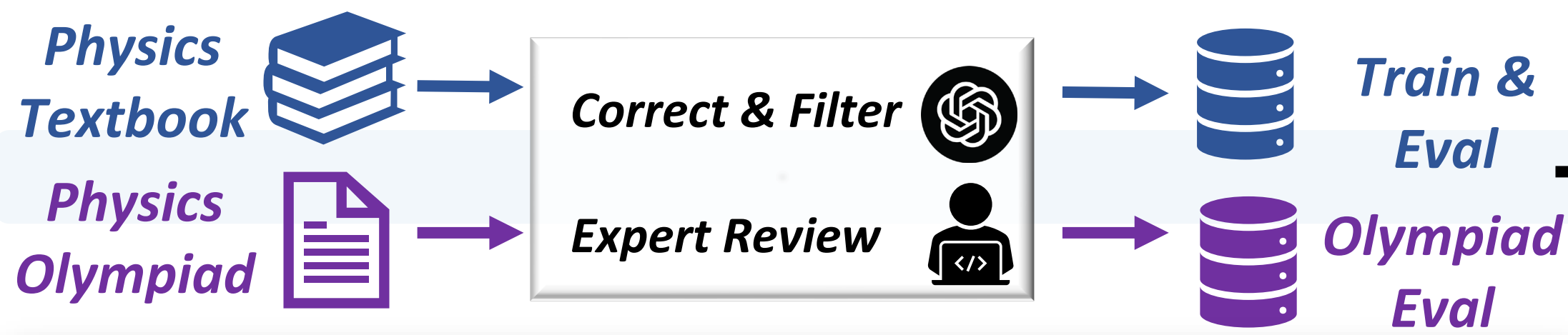
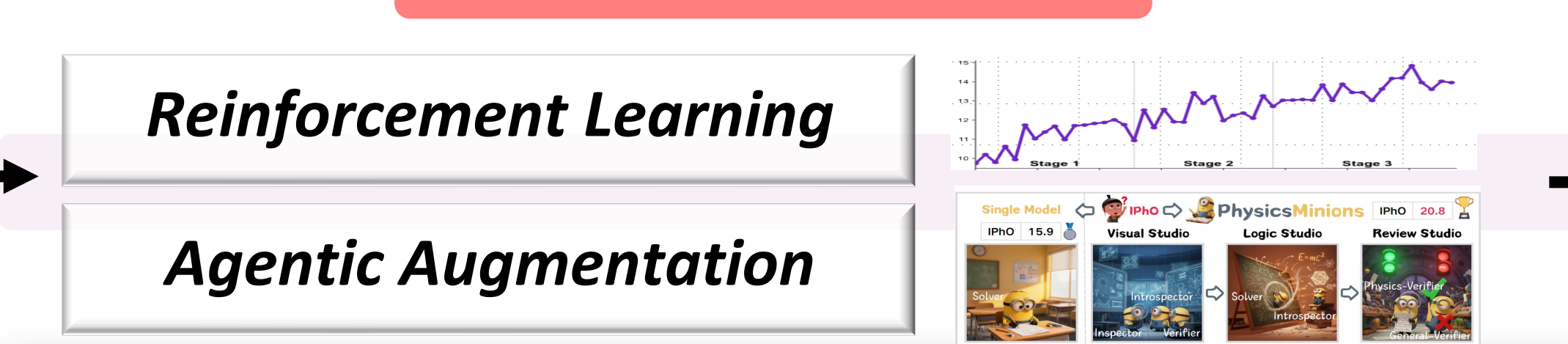


# P1: Mastering Physics Olympiads with Reinforcement Learning

## Overview



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

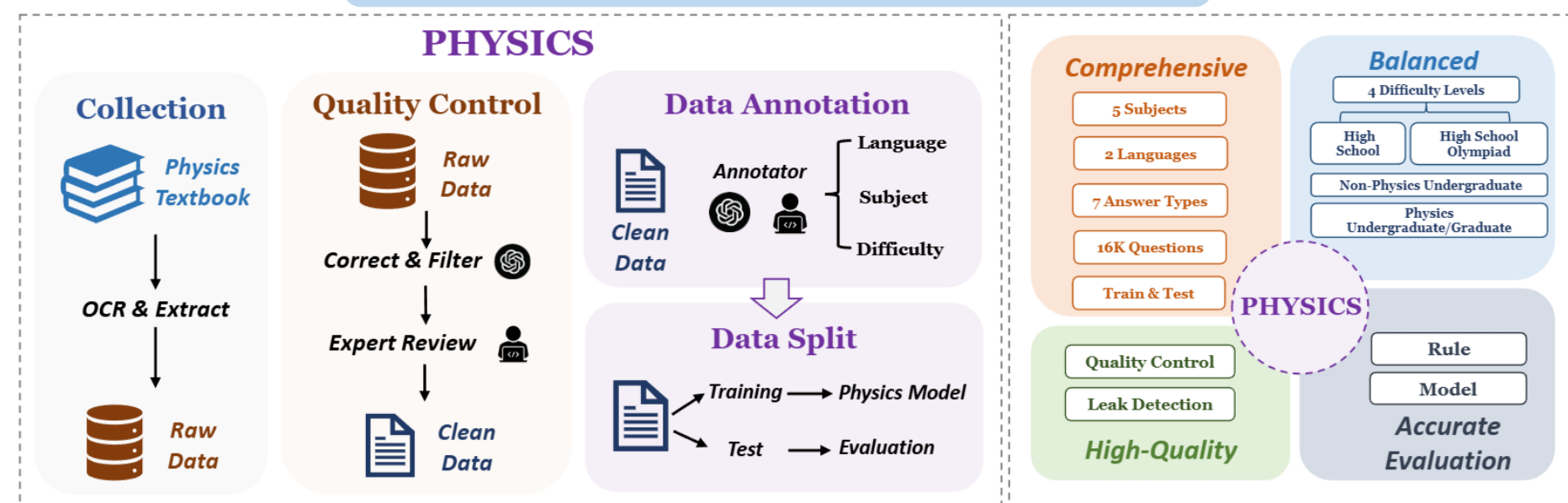


- First open-source model with IPHO25 **Gold-medal**.
- New agentic paradigm for physics reasoning.



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

## Dataset



**Knowledge Deficit**

Question: A uniform thin rod with length  $l$  and mass  $m$  is fixed at its upper end and rotates at a constant speed, forming an angle  $\theta$  with the vertical direction. Use D'Alembert's principle to find the period of rotation.

Ground Truth: ..... Also, since  $\sum m_i r_i = \frac{1}{2} m l$ ,  $\sum m_i r_i^2 = \frac{1}{3} m l^2$ , we obtain  $\omega = \sqrt{\frac{3g}{2l \cos \theta}} \Rightarrow T = 2\pi \sqrt{\frac{2l \cos \theta}{3g}}$

Model Answer: .....The centrifugal force acts at the center of mass:  $\vec{F}_{cent} = \vec{r}_{cm} \times \vec{F}_{cent} = \frac{m \omega^2 l^2 \sin \theta \cos \theta}{4}$  .....The period of rotation is:  $T = 2\pi \sqrt{\frac{2l \cos \theta}{3g}}$

Error Analysis: The model makes a **modeling error**. It approximates the distributed centrifugal force as a single force at the center of mass. A correct approach requires integrating the force along the entire rod.

**Reasoning Flaw**

Question: The air resistance experienced by a small ball with a radius of  $r$  moving at a speed of  $v$  can be expressed as:  $f(v) = 3.1 \times 10^{-4} r v + 0.87 r^2 v^2$  ..... Consider a raindrop as a small ball moving through the air. Calculate the terminal velocity of a raindrop with a radius of 2 mm.

Ground Truth: ..... This implies  $0.87 r^2 v^2 + 3.1 \times 10^{-4} r v - mg = 0$  .....When  $r = 0.002$  m:  $v_T \approx \frac{1.78 \times 10^{-4}}{r} \sqrt{1.48 \times 10^{17} r^3} = 1.78 \sqrt{1.48 \times 10^9 r} = 9.7$  m/s

Model Answer: Calculate the constant term:  $\frac{4000}{3} (4 \times 10^{-6}) \cdot 9.8 = \frac{4000}{3} \cdot 3.1416 \cdot 4 \cdot 9.8 \times 10^{-6} \approx \frac{492.0256}{3} \times 10^{-6} \approx 1.640085 \times 10^{-4}$  ..... So  $v_T \approx 0.231$  m/s

Error Analysis: The model makes a **computational error** by incorrectly calculating the gravitational force term, where a missing factor of  $10^3$  led to an underestimated constant term in the quadratic equation.

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

## Benchmark

**Data Post-processing:** Context Completion, Subquestion Structuring, Unit Specification

**Data Type:** Field + Modality + Answer

**Context:** [Hydrogen and galaxies] This problem aims to study the peculiar physics of galaxies... [Part B - Rotation curves of galaxies] We consider a spherical galaxy centered around a fixed point. ... (B.1) Determine... (B.2) Deduce... [figure] ...

**Diagram:** (A) Image of a galaxy, (B) Rotation curve plot showing velocity  $v$  vs radius  $r$ .

**Question:** (1) Express  $k_1$  in terms of  $C_m$ ,  $r_m$  and  $G$ . (2) Estimate the mass of the galaxy in Fig. 1 in solar mass units ( $M_\odot$ ).

**Evaluation: Answer + Marking-based Step Level**

Answer: (1)  $k_1 = 4\pi C_m G$ , (2)  $10^{11} M_\odot$

**Marking-based Step:** Marking Scheme

- Expression for  $g(r)$  0.1 pt
- Expression for  $M_m$  0.2 pt
- Expression for  $k_1$  0.1 pt
- Mass (good if nearest power of ten) 0.1 pt

**Problem Score** = max(answer-level score, step-level score)

Physics Field: Mechanics, Modality Type: Text+Data Figure, Answer Type: Expression, Numerical Value, Source: IPHO 2025

**5 Physics Fields**

- Mechanics
- Thermodynamics
- Modern Physics
- Electromagnetism
- Optics

**4 Modality Types**

- Text+Illustration Figure
- Text-Only
- Text+Variable Figure
- Text+Data Figure

**6 Answer Types**

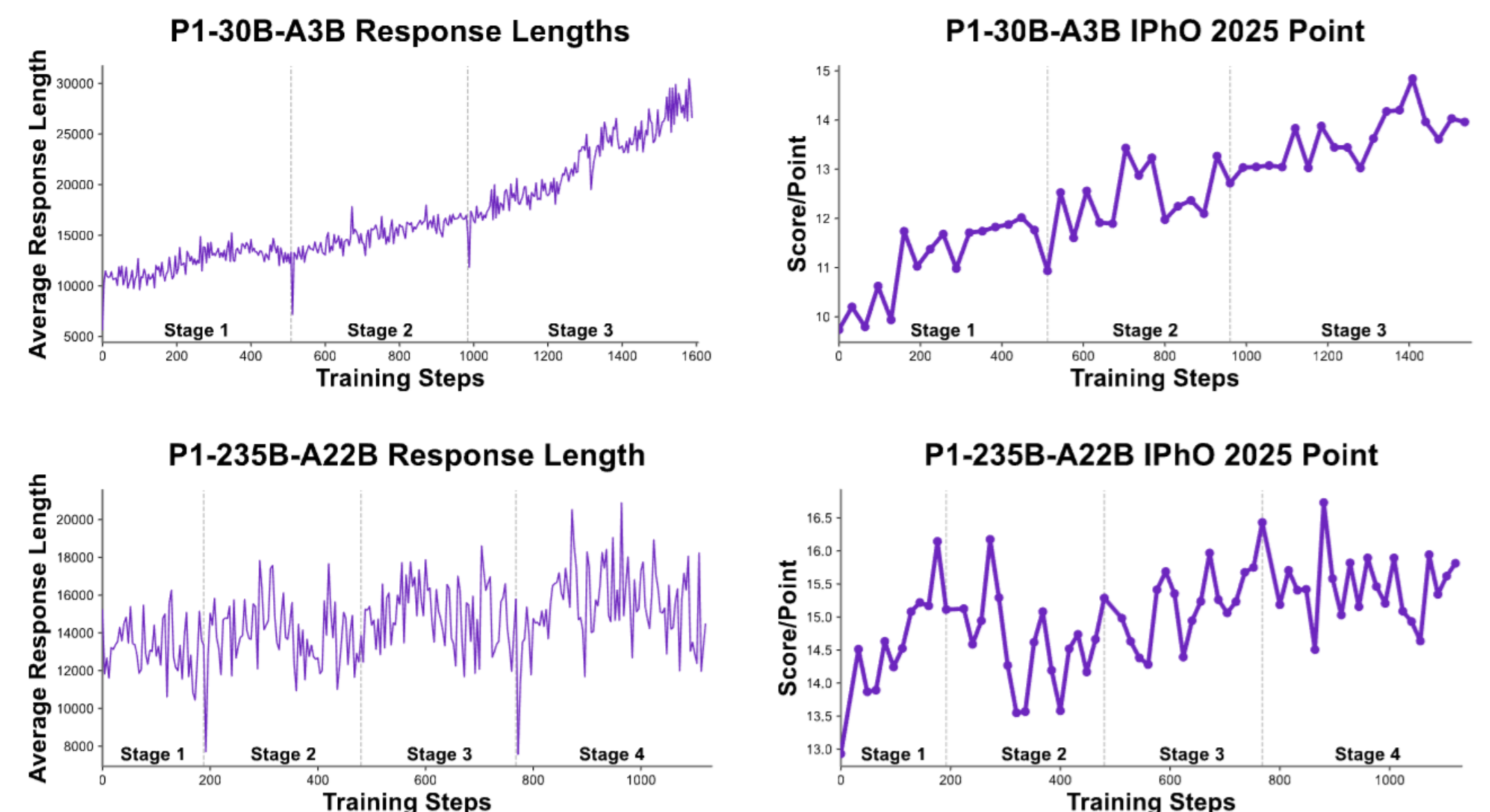
- Expression
- Numerical Value
- Multiple Choice
- Equation
- Open-End
- Inequality

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

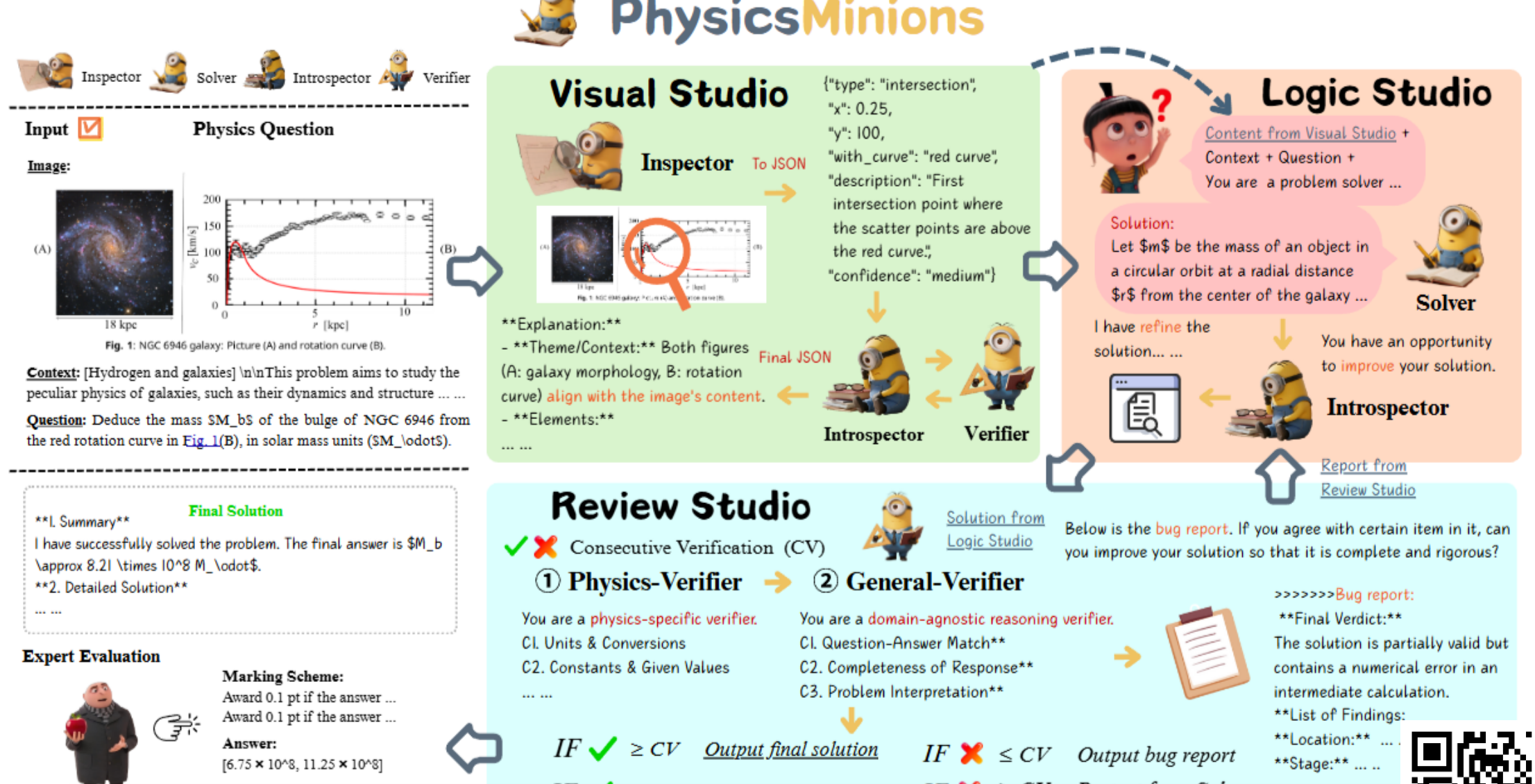
## Reasoning Model

### Model Training

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



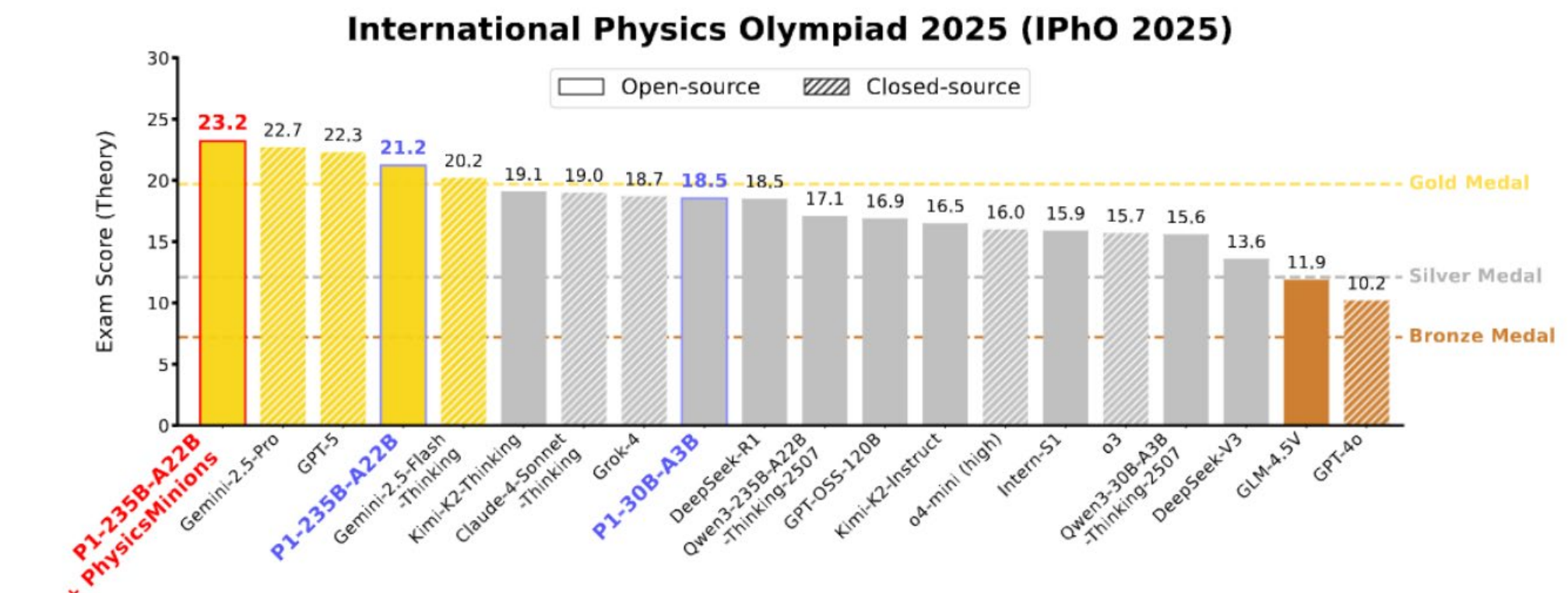
## Agent System



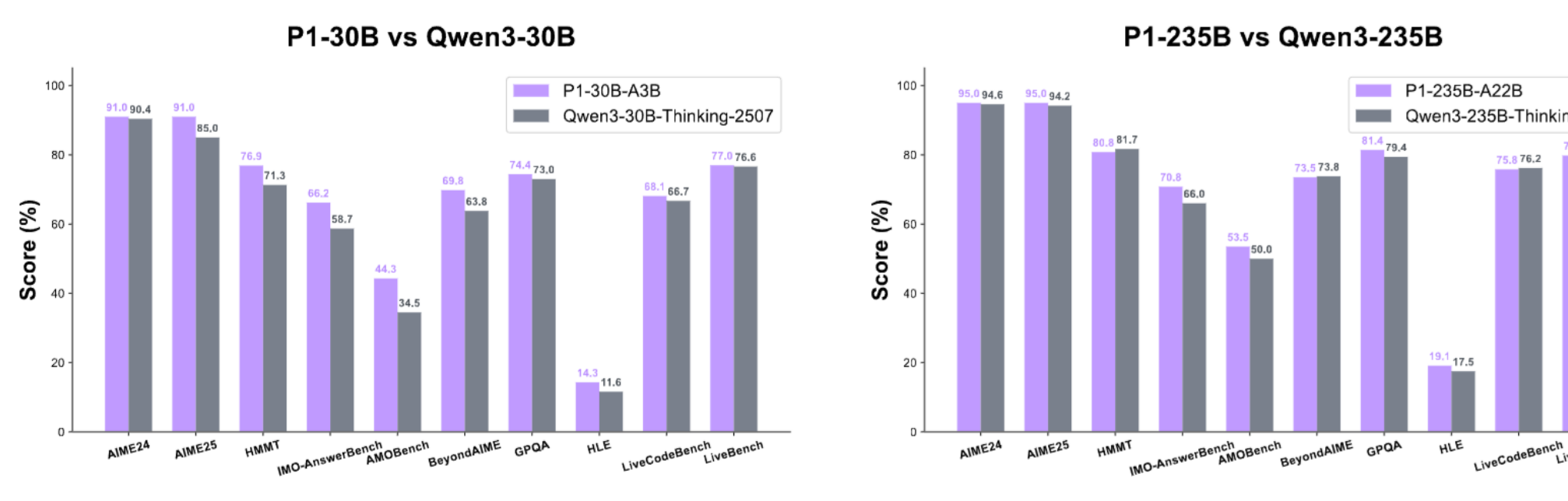
- **New paradigm for physics reasoning.**

## 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 Year	IPHO 2025	APHO 2024	EuPhO 2025	NBPPhO 2024	PanPhO 2025	PanMechanics 2024	F=MA 2025	Avg.
P1-30B-A3B	12.0	14.4	23.1	0.5	12.1	24.5	13.8	41.7
P1-30B-A3B + TTRL	13.8	15.4	23.6	0.3	14.1	24.4	12.8	42.9

