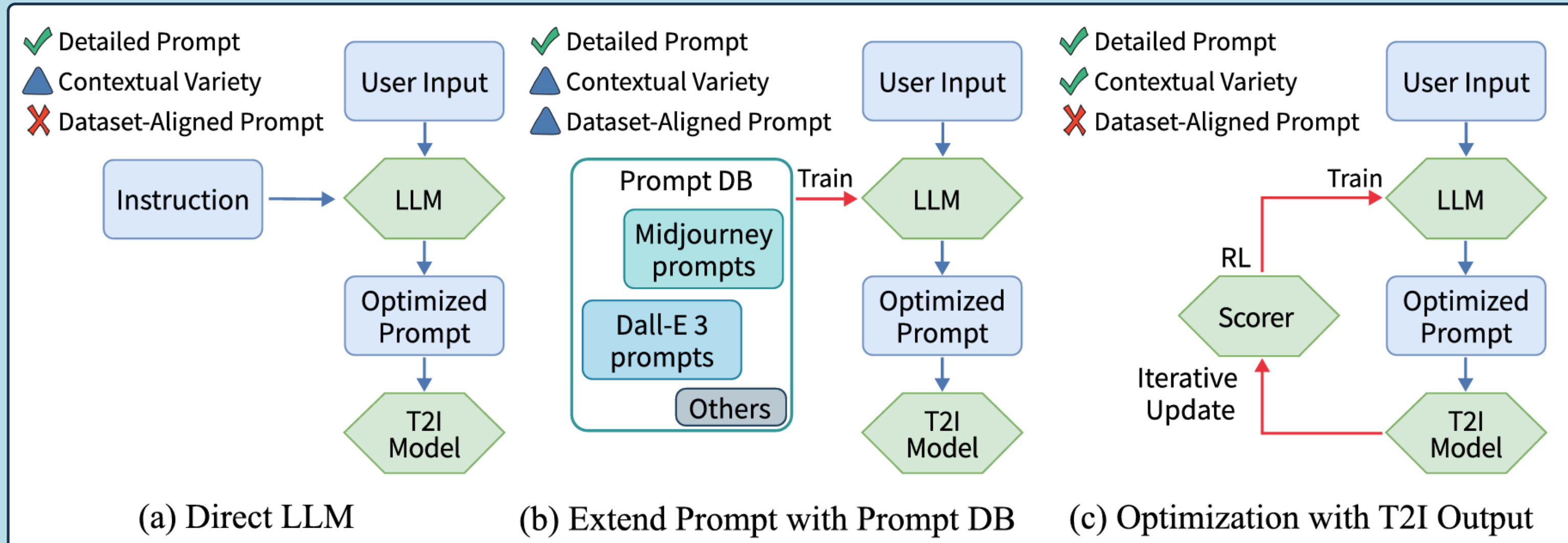


TIPO: Text-to-Image with Text Pre-sampling for Optimal Prompting

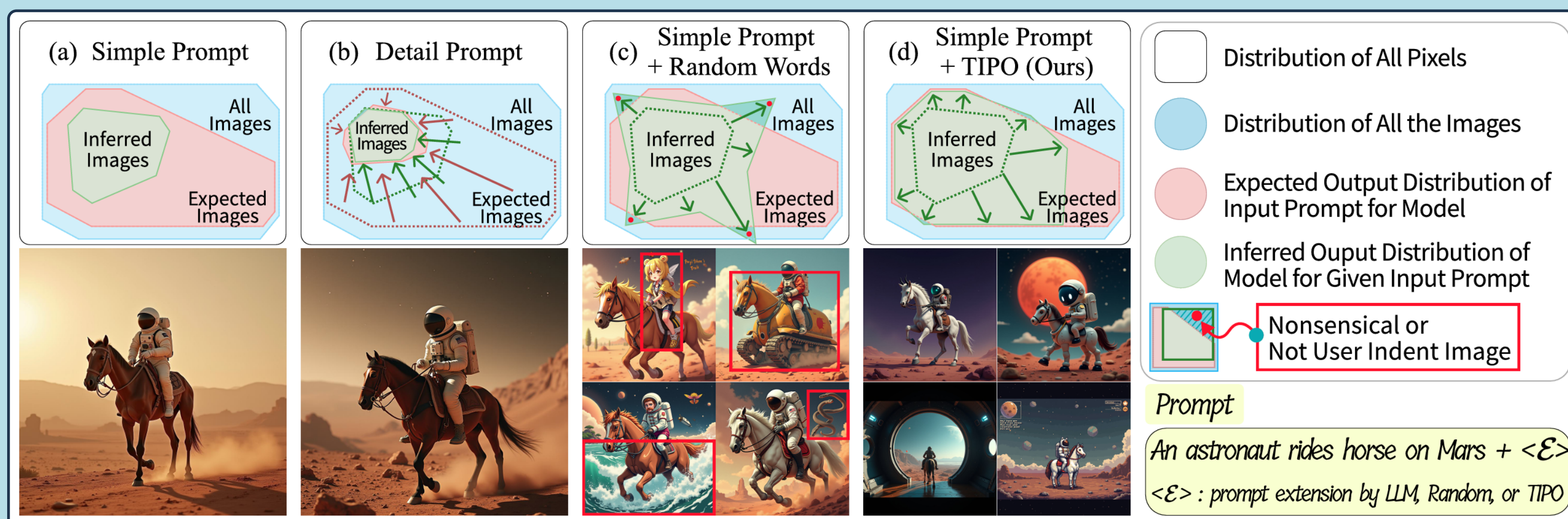
EECS06 葉適穎 Shih-Ying Yeh

Existing Prompting Method



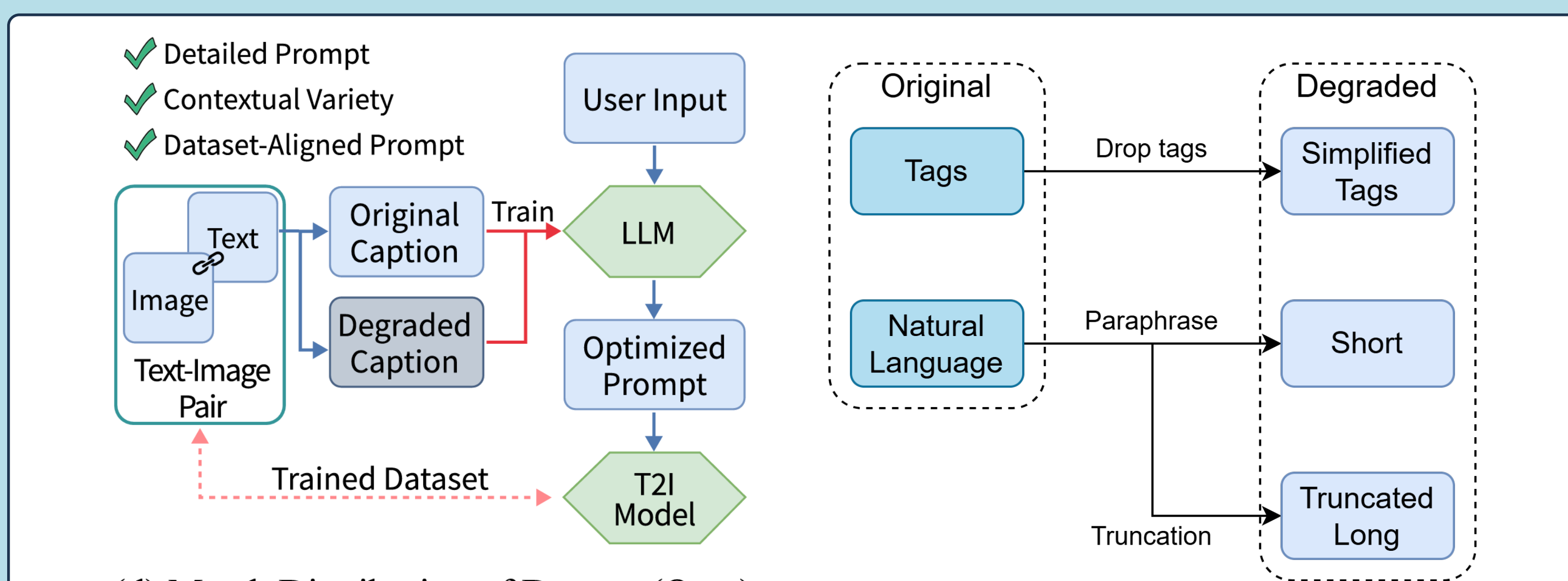
- Previous method rely on LLM's ability or can only work with specific Text-to-Image model
- Hard to generalize or transfer between different text-to-image models

The idea of Text Pre-sampling

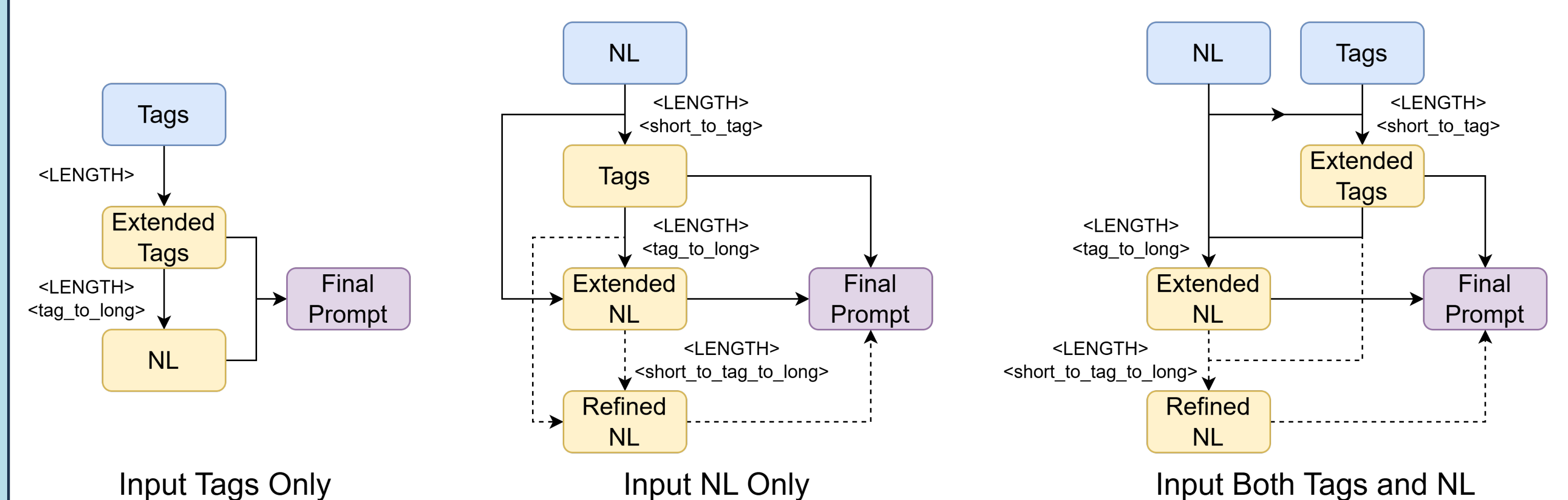


- We reveal the reason why "detail prompt" helps in Text-to-image
- Simple prompt require model to have higher capacity

TIPO Framework

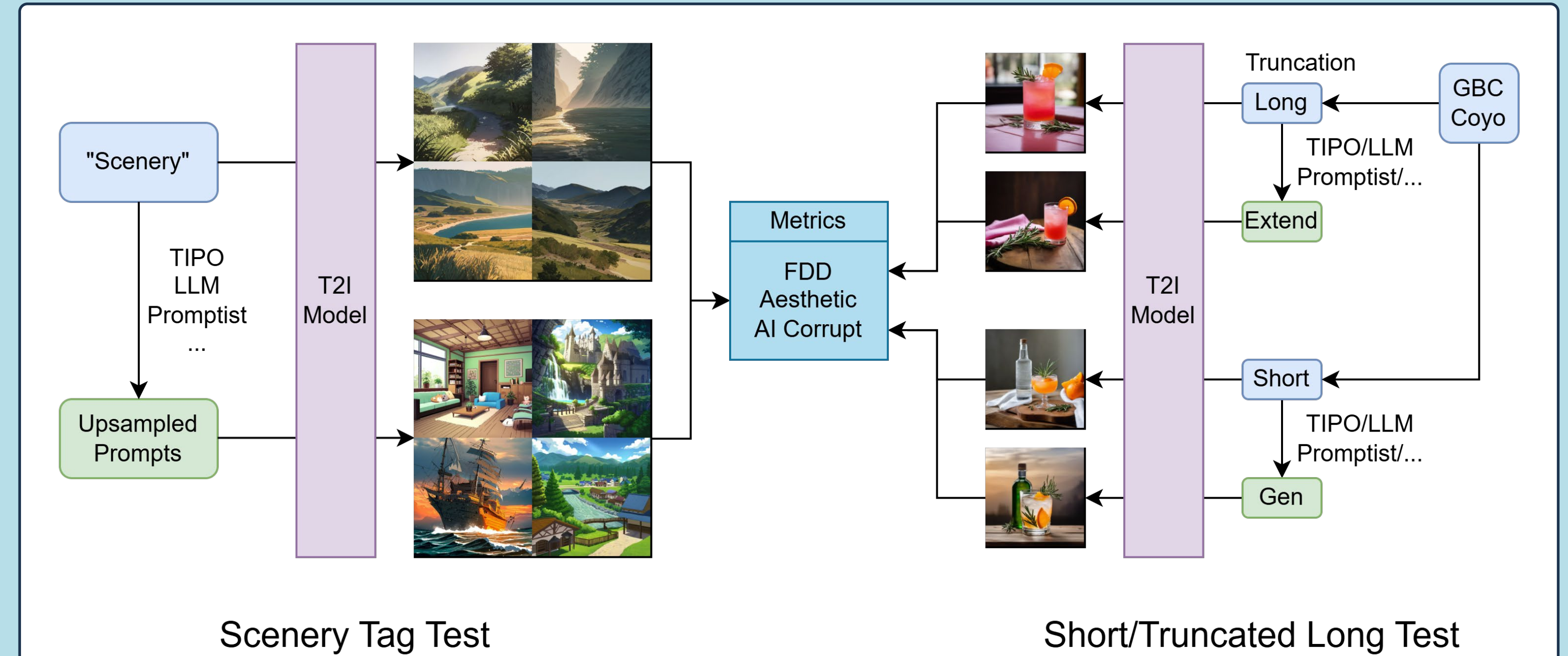


- Use the dataset we used for training Text-to-Image
- Use degraded caption to simulate user input
- We trained 2 models based on Danbooru2023, GBC10M, CoyoHD11M dataset
 - TIPO-200M: 200M LLaMA arch, 40B token seen
 - TIPO-500M: 500M LLaMA arch, 30B token seen



- We have two type of input, so we need to process them separately
- Use user input as condition to generate/extend final prompt
 - With some manually choose property such as length or output format

Experiments



- Metrics
 - FDD, Aesthetic Score, AI Corrupt Score
 - Compare with other methods
- Scenery Tag Test
 - Use "scenery" tag only as input to simulate extremely simple prompts
 - Use Kohaku-XL-Zeta Model which is trained on Danbooru2023 to test
- Short/Truncated Long Test
 - Use caption from GBC10M and CoyoHD11M
 - Use SDXL Model which is trained on general dataset

Results

Result of Scenery Tag Test



(a) Direct LLM (b) Prompt DB (c) RL (Promptist) (d) TIPO (Ours)

Metric	Original	LLM	Prompt DB	Promptist	TIPO (Ours)
FDD ↓	0.3558	0.5414	0.3247	0.2350	0.2282
Aesthetic Score ↑	5.0569	6.3676	6.1609	5.9468	6.2571
AI Corrupt Score ↑	0.4257	0.7490	0.5024	0.5669	0.9195



Result of Short/Truncated Long Test

Short	Original	LLM	Prompt DB	Promptist	TIPO(Ours)
FDD ↓	0.0957	0.1668	0.0980	0.1783	0.1168
Aesthetic Score ↑	5.8370	6.0589	5.8213	5.7963	<u>5.8531</u>
AI Corrupt Score ↑	0.7113	0.6985	0.7064	0.6314	0.7131
Truncated Long	Original	LLM	Prompt DB	Promptist	TIPO(Ours)
FDD ↓	0.0955	0.1683	0.1247	0.2096	<u>0.1210</u>
Aesthetic Score ↑	5.7497	6.0168	5.8191	5.7759	<u>5.8364</u>
AI Corrupt Score ↑	0.6868	0.6712	0.6741	0.5925	0.7130

