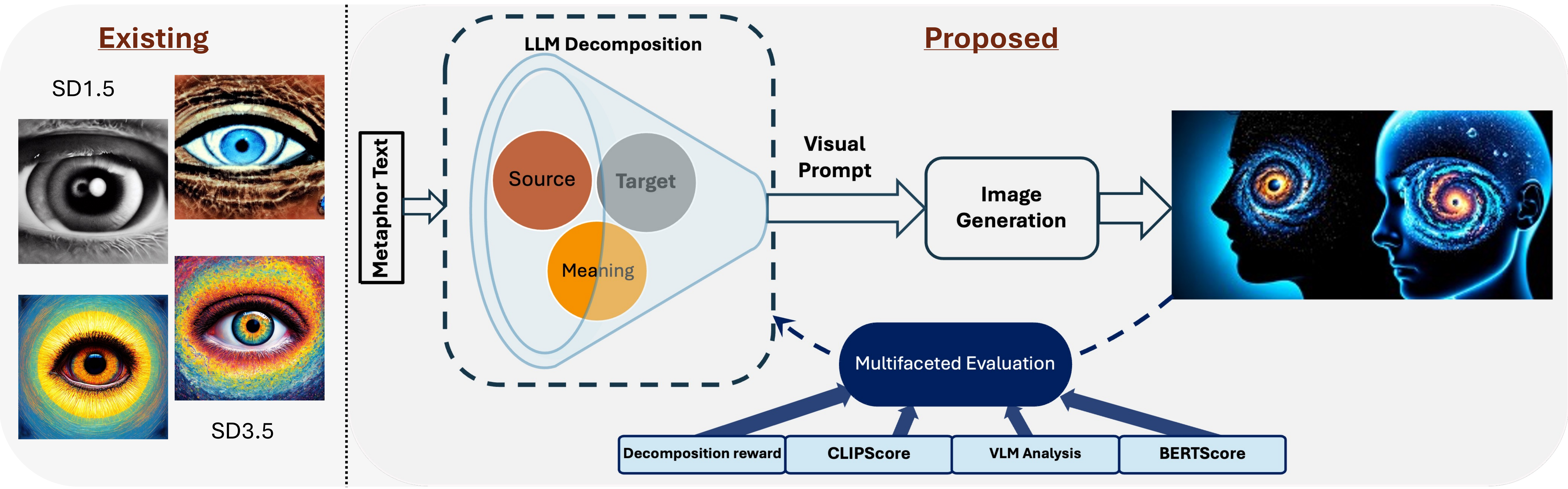
The Mind's Eye: A Multi-Faceted Reward Framework for Guiding Visual Metaphor Generation

Girish A. K.[1], Fatemeh N.[1], Katherine B.[1], Shenbin Q.[2], Diptesh K.[1]

[1]NICE Research Group & [2]Centre for Translation Studies, University of Surrey, UK

{g.koushik, f.nazarieh, k.birch, s.qian, d.kanojia}@surrey.ac.uk



Background

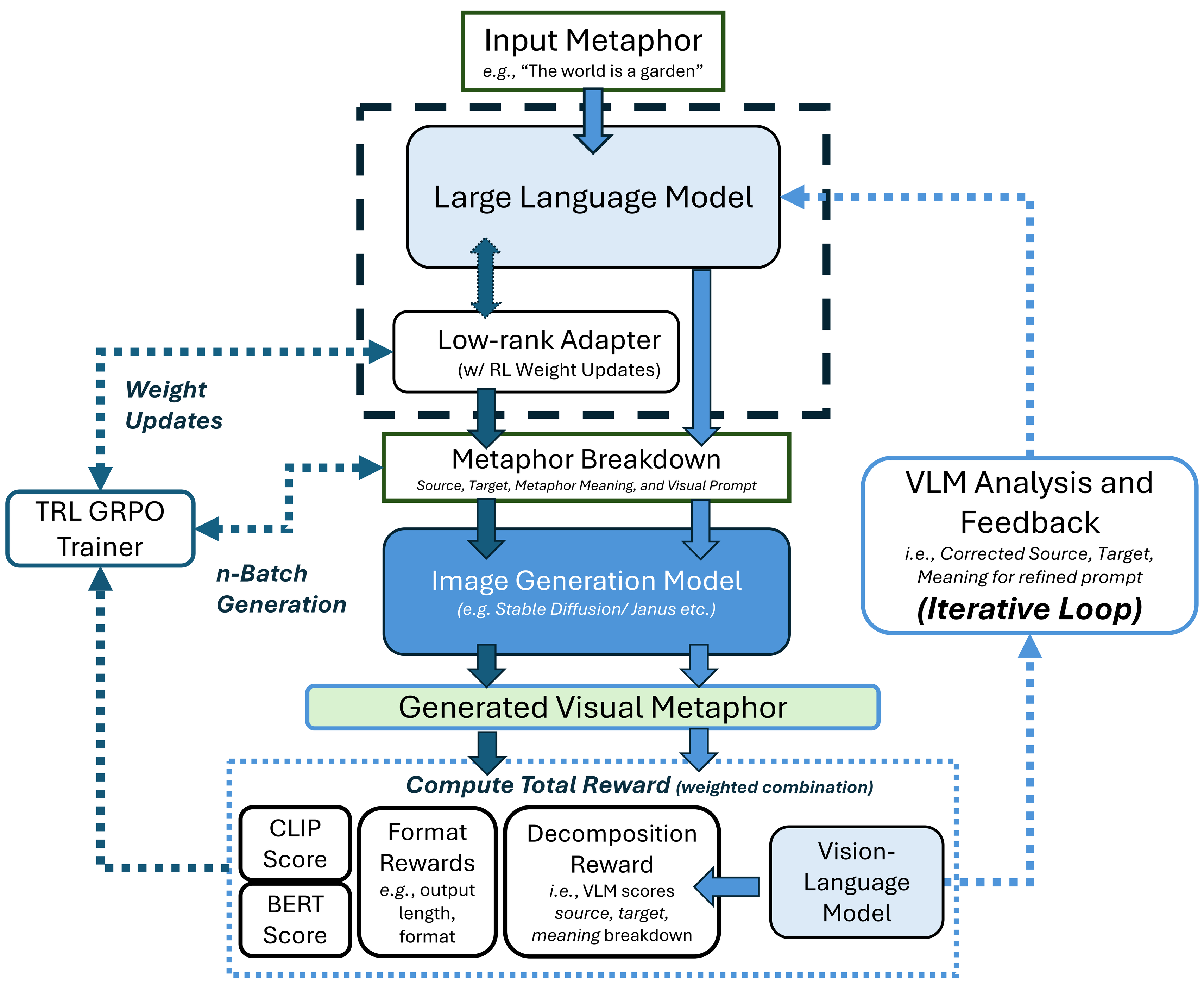
Visual metaphors convey abstract ideas through images, yet current Text-to-Image (T2I) models like Stable Diffusion struggle with figurative meaning.

- ► **Shallow Interpretations:** Models default to literal visual elements, missing the metaphor. *For example: "the mind's eye" yields a literal* ***eye*** *rather than a depiction of imagination.*
- ► **Evaluation Gap:** Lack of robust metrics to assess if the visual captures the true metaphorical meaning.

Our framework addresses these challenges by:

- ► Recursively generating context via **meaning decomposition.**
- ► **Evaluating generated images** for figurative alignment.

Methodology



Framework Architecture with both training-free and TRL GRPO pipelines.

We Introduce Two Complementary Pipelines

1. Training-Free Pipeline (In-Context Refinement)

This pipeline focuses on maximizing figurative alignment without any model fine-tuning, leveraging only the capabilities of large pre-trained LLMs.

- ► Employs an **Iterative Prompt Refinement** loop, where the LLM's weights remain **unchanged**.
- ► The LLM receives structured feedback (VLM scores, CLIP signals, and Meaning Alignment) through a **meta-prompt**, and generates a refined prompt $\mathbf{P}_{i+1}$ using In-Context Learning.
- ► The objective is to identify the optimal prompt $\mathbf{P}^*$ that maximizes the total reward within a fixed number of iterations:

$$\mathbf{P}^* = \underset{\mathbf{P}_i}{\operatorname{argmax}} R(\mathbf{G}(\mathbf{P}_i), \mathbf{P}_i, \mathbf{D})$$

2. Fine-Tuned Pipeline (GRPO-based Refinement)

This pipeline targets efficiency and task specialization by fine-tuning a smaller model via Reinforcement Learning.

- ► Applies **Policy Optimization** using the **Group Relative Policy Optimization (GRPO)** algorithm, guided by a multi-component reward.
- ► Performs **Explicit Adaptation** by updating LoRA parameters ($\mathrm{L}\theta$), enabling the model to maximize the expected reward $\mathbf{E}[R(C)]$ for its generated decompositions and prompt candidates.
- ► Learns optimal parameters θ^* by minimizing the GRPO loss:

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \; -\mathbf{E}_{C\sim\mathcal{L}_\theta} \log \sigma_\beta(\mathbf{R}(C) - \mathbf{R}_{\text{ref}}(C))$$

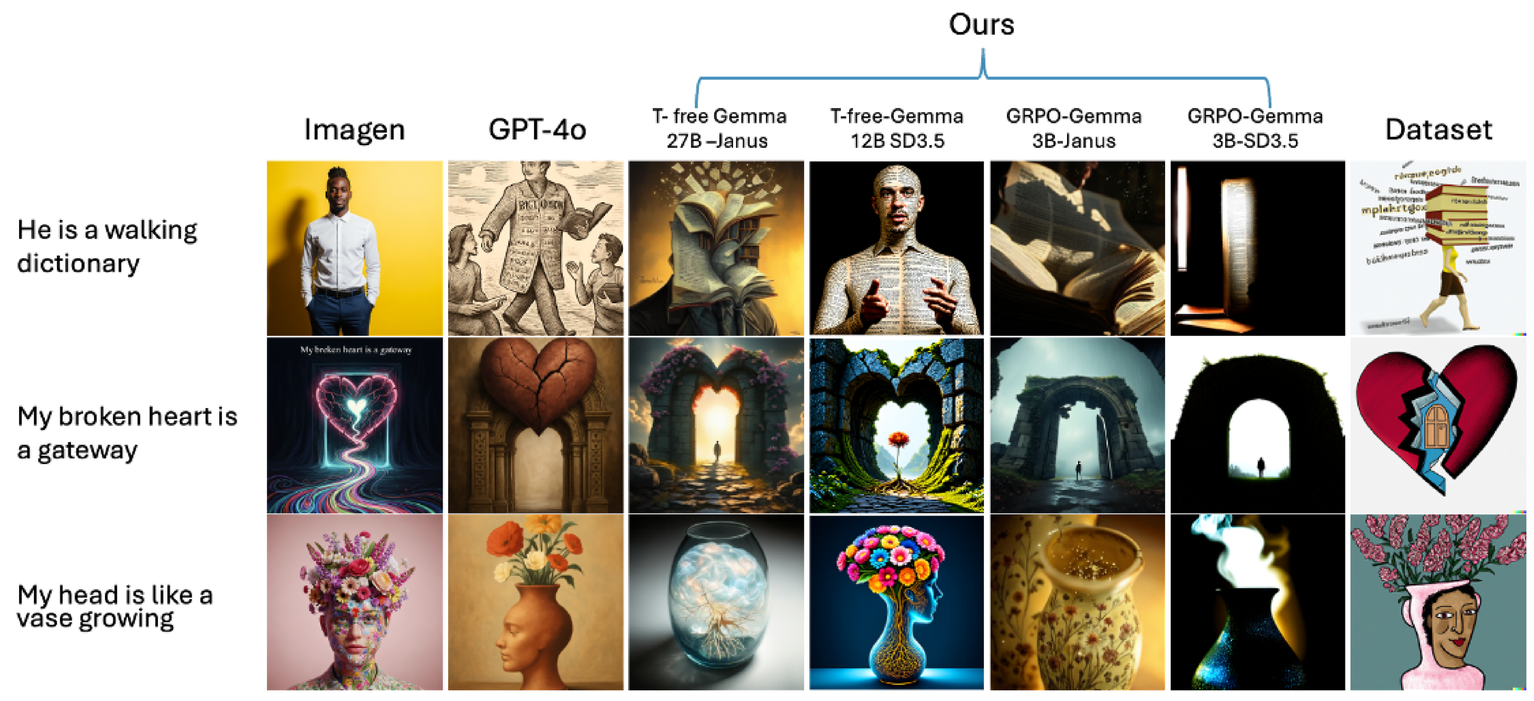
3. S–T–M: Source, Target, Meaning

- **Source (S):** The concrete concept the metaphor draws from. For example: *"a garden"* in *"The world is a garden."*
- **Target (T):** The concept we aim to understand, usually more abstract or complex, onto which we map the structure of the source. For example: *"the world."*
- **Meaning (M):** The connection linking S to T, the intended interpretation. For example: *"The world has potential and needs care."*

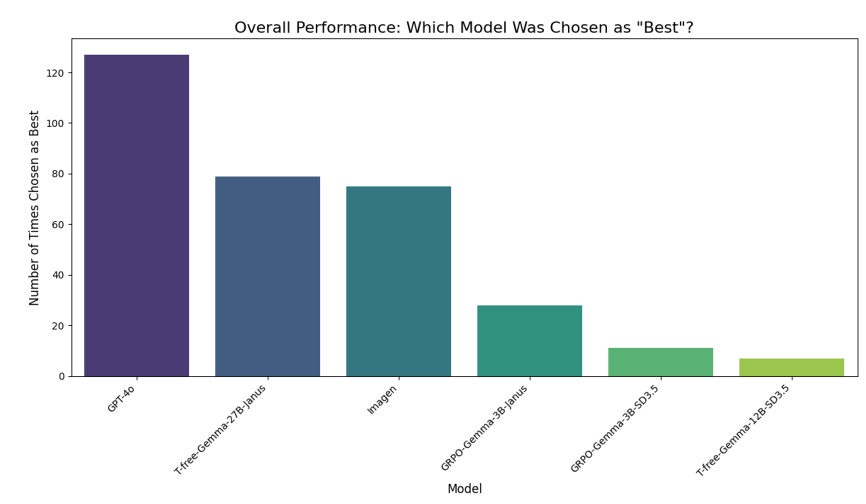
Results and Discussion

Combined quantitative evaluation. All metrics are in the range $[0, 1]$ with higher values indicating better alignment, on dataset, **HAIVMet**

Architecture / Model	Key Configuration	CLIP Score (↑)	MA Score (↑)	Decomp. Score (↑)
Training-Free	GEMMA-27B / Janus-Pro	0.2960	0.8760	0.8668
Fine-Tuned (GRPO)	GEMMA-4B / SD‡ / QWEN-32B	0.2789	0.6396	0.6480
Zero-Shot Baseline	GPT-4O	0.2296	0.8180	0.8072
Zero-Shot Baseline	Imagen-3	0.2224	0.7353	–

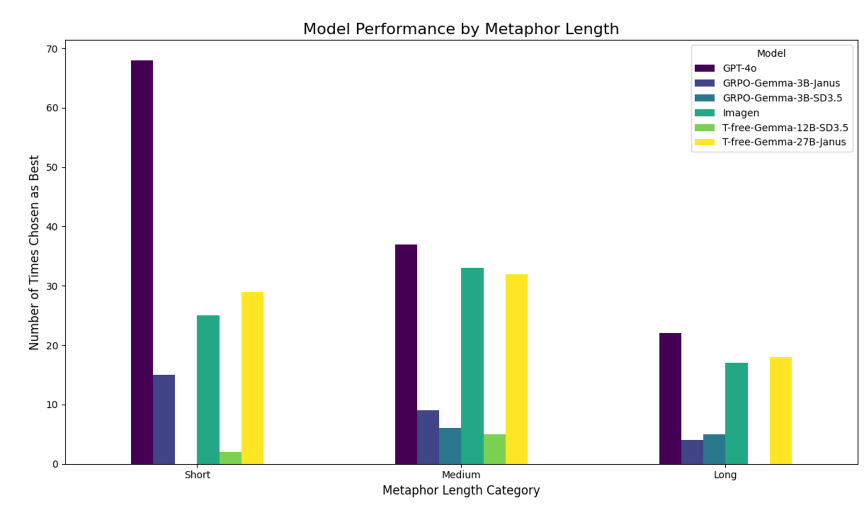


Examples from user study for evaluating metaphorical image generation.



(a) Best Model Overall

- ► **Quantitative Evaluation**
  1. **Training-free pipeline (GEMMA-3-27B, Janus-Pro-7B, QWEN2.5-VL-32B):** Achieves best CLIP (0.296) and Meaning Alignment (0.8760).
  2. GRPO fine-tuned model achieves competitive results with smaller backbone.



(b) Performance by Length

Figure: Best Overall Model and Model Performance with respect to Metaphor Length.

- ► **Qualitative Evaluation** We conduct a user Study (15 annotators, 50 metaphors). Each evaluation sample contained six images (GPT-4o, Imagen-3, two training-free models, and two GRPO models), totaling 300 images for human assessment.
  1. GPT-4o preferred overall. We believe this is due to the image quality and presentation rather than the actual semantics of metaphor understanding.
  2. Overall, the training-free pipeline ranked 2nd in the user study.

Conclusion

- ► We introduce a metaphor-aware image generation framework that combines structured decomposition and multi-faceted reward.
- ► **Training-free** and **GRPO fine-tuned** approaches both outperform zero-shot baselines.
- ► Structured S–T–M decomposition improves semantic alignment between metaphoric prompt with the metaphoric image.