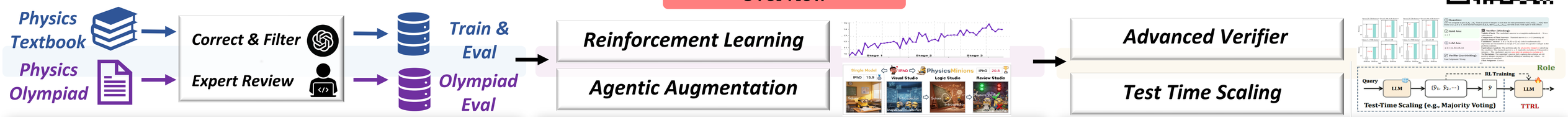
P1: Mastering Physics Olympiads with Reinforcement Learning

上海人工智能实验室
Shanghai Artificial Intelligence Laboratory

NEURAL INFORMATION PROCESSING SYSTEMS

Overview

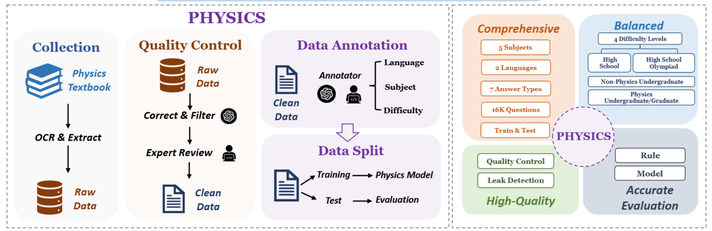


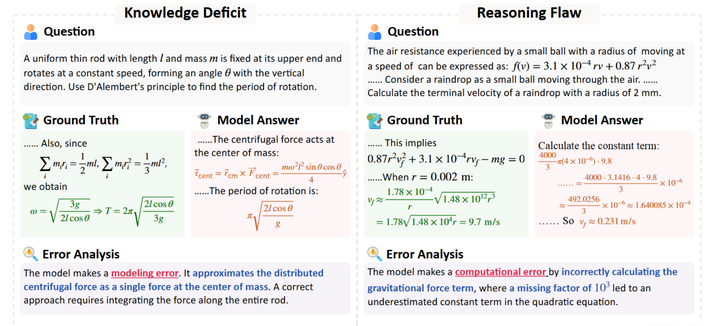
- **Large-scale, high-quality** physics train/test dataset.
- **Olympiad-focused** Physics Benchmark.

- First open-source model with IPHO25 **Gold-medal**.
- New agentic paradigm for physics reasoning.

- **Verifier with COT** for scientific LLM evaluation and training.
- Test time training with unsupervised data for physics reasoning.

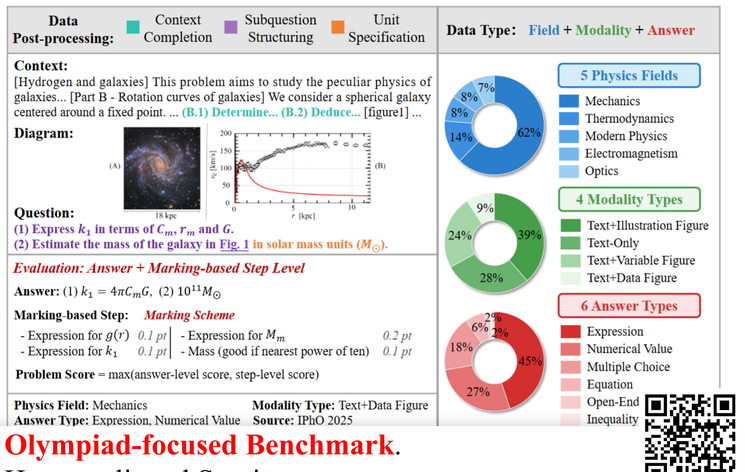
Dataset





- **Large-scale, high-quality** physics train/test dataset.
- Rule+Model **evaluation framework** for physics.

Benchmark

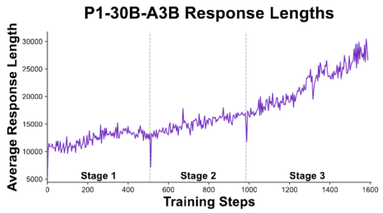


- **Olympiad-focused Benchmark**.
- Human-aligned Scoring.

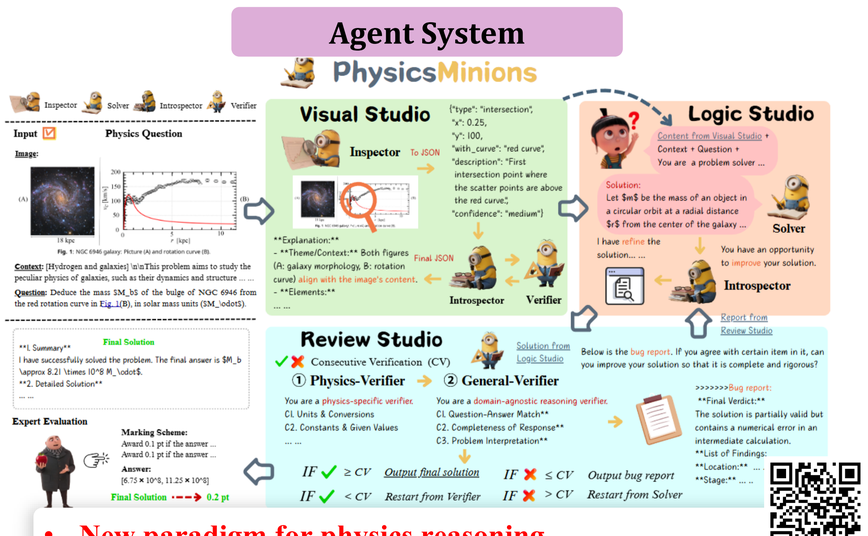
Reasoning Model

Model Training

- Adaptive Learnability Adjustment
  - Preliminary Pass Rate Filtering
  - Adaptive Exploration Space Expansion
- Training Stabilization Mechanism



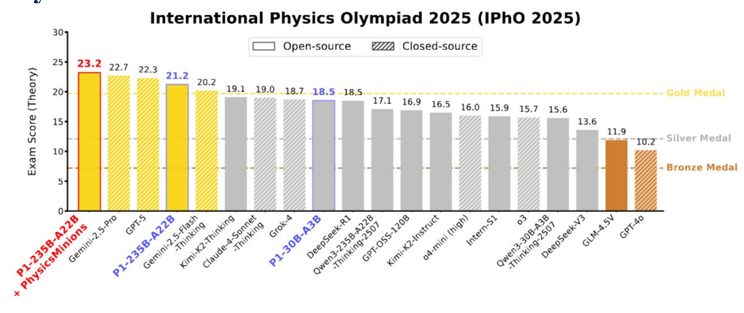
Agent System



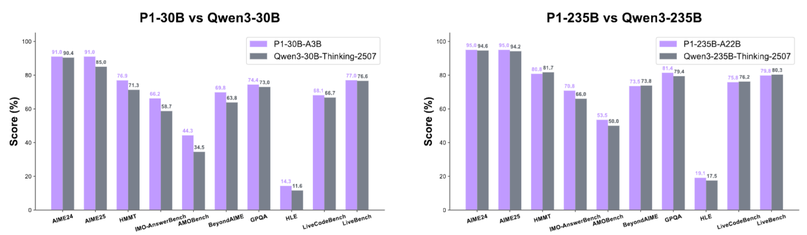
- **New paradigm for physics reasoning.**

Experiments

Physics Results



Math & Code Results



Discussions

Test-time Training & Verifier

Table 5 | Performance of our P1-30B-A3B model combined with TTRL on the HiPhO benchmark (note that results presented here are evaluated only in answer level, leveraging Qwen3-30B-A3B-Instruct as judge).

Physics Olympiad	IPhO		APhO	EuPhO		NBPhO		PanPhO		PanMechanics		F=MA		Avg.
Year	2025	2024	2025	2025	2024	2025	2024	2025	2024	2025	2024	2025	2024	
P1-30B-A3B	12.0	14.4	23.1	0.5	12.1	24.5	13.8	41.7	49.5	58.1	73.1	17.9	17.8	27.6
P1-30B-A3B + TTRL	13.8	15.4	23.6	0.3	14.1	24.4	12.8	42.9	50.4	61.1	75.2	18.0	17.8	28.4