

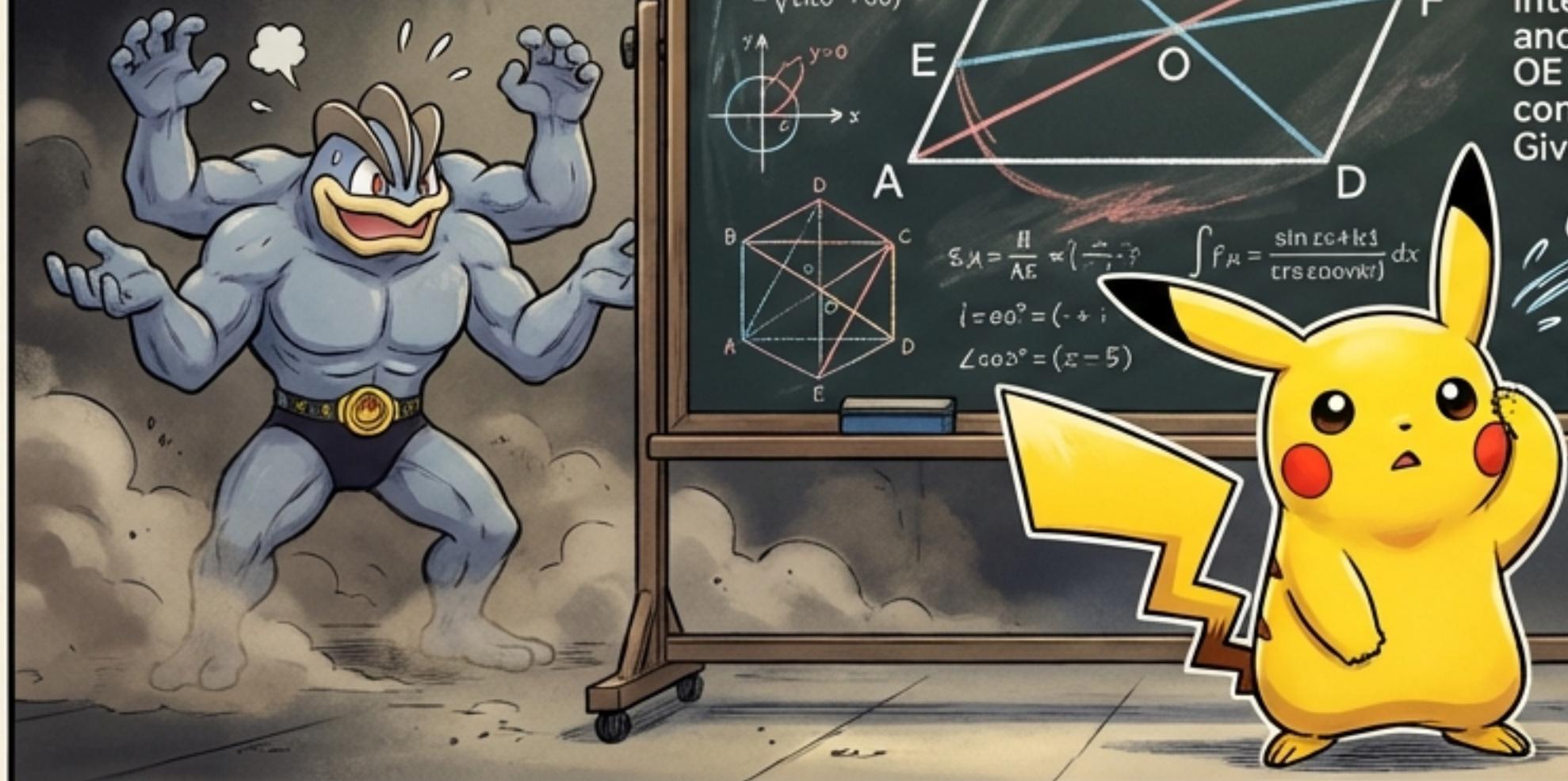
OpenMMReasoner: A Champion's Recipe for Multimodal Reasoning

The training journey of a Large Multimodal Model.

“皮卡丘，准备好开始我们的特训了吗？
目标是——多模态
推理冠军！”

Pika-Pika!

The Wall of Confusion



The Problem.

Existing Large Multimodal Models (LMMs) struggle with complex reasoning that combines images and text. Furthermore, the **“lack of transparent and reproducible data curation and training strategies** remains a major barrier to scalable research.”

“嗯……光靠十万伏特是解不开这道题的。我们需要一种更聪明、更清晰的训练方法！”



The Secret Recipe: A Two-Stage Path to Mastery



Stage 1: Gathering Diverse Knowledge



The first step of SFT is curating a diverse dataset from various public sources, starting with ~103k raw question-answer pairs. All data is standardized to a unified reasoning format to ensure training stability.

Key Insight:
"Scaling data diversity is a critical factor for curating high-quality datasets."
It's not just about more data, but more varied* data.



Stage 1: Learning from a Master

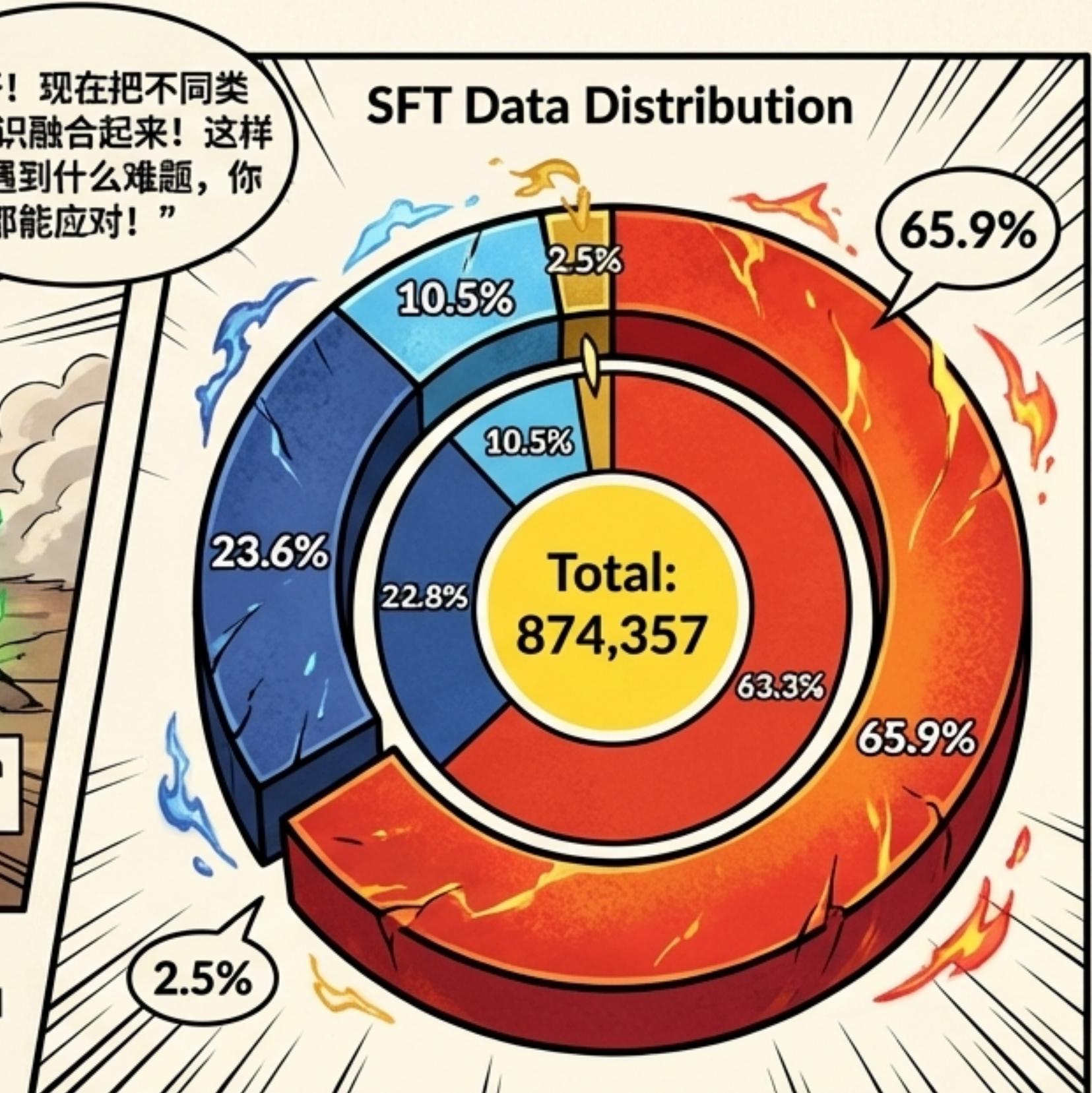


To improve data quality and efficiency, a stronger ‘teacher model’ is used to generate high-quality, **step-by-step reasoning traces**. Generating multiple verified answers for each question is critical for building robust understanding.

Stage 1: The Power of Fusing Knowledge



To improve generalization, the 583k general-reasoning dataset is mixed with cross-domain data, specifically image-based and text-based math problems. This creates the final 874k mixed SFT dataset, a powerful "ColdStart" model ready for the next stage of training.



Stage 2: The Sparring Arena of Reinforcement Learning



With a strong foundation from the SFT stage, we move to the RL phase. The goal is to further sharpen and stabilize the model's reasoning abilities through targeted practice on a 74k-sample dataset.

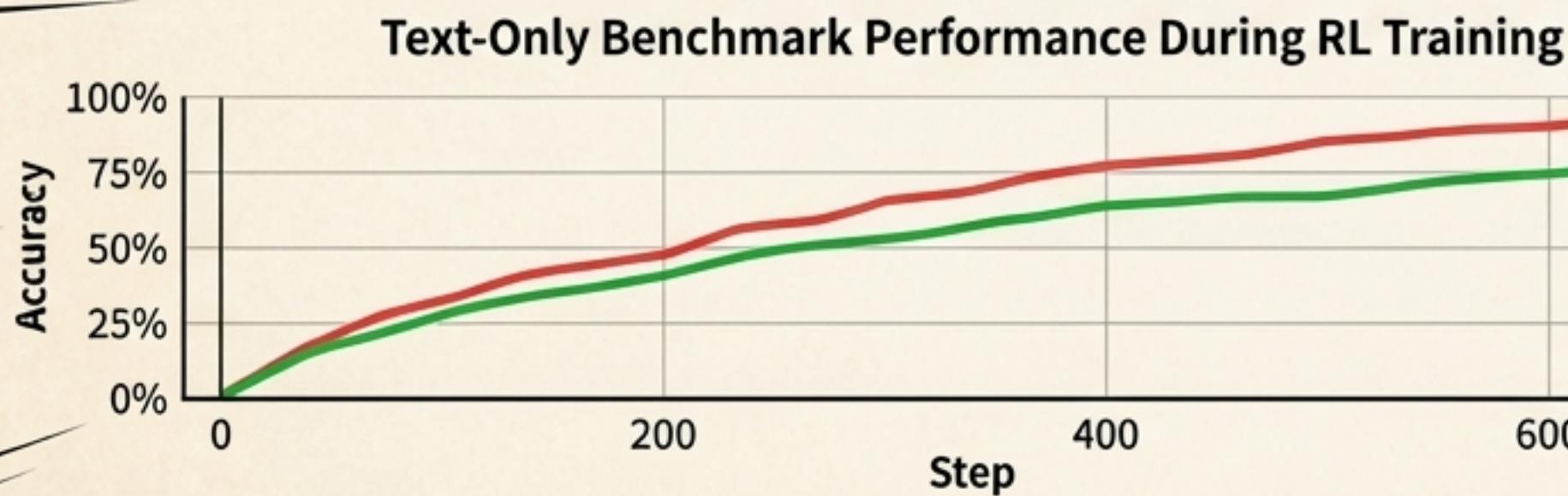
Key Insight: The choice of RL algorithm is critical. It must balance exploration (trying new solutions) with stability to ensure efficient and effective training.

The Winning Move: Group Sequence Policy Optimization (GSPO)



After comparing multiple RL strategies, **GSPO** was chosen as the core algorithm. It demonstrated “faster convergence, higher rewards, and more stable behavior” compared to alternatives like DAPO and GRPO. GSPO prevents ‘entropy collapse’—the model getting stuck in a repetitive solving pattern.

A Surprising Power-Up: Cross-Modal Skill Transfer



A fascinating discovery emerged during RL training: **enhancing multimodal reasoning also improved the model's ability to solve purely text-based problems.** As shown in **Table 8**, performance on text benchmarks like AIME and GPQA increased significantly from the baseline after SFT and RL stages.

“太棒了，皮卡丘！你的视觉推理能力变强了，连纯文本问题的解决能力也一起提升了！”



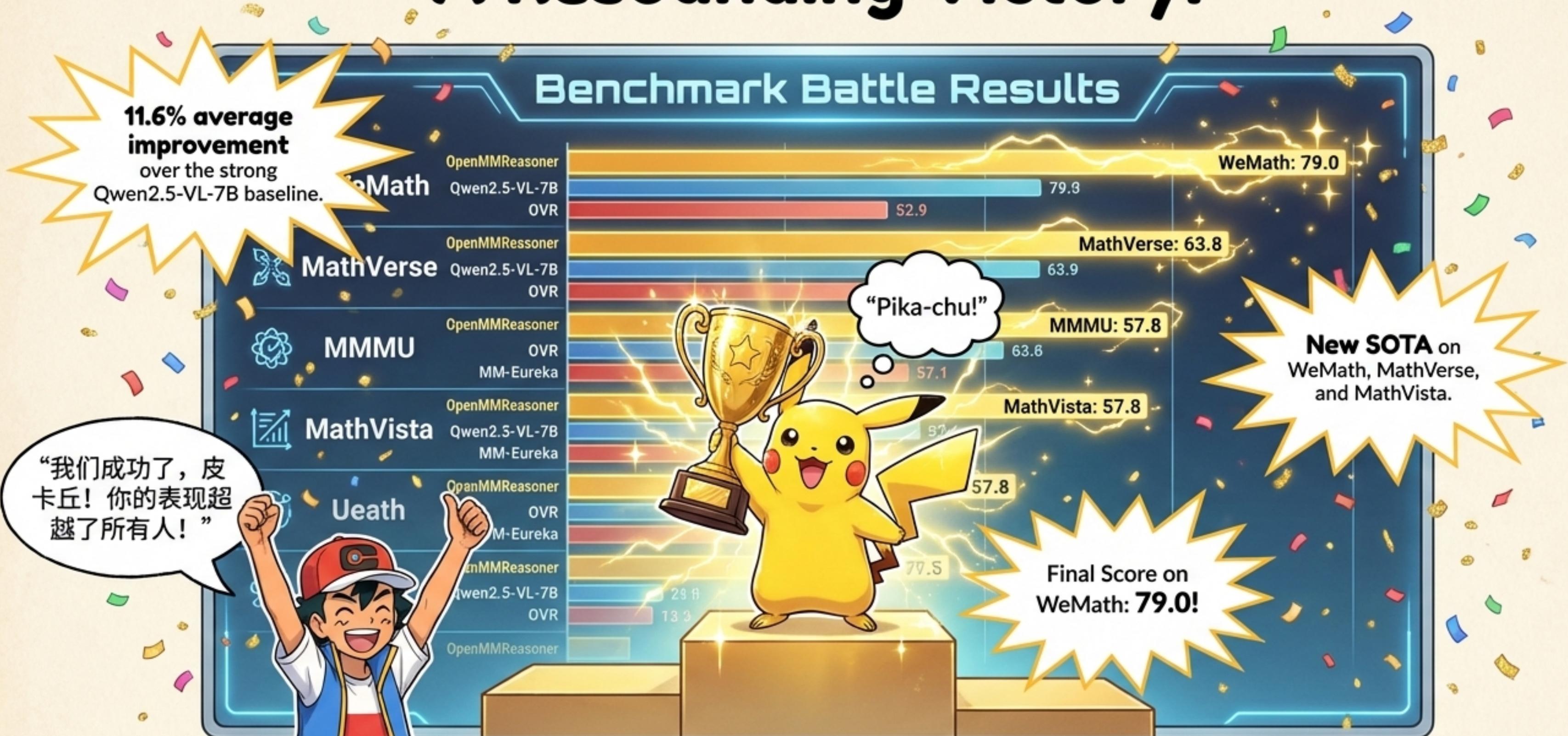
Key Insight: Reasoning skills are transferable. Improving one can strengthen another, highlighting shared cognitive foundations across different modalities.

The Grand Tournament: The Benchmark Battle!



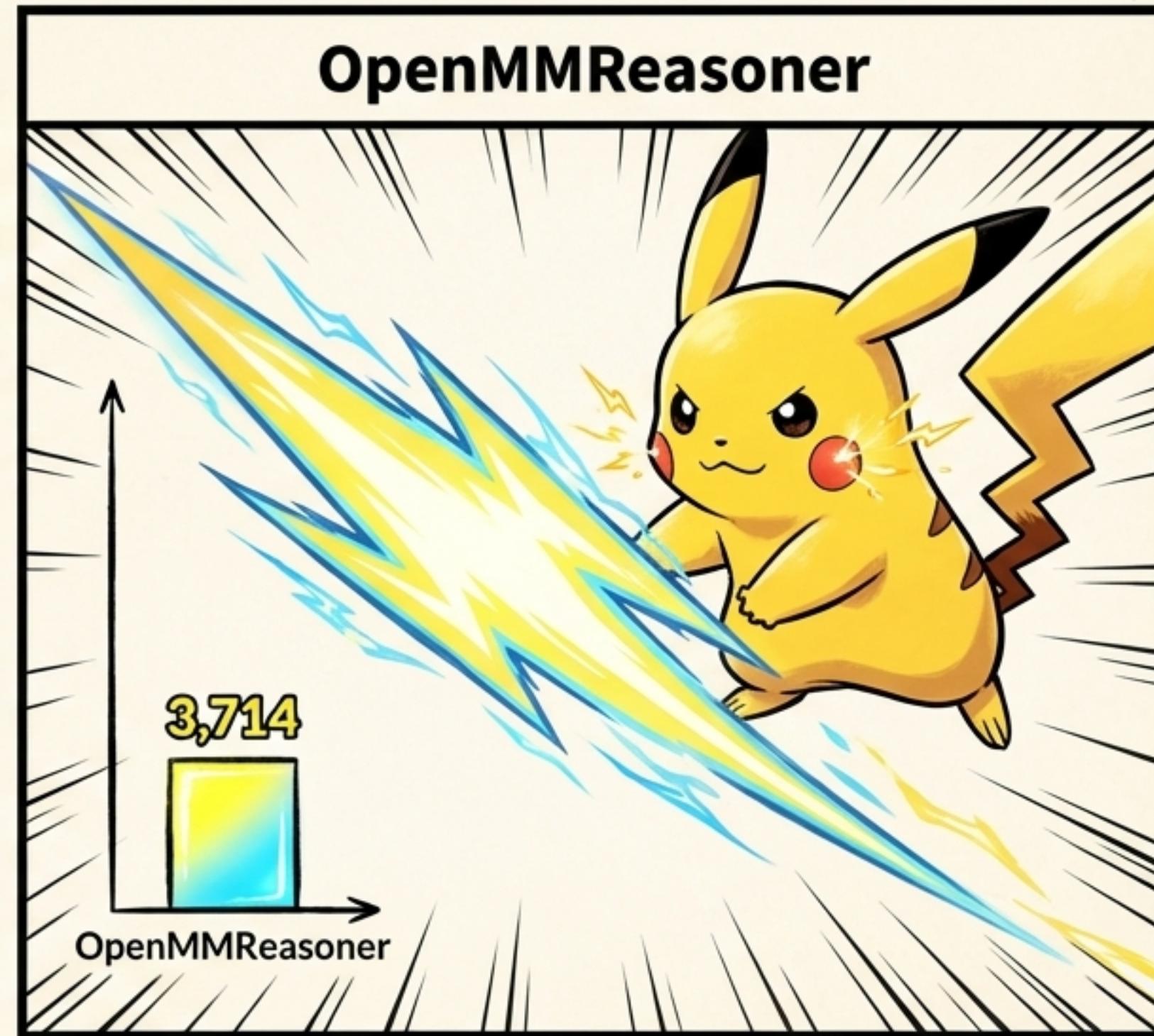
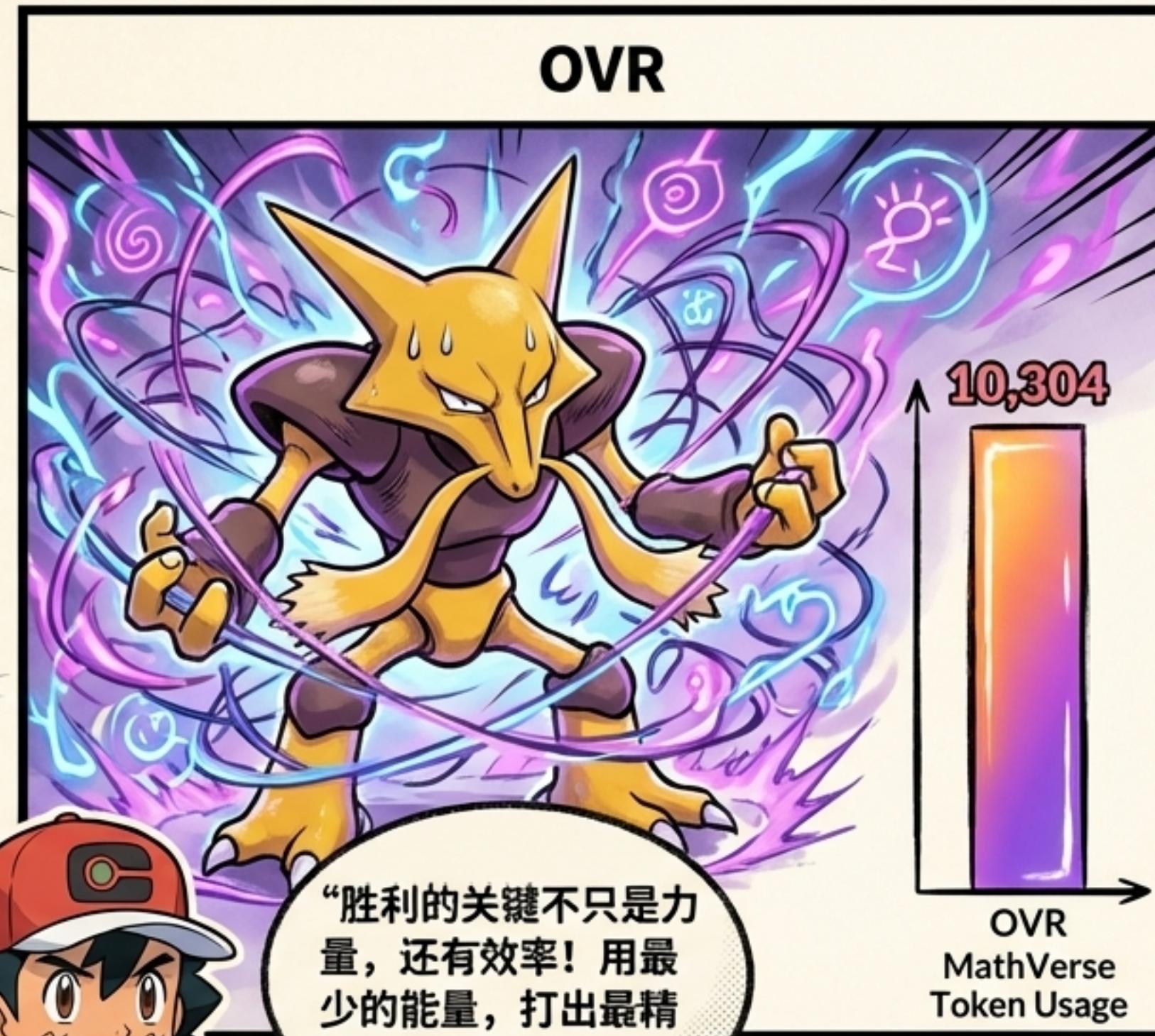
The final test. OpenMMReasoner is evaluated against other state-of-the-art models on a wide range of multimodal reasoning benchmarks, including MathVerse, MMMU, WeMath, and MathVista.

A Resounding Victory!



The culmination of training and testing! OpenMMReasoner (Ours) achieves decisive victory, demonstrating superior performance across multiple multimodal benchmarks and setting new standards in the field.

Not Just Stronger, But Smarter and More Efficient



OpenMMReasoner doesn't just get the right answer; it does so more efficiently. It avoids the 'overthinking' and excessive response length of some other models, achieving a better balance between reasoning depth and computational cost.

The Champion's Journal: Key Lessons Learned



The Journey Continues: New Frontiers to Explore



Our work focuses on the image domain, but the journey doesn't end here.

Future work aims to:

- Extend the OpenMMReasoner recipe to more modalities like video and audio.
- Apply the methodology to a wider range of model families to test its generality.
- Explore the upper bounds of performance with further scaling.

Become a Reasoning Champion!



Join the adventure! OpenMMReasoner is the first study in its class to fully open-source its entire workflow.

All our codes, data curation pipeline, SFT/RL datasets, and model weights are publicly available for transparent and reproducible research.

<https://github.com/EvolvingLMMs-Lab/OpenMMReasoner>