LMM-R1: empowering 3B LMMs with strong reasoning abilities with two-stage rule-based RL

Yingzhe PENG, Gongrui ZHANG, Miaosen ZHANG, Zhiyuan YOU,Jie LIU, Qipeng ZHU, Kai YANG, Xingzhong XU, Xin GENG, Xu YANG

Frontiers of Computer Science DOI: 10.1007/s11704-026-51776-8

Problems & Ideas

Problems

- Multimodal reasoning data is far scarcer than text data, limiting direct RL training.
- Vision-language alignment degrades pretrained language reasoning, especially in 3B models.
- Direct rule-based RL on multimodal tasks yields limited reasoning improvement.

Our Idea: LMM-R1

- Stage 1 (FRE): Restore foundational reasoning with abundant text-only data via rule-based RL.
- Stage 2 (MGT): Generalize enhanced reasoning to multimodal domains with limited multimodal data.
- Key insight: Text reasoning first creates a robust foundation for multimodal generalization.

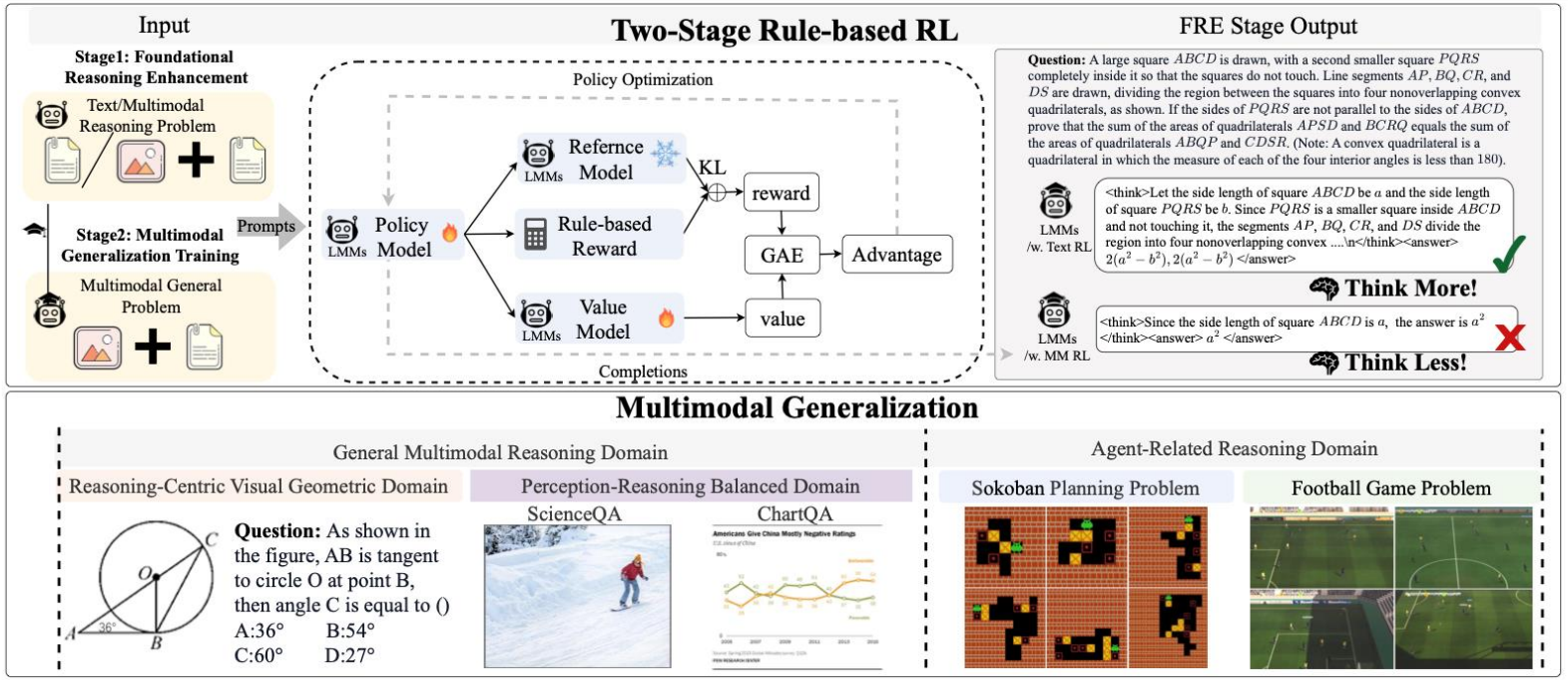


Fig. 2: LMM-R1 two-stage framework. FRE trains on text-only data; MGT generalizes to multimodal domains.

Main Results

- LMM-R1 achieves 40.95% MM Avg (+4.69%) and 51.35% Text Avg (+4.50%) over baseline, comparable to GPT-4o.
- Text-RL alone significantly boosts reasoning; Multimodal-RL alone improves perception but hurts text reasoning.

