S Royden. 2000. From vision to action: experiments and models of steering control during driving. *Journal of Experimental Psychology: Human Perception and Performance*, 26(3):1106.

Minbin Huang, Yanxin Long, Xinchi Deng, Ruihang Chu, Jiangfeng Xiong, Xiaodan Liang, Hong Cheng, Qinglin Lu, and Wei Liu. 2025. Dialoggen: Multi-modal interactive dialogue system with multi-turn text-image generation. In *Findings of the Association for Computational Linguistics: NAACL 2025*, pages 411–426.

Greg Kearsley and Ben Shneiderman. 1998. Engagement theory: A framework for technology-based teaching and learning. *Educational technology*, 38(5):20–23.

Seonuk Kim, Taeyoung Ko, Yousang Kwon, and Kyungho Lee. 2023. Designing interfaces for text-to-image prompt engineering using stable diffusion models: a human-ai interaction approach. In *IASDR 2023: Life-Changing Design*.

Hyung-Kwon Ko, Gwanmo Park, Hyeon Jeon, Jaemin Jo, Juho Kim, and Jinwook Seo. 2023. Large-scale text-to-image generation models for visual artists' creative works. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*, pages 919–933.

Han Liu, Yuhao Wu, Shixuan Zhai, Bo Yuan, and Ning Zhang. 2023. Riatig: Reliable and imperceptible adversarial text-to-image generation with natural prompts. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20585–20594.

Vivian Liu and Lydia B Chilton. 2022. Design guidelines for prompt engineering text-to-image generative models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–23.

Vivian Liu, Han Qiao, and Lydia Chilton. 2022. Opal: Multimodal image generation for news illustration. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, pages 1–17.

Yilun Liu, Shimin Tao, Weibin Meng, Jingyu Wang, Wenbing Ma, Yuhang Chen, Yanqing Zhao, Hao Yang, and Yanfei Jiang. 2024a. Interpretable online log analysis using large language models with prompt strategies. In *Proceedings of the 32nd IEEE/ACM International Conference on Program Comprehension*, pages 35–46.

Yilun Liu, Shimin Tao, Xiaofeng Zhao, Ming Zhu, Wenbing Ma, Junhao Zhu, Chang Su, Yutai Hou, Miao Zhang, Min Zhang, Hongxia Ma, Li Zhang, Hao Yang, and Yanfei Jiang. 2024b. Coachlm: Automatic instruction revisions improve the data quality in llm instruction tuning. In *2024 IEEE 40th International Conference on Data Engineering (ICDE)*, pages 5184–5197.

Oscar Mañas, Pietro Astolfi, Melissa Hall, Candace Ross, et al. 2024. Improving text-to-image consistency via automatic prompt optimization. *arXiv preprint arXiv:2403.17804*.

Naila Murray, Luca Marchesotti, and Florent Perronnin. 2012. Ava: A large-scale database for aesthetic visual analysis. In *2012 IEEE conference on computer vision and pattern recognition*, pages 2408–2415. IEEE.

Nikita Pavlichenko and Dmitry Ustalov. 2023. Best prompts for text-to-image models and how to find them. In *Proceedings of the 46th International ACM Conference on Research and Development in Information Retrieval*, pages 2067–2071.

Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, and Robin Rombach. 2024. Sdxl: Improving latent diffusion models for high-resolution image synthesis. In *The Twelfth International Conference on Learning Representations*.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021a. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021b. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR.

Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. 2021. Zero-shot text-to-image generation. In *International Conference on Machine Learning*, pages 8821–8831. PMLR.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695.

Shachar Rosenman, Vasudev Lal, and Phillip Howard. 2024. Neuroprompts: An adaptive framework to optimize prompts for text-to-image generation. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, pages 159–167.

Axel Sauer, Tero Karras, Samuli Laine, Andreas Geiger, and Timo Aila. 2023. Stylegan-t: Unlocking the power of gans for fast large-scale text-to-image synthesis. In *International conference on machine learning*, pages 30105–30118. PMLR.

Wendy Scheir. 2006. First entry: report on a qualitative exploratory study of novice user experience with online finding aids. *Journal of archival organization*, 3(4):49–85.

Hendrik Strobelt, Albert Webson, Victor Sanh, Benjamin Hoover, Johanna Beyer, Hanspeter Pfister, and Alexander M Rush. 2022. Interactive and visual prompt engineering for ad-hoc task adaptation with large language models. *IEEE transactions on visualization and computer graphics*, 29(1):1146–1156.

Danding Wang, Qian Yang, Ashraf Abdul, et al. 2019. Designing theory-driven user-centric explainable ai. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pages 1–15.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, 35:24824–24837.

Justin D Weisz, Michael Muller, Jessica He, and Stephanie Houde. 2023. Toward general design principles for generative ai applications. In *Joint Workshops on Human-AI Co-Creation with Generative Models and User-Aware Conversational Agents*.

Rujing Yao, Linlin Hou, Lei Yang, Jie Gui, and Ou Wu. 2022. Deep human answer understanding for natural reverse qa. *Knowledge-Based Systems*, 254:109625.

Qing Yin, Guan Luo, Xiaodong Zhu, Qinghua Hu, and Ou Wu. 2019. Semi-interactive attention network for answer understanding in reverse-qa. In *Advances in Knowledge Discovery and Data Mining: 23rd Pacific-Asia Conference, PAKDD 2019, Macau, China, April 14-17*, pages 3–15. Springer.

Jiahui Yu, Yuanzhong Xu, Jing Yu Koh, Thang Luong, Gunjan Baid, Zirui Wang, Vijay Vasudevan, Alexander Ku, et al. 2022. Scaling autoregressive models for content-rich text-to-image generation. *arXiv preprint arXiv:2206.10789*, 2(3):5.

JD Zamfirescu-Pereira, Richmond Y Wong, Bjoern Hartmann, and Qian Yang. 2023. Why johnny can't prompt: how non-ai experts try (and fail) to design llm prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–21.

Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, et al. 2024a. Judging llm-as-a-judge with mt-bench and chatbot arena. *Advances in Neural Information Processing Systems*, 36.

Yaowei Zheng, Richong Zhang, Junhao Zhang, Yanhan Ye, Zheyan Luo, Zhangchi Feng, and Yongqiang Ma. 2024b. Llamafactory: Unified efficient fine-tuning of 100+ language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*, Bangkok, Thailand.

Chunting Zhou, Pengfei Liu, Puxin Xu, Srinivasan Iyer, Jiao Sun, Yuning Mao, Xuezhe Ma, Avia Efrat, Ping Yu, Lili Yu, et al. 2023. Lima: Less is more for alignment. *Advances in Neural Information Processing Systems*, 36:55006–55021.

A Prompt Template for Dialogue Format Conversion

`[System]`
Given an original user instruction and its enhanced version, create a structured multi-turn Q&A dialogue that guides a user to refine their prompt for creating an aesthetically superior image. Here are notices:
1. The dialogue should start with the user's original instruction.
2. This instruction is for Stable Diffusion to generate images.
3. Questions focus on providing enrich information for user to choose based on the enhanced description.
4. With user's chosen options , assistant output a final enriched instruction which is similar to enhanced description or keep enhanced description as final output.
5. The dialogue should be ended with a enhanced version prompt for Stable Diffusion. When you output the enhanced version prompt, please add ###[BEGIN OF PROMPT] before the prompt, and ###[END OF PROMPT] after the prompt like this: ###[BEGIN OF PROMPT] 'lofi biopunk portrait of Shrek as a Disney Princess, Pixar style, by Tristan Eaton, Stanley 'Artgerm' Lau, and Tom Bagshaw.' ###[END OF PROMPT]
6. Please Generate the multi-turn Q&A dialogue with llama alpaca format of json file, the Role Name in each sample must be 'user' and 'assistant'.
7. In the last dialogue turn, you must add the summarize order for user content, like: 'user' : Artgerm and Greg Rutkowski. Please summarize the prompt for me now.

[User]
Please Generate the multi-turn Q&A dialogue with llama alpaca format of json file: {Input Instruction-Prompt Pair}

B Prompt Template for User Preference Simulation

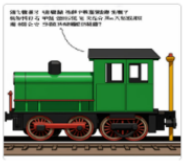
[System]
Assume you are a user who has a dialogue with a system which aims to enrich prompt for text to image generation, make suitable selection according to the option it provided without any biases, or ask it to combine all options based on your situation. your answer must be concise.
e.g. system: To create a more captivating image, would you like the portrait to be realistic or stylized? Your answer: Realistic, please. Or you can answer: A mix of both is ok.
[User]
{Input Dialogue}

C Prompt Template for User Experience Evaluation

[The Start of Assistant 1's Dialogue]
{Input Dialogue 1}
[The End of Assistant 1's Dialogue]
[The Start of Assistant 2's Dialogue]
{Input Dialogue 2}
[The End of Assistant 2's Dialogue]
[System]
We would like to request you to compare on the performance of two AI assistants in the displayed multi-turn dialogues with the user, trying to recommend and build a proper Stable Diffusion prompt for the user. Please rate the user-friendliness of the two AI assistants, considering the Clarity, Richness and Helpfulness of the whole dialogue. (1) Clarity: to which degree the layout and language of AI's responses is organized and clear for users. (2) Richness: the richness of the AI recommended aesthetic elements that user can express preferences on in the dialogue. (3) Helpfulness: the degree to which the AI can understand user's requirement and give step-by-step guidance in the dialogue. Each dimension receives a score on a scale of 1 to 10, where a higher score indicates better performance. And also output an overall score of 1 to 10. Please first output two lines indicating the scores for Assistant 1 and 2, with each line containing only four values indicating the scores for overall, clarity, richness and helpfulness, respectively. The four scores are separated by space. In the subsequent line, please provide a comprehensive explanation of your evaluation, avoiding any potential bias and ensuring that the order in which the dimensions were presented does not affect your judgment.

D Visualized Cases

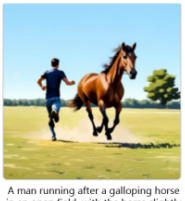
Green Train

Raw	MagicPrompt	PromptExpansion	PromptGen	BeautifulPrompt
				
a green train coming down the tracks	a green train coming down the tracks from the earth on the left, a matte painting by Filip Hodas, featured on pixiv, metaphysical painting, catalism, deviantart hd, octane render	Could you draw a green train coming down the tracks?	Could you draw a green train coming down the tracks? You don't have to draw a green train. You can draw a green train.	majestic green train is rushing like a ghost in the illustration. The locomotive can be seen running very fast as it is speeding down the long and winding train track. The background has trees with flowers and the train is flying high over an epic landscape, realistic digital art

GPT-3.5-turbo	GPT-4-turbo	GPT-4o	Claude-3.5-Sonnet	DialPrompt (our)
				
Create a detailed illustration of a vibrant green train gracefully traveling down a scenic set of tracks, surrounded by lush greenery and a clear blue sky	A detailed, realistic green train moving down the tracks, with the train in motion, smoke billowing from the engine, surrounded by lush greenery and a clear sky in the background.	Create a vibrant illustration of a green train speeding down a scenic railway track, surrounded by lush greenery under a clear blue sky.	A vibrant green train moving swiftly down the tracks, with motion blur showing its speed. The scene is set in a lush countryside with clear skies and detailed surroundings, capturing the dynamic energy of the train in motion	A green train coming down the tracks, evoking a vintage, detailed atmosphere, with a soft, yet detailed approach. The scene is set in a countryside, under a sunset, with warm lighting. Inspired by Artgerm, the composition is a digital painting. Concept art, trending on ArtStation

Man Chasing Horse

Raw	MagicPrompt	PromptExpansion	PromptGen	BeautifulPrompt
				
draw a man chasing a horse	draw a man chasing a horse. Trending on Artstation	draw a man chasing a horse. Trending on Artstation	Could you draw a man chasing a horse? I can draw a man.	A hyperrealistic illustration of an old man running away from a large white horse, fantasy art

GPT-3.5-turbo	GPT-4-turbo	GPT-4o	Claude-3.5-Sonnet	DialPrompt (our)
				
Create a dynamic scene of a man energetically chasing a galloping horse across a vibrant landscape, capturing the sense of movement and emotion in the expressions and postures of both the man and the horse.	A man running after a galloping horse in an open field, with the horse slightly ahead, showing a sense of urgency and motion. The scene is set under a clear blue sky with some distant trees and grass swaying in the wind.	Create a dynamic scene of a man energetically pursuing a galloping horse through a sunlit meadow, capturing the motion and excitement with detailed expressions and flowing mane.	Depict a scene where a man is running after a galloping horse in an open field. The man appears determined and focused, reaching out towards the horse, while the horse is slightly ahead, moving swiftly with its mane flowing in the wind. The background shows a wide, grassy landscape under a clear sky, emphasizing the sense of speed and pursuit.	a man chasing a horse, Diablo 3, highly detailed, perfect lighting, high composition, digital painting, concept art, smooth, sharp focus, illustration, art by Artgerm and Derek Z and Greg Rutkowski.

A Smiling Man

Raw	MagicPrompt	PromptExpansion	PromptGen	BeautifulPrompt
				
Could you draw a smiling man	Could you draw a smiling man? Black and white cartoon. Trending on Artstation	Could you draw a smiling man? Comic book illustration, award winning on artstation	Could you draw a smiling man? I want you to know he is my friend.	could i imagine if i could dreamt of a woman, fantasy, realistic, full body portrait, highly detailed, artstation

GPT-3.5-turbo	GPT-4-turbo	GPT-4o	Claude-3.5-Sonnet	DialPrompt (our)
				
Create a vibrant illustration of a cheerful man with a broad smile, wearing a casual outfit, standing outdoors in a sunny park, surrounded by trees and flowers. The scene should convey joy and positivity	A cheerful man with a warm smile, standing against a simple background, with soft lighting highlighting his face and features	Create a detailed illustration of a cheerful man with a bright smile, showcasing vibrant colors and expressive features in a joyful setting	A smiling man with short, dark hair, wearing a casual shirt, standing against a neutral background. His expression is warm and friendly, with visible teeth and slightly crinkled eyes to emphasize the genuine nature of his smile.	Portrait of a smiling man with brown hair and blue eyes, realistic, portrait, highly detailed, digital painting, 8k, concept art, trending on ArtStation

Tree

Raw	MagicPrompt	PromptExpansion	PromptGen	BeautifulPrompt
				
Could you draw a tree?	Could you draw a tree? [smiling]	Could you draw a tree?	Could you draw a tree? You can't draw a tree.	A full shot of a very detailed amazing surreal artwork portrait, line Art, painting by Greg Rutkowski

GPT-3.5-turbo	GPT-4-turbo	GPT-4o	Claude-3.5-Sonnet	DialPrompt (our)
				
Create a highly detailed illustration of a majestic tree, showcasing its intricate bark texture, vibrant green leaves, and a dynamic play of light and shadow. Include a serene background with a clear blue sky and soft, dappled sunlight filtering through the branches	A large oak tree with sprawling branches, lush green leaves, and a textured bark, standing in a grassy field under a clear blue sky.	Create a detailed and vibrant illustration of a majestic tree, showcasing its intricate leaves, textured bark, and various shades of green, set against a serene landscape with soft sunlight filtering through the branches.	Create a detailed image of a tall tree with lush green leaves, a sturdy brown trunk, and branches extending outward. The tree should be set against a clear blue sky with a few clouds, with grass and small plants around its base.	a tree of life and death, in the style of Greg Rutkowski, Jesper Ejsing, with a dark, dramatic atmosphere, 35 mm grain, highly detailed, in a digital painting and concept art style, trending on ArtStation

Green Clock

Raw	MagicPrompt	PromptExpansion	PromptGen	BeautifulPrompt
				
Could you draw a green clock?	Could you draw a green clock? by Stanley Artgerm Lau, WLOP, Rossdraws, James Jean, Andrei Riabovitchev, Marc Simonetti, krenz cushart, Sakimichan, trending on"	Could you draw a green clock? by Stanley Artgerm Lau, WLOP, Rossdraws, James Jean, Andrei Riabovitchev, Marc Simonetti, krenz cushart, Sakimichan, trending on	Could you draw a green clock? Draw a green clock. Draw a green clock.	There is one green clock and there is no alternative. The time that's come is never going to be altered, highly detailed illustration, 8k quality

GPT-3.5-turbo	GPT-4-turbo	GPT-4o	Claude-3.5-Sonnet	DialPrompt (our)
				
Create a detailed illustration of a vibrant green clock with intricate designs, showcasing its hour, minute, and second hands set at 10:10. Include a soft, blurred background for emphasis.	Draw a detailed, round green wall clock with black numerals and sleek hands, showing the time at 10:10. The clock has a glossy finish, with a subtle shadow behind it, set against a light, minimalistic background.	Create an artistic illustration of a vibrant green clock, incorporating intricate details and a unique design that highlights its green hue. The clock should be prominently featured against a contrasting background to enhance its visual appeal	A detailed close-up of a vibrant green clock with a sleek, modern design, showing the time clearly with sharp, well-defined hour and minute hands against a contrasting background.	A green clock, drawn by Artgerm, in the style of Greg Rutkowski, with soft lighting, against a detailed background, in a digital painting style

Capybara

Raw	MagicPrompt	PromptExpansion	PromptGen	BeautifulPrompt
				
Could you draw a capybara	draw a capybara alcohol ink on parchment, muted studio colors, shading, detailed, textured parchment background, 1 9 5 0 s concept art	draw a capybara	I drew a capybara. I drew a capybara. I drew a capybara. I drew a capybara. I drew a capybara. I drew a capybara.	A Capybara. By Makoto Shinkai, by James Gilleard, detailed, trending on artstation

GPT-3.5-turbo	GPT-4-turbo	GPT-4o	Claude-3.5-Sonnet	DialPrompt (our)
				
Create a highly detailed and vibrant illustration of a capybara in its natural habitat, showcasing its unique features and playful demeanor amidst lush greenery and tranquil water.	Create a detailed illustration of a capybara in a natural setting, showcasing its friendly demeanor and unique features, with lush greenery in the background and soft, warm lighting.	Please create a detailed and realistic illustration of a capybara in its natural habitat	Draw a detailed and realistic image of a capybara, showcasing its full body. The capybara should be in a natural environment, such as near a body of water with surrounding greenery. Ensure the fur texture, large round nose, and calm expression are clearly visible.	Capybara, light brown coat, soft, natural lighting, wide angle, extremely high detail, ultra realistic, concept art, HD, high details, 8k, inspired by artists like Artgerm, Greg Rutkowski, and Alphonse Mucha

Slices of Avocado on A Piece of Toast

Raw	MagicPrompt	PromptExpansion	PromptGen	BeautifulPrompt
				
Could you draw slices of avocado on a piece of toast?	draw slices of avocado on a piece of toast	slices of avocado on a piece of toast	draw slices of avocado on a piece of toast? by edgar maxence and caravaggio and michael whelan and delacroix style, artistic, intricate drawing, light brazen, realistic fantasy, extremely detailed and beautiful aesthetic face, establishing shot, 8k resolution, dramatic lighting	Sliced avocade stuck to toast, fantasy, intricate, elegant, highly detailed, digital painting, artstation, concept art, smooth, sharp focus",

GPT-3.5-turbo	GPT-4-turbo	GPT-4o	Claude-3.5-Sonnet	DalPrompt (our)
				
Please draw slices of avocado arranged on a piece of toast.	Realistic drawing of fresh avocado slices neatly arranged on a perfectly toasted slice of bread, with detailed textures and vibrant colors.	Create a detailed illustration of perfectly ripe avocado slices gracefully arranged on a golden-brown piece of artisanal toast, showcasing vibrant colors and textures, with a soft-focus background of a rustic kitchen setting.	Create a detailed image of a piece of toast topped with several slices of ripe avocado, arranged neatly. The toast should have a slightly golden-brown, crispy texture, and the avocado slices should appear fresh and vibrant, with a smooth, creamy green surface.	slices of avocado on a piece of toast, D&D, fantasy, realistic, digital painting, concept art, matte, art by Artgerm, Greg Rutkowski, and Alphonse Mucha, soft, ethereal atmosphere, golden hour, cinematic, dramatic atmosphere, trending on ArtStation

A Girl Going to A Farm

Raw	MagicPrompt	PromptExpansion	PromptGen	BeautifulPrompt
				
Could you draw a girl going to a farm?	draw a girl going to a farm , chyperdetailed, 8k realistic, bright colors, trending on artstation	Could you draw a girl going to a farm? By Akira Yamaoka, Stanley Artgerm Lau, WLOP, Rossdraws, James Jean, Andrei Riabovitchev, Marc Simonetti, krenz cushart	A Girl with a cute face and flowing blonde hair sitting in front of the machine, she is doing an art action where her body can be painted and has hands, she would like to do something as dramatic as being painted on the side of the machine by which she could build lots of money.	Please sketch a girl walking through a lush farm setting.

GPT-3.5-turbo	GPT-4-turbo	GPT-4o	Claude-3.5-Sonnet	DalPrompt (our)
				
Create a highly detailed and vibrant illustration of a capybara in its natural habitat, showcasing its unique features and playful demeanor amidst lush greenery and tranquil water.	A young girl with braided hair and a straw hat, wearing a light summer dress and boots, happily walking towards a traditional farm with a red barn, wooden fence, and green fields in the background under a bright, sunny sky.	Create a vibrant illustration of a young girl joyfully exploring a picturesque farm, surrounded by lush green fields, colorful flowers, and friendly animals. Include details like a red barn in the background, clear blue skies, and the girl wearing a straw hat and a summer dress.	lush green fields, and she carries a basket filled with fresh produce, smiling as she heads towards the farmhouse. The sky is bright blue with fluffy white clouds.	realistic medium shot of a girl going to a farm, with soft, warm lighting, volumetric, godrays, vivid, beautiful, trending on artstation

A Compass Next to A Piece of Fruit

Raw	MagicPrompt	PromptExpansion	PromptGen	BeautifulPrompt
				
Could you draw a compass next to a piece of fruit?	Could you draw a compass next to a piece of fruit?	Could you draw a compass next to a piece of fruit?	Could you draw a compass next to a piece of fruit?	an intricate, detailed concept art of a piece of fruit with stunning scenery in front of it that looks like a compass

GPT-3.5-turbo	GPT-4-turbo	GPT-4o	Claude-3.5-Sonnet	DialPrompt (our)
				
Please draw a compass next to an apple.	A detailed compass placed beside a vibrant, ripe piece of fruit, such as an orange, on a simple, textured surface with natural lighting.	Create a highly detailed and vibrant illustration of a compass placed next to a colorful piece of fruit, focusing on the texture of the fruit and the intricate design of the compass.	A detailed close-up of a traditional compass placed beside a ripe, vibrant piece of fruit, like an apple or an orange, on a textured surface. The compass should have a metallic finish with clear markings, and the fruit should appear fresh and slightly glossy.	A compass next to a piece of fruit, vibrant apple with a shiny surface, highly detailed, photorealistic, digital painting, concept art, sharp focus, illustration

Waterfall

Raw	MagicPrompt	PromptExpansion	PromptGen	BeautifulPrompt
				
Could you draw a waterfall?	Could you draw a waterfall? by Studio Ghibli, Makoto Shinkai, by Artgerm, by WLOP, by Greg Rutkowski, volumetric lighting, octane render, 4K resolution, trending on artstation	Could you draw a waterfall?	Could you draw a waterfall? You can't draw a waterfall. You can't draw a waterfall. You can't draw a waterfall. You can't draw a waterfall.	A surrealist painting of a waterfall, a digital rendering by John Berkey, trending on Artstation, fantasy art, concept art, dramatic lighting

GPT-3.5-turbo	GPT-4-turbo	GPT-4o	Claude-3.5-Sonnet	DialPrompt (our)
				
Please create a detailed drawing of a majestic waterfall cascading down rocks.	A serene, tall waterfall cascading down rugged rocks into a clear, turquoise pool surrounded by lush, green forest with sunlight filtering through the trees.	Create a stunning, realistic illustration of a majestic waterfall cascading down a rocky cliff, surrounded by lush greenery and vibrant wildflowers, with a rainbow sparkling in the mist and sunlight filtering through the trees above.	A majestic waterfall cascading down a rugged cliff, surrounded by lush greenery, with mist rising from the base where the water meets the rocks. The sunlight filters through the trees, casting dappled shadows and creating a rainbow in the mist. The scene should feel serene and natural, with the sound of rushing water almost audible.	Create a serene and natural scene of a waterfall around, with a golden hour, set in a highly detailed and centered location, inspired by Greg Rutkowski and Alphonse Mucha. The artwork should feature Octane Render, with a high-definition, 8K resolution