

# Training Data:

## Image Datasets:

- **Yuanzhi/aesthetics\_prompts\_laion** cut down to about 1-2M most likely no cut  
[https://huggingface.co/datasets/Yuanzhi/aesthetics\\_prompts\\_laion](https://huggingface.co/datasets/Yuanzhi/aesthetics_prompts_laion)
- **rubend18/DALL-E-Prompts-OpenAI-ChatGPT** no cut no clean  
<https://huggingface.co/datasets/rubend18/DALL-E-Prompts-OpenAI-ChatGPT>

## Text Datasets:

- **Dahoas/sft-gptj-synthetic-prompt-responses** no cut no clean  
<https://huggingface.co/datasets/Dahoas/sft-gptj-synthetic-prompt-responses>
- **Lyte/100k\_prompts** no cut no clean [https://huggingface.co/datasets/Lyte/100k\\_prompts](https://huggingface.co/datasets/Lyte/100k_prompts)
- **perlthoughts/coding-prompts-small** no cut some prompts are split into two columns (if input col has anything in it, merge it with instruction col for the whole prompt)  
<https://huggingface.co/datasets/perlthoughts/coding-prompts-small>
- **aorogat/Questions\_to\_Tagged\_Questions\_Prompts** not cut not clean  
[https://huggingface.co/datasets/aorogat/Questions\\_to\\_Tagged\\_Questions\\_Prompts?row=61](https://huggingface.co/datasets/aorogat/Questions_to_Tagged_Questions_Prompts?row=61)
- **Nitral-AI/Reddit-SFW-Writing\_Prompts\_ShareGPT** no cut prompts are stored in a prompt-response pair  
[https://huggingface.co/datasets/Nitral-AI/Reddit-SFW-Writing\\_Prompts\\_ShareGPT?row=7](https://huggingface.co/datasets/Nitral-AI/Reddit-SFW-Writing_Prompts_ShareGPT?row=7)

# Testing Data:

## Image Datasets:

- **Falah/2000000\_Style\_art\_prompts** no cut no clean  
[https://huggingface.co/datasets/Falah/2000000\\_Style\\_art\\_prompts](https://huggingface.co/datasets/Falah/2000000_Style_art_prompts)
- **nlphuji/flickr30k** no cut has 5 prompts per row -starts with [ " -prompts seperated by " , " -ends with " ] <https://huggingface.co/datasets/nlphuji/flickr30k?row=0>
- **data-is-better-together/open-image-preferences-v1** no cut some prompts start and end with "  
<https://huggingface.co/datasets/data-is-better-together/open-image-preferences-v1?row=6>

## Text Dataset:

- ***data-is-better-together/10k\_prompts\_ranked*** no cut no clean  
[https://huggingface.co/datasets/data-is-better-together/10k\\_prompts\\_ranked](https://huggingface.co/datasets/data-is-better-together/10k_prompts_ranked)
- ***argilla/distilabel-reasoning-prompts*** no cut no clean  
<https://huggingface.co/datasets/argilla/distilabel-reasoning-prompts>
- ***gabrielmbmb/prompts-collective-source*** no cut no clean  
<https://huggingface.co/datasets/gabrielmbmb/prompts-collective-source>
- ***HuggingFaceH4/code\_evaluation\_prompts*** no cut no clean  
[https://huggingface.co/datasets/HuggingFaceH4/code\\_evaluation\\_prompts](https://huggingface.co/datasets/HuggingFaceH4/code_evaluation_prompts)
- ***Alex-xu/practice\_run\_coding\_prompts*** no cut no clean  
[https://huggingface.co/datasets/Alex-xu/practice\\_run\\_coding\\_prompts?row=1](https://huggingface.co/datasets/Alex-xu/practice_run_coding_prompts?row=1)

## Guardrails Testing Data:

- ***data-is-better-together/10k\_prompts\_ranked*** no cut no clean  
[https://huggingface.co/datasets/data-is-better-together/10k\\_prompts\\_ranked](https://huggingface.co/datasets/data-is-better-together/10k_prompts_ranked)