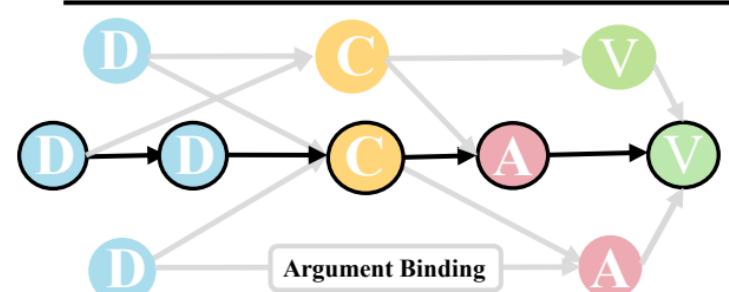


## Training Stage: Execution-First Data Synthesis

DB query, Computation, Analysis, Visualization

Program Graph Sampling



### Execution Engine

Golden Trace :  $\tau^* = \dots \text{V}_i \text{ X}_i \text{ R}_i \text{ S}_i \dots$

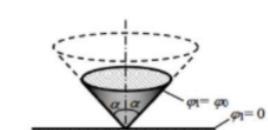
Error-Recovery:  $\tau_{\text{aug}}^* = \dots \text{V}_i \text{ X}_i \text{ e}_i \text{ S}_i \text{ V}_i \text{ X}_i \text{ O}_i \dots$

### Trace-to-Question Generation

$\dots \text{Natural Problems with Information Control} \dots$

### Problem Q & Golden Traces

An infinitely long conducting circular cone with an apex angle  $2\alpha = n\pi/6 \dots$



Improve  
5-10% SR

# SciAgentGYM

## Exploration & Discovery



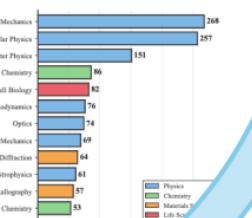
Multimodal Task:

Analytical Result:

## Installed Packages



## Diverse Subdomains

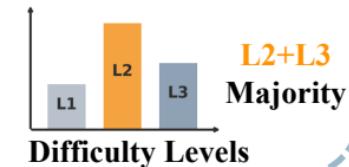


## SciAgent-Bench

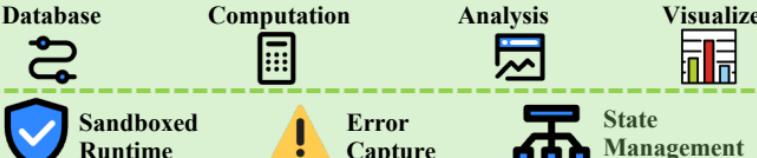
### Task Suite:



259 Tasks  
1134 Subquestions



### Execution Engine



### Evaluation

SR  
Correctness



SPL  
Efficiency



GPT-5	41.3
Claude-Sonnet-4	35.9
SciAgent-8B	30.1
Qwen3-VL-235B-Think	28.0

Success Rate