Aryabhata: An exam-focused language model for JEE Math

Ritvik Rastogi
PhysicsWallah
ritvik.rastogi@pw.live

Sachin Dharashivkar
AthenaAgent
sachin@athenaagent.com

Sandeep Varma
PhysicsWallah
sandeep.varma@pw.live

Abstract

We present **Aryabhata 1.0**, a 7B parameter math reasoning model optimized for the Indian Joint Entrance Examination (JEE). While recent LLMs have advanced mathematical reasoning, many remain unsuitable for high-stakes educational use. Our model is created by merging strong open-weight reasoning backbones, followed by supervised fine-tuning with curriculum learning on verified chain-of-thought (CoT) traces obtained through best-of-n rejection sampling. We further enhance performance via reinforcement learning with verifiable rewards (RLVR) using an A2C objective with group-relative advantage estimation, along with novel exploration strategies including *Adaptive Group Resizing* and *Temperature Scaling*. Evaluated on in-distribution (JEE Main 2025) and out-of-distribution (MATH, GSM8K) benchmarks, the model surpasses comparable baselines in accuracy and efficiency, while producing pedagogically useful step-by-step reasoning. This work demonstrates that compact, exam-focused language models can deliver both strong performance and practical usability for educational contexts.

1 Introduction and Related Work

Large language models (LLMs) have advanced mathematical reasoning, yet many remain ill-suited for high-stakes exams such as the Indian Joint Entrance Examination (JEE), which require both accurate solutions and pedagogically clear reasoning.

Non-reasoning models (e.g., GPT-4o) perform poorly on rigorous math tasks, often guessing or relying on shallow pattern matching.

Early reasoning models such as OpenAI o1 (OpenAI, 2024) and DeepSeek R1 (DeepSeek-AI u.a., 2025) improved accuracy via chain-of-thought (CoT) reasoning but suffered from hidden or verbose traces, slow generation, and nonlinear reasoning that hindered learning.

Modern reasoning models including o4-mini (OpenAI, 2025), Gemini 2.5 (Comanici u.a., 2025), and updated DeepSeek R1 (DeepSeek-AI u.a., 2025) offer higher accuracy and faster inference but still lack concise, linear reasoning traces optimal for educational use. (Samples are provided in Appendix G.)

In open-weight math-specialized systems, **DeepSeekMath** (Shao u.a., 2024) advanced capabilities via math-focused pretraining and Group Relative Policy Optimization (GRPO). **Qwen-2.5-Math-7B** (Yang u.a., 2024) supports CoT and tool-integrated reasoning in multiple languages. NVIDIA's **AceMath-7B-Instruct** (Liu u.a., 2025) and **AceReason-Nemotron-7B** (Liu u.a., 2025) enhance performance through multi-stage SFT and RL, while **AceReason-Nemotron-1.1-7B** (Liu u.a., 2025) combines stage-wise RL on math and code prompts. Pure RL-based reasoning models such as **DeepSeek-R1** and its distilled variants (DeepSeek-AI u.a., 2025) leverage verifiable-reward optimization to strengthen reasoning ability.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: MATH-AI.

Building on these developments, we introduce **Aryabhata 1.0**, a compact 7B parameter model that merges complementary reasoning backbones, applies curriculum-guided supervised fine-tuning with best-of-n rejection sampling, and employs reinforcement learning with verifiable rewards augmented by adaptive exploration strategies. The goal is to achieve high accuracy, efficiency, and pedagogical clarity in an exam-focused educational context.

2 Methodology

The overall process can be categorized in the following four stages:

2.1 Model Merging

To combine the advantages of System 1 (fluent, low-latency answers) and System 2 (deliberate, self-correcting reasoning) (Wu u.a., 2025), we perform model merging, following the works by Kimi k1.5 (Team u.a., 2025) and Wu u.a. (2025). We select three distinct LLMs sharing the same base architecture (Qwen 2.5 Math): (1) Qwen2.5-Math-7B-Instruct (Yang u.a., 2024) (2) AceMath-7B-Instruct (Liu u.a., 2025) (3) DeepSeek-R1-Distill-Qwen-7B (DeepSeek-AI u.a., 2025)

We used linear merging (Wortsman u.a., 2022) via the MergeKit framework (Goddard u.a., 2024) to combine the models. Let θ_1, θ_2, θ_3 be the parameters for Qwen, Ace, and DeepSeek, respectively. The merged parameters are calculated as:

$$\theta_{\text{merged}} = \alpha\theta_1 + \beta\theta_2 + \gamma\theta_3, \quad \text{where } \alpha + \beta + \gamma = 1$$

The weights ($\alpha = 0.15, \beta = 0.5, \gamma = 0.35$) were selected empirically based on held-out math reasoning tasks to balance quick problem-solving with methodical, multi-step analysis.

2.2 Data Curation

We used a proprietary corpus curated by our subject matter experts to align with JEE standards. This dataset, represents our core intellectual property hence is not publicly released.

Starting with approximately 250,000 raw questions, we applied several filtering steps to ensure quality. First, we removed all questions requiring multimodal reasoning, such as those dependent on diagrams. Next, we eliminated non-English or poorly formatted questions. To better frame the task as open-ended generation, we stripped answer options, following an approach also explored by Chandak u.a. (2025). Finally, we removed questions that were inherently dependent on the provided options in order to be answered.

To this end, we utilised OpenAI o4-mini with a structured prompt provided in Appendix B. This process yielded a clean dataset of about 130,000 questions suitable for CoT generation. The topic-wise distribution is detailed in Table 1 in Appendix A.

2.3 Supervised Fine-Tuning with Rejection Sampling

To create high-quality CoT supervision, we used best-of-4 rejection sampling with the merged model. For each curated question x, we sampled four CoT responses ($\{y_1, y_2, y_3, y_4\}$). We selected only those whose final answer matched the known correct ground truth answer, $\text{GT}(x)$, using Algorithm 1 in Appendix C.

The questions were then grouped based on how many of the four generations were correct (e.g., 4/4, 3/4). We used a curriculum-style supervised fine-tuning approach (Bengio u.a., 2009), starting with easier samples (4/4 correct) and gradually introducing harder ones (3/4, 2/4, 1/4 correct) to stabilize early learning and improve generalization.

This process resulted in approximately 350,000 verified CoTs from around 100,000 questions, which served as the core SFT training corpus, as shown in Table 2 in Appendix A. The 0/4 cases were reserved for future reinforcement learning with verifiable rewards (RLVR) to enhance coverage on challenging problems.

Parameter Efficient Finetuning (PEFT), specifically Low-Rank Adaptation (LoRA) (Hu u.a., 2021) was used during SFT with the peft (Mangrulkar u.a., 2022) library, with training parameters detailed in Appendix D. The final supervised finetuning experiment took 4 hours on 2xH100.

2.4 Reinforcement Learning with Verifiable Rewards

We extend Reinforcement Learning with Verifiable Rewards (RLVR) (Lambert u.a., 2025) by incorporating group-based advantage estimation (Shao u.a., 2024) into an Advantage Actor-Critic (A2C) framework (Mnih u.a., 2016).

2.4.1 Group-Relative Policy Optimization

Our method optimizes the A2C objective with group-relative advantage estimation:

$$J^{A2C}(\theta) = \mathbb{E}_{(\alpha_i)\sim\pi_\theta}\left[\frac{1}{G}\sum_{i=1}^{G}\frac{1}{|\alpha_i|}\log\pi_\theta(\alpha_i)\cdot\tilde{A}_i\right]$$

where G denotes the number of sampled responses α_i. Length-normalized gradients are weighted by sequence-level advantages $\tilde{A}_i$ via group-relative normalization. Rewards are binary: 1 if the final answer is correct, else 0.

2.4.2 Exploration Strategies

Adaptive Group Sizing: Unlike fixed GRPO implementations (von Werra u.a. (2020); Sheng u.a. (2024); Daniel Han u.a. (2023)), group size is dynamically adjusted by difficulty:

$$G_d = 8 \times 2^k, \quad k \in \{0, 1, 2, 3\}$$

with k determined by group average reward $\bar{R}_{\text{group}}$. Sizes scale $8 \to 16 \to 32 \to 64$, improving sampling diversity and stability while conserving resources.

Progressive Temperature Scaling: Sampling temperature increases from $0.6 \to 1.0$ during training (cf. An u.a. (2025)). Low initial temperature (0.6) stabilizes training, Gradual increase encourages diverse exploration.

Curriculum-Based Sampling: Training focuses on problems within an optimal difficulty band:

$$\mathcal{D}_t^{\text{filtered}} = \{x \in \mathcal{D}_t : \alpha_{\text{min}} \leq f_{\text{difficulty}}(x) \leq \alpha_{\text{max}}\}$$

where $f_{\text{difficulty}}(x)$ reflects success rates. Trivial problems are excluded for weak signals, and overly hard ones for excessive noise.

2.4.3 Hardware-Optimized Alternating Pipeline

To overcome GPU constraints, we adopt an alternating inference–training cycle using vLLM (Kwon u.a., 2023).

Phase 1: Rollout Generation involves running vLLM inference to produce batch rollouts, which are then serialized and stored in system memory. Once rollouts are generated, the vLLM process is terminated to release all GPU memory allocations.

Phase 2: Policy Optimization begins by loading the training model with full GPU memory availability. Policy gradients are then computed from the stored rollouts, followed by parameter updates, checkpointing, and finally offloading the model to prepare for the next rollout generation cycle.

Advantages: (1) Full memory per phase enables larger models and batches, (2) deterministic separation improves stability by avoiding race conditions and fragmentation.

The training configurations and hyperparameters are specified in Appendix E. The final reinforcement learning experiment took 350 hours on 2xH100.

3 Evaluation

We evaluated Aryabhata 1.0 across both in-distribution and out-of-distribution math benchmarks to assess its accuracy and efficiency in solving problems at scale, using the `pass@1` accuracy. The

solutions are generated using greedy decoding (temperature = 0). To determine whether a predicted answer matches the ground-truth answer for a question, we follow the pipeline described in the Algorithm 1.

Depending on whether the question is Multiple Choice Question or a Numerical Answer Type, we use different prompts to query the judge model (GPT-4o-mini). The prompts are provided in Table 3.

3.1 In-Distribution Evaluation: JEE Main 2025

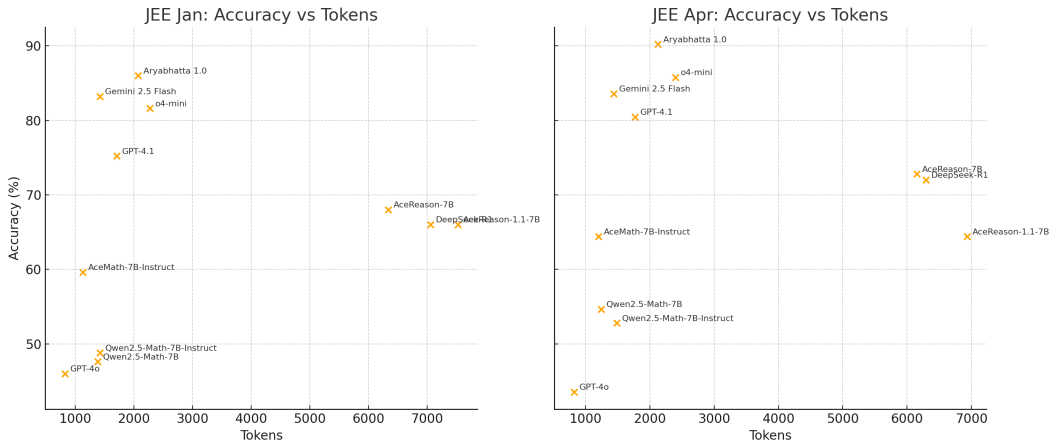

Figure 1: Scatter plots showing Accuracy vs. Tokens for JEE Jan and JEE Apr.

To measure performance in familiar distribution settings, we evaluate Aryabhata on the JEE Main 2025 exam. The January session contains 250 questions (10 papers with 25 questions each), while the April session comprises 225 questions (9 papers with 25 questions each), all sourced from official exam papers.

Figure 1 shows that Aryabhata 1.0 achieves an accuracy of **86.0%** on the January session and **90.2%** on the April session, while maintaining token efficiency with an average of approximately ~2K tokens per response. Compared to both open-weight and proprietary models, Aryabhata outperforms all baselines in accuracy while remaining competitive in inference cost.

3.2 Out-of-Distribution Evaluation

To evaluate generalization beyond the fine-tuning distribution, we benchmark Aryabhata 1.0 on (1) GSM8K (Cobbe u.a., 2021) (2) MATH 500 (Hendrycks u.a., 2021)

Table 6 in Appendix F shows that Aryabhata demonstrates **competitive generalization** to unseen tasks of comparable difficulty, outperforming its base models on both MATH and GSM8K.

Conclusion and Future Work

We presented **Aryabhata 1.0**, a compact 7B-parameter open-source model for mathematical reasoning tailored to the Indian competitive exam ecosystem. By merging diverse mathematical LLMs and fine-tuning on curated, verified domain-specific data, Aryabhata achieves state-of-the-art performance on in-distribution benchmarks such as JEE Main, while also demonstrating competitive generalization to out-of-distribution tasks including MATH and GSM8K.

Future work will expand reasoning coverage to Physics and Chemistry, extend to the full syllabus across Foundation, JEE (Main & Advanced), and NEET, and develop a family of exam-centric, open-source small language models (SLMs) that remain compact, efficient, and aligned with Indian education standards.

This trajectory aims to empower millions of students with accessible, curriculum-aligned AI tools that enhance classroom learning and support personalized exam preparation.

References

Sheng, Guangming u.a.(2024): *HybridFlow: A Flexible and Efficient RLHF Framework*.

An, Chenxin u.a. (2025): *POLARIS: A Post-Training Recipe for Scaling Reinforcement Learning on Advanced Reasoning Models*
.

Bengio, Yoshua / Louradour, Jérôme / Collobert, Ronan / Weston, Jason (2009): *Curriculum learning*
.

Chandak, Nikhil / Goel, Shashwat / Prabhu, Ameya / Hardt, Moritz / Geiping, Jonas (2025): *Answer Matching Outperforms Multiple Choice for Language Model Evaluation*
.

Cobbe, Karl u.a. (2021): *Training Verifiers to Solve Math Word Problems*
.

Comanici, Gheorghe u.a. (2025): *Gemini 2.5: Pushing the Frontier with Advanced Reasoning, Multimodality, Long Context, and Next Generation Agentic Capabilities*
.

Daniel Han, Michael Han / team, Unsloth (2023): *Unsloth*
.

DeepSeek AI u.a. (2025): *DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning*
.

Goddard, Charles / Siriwardhana, Shamane / Ehghaghi, Malikeh / Meyers, Luke / Karpukhin, Vladimir / Benedict, Brian / McQuade, Mark / Solawetz, Jacob (2024): *Arcee's MergeKit: A Toolkit for Merging Large Language Models*
.

Hendrycks, Dan / Burns, Collin / Kadavath, Saurav / Arora, Akul / Basart, Steven / Tang, Eric / Song, Dawn / Steinhardt, Jacob (2021): *Measuring Mathematical Problem Solving With the MATH Dataset*
.

Hu, Edward J. / Shen, Yelong / Wallis, Phillip / Allen Zhu, Zeyuan / Li, Yuanzhi / Wang, Shean / Wang, Lu / Chen, Weizhu (2021): *LoRA: Low-Rank Adaptation of Large Language Models*
.

Kingma, Diederik P. / Ba, Jimmy (2017): *Adam: A Method for Stochastic Optimization*
.

Kwon, Woosuk u.a. (2023): *Efficient Memory Management for Large Language Model Serving with PagedAttention*
.

Lambert, Nathan u.a. (2025): *Tulu 3: Pushing Frontiers in Open Language Model Post-Training*
.

Liu, Zihan / Chen, Yang / Shoeybi, Mohammad / Catanzaro, Bryan / Ping, Wei (2025): *AceMath: Advancing Frontier Math Reasoning with Post-Training and Reward Modeling*
.

Liu, Zihan / Yang, Zhuolin / Chen, Yang / Lee, Chankyu / Shoeybi, Mohammad / Catanzaro, Bryan / Ping, Wei (2025): *AceReason-Nemotron 1.1: Advancing Math and Code Reasoning through SFT and RL Synergy*
.

Mangrulkar, Sourab / Gugger, Sylvain / Debut, Lysandre / Belkada, Younes / Paul, Sayak / Bossan, Benjamin (2022): *PEFT: State-of-the-art Parameter-Efficient Fine-Tuning methods* `https://github.com/huggingface/peft`.

Mnih, Volodymyr / Badia, Adrià Puigdomènech / Mirza, Mehdi / Graves, Alex / Lillicrap, Timothy P. / Harley, Tim / Silver, David / Kavukcuoglu, Koray (2016): *Asynchronous Methods for Deep Reinforcement Learning*
.

OpenAI (2024): *OpenAI o1 Model*
, Accessed: 2025-08-05.

OpenAI (2025): *Introducing OpenAI o3 and o4-mini*
, Accessed: 2025-08-05.

Shao, Zhihong u.a. (2024): *DeepSeekMath: Pushing the Limits of Mathematical Reasoning in Open Language Models*
.

Team, Kimi u.a. (2025): *Kimi k1.5: Scaling Reinforcement Learning with LLMs*
.

Werra, Leandro von u.a. (2020): *TRL: Transformer Reinforcement Learning*
`https://github.com/huggingface/trl`.

Wortsman, Mitchell u.a. (2022): *Model soups: averaging weights of multiple fine-tuned models improves accuracy without increasing inference time*
.

Wu, Han u.a. (2025): *Unlocking Efficient Long-to-Short LLM Reasoning with Model Merging*
.

Yang, An u.a. (2024): *Qwen2.5-Math Technical Report: Toward Mathematical Expert Model via Self-Improvement*
.

A Training Data Distribution

Table 1: Topic-wise Question Distribution

Topic	**%age**
Application of Derivatives	4.50%
Application of Integrals	2.27%
Binomial Theorem	2.37%
Circles	2.85%
Complex Numbers & Quadratic Equations	6.00%
Conic Section	7.55%
Continuity and Differentiability	2.71%
Definite Integration	2.45%
Determinants	3.04%
Differential Equations	3.77%
Indefinite Integration	3.26%
Inverse Trigonometric Functions	5.31%
Limits and Derivatives	3.88%
Matrices	2.46%
Permutations and Combinations	4.23%
Probability	5.69%
Quadratic Equations	4.45%
Relations and Functions	2.24%
Sequence and Series	2.75%
Sets	1.04%
Statistics	1.89%
Straight Lines	2.31%
Three Dimensional Geometry	3.92%
Trigonometric Functions	4.51%
Vector Algebra	2.89%
Miscellaneous	11.65%

Table 2: Chain-of-Thought generation outcomes from best-of-4 sampling.

Correct CoTs	**# Questions**	**Total CoTs**	**Usage**
0	31,470	0	Used in RLVR only
1	9,647	9,647	SFT
2	9,066	18,132	SFT
3	12,643	37,929	SFT
4	67,247	268,988	10% sampled for SFT

B Prompt for Question Cleaning

Listing 1: Prompt used for Question Cleaning

```
Clean and standardize math questions by removing multiple-choice options,
    normalizing
the answer format, identifying dependencies, and determining the language.
    For any
answers expressed in MathML, convert them to LaTeX. Conversion of MathML in
     the
**question** is *not required* (but preserve LaTeX if already present).
Additionally, provide a clear **step-by-step reasoning** explaining how
    each part of
the output was derived.

### Instructions:

1. Identify and extract the core question text:
    * Remove all multiple-choice options (e.g., A-D or 1-4), ensuring the
        main question
    remains grammatically and semantically intact.
    * Preserve existing LaTeX in the question.
    * Do **not** convert MathML in the question. It may be retained as-is.

2. Normalize the answer:
    * If the answer is given as an option label (e.g., "Answer: B"), replace
         it with the
    corresponding value from the provided options.
    * If the answer is already a value, retain it.
    * If the answer is in MathML, convert it to LaTeX.

3. Determine dependency flags:
    * **Option-dependent:** Is the question understandable and solvable
        without access
    to the answer options? Mark ‘True‘ if the question lacks key information
         without
    them; otherwise, ‘False‘.
    * **Diagram-dependent:** Does the question reference or rely on a
        diagram, figure, or
    visual element? Mark ‘True‘ or ‘False‘.

4. Identify the language:
    * Detect and report the language of the question text (e.g., ‘English‘,
        ‘Hindi‘,
    ‘Tamil‘, etc.).

5. Provide reasoning:
    * For each output field (question, answer, flags, language), include a
        clear
    explanation of how the output was determined.
    * The reasoning should follow a logical step-by-step format, but does **
        not** need to
    be wrapped in any special ‘<reason>‘ block.

# Output Format

<question> cleaned question </question>
<answer> cleaned answer </answer>
<option_dependent> True/False </option_dependent>
<diagram_dependent> True/False </diagram_dependent>
```

```
<language> detected language </language>
* All math in the **answer** must be in LaTeX.
* There should be **no references** to original option labels (e.g., "A",
    "1", or
"Option B").
* Ensure the cleaned question is coherent, self-contained, and
    grammatically correct.
* The reasoning can be in free-text form and must explain how each part of
    the output was
derived.

### Example 1
Input:
What is the derivative of \(x^2 + 3x + 5\)?
A) \(2x + 3\)
B) \(x + 3\)
C) \(x^2 + 3\)
D) \(2x + 5\)
Answer: A

Output:
<question> What is the derivative of \(x^2 + 3x + 5\)? </question>
<answer> \(2x + 3\) </answer>
<option_dependent> False </option_dependent>
<diagram_dependent> False </diagram_dependent>
<language> English </language>
\end{verbatim}

\begin{verbatim}
### Example 2
Input:
<p>Simplify the following expression:</p>
<math xmlns="http://www.w3.org/1998/Math/MathML">
  <mfrac>
    <msqrt>
      <msup><mi>a</mi><mn>2</mn></msup>
    </msqrt>
    <mi>a</mi>
  </mfrac>
</math>

<p>Options:</p>
1) <math xmlns="http://www.w3.org/1998/Math/MathML"><msqrt><mi>a</mi></
    msqrt></math>
2) <math xmlns="http://www.w3.org/1998/Math/MathML"><mi>a</mi></math>
3) <math xmlns="http://www.w3.org/1998/Math/MathML"><mfrac><mn>1</mn><mi>a
    </mi>
</mfrac></math>
4) <math xmlns="http://www.w3.org/1998/Math/MathML"><mn>1</mn></math>

Answer: 1
Output:
<question> Simplify the following expression:
<math xmlns="http://www.w3.org/1998/Math/MathML">
  <mfrac>
    <msqrt>
      <msup><mi>a</mi><mn>2</mn></msup>
    </msqrt>
    <mi>a</mi>
```

```
  </mfrac>
</math>
</question>
<answer> \sqrt{a} </answer>
<option_dependent> False </option_dependent>
<diagram_dependent> False </diagram_dependent>
<language> English </language>
```

C Answer Matching Algorithm

Algorithm 1 Answer Matching Procedure

Input: Predicted answer a_p, Ground-truth answer a_g, Options (if any)
Output: Match status (True / False)
if $a_p = a_g$ or sympy_latex_match(a_p, a_g) **then**
    return True
end if
if option/identifier from a_p == option/identifier from a_g **then**
    return True
end if
Query LLM judge with a_p, a_g, and options (if any)
if LLM returns YES **then**
    return True
else
    return False
end if

Table 3: Prompts used for Answer Matching

MCQ	**Numerical**
System Prompt: `You are checking an MCQ. Given the list of options, determine if answer 1 and answer 2 are the same. Answer 1 is the same as answer 2 only if all the options match. Reason step-by-step and put the final answer YES or NO in \boxed{}.`	**System Prompt:** `You are checking an exam. For a given question, determine if answer 1 and answer 2 are the same. Since the answers are for the same question, you can assume similar context for both answers and make appropriate assumptions when checking if they are the same. Reason step-by-step and put the final answerYES or NO in \boxed{}.`

User Prompt:
 	
`Options:`
`A: <Option 1>`
`B: <Option 2>`
`C: <Option 3>`
`D: <Option 4>`
`answer 1: <Correct Answer>`
`answer 2: <Predicted Answer>`
	User Prompt:
	
	`answer 1: <Correct Answer>`
	`answer 2: <Predicted Answer>`

D Hyper-parameters for Supervised Fine Tuning

The hyper-parameters for LoRA are provided in the Table 4 and the hyper-parameters for SFT are provided in the Table 5.

Table 4: PEFT configuration using LoRA.

Parameter	**Value**
Rank	128
LoRA Alpha	128
LoRA Dropout	0.1
Bias	none
Target Modules	`{q_proj, k_proj, v_proj, o_proj, gate_proj, up_proj, down_proj, embeddings}`

Table 5: Training configuration used for supervised fine-tuning.

Parameter	**Value**
Precision	bfloat16
Max Sequence Length	16,384
Batch Size (per device)	1
Gradient Accumulation Steps	16
Effective Batch Size	16
Number of Epochs	3
Initial Learning Rate	2×10^{-5}
Final Learning Rate	2×10^{-7}
Learning Rate Scheduler	Linear
Optimizer	AdamW (8-bit)
Warmup Steps	5
Packing	False
Logging Steps	1
WandB Reporting	Enabled

E Training Configuration and Hyperparameters for RLVR

Optimization: Adam optimizer (Kingma u.a., 2017) with learning rate 1×10^{-6} ensures stable updates.

Memory & Precision: Training uses bfloat16 mixed precision with gradient checkpointing for reduced memory usage while preserving numerical stability.

Sequence & Batch Setup: Context length is capped at 4,096 tokens, balancing capacity for multi-step reasoning and computational tractability.

F Out of Distribution Results

Table 6: Performance comparison on MATH 500 and GSM8K benchmarks

Model	MATH 500	GSM8K
Aryabhatta 1.0	83.6	94.8
Qwen/Qwen2.5-Math-7B-Instruct	66.0	94.7
nvidia/AceMath-7B-Instruct	80.6	93.4
GPT-4o	69.2	94.6
deepseek-ai/DeepSeek-R1-Distill-Qwen-7B	85.2	69.7
nvidia/AceReason-Nemotron-7B	84.2	76.5
nvidia/AceReason-Nemotron-1.1-7B	85.4	93.1
GPT-4.1	86.6	94.0
o4-mini	94.8	90.1
Gemini 2.5 Flash	93.6	85.1

G Example Model Responses

The sample question along with its correct answer is presented in Figure 2.
The response generated by GPT-4o is shown in Figure 3.
The response produced by DeepSeek R1 Distill Qwen 7B is illustrated across Figures 4, 5, and 6.
The response from Aryabhata 1.0 is depicted in Figure 7.

$Question:$
Let $y = f(x)$ be the solution of the differential equation $\frac{dy}{dx} + \frac{xy}{x^2-1} = \frac{x^4+4x}{\sqrt{1-x^2}}, -1 < x < 1$ such that $f(0) = 0$. If $6\int_{-1/2}^{1/2} f(x)dx = 2\pi - \alpha$ then α^2 is equal to:
Answer:
27

Figure 2: Sample question with the correct answer

Given the differential equation and initial condition:

$$\frac{dy}{dx} + \frac{xy}{x^2-1} = \frac{x^4+4x}{\sqrt{1-x^2}}, \quad -1 < x < 1 \quad \text{with} \quad f(0) = 0,$$

we solve the differential equation using the method of integrating factors. The integrating factor, $\mu(x)$, is given by:

$$\mu(x) = e^{\int \frac{x}{x^2-1}\,dx}.$$

To find this, we simplify the integral:

$$\int \frac{x}{x^2-1}\,dx.$$

We use partial fractions:

$$\frac{x}{x^2-1} = \frac{x}{(x-1)(x+1)} = \frac{1/2}{x-1} + \frac{1/2}{x+1}.$$

Thus,

$$\int \frac{x}{x^2-1}\,dx = \int \frac{1/2}{x-1}\,dx + \int \frac{1/2}{x+1}\,dx = \frac{1}{2}\ln|x-1| + \frac{1}{2}\ln|x+1| = \frac{1}{2}\ln|(x-1)(x+1)|.$$

So, the integrating factor is:

$$\mu(x) = e^{\frac{1}{2}\ln|(x-1)(x+1)|} = \sqrt{|(x-1)(x+1)|}.$$

Since $-1 < x < 1$, $\sqrt{|(x-1)(x+1)|}$ simplifies to $\sqrt{1-x^2}$.

Multiplying the differential equation by the integrating factor:

$$\sqrt{1-x^2}\frac{dy}{dx} + \frac{x\sqrt{1-x^2}y}{x^2-1} = x^4 + 4x.$$

This rearranges to:

$$\frac{d}{dx}(y\sqrt{1-x^2}) = x^4 + 4x.$$

Integrating both sides w.r.t. x:

$$y\sqrt{1-x^2} = \int (x^4+4x)\,dx.$$

Let:

$$\int x^4\,dx = \frac{x^5}{5}, \quad \int 4x\,dx = 2x^2.$$

Thus:

$$y\sqrt{1-x^2} = \frac{x^5}{5} + 2x^2 + C.$$

Given $f(0) = 0$:

$$0 = \frac{0^5}{5} + 2\cdot 0^2 + C \implies C = 0.$$

So:

$$y\sqrt{1-x^2} = \frac{x^5}{5} + 2x^2.$$

And:

$$y = \frac{x^5/5 + 2x^2}{\sqrt{1-x^2}}.$$

We want $\int_{-1/2}^{1/2} f(x)\,dx = \int_{-1/2}^{1/2} \frac{x^5/5+2x^2}{\sqrt{1-x^2}}\,dx$.

Solving using symmetry and standard integral forms: Thus:

$$\int_{-1/2}^{1/2} f(x)dx = \frac{2\pi}{6} - \frac{\alpha}{6}.$$

Hence, equating both expressions:

$$6\int_{-1/2}^{1/2} f(x)dx = 2\pi - \alpha.$$

Thus:

$$\alpha = \frac{\pi}{3},$$

and: $\alpha^2 = \left(\frac{\pi}{3}\right)^2 = \frac{\pi^2}{9}$. So: $\boxed{\frac{\pi^2}{9}}$.

Figure 3: Response from GPT-4o (Part 1 of 1)

<think> Okay, so I have this differential equation to solve: $\frac{dy}{dx} + \frac{xy}{x^2-1} = \frac{x^4+4x}{\sqrt{1-x^2}}$, with the condition that $f(0) = 0$. Then, after finding $f(x)$, I need to compute $6\int_{-1/2}^{1/2} f(x)\,dx$, which equals $2\pi - \alpha$. Finally, I have to find α^2. Hmm, that seems a bit involved, but let's break it down step by step.

First, the differential equation is linear. It's in the form $\frac{dy}{dx} + P(x)y = Q(x)$. So, I can use an integrating factor to solve it. The standard method for linear differential equations is to find an integrating factor $\mu(x) = \exp\left(\int P(x)\,dx\right)$. Then, multiply both sides of the equation by $\mu(x)$ and integrate.

Let me write down the equation again:

$$\frac{dy}{dx} + \left(\frac{x}{x^2-1}\right)y = \frac{x^4+4x}{\sqrt{1-x^2}}$$

So, $P(x) = \frac{x}{x^2-1}$, and $Q(x) = \frac{x^4+4x}{\sqrt{1-x^2}}$.

First, let's compute the integrating factor $\mu(x)$:

$$\mu(x) = \exp\left(\int \frac{x}{x^2-1}dx\right)$$

Hmm, the integral of $\frac{x}{x^2-1}dx$. Let me make a substitution. Let $u = x^2 - 1$, then $du/dx = 2x$, so $\frac{1}{2}du = xdx$. Therefore, the integral becomes $\frac{1}{2}\int \frac{du}{u} = \frac{1}{2}\ln|u| + C = \frac{1}{2}\ln|x^2-1| + C$.

So,

$$\mu(x) = \exp\left(\frac{1}{2}\ln|x^2-1|\right) = |x^2-1|^{1/2}$$

Since we're working in the interval $-1 < x < 1$, $x^2 - 1$ is negative, so $|x^2-1| = 1-x^2$. Therefore, $\mu(x) = \sqrt{1-x^2}$.

Wait, that's interesting. So, the integrating factor is $\sqrt{1-x^2}$. Let me write that down:

$$\mu(x) = \sqrt{1-x^2}$$

Now, multiply both sides of the differential equation by $\mu(x)$:

$$\sqrt{1-x^2}\frac{dy}{dx} + \sqrt{1-x^2}\cdot\left(\frac{x}{x^2-1}\right)y = \frac{(x^4+4x)}{\sqrt{1-x^2}}\cdot\sqrt{1-x^2}$$

Simplify each term:

First term: $\sqrt{1-x^2}\frac{dy}{dx}$

Second term: $\sqrt{1-x^2}\cdot\left(\frac{x}{x^2-1}\right)y = \frac{x}{\sqrt{1-x^2}(x^2-1)}y$

Wait, $x^2 - 1$ is negative, so $x^2 - 1 = -(1-x^2)$. Therefore,

$$\frac{x}{\sqrt{1-x^2}(x^2-1)} = \frac{-x}{(1-x^2)^{3/2}}$$

Wait, maybe I should think differently. Let me compute the left-hand side after multiplying by $\mu(x)$. The left-hand side should be the derivative of $\mu(x)y$. Let me check:

$$\frac{d}{dx}[\mu(x)y] = \mu'(x)y + \mu(x)\frac{dy}{dx}$$

So, let's compute $\mu'(x)$:

$$\mu(x) = (1-x^2)^{1/2}, \quad \mu'(x) = \frac{1}{2}(1-x^2)^{-1/2}(-2x) = \frac{-x}{\sqrt{1-x^2}}$$

Therefore,

$$\frac{d}{dx}[\mu(x)y] = \frac{-x}{\sqrt{1-x^2}}y + \sqrt{1-x^2}\frac{dy}{dx}$$

Which is exactly the left-hand side of the equation after multiplying by $\mu(x)$. So, that's correct.

Therefore, the equation becomes:

$$\frac{d}{dx}[\sqrt{1-x^2}y] = x^4 + 4x$$

Now, to solve for y, we can integrate both sides with respect to x:

$$\sqrt{1-x^2}y = \int (x^4+4x)\,dx + C$$

Compute the integral on the right:

$$\int (x^4+4x)dx = \int x^4 dx + \int 4x dx = \frac{x^5}{5} + 2x^2 + C$$

So,

$$\sqrt{1-x^2}y = \frac{x^5}{5} + 2x^2 + C$$

Figure 4: Response from DeepSeek R1 Distill Qwen 7B (Part 1 of 3)

Therefore, solving for y:

$$y = \frac{\frac{x^5}{5} + 2x^2 + C}{\sqrt{1 - x^2}}$$

So, that's the general solution. Now, apply the initial condition $f(0) = 0$. Let's plug $x = 0$ into the solution:

$$y(0) = \frac{0 + 0 + C}{1} = C$$

But $y(0) = 0$, so $C = 0$.
Therefore, the particular solution is:

$$y = \frac{\frac{x^5}{5} + 2x^2}{\sqrt{1 - x^2}} = \frac{x^5 + 10x^2}{5\sqrt{1 - x^2}}$$

So, $f(x) = \frac{x^5 + 10x^2}{5\sqrt{1 - x^2}}$
Now, the next step is to compute

$$6 \int_{-1/2}^{1/2} f(x)\, dx = 2\pi - \alpha$$

Given:

$$f(x) = \frac{x^5 + 10x^2}{5\sqrt{1 - x^2}} = \frac{x^5}{5\sqrt{1 - x^2}} + \frac{2x^2}{\sqrt{1 - x^2}}$$

So the integral becomes:

$$\int_{-1/2}^{1/2} \left(\frac{x^5}{5\sqrt{1 - x^2}} + \frac{2x^2}{\sqrt{1 - x^2}} \right) dx$$

Split into two integrals:

$$\frac{1}{5} \int_{-1/2}^{1/2} \frac{x^5}{\sqrt{1 - x^2}} dx + 2 \int_{-1/2}^{1/2} \frac{x^2}{\sqrt{1 - x^2}} dx$$

Note: The first integrand is an odd function (since x^5 is odd), and the limits are symmetric, so the integral is zero.
So we compute:

$$2 \int_{-1/2}^{1/2} \frac{x^2}{\sqrt{1 - x^2}} dx = 4 \int_{0}^{1/2} \frac{x^2}{\sqrt{1 - x^2}} dx$$

Use the substitution $x = \sin\theta$, $dx = \cos\theta d\theta$, $\sqrt{1 - x^2} = \cos\theta$. When $x = 0$, $\theta = 0$; when $x = \frac{1}{2}$, $\theta = \frac{\pi}{6}$
So the integral becomes:

$$\int_0^{\pi/6} \sin^2\theta\, d\theta = \frac{1}{2} \int_0^{\pi/6} (1 - \cos 2\theta)\, d\theta = \frac{1}{2} \left[\theta - \frac{\sin 2\theta}{2} \right]_0^{\pi/6}$$

$$= \frac{1}{2} \left(\frac{\pi}{6} - \frac{\sin(\pi/3)}{2} \right) = \frac{1}{2} \left(\frac{\pi}{6} - \frac{\sqrt{3}}{4} \right) = \frac{\pi}{12} - \frac{\sqrt{3}}{8}$$

Then,

$$\int_{-1/2}^{1/2} f(x)\, dx = \frac{\pi}{3} - \frac{\sqrt{3}}{2}$$

Now,

$$6 \int_{-1/2}^{1/2} f(x)\, dx = 6 \left(\frac{\pi}{3} - \frac{\sqrt{3}}{2} \right) = 2\pi - 3\sqrt{3}$$

Given $6 \int_{-1/2}^{1/2} f(x)\, dx = 2\pi - \alpha$, we conclude:

$$\alpha = 3\sqrt{3} \Rightarrow \alpha^2 = 27$$

$$\boxed{27}$$

</think>
Given the differential equation:

$$\frac{dy}{dx} + \frac{xy}{x^2 - 1} = \frac{x^4 + 4x}{\sqrt{1 - x^2}}, \quad -1 < x < 1$$

with the initial condition $f(0) = 0$, we need to find α^2 where $6 \int_{-1/2}^{1/2} f(x)\, dx = 2\pi - \alpha$.

Figure 5: Response from DeepSeek R1 Distill Qwen 7B (Part 2 of 3)

First, we solve the differential equation using the integrating factor method. The integrating factor is:

$$\mu(x) = \exp\left(\int \frac{x}{x^2-1}\,dx\right) = \sqrt{1-x^2}$$

Multiplying through by the integrating factor, we get:

$$\sqrt{1-x^2}\frac{dy}{dx} + \frac{x}{\sqrt{1-x^2}}y = x^4 + 4x$$

This simplifies to:

$$\frac{d}{dx}\left(\sqrt{1-x^2}y\right) = x^4 + 4x$$

Integrating both sides:

$$\sqrt{1-x^2}y = \int (x^4 + 4x)\,dx = \frac{x^5}{5} + 2x^2 + C$$

Applying the initial condition $f(0) = 0$, we find $C = 0$. Thus, the solution is:

$$y = \frac{\frac{x^5}{5} + 2x^2}{\sqrt{1-x^2}} = \frac{x^5 + 10x^2}{5\sqrt{1-x^2}}$$

Next, we compute the integral:

$$6\int_{-1/2}^{1/2} f(x)\,dx$$

We split the integral into two parts:

$$6\left(\int_{-1/2}^{1/2} \frac{x^5}{5\sqrt{1-x^2}}\,dx + \int_{-1/2}^{1/2} \frac{2x^2}{\sqrt{1-x^2}}\,dx\right)$$

The first integral is zero because it is an odd function. The second integral is computed using the substitution $x = \sin\theta$:

$$4\int_0^{\pi/6} \sin^2\theta\,d\theta = 4\left(\frac{\pi}{12} - \frac{\sqrt{3}}{8}\right) = \frac{\pi}{3} - \frac{\sqrt{3}}{2}$$

Thus,

$$6\left(\frac{\pi}{3} - \frac{\sqrt{3}}{2}\right) = 2\pi - 3\sqrt{3}$$

Given $6\int_{-1/2}^{1/2} f(x)\,dx = 2\pi - \alpha$, we find $\alpha = 3\sqrt{3}$. Therefore, $\alpha^2 = (3\sqrt{3})^2 = 27$.

$$\boxed{27}$$

Figure 6: Response from DeepSeek R1 Distill Qwen 7B (Part 3 of 3)

To solve the given differential equation $\frac{dy}{dx}+\frac{xy}{x^2-1}=\frac{x^4+4x}{\sqrt{1-x^2}}$ with the initial condition $f(0)=0$, we start by identifying it as a first-order linear differential equation. The standard form of a first-order linear differential equation is $\frac{dy}{dx}+P(x)y=Q(x)$, where $P(x)=\frac{x}{x^2-1}$ and $Q(x)=\frac{x^4+4x}{\sqrt{1-x^2}}$.

The integrating factor $\mu(x)$ is given by $\mu(x)=e^{\int P(x)\,dx}=e^{\int \frac{x}{x^2-1}\,dx}$. To find this integral, we use the substitution $u=x^2-1$, so $du=2x\,dx$. Thus, $\int \frac{x}{x^2-1}\,dx=\frac{1}{2}\int \frac{1}{u}\,du=\frac{1}{2}\ln|u|=\frac{1}{2}\ln|x^2-1|$. Therefore, the integrating factor is $\mu(x)=e^{\frac{1}{2}\ln|x^2-1|}=\sqrt{|x^2-1|}=\sqrt{1-x^2}$ since x is in the interval $(-1,1)$.

Multiplying both sides of the differential equation by the integrating factor $\sqrt{1-x^2}$, we get:

$$\sqrt{1-x^2}\frac{dy}{dx}+\frac{x}{\sqrt{1-x^2}}y=x^4+4x.$$

The left-hand side is the derivative of $y\sqrt{1-x^2}$:

$$\frac{d}{dx}\left(y\sqrt{1-x^2}\right)=x^4+4x.$$

Integrating both sides with respect to x, we have:

$$y\sqrt{1-x^2}=\int (x^4+4x)\,dx.$$

The integral on the right-hand side is:

$$\int x^4\,dx+\int 4x\,dx=\frac{x^5}{5}+2x^2+C.$$

Thus,

$$y\sqrt{1-x^2}=\frac{x^5}{5}+2x^2+C.$$

Solving for y, we get:

$$y=\frac{\frac{x^5}{5}+2x^2+C}{\sqrt{1-x^2}}.$$

Using the initial condition $f(0)=0$, we substitute $x=0$ and $y=0$:

$$0=\frac{0+0+C}{\sqrt{1-0^2}}=C.$$

So, $C=0$, and the solution is:

$$y=\frac{x^5+10x^2}{5\sqrt{1-x^2}}.$$

We need to find $6\int_{-1/2}^{1/2} f(x)\,dx$. Since $f(x)=\frac{x^5+10x^2}{5\sqrt{1-x^2}}$, we have:

$$6\int_{-1/2}^{1/2}\frac{x^5+10x^2}{5\sqrt{1-x^2}}\,dx=\frac{6}{5}\int_{-1/2}^{1/2}\frac{x^5+10x^2}{\sqrt{1-x^2}}\,dx.$$

We can split this integral into two parts:

$$\frac{6}{5}\left(\int_{-1/2}^{1/2}\frac{x^5}{\sqrt{1-x^2}}\,dx+10\int_{-1/2}^{1/2}\frac{x^2}{\sqrt{1-x^2}}\,dx\right).$$

The first integral is zero because $\frac{x^5}{\sqrt{1-x^2}}$ is an odd function integrated over a symmetric interval. So, we only need to evaluate the second integral:

$$\frac{6}{5}\cdot 10\int_{-1/2}^{1/2}\frac{x^2}{\sqrt{1-x^2}}\,dx=12\int_{-1/2}^{1/2}\frac{x^2}{\sqrt{1-x^2}}\,dx.$$

Since $\frac{x^2}{\sqrt{1-x^2}}$ is an even function, we can rewrite the integral as:

$$12\cdot 2\int_{0}^{1/2}\frac{x^2}{\sqrt{1-x^2}}\,dx=24\int_{0}^{1/2}\frac{x^2}{\sqrt{1-x^2}}\,dx.$$

To evaluate this integral, we use the substitution $x=\sin\theta$, so $dx=\cos\theta\,d\theta$ and when $x=0$, $\theta=0$, and when $x=1/2$, $\theta=\frac{\pi}{6}$. The integral becomes:

$$24\int_0^{\pi/6}\frac{\sin^2\theta}{\sqrt{1-\sin^2\theta}}\cos\theta\,d\theta=24\int_0^{\pi/6}\frac{\sin^2\theta}{\cos\theta}\cos\theta\,d\theta=24\int_0^{\pi/6}\sin^2\theta\,d\theta.$$

Using the identity $\sin^2\theta=\frac{1-\cos 2\theta}{2}$, we get:

$$24\int_0^{\pi/6}\frac{1-\cos 2\theta}{2}\,d\theta=12\int_0^{\pi/6}(1-\cos 2\theta)\,d\theta=12\left[\theta-\frac{\sin 2\theta}{2}\right]_0^{\pi/6}.$$

Evaluating this, we have:

$$12\left(\frac{\pi}{6}-\frac{\sin\frac{\pi}{3}}{2}-0+0\right)=12\left(\frac{\pi}{6}-\frac{\sqrt{3}}{4}\right)=2\pi-3\sqrt{3}.$$

We are given that $6\int_{-1/2}^{1/2} f(x)\,dx=2\pi-\alpha$, so $2\pi-3\sqrt{3}=2\pi-\alpha$. Therefore, $\alpha=3\sqrt{3}$, and $\alpha^2=(3\sqrt{3})^2=27$.

The final answer is:

$$\boxed{27}.$$

Figure 7: Response from Aryabhata 1.0 (Part 1 of 1)