# Holistic Capability Preservation: Towards Compact Yet Comprehensive Reasoning Models

**Ling Team, AI@Ant Group**

## Abstract

This technical report presents Ring-Lite-Distill, a lightweight reasoning model derived from our open-source Mixture-of-Experts (MoE) Large Language Models (LLMs) Ling-Lite. This study demonstrates that through meticulous high-quality data curation and ingenious training paradigms, the compact MoE model Ling-Lite can be further trained to achieve exceptional reasoning capabilities, while maintaining its parameter-efficient architecture with only 2.75 billion activated parameters, establishing an efficient lightweight reasoning architecture. In particular, in constructing this model, we have not merely focused on enhancing advanced reasoning capabilities, exemplified by high-difficulty mathematical problem solving, but rather aimed to develop a reasoning model with more comprehensive competency coverage. Our approach ensures coverage across reasoning tasks of varying difficulty levels while preserving generic capabilities, such as instruction following, tool use, and knowledge retention. We show that, Ring-Lite-Distill's reasoning ability reaches a level comparable to DeepSeek-R1-Distill-Qwen-7B, while its general capabilities significantly surpass those of DeepSeek-R1-Distill-Qwen-7B, as shown in Figure 1. The models are accessible at https://huggingface.co/inclusionAI/Ring-lite-distill-preview.

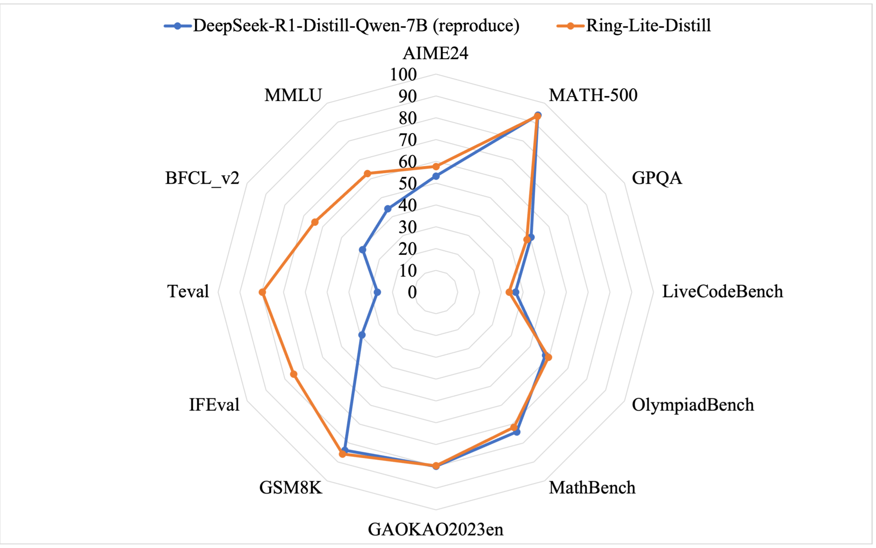

Figure 1: Performance comparison between Ring-Lite-Distill and Deepseek-R1-Distill-Qwen-7B.

## 1 Introduction

Recent advances in long chain-of-thought reasoning models have attracted growing research interest, particularly since the release of the DeepSeek-R1 series of models DeepSeek-AI (2025). Facing the exorbitant computational costs inherent in training and deploying ultra-large language models, the research community has increasingly focused on developing compact yet capable alternatives. While most current efforts build upon the Qwen open-source model family Qwen (2025) with a predominant focus on enhancing advanced reasoning capabilities, such as solving complex mathematical problems, significantly less attention has been paid to developing models with comprehensive capability coverage using alternative base models, which will significantly impact the overall usability of the model. Our work, originating from our self-developed lightweight model Ling-Lite Team (2025) and aiming to construct a reasoning model with more comprehensive capability coverage, can be seen as a contribution to filling this gap.

This work is driven by two principal motivations. Primarily, we aim to investigate the viability of developing a powerful reasoning model from our self-developed lightweight MoE LLM foundation. The resultant model achieves remarkable reasoning performance while maintaining extreme parameter efficiency (2.75B activated parameters). **Furthermore, we observe that the improvement of pure reasoning ability alone can greatly affect other general capabilities such as instruction following, which in turn affects the real application of the model. This issue clearly needs to be considered and addressed in subsequent training, such as in reinforcement learning.** In this report, we seek to systematically preserve general competencies (e.g. instruction following, tool use, and commonsense reasoning) during specialized capability enhancement, a critical yet challenging balance to strike, particularly within a relatively lightweight model.

To accomplish these objectives, we developed a data taxonomy system, and implemented a rigorous reasoning data curation pipeline that integrates open-source data collection, reasoning data synthesis and distillation, as well as data purification. To prevent significant degradation in other capabilities, we carefully curated a general-purpose dataset, combining it with a sampled subset of reasoning data to form a balanced supervised fine-tuning (SFT) dataset that maintains proficiency across all key competencies. **It should be noted that dedicated data distillation for the corresponding data is not essential to recover general capabilities.** We adopted a systematic training approach to maximize the value of the dataset, improving reasoning ability while preserving other skills. Specifically, we executed a two-stage SFT process: the first stage focused on improving reasoning capabilities, while the second stage emphasized balancing general abilities. Subsequently, Direct Preference Optimization (DPO) Rafailov et al. (2023) training with specifically enhanced data further contributed to improvements such as format refinement, helping to further refine the model's performance.

The contributions of this paper can be summarized as follows:

- We introduce Ring-Lite-Distill, a lightweight yet powerful reasoning model built upon our open source MoE LLM Ling-Lite, achieving state-of-the-art performance with only 2.75 billion activated parameters.
- We detail our comprehensive data curation framework. Based on our data taxonomy system, we analyzed the collected data and further systematically carried out multi-disciplinary data synthesis, data distillation, and data purification, thereby constructing a robust reasoning dataset.
- We present our principled training recipe for preserving general competencies during specialized reasoning enhancement, addressing a critical challenge in large reasoning model development.

## 2 Data Curation

In this section, we begin with a brief introduction to our proposed data taxonomy system, which helps systematically analyze collected data and guide subsequent data synthesis and utilization. Building upon this foundation, we meticulously crafted a data production pipeline (Figure 2) that incorporates open-source data collection, reasoning data synthesis, and data purification. During the synthesis phase, we focused on constructing comprehensive and effective reasoning prompts, combined with knowledge distillation and multi-stage validation filtering, to enhance Ring-Lite-Distill's reasoning capabilities. This end-to-end pipeline yielded approximately 2.59 million high-quality SFT samples, specifically optimized to advance Ring-Lite-Distill's reasoning performance. To further ensure model versatility, we curated a general-purpose dataset (890K samples) for cross-domain competency preservation. Additionally, we developed a 23K-sample DPO dataset to strengthen instruction adherence and answer formatting precision while minimizing redundant outputs.

### 2.1 Data Taxonomy System

To support the data construction pipeline, we developed a Data Taxonomy System that enables systematic analysis of dataset distribution and structured categorization. This framework serves two primary purposes: (1) identifying data strengths and weaknesses to guide subsequent collection, generation, and model training strategies; and (2) facilitating iterative model optimization through retrospective data analysis, which reveals correlations between data composition and capability development. By establishing this taxonomy, we aim to decode the critical link between data attributes and emergent model competencies in large-scale AI training.

We introduce a hierarchical taxonomy to ensure systematic and interpretable analysis. The first-level classification adheres to the Chinese Library Classification (CLC) framework, covering major domains such as Math, Code, Physics, Chemistry, and Biology, among others. Since we focuses on the reasoning tasks, we expand the classification on Math, code and STEM (Physics, Chemistry, and Biology) domains to improve granularity. In detail, 1) Physics, Chemistry, and Biology are sub-classified according to the BISAC Subject Headings Book Industry Study Group (BISG) (2024);

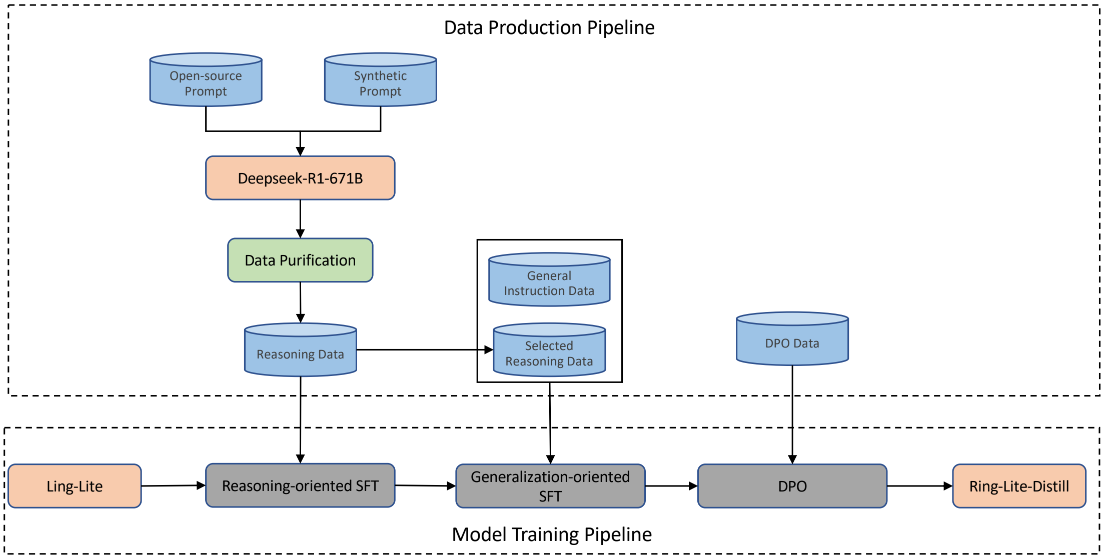

Figure 2: Data production and model training pipeline.

2) Code is categorized based on task types like code generation and code translation; 3) Math is classified using a fine-grained taxonomy derived from standard Chinese mathematics textbooks and National Mathematics Subject Classification (China). In Table 1, we present an overview of the hierarchical taxonomy.

Table 1: The overview of the hierarchical taxonomy.

| **Domain** | **Subtopics** |
|---|---|
| Math | • Primary School<br>– Algebra, Others<br>• Junior High School<br>– Algebra, Geometry, Probability and Statistics, Others<br>• High School<br>– Algebra, Functions, Geometry, Analytic Geometry, Sets and Functions, . . .<br>• University<br>– Number Theory, Calculus, Operations Research, . . . |
| Code | • Code Generation, Code Translation, Code Testing, Code Review, . . . |
| Physics | • Electricity, Geophysics, Magnetism, Nuclear Physics, . . . |
| Chemistry | • Electrochemistry, Environmental Chemistry, Organic Chemistry, . . . |
| Biology | • Genetics, Zoology, Botany, Entomology, . . . |
| . . . | . . . |

## 2.2 Reasoning Data

### 2.2.1 Open Source Data Collection

Our process commenced with the collection and refinement of open-source datasets spanning mathematics, science, and code domains, including OpenR1-Math-220k OpenR1 (2025b), numina-cot AxolotlAI (2025), SYNTHETIC-1 PrimeIntellect (2025), SCP-116K EricLu (2025), Bespoke-Stratos-17k Bespokelabs (2025), GeneralThought-430K GeneralReasoning (2025), OpenR1-codeforces-cots OpenR1 (2025a), etc. If the dataset includes outputs distilled from

DeepSeek-R1 DeepSeek-AI (2025), we implement selective retention based on quality criteria. For those without DeepSeek-R1 distilled outputs, query quality is first evaluated, with only higher-quality queries being retained for subsequent distillation. The generation parameters were configured following DeepSeek-R1's specifications, with temperature set to 0.6 and top-p to 0.95. The samples were then decontaminated through conducting N-gram overlap analysis and semantic similarity checks against established benchmarks including AIME24, GPQA Diamond, MATH-500, LiveCodeBench, and other evaluation benchmarks to ensure data purity. Thus far, we have screened and obtained approximately 1.04 million samples.

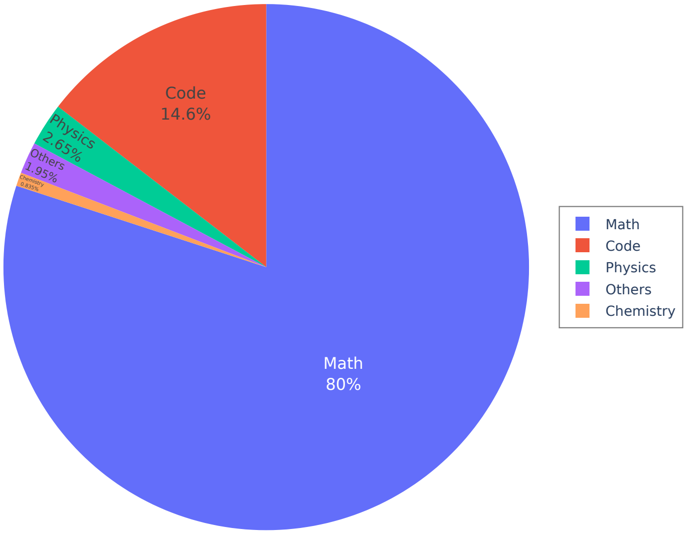

Figure 3: Overall domain distribution of the processed open source data.

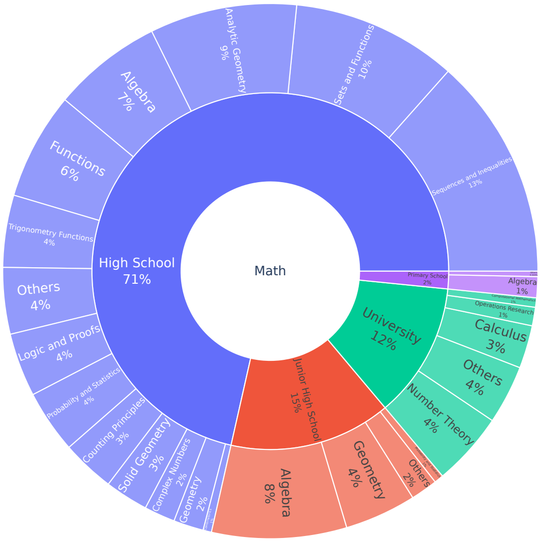

Figure 4: Detailed distribution of the processed open source math data.

We further employed our data taxonomy system to analyze the processed open-source data. Figure 3 and Figure 4 respectively illustrate the overall domain distribution of the processed open-source data and the detailed distribution statistics for the mathematical data among them. It can be observed that the current dataset lacks data from the biological sciences, and the volume of chemistry data is a mere 8K plus $(0.835\%)$. Moreover, data for some subtopics within mathematics domain may also be insufficient. We attempted to further refine and diversify our training data through the following data synthesis procedure.

### 2.2.2 Reasoning Data Synthesis

To further enhance the diversity of reasoning data and supplement open-source resources, we have developed a systematic, multi-disciplinary pipeline for reasoning data synthesis, with the help of the data taxonomy system and advanced LLMs. This process first generates problem prompts, and then distills answers from DeepSeek-R1 DeepSeek-AI (2025) based on the generated prompts. The prompt synthetic pipeline iteratively leverages open-source and the already-generated seed problems, guided by topic categorization from our data taxonomy system, to progressively expand and perfect the prompt set. For mathematics and code problems—where data is abundant—we employ LLMs to generate novel problems across diverse topics using existing seeds, while for undersupplied scientific domains, we leverage the hierarchical structure of the topics and their interdependencies to facilitate data generation. To ensure accurate Long-CoT answers for these synthesized prompts, DeepSeek-R1 is leveraged for answer distillation, after which basic validation checks are performed. Details of the prompt generation are elaborated subsequently.

**Math Problem** We started by using the training problems of MATH, GSM8K, as well as AoPS competition problems AoPS contributors (2025) as the seed prompts. For each taxonomy topic, we sample its associated seed prompts and employ LLMs to generate novel, difficulty-stratified prompts that maintain reasonableness and completeness. The synthesized prompts undergo rigorous validation where LLMs provides 10 independent solutions per problem, with only those prompts that exhibits high response consistency (over 5 consistent answers) are retained. This filtered set of new prompts then serve as refreshed seed prompts for iterative synthesis cycles, ultimately yielding a large-scale collection of high-quality mathematical prompts with systematically varied difficulty levels.

**Code Problem** We adopt a similar iterative framework as described above to generate code prompts. Initially, we collect seed problem prompts from open-source resources and label these problems with different taxonomy topics to form an initial pool of candidate prompts. For each topic, based on the existing seed prompts, we utilize LLMs to generate new problems, solutions, and test cases. Problems that fail to pass the test cases are initially discarded. Furthermore, problems with higher difficulty are retained to more effectively refine reasoning capabilities. We conduct N-gram and semantic similarity checks on these newly generated prompts, expanding the pool by incorporating prompts that exhibit significant differences.

**Science Problem** Unlike mathematical and coding data, scientific data is relatively scarce. To address this challenge, we fully utilize the hierarchical structure of topics to facilitate prompt generation, even when there are insufficient seed prompts to guide the LLMs. Additionally, to ensure alignment with real-world reasoning benchmarks and address disciplinary complexity, we implement an LLM-judge to retain topics conducive to computational or analytical problem-solving (e.g., mathematics, computer science, applied mechanics) while eliminating those skewed toward factual knowledge recall. Subsequently, we incorporate these retained topics and predefined difficulty prompts (e.g., Olympic-level, Postgraduate Level, etc.) into synthetic data prompts, and leverage LLMs to generate new problems and answers. Through this approach, we achieve the generation of problems across different topics and difficulty levels, thereby significantly enhancing our scientific dataset.

### 2.2.3 Data Purification

With the reasoning data constructed, we first perform data decontamination against established benchmarks including AIME24, GPQA Diamond, MATH-500, LiveCodeBench, and other evaluation benchmarks to ensure data purity. In addition, we implemented a comprehensive set of rigorous filtering and quality control measures to further ensure the quality of the data. These measures mainly included the detection of: (1) Repetitive content patterns; (2) Incomplete or improper truncation; (3) Anomalous mathematical formulas; (4) Image/Figure contained. Specifically,

- **Repetitive content patterns**: Detect abnormal repetitions in the response, such as:

  ```
  Oh right, the date.weekday() function returns 0 for Monday through 6 for Sunday.
  Wait, no, wait. Wait, no, Wait, no, wait. Wait, no ...
  ```

- **Incomplete or improper truncation**: Detect the incomplete answer with abrupt terminations:

  ```
  $$\tan 2\alpha = \frac{-6}{-8} = \frac{3}{4}$$. Thus,
  ```

- **Anomalous mathematical formulas**: Detect the invalid LaTeX formula syntax:

  ```
  \( Z_m \subsetneq G \)
  ```

- **Image/Figure contained**: Check if the question contains image content like:

  ```
  In the figure, triangle  $ADE$  is produced from triangle ...
  ![Image](https:xxx.png)
  ```

After applying these filtering measures, 39K low-quality entries were removed, leaving 2.59 million high-quality entries.

Combining our data taxonomy system, we present the data distribution of our final version of reasoning data. Figure 5 illustrates the overall domain distribution of the dataset. The results indicate that reasoning-related content is mainly concentrated in the fields of mathematics, coding, and STEM (Physics, Biology, and Chemistry), with mathematics constituting the majority (84.2%), followed by coding (8.33%) and STEM (6.1%). The biology and chemistry questions have been significantly supplemented compared to before. A finer-grained analysis of the mathematics domain, presented in Figure 6, categorizes the data by educational level: Primary School (3%), Junior High School (16%), High School (68%), and University (14%). Compared to the previous versions of the open-source data, there have been some adjustments in the distribution. This distribution reflects the varying complexity of mathematical questions. Additionally, each level is further subdivided by topic, such as Geometry and Algebra, providing a detailed breakdown of subject coverage.

## 2.3 General Data

To improve the LLM capabilities of tool use, instruction following, and knowledge, we reuse portions of the general SFT dataset of Ling Team (2025). For tool use, high-quality datasets are synthesized from various sources, including open-source resources, APIs, task templates, and subgraph patterns, supported by adaptive learning mechanisms like policy agents and reference models for enhanced accuracy and reliability. Instruction following relies on structured strategies, direct and indirect synthesis, using seed instructions processed by teacher, judgment, and reference models to

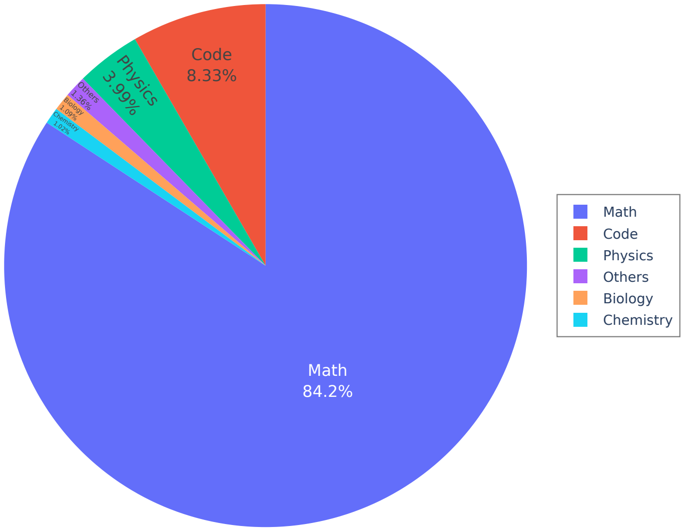

Figure 5: Overall domain distribution of the final dataset.

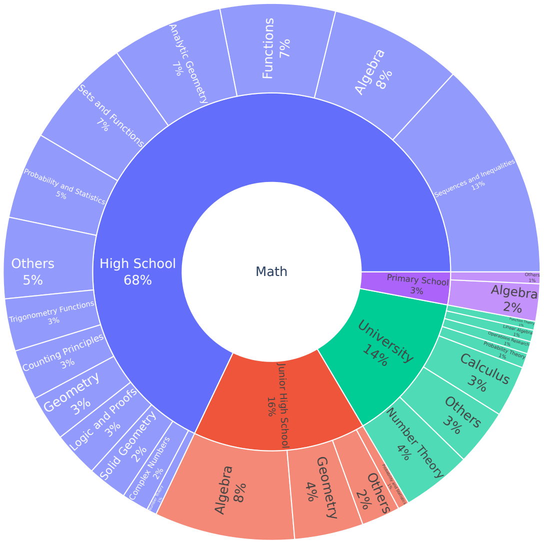

Figure 6: Detailed distribution of the final math data.

produce diverse dialog types (e.g., Instruction SFT, DPO, PPO). Additionally, knowledge data is collected from social media and assessed for clarity, specificity, complexity, and dependency to ensure quality. **Note that dedicated data distillation for the corresponding data is not essential to recover general capabilities.** The general data comprised approximately 437K samples, including around 197K samples of general instruction following and QA data, as well as 240K samples of tool-usage data.

In subsequent Generalization-oriented SFT 3.2 phases, in order to balance the model's reasoning ability and general capability, we further sampled around 645K of the aforementioned reasoning samples, including approximately 195K code-related reasoning examples, 400K mathematical reasoning examples, and around 50K science-related reasoning examples, and combined them with the general data, resulting in a hybrid dataset totaling 1082K samples.

### 2.4 DPO Data

Building on experimental analysis of model behavior and SFT-stage data characteristics, we constructed a 20K-sample DPO dataset to enhance response quality and complex query handling through comparative feedback. Specifically, chosen responses prioritized prompts with lower n-gram overlap to mitigate repetitive outputs, as higher overlap empirically correlated with increased redundancy, while rejected responses were sampled via rule-based selection (e.g., favoring longer outputs) from the SFT model to sharpen preference optimization. To ensure robust model improvement, the dataset blended prompts from both SFT training data and a proprietary subjective dataset, guaranteeing diversity and representativeness.

Additionally, to address observed structural issues for the obtained models, specifically the failure to generate properly paired <think>-</think> tags when required and the production of redundant outputs, we augmented the DPO dataset with an additional 3K human-curated chosen/rejected response pairs. The complete 23K samples were utilized for DPO training, resulting in significant model refinement, particularly in output formatting, yielding further performance improvements.

## 3 Training Recipe & Results

Ring-Lite-Distill is derived from our open source model Ling-Lite through further training. Leveraging our meticulously processed dataset, the development primarily involved a two-stage SFT process, followed by subsequent DPO training, as illustrated in Figure 2. The primary goal of the first stage of SFT is to enhance the reasoning ability of Ring-Lite-Distill, while the second stage of STF focuses on improving its general capabilities, including tool use, instruction following, and general knowledge, etc. After completing SFT, we perform DPO to optimize the format of long chain-of-thought answers and alleviate issues of repetition in responses. Following the two-stage SFT and DPO, Ring-Lite-Distill's reasoning ability reaches a level comparable to DeepSeek-R1-Distill-Qwen-7B, while its general capabilities significantly surpass those of DeepSeek-R1-Distill-Qwen-7B, as shown in Table 2 and Figure 1.

### 3.1 Reasoning-oriented SFT

To systematically enhance the model's reasoning capabilities, we first performed a SFT process using a high-quality dataset comprising approximately 2.59 million samples, as introduced in Subsection 2.2. This dataset was meticulously curated to encompass a broad spectrum of reasoning tasks, including mathematical problem-solving, rigorous science question-answering, and coding challenges as discussed in Section 2. The model was trained on Ling-Lite Team (2025) with a learning rate of $4 \times 10^{-5}$, a batch size of 256, and a maximum sequence length of 16K tokens.

In Table 2, we summarize the results of different models. Among them, *R1-Distill-Qwen-7B (reported)* refers to the evaluation results of DeepSeek-R1-Distill-Qwen-7B reported in the original paper, while *R1-Distill-Qwen-7B (reproduce)* refers to the evaluation results of DeepSeek-R1-Distill-Qwen-7B obtained through our evaluation system. Evaluation of the resulted model of this stage (*Ring-Lite-Distill-Stage-1*) demonstrated substantial performance improvements across multiple benchmarks. Specifically, the model achieved competitive results in AIME-24 (+1.0)MAA (2024) (mathematical olympiad problems), MATH-500 (-0.2) (challenging competition mathematics problems) Hendrycks et al. (2021), GPQA (-2.9) Rein et al. (2024)(biology, physics, and chemistry problems), and LiveCodeBench (-3.6) Jain et al. (2024) (code related problems). In addition,our model demonstrates more competitive performance on most other reasoning-related evaluation benchmarks. Compared to DeepSeek-R1-Distill-Qwen-7B, it achieves superior results on OlympiadBench (+1.5), GAOKAO2023en (+0.3), and GSM8K (+5.7). Unlike datasets like AIME that focus on more complex mathematical problems, these benchmarks assess mathematical capabilities across varying difficulty levels. Considering all reasoning evaluation metrics collectively, our model exhibits more comprehensive improvements in reasoning ability. This can be attributed to our thorough and systematic approach in dataset construction.

However, when shift our focus to more general capabilities beyond reasoning, we observe that both our further-trained model and DeepSeek-R1-Distill-Qwen-7B exhibit significant trade-offs in other abilities while improving reasoning performance. As demonstrated in evaluation benchmarks such as IFEval Zhou et al. (2023), Teval Chen et al. (2023), BFCL_v2 Yan et al. (2024), and MMLU (shown in the Table 2), **models focused on enhancing reasoning capabilities perform poorly on these metrics. In practical applications, the absence of these capabilities substantially impacts the model's overall utility. This issue clearly needs to be addressed in the subsequent training procedure.**

### 3.2 Generalization-oriented SFT

The primary objective of the second-stage SFT is to comprehensively restore the model's general capabilities, including critical foundational competencies such as instruction following, tool use, and linguistic comprehension, while preserving the reasoning abilities acquired during the previous stage. This dual focus forms the essential foundation for enabling reasoning models to function effectively across diverse applications. During this phase, we constructed a hybrid dataset containing approximately 1082K samples, which includes around 437K general task samples and 645K reasoning task samples sampled from the dataset used in the first-stage SFT, as described in Subsection 2.3. **Note that, for the construction of general data, dedicated data distillation is not essential.** Based on the model obtained in the previous stage, we performed a second-stage SFT for two epochs using this dataset with a learning rate of $1 \times 10^{-5}$, a batch size of 128, and a maximum sequence length of 16K tokens.

Evaluation results for the obtained model in this stage (*Ring-Lite-Distill-Stage-2*) indicate that the model demonstrates superior performance across multiple benchmark tasks, as shown in Table 2. Specifically, *Ring-Lite-Distill-Stage-2* exhibits significant improvements compared to the DeepSeek-R1-Distill-Qwen-7B in benchmarks related to instruction following (e.g., IFEval), tool usage (e.g., Teval, BFCL_v2), and language understanding (e.g., MMLU). For these general metrics aggregating broader aspects of performance, DeepSeek-R1-Distill-Qwen-7B obtains an average score of 37.3, whereas *Ring-Lite-Distill-Stage-2* significantly outperforms it with an average score of 70.5. Furthermore, in benchmarks focused on reasoning abilities, such as AIME24, MATH-500 and GPQA, the model's performance remains comparable to that of DeepSeek-R1-Distill-Qwen-7B, with the same average score of 66.2. The comparison between *Ring-Lite-Distill-Stage-1* and *Ring-Lite-Distill-Stage-2* reveals substantial gains in general capabilities, with the average metric score increasing from 42.9 to 70.5. Meanwhile, reasoning performance remains stable, with a slightly improvement from 66.0 to 66.2.

These results demonstrate that through meticulous data design coupled with appropriate training strategies, it is possible to significantly enhance reasoning capabilities while maintaining other core competencies without substantial compromise—even when using a relatively lightweight model architecture.

### 3.3 DPO

After the two-stage SFT training, we observed a significant improvement in the model's reasoning capabilities, while its general abilities showed no notable degradation. However, upon further usage and analysis, we identified several

Table 2: Performance Comparison Between Deepseek-R1-Distill-Qwen-7B and Ring-Lite-Distill at Various Stages.

| | **Benchmark** | **R1-Distill-Qwen 7B (reported)** | **R1-Distill-Qwen 7B (reproduce)** | **Ling-Lite** | **Ring-Lite-Distill-Stage-1** | **Ring-Lite-Distill-Stage-2** | **Ring-Lite-Distill-DPO** |
|---|---|---|---|---|---|---|---|
| Reasoning | AIME24 | 55.5 | 53.2 | 6.7 | 54.2 | **57.5** | 53.44 |
| | Math-500 | 92.8 | **93.7** | 71.8 | 93.5 | 93.2 | 92.90 |
| | GPQA | 49.1 | **50.4** | 30.3 | 47.5 | 48.2 | 46.59 |
| | LiveCodeBench | 37.6 | **36.5** | 15.2 | 32.9 | 33.6 | 32.01 |
| | OlympiadBench | - | 58.1 | 34.4 | 59.6 | **59.7** | 56.44 |
| | MathBench | - | **74.1** | 69.2 | 70.8 | 71.6 | 69.69 |
| | GAOKAO2023en | - | 80.0 | 61.0 | **80.3** | 79.7 | 80.26 |
| | GSM8K | - | 83.8 | 86.9 | **89.5** | 85.8 | 84.53 |
| | **Avg.** | - | **66.2** | 46.9 | 66.0 | **66.2** | 64.68 |
| General | IFEval | - | 39.3 | **79.4** | 34.9 | 75.4 | 77.22 |
| | Teval | - | 26.9 | **85.6** | 28.0 | 79.7 | 80.08 |
| | BFCL_v2 | - | 38.9 | **67.9** | 44.2 | 64.2 | 63.22 |
| | MMLU | - | 44.1 | **71.2** | 64.4 | 62.8 | 62.91 |
| | **Avg.** | - | 37.3 | **76.0** | 42.9 | 70.5 | 70.86 |

areas for further improvement. These include issues related to structural integrity, such as failing to generate properly paired <think>-</think> tags when required or producing redundant outputs, particularly when processing subjective queries. To better demonstrate the model's performance in handling structural issues, we conduct a focused analysis on its ability to generate properly formatted <think>-</think> tag pairs. Specifically, a total of 418 subjective queries were evaluated across the obtained models. We quantified the frequency of incorrect usage (either missing or redundant) of paired <think>-</think> tags in the model responses. As shown in Table 3, '**#Missing Error**' denotes the number of occurrences where the <think>-</think> tags were missing, '**#Redundancy Error**' counts instances of unnecessary tag duplication, and '**#All Error**' represents the combined total of all incorrect tag usages. Our analysis revealed that errors predominantly involved missing <think>-</think> tags. This likely stems from the model's strong specialization in reasoning tasks during training, which may have limited its ability to generalize structured formatting to subjective or open-ended scenarios.

Table 3: Frequency Analysis of Paired <think>-</think> in Subjective Queries.

| | **#Missing Error** | **#Redundancy Error** | **#All Error** |
|---|---|---|---|
| **Ring-Lite-Distill-Stage-2** | 90 | 4 | 94 |
| **Ring-Lite-Distill-DPO** | 44 | 5 | 49 |

To mitigate this issue, we augmented the dataset with an additional 3K human-curated chosen/rejected response pairs, extending the original 20K DPO foundation dataset, as introduced in Subsection 2.4. These supplementary response pairs were constructed in accordance with Team (2025), ensuring consistent reasoning chains while reflecting varying degrees of structural fidelity. This approach effectively enhances the model's ability to follow formatting instruction during DPO training, mitigating the issue where excessively long CoT responses weaken the representation of formatting-related tokens in the DPO loss. As shown in Table 3, following the application of DPO, the model shows a notable improvement, reducing the total number of errors to 49, with 44 instances involving missing error. These results demonstrate that DPO training yields additional refinements for the model, particularly in output formatting, leading to further performance enhancements of the model.

## 4 Conclusion

In this report, we present Ring-Lite-Distill, a compact reasoning model with only 2.75B parameters (16.8B in total). Through carefully constructed SFT and DPO datasets combined with distillation from more powerful models, we demonstrate that this architecture - based on our open source model Ling-Lite - can achieve notable reasoning capabilities. To address the common performance degradation in general abilities (e.g., tool use and instruction following) observed in reasoning-focused models, we implement a second-stage SFT process that successfully recovers these competencies without compromising the model's core reasoning performance, results show that at comparable sizes, the proposed model achieves competitive performance with existing reasoning-focused models while demonstrating enhanced general capabilities, including instruction following and tool utilization, etc. These enhancements substantially increase the practical applicability of reasoning models in real-world scenarios. Our future work will focus on further improving the model's generalization and reasoning abilities through reinforcement learning training. We also plan to explore scaling these approaches to larger model architectures.

## References

AoPS contributors. 2025. AoPS Online. `https://artofproblemsolving.com/` [Online; accessed 7-April-2025].

AxolotlAI. 2025. *numina-cot-logprobs-859k-8b-sft dataset*. `https://huggingface.co/datasets/axolotl-ai-co/numina-cot-logprobs-859k-8b-sft/`.

Bespokelabs. 2025. *Bespoke-Stratos-17k dataset*. `https://huggingface.co/datasets/bespokelabs/Bespoke-Stratos-17k/`.

Book Industry Study Group (BISG). 2024. BISAC Subject Headings List. `https://www.bisg.org/bisac-subject-headings-list` Accessed: 2024.

Zehui Chen, Weihua Du, Wenwei Zhang, Kuikun Liu, Jiangning Liu, Miao Zheng, Jingming Zhuo, Songyang Zhang, Dahua Lin, Kai Chen, et al. 2023. T-Eval: Evaluating the Tool Utilization Capability Step by Step. *arXiv preprint arXiv:2312.14033* (2023).

DeepSeek-AI. 2025. DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning. *CoRR* abs/2501.12948 (2025). `https://doi.org/10.48550/ARXIV.2501.12948` arXiv:2501.12948

EricLu. 2025. *SCP-116K dataset*. `https://huggingface.co/datasets/EricLu/SCP-116K/`.

GeneralReasoning. 2025. *GeneralThought-430K dataset*. `https://huggingface.co/datasets/GeneralReasoning/GeneralThought-430K/`.

Dan Hendrycks, Collin Burns, Saurav Kadavath, Akul Arora, Steven Basart, Eric Tang, Dawn Song, and Jacob Steinhardt. 2021. Measuring Mathematical Problem Solving With the MATH Dataset. arXiv:2103.03874 [cs.LG] `https://arxiv.org/abs/2103.03874`

Naman Jain, King Han, Alex Gu, Wen-Ding Li, Fanjia Yan, Tianjun Zhang, Sida Wang, Armando Solar-Lezama, Koushik Sen, and Ion Stoica. 2024. LiveCodeBench: Holistic and Contamination Free Evaluation of Large Language Models for Code. arXiv:2403.07974 [cs.SE] `https://arxiv.org/abs/2403.07974`

MAA. 2024. American invitational mathematics examination - aime. *In American Invitational Mathematics Examination - AIME 2024, February 2024.* (2024). `URLhttps://maa.org/math-competitions/american-invitational-mathematics-examination-aime`.

OpenR1. 2025a. *OpenR1-codeforces-cots dataset*. `https://huggingface.co/datasets/open-r1/codeforces-cots/`.

OpenR1. 2025b. *OpenR1-Math-220k dataset*. `https://huggingface.co/datasets/open-r1/OpenR1-Math-220k/`.

PrimeIntellect. 2025. *SYNTHETIC-1 dataset*. `https://huggingface.co/datasets/PrimeIntellect/SYNTHETIC-1/`.

Qwen. 2025. Qwen2.5 Technical Report. arXiv:2412.15115 [cs.CL] `https://arxiv.org/abs/2412.15115`

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D. Manning, Stefano Ermon, and Chelsea Finn. 2023. Direct Preference Optimization: Your Language Model is Secretly a Reward Model. In *NeurIPS*.

David Rein, Betty Li Hou, Asa Cooper Stickland, Jackson Petty, Richard Yuanzhe Pang, Julien Dirani, Julian Michael, and Samuel R Bowman. 2024. Gpqa: A graduate-level google-proof q&a benchmark. In *First Conference on Language Modeling*.

Ling Team. 2025. Every FLOP Counts: Scaling a 300B Mixture-of-Experts LING LLM without Premium GPUs. *arXiv preprint arXiv:2503.05139* (2025).

Fanjia Yan, Huanzhi Mao, Charlie Cheng-Jie Ji, Tianjun Zhang, Shishir G. Patil, Ion Stoica, and Joseph E. Gonzalez. 2024. Berkeley Function Calling Leaderboard. `https://gorilla.cs.berkeley.edu/blogs/8_berkeley_function_calling_leaderboard.html`

Jeffrey Zhou, Tianjian Lu, Swaroop Mishra, Siddhartha Brahma, Sujoy Basu, Yi Luan, Denny Zhou, and Le Hou. 2023. Instruction-following evaluation for large language models. *arXiv preprint arXiv:2311.07911* (2023).