Reasoning with Less: Distilling Reasoning Models with Small Datasets

Jonathan Yin
Department of Computer Science
Yale University
New Haven, CT 06520
jonathan.yin@yale.edu

Arman Cohan
Department of Computer Science
Yale University
New Haven, CT 06520
arman.cohan@yale.edu

Abstract

While large language models (LLMs) have made substantial progress in reasoning through large-scale reinforcement learning techniques, their size and compute demands limit practical deployment. In this work, we investigate how to effectively distill reasoning models on small-scale, high-quality datasets. We use a dataset of approximately 1,000 challenging high-school competition math problems as our primary training corpus. Our experiments span student models from 1.5B to 32B parameters and teacher–student pairings drawn from Grok-3 Mini, QwQ-32B, DeepSeek-R1, and DeepSeek-R1-Distill models. We find that larger student models can develop strong reasoning skills even from extremely limited data, while smaller models struggle to benefit meaningfully. Moreover, as model scale increases, student performance becomes increasingly sensitive to teacher quality, with 32B students showing substantial gains when distilled from stronger teachers. Notably, we observe that reasoning abilities acquired from a narrow domain generalize effectively to out-of-domain reasoning tasks, with larger student models distilled from stronger teachers performing the best. Together, these results highlight that both student capacity and teacher quality are critical for effective reasoning distillation in data-constrained settings, offering practical guidance for building lightweight yet capable reasoning models. Our models, data, and code are open-source at https://github.com/jonathanyin12/distilled-reasoning.

1 Introduction

Recent advancements in large language models (LLMs) have shifted the focus from merely scaling model parameters to explicitly cultivating reasoning abilities. OpenAI's o1 series exemplified this paradigm shift by employing reinforcement learning to generate extended chains-of-thought (CoT), achieving state-of-the-art performance on reasoning benchmarks.

Building upon this foundation, DeepSeek introduced the DeepSeek-R1 model [1], demonstrating that high-performance reasoning could be attained through a multi-stage large-scale reinforcement learning pipeline. In parallel, DeepSeek's R1-Distill series demonstrated the feasibility of distilling these reasoning capabilities into smaller non-reasoning models via supervised fine-tuning [1].

These developments have spurred a wave of open-source efforts [4, 9, 2, 6, 5, 14] to replicate and extend this line of work, reflecting a growing interest in understanding and enhancing reasoning across diverse model scales and architectures. While some reproductions aim to match the scale of R1 by collecting large volumes of reasoning traces (e.g., OpenThoughts [9], Open-R1 [2]), others, like LIMO [14] and s1 [6], take the opposite approach, curating small, high-quality datasets.

This growing body of work coincides with an increasing need for smaller, more efficient language models. Real-world deployment scenarios often impose strict constraints on latency, memory, and

Preprint. Under review.

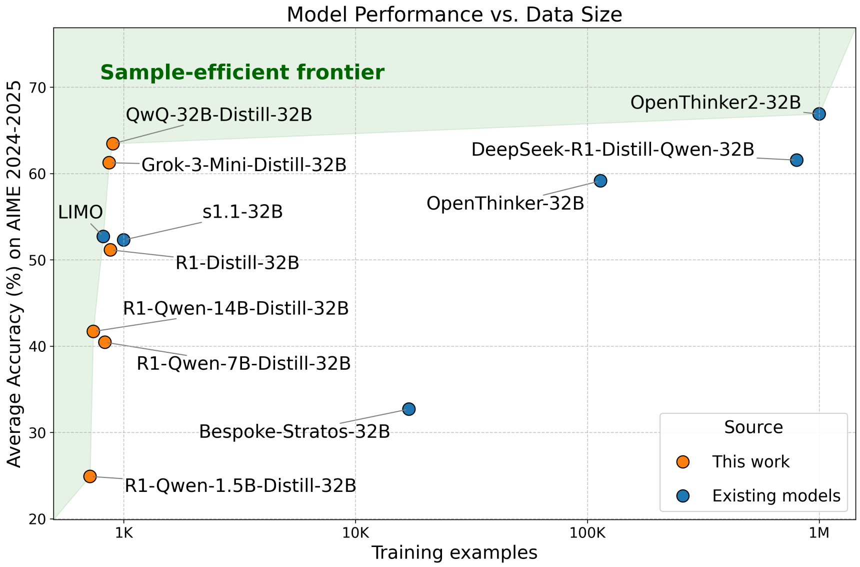

Figure 1: Pass@1 accuracy on AIME 2024–2025 vs. the number of SFT examples for various 32B parameter distilled reasoning models. Notably, QwQ-32B-Distill-32B from our work reaches 63.5% average accuracy using <1K examples, rivaling or exceeding many state-of-the-art large-data baselines. See Table 2 for more details.

energy consumption, making large-scale models impractical in many settings. At the same time, applications like problem-solving, coding, and scientific reasoning demand models that go beyond surface-level pattern matching to exhibit robust, multi-step reasoning behavior. These dual pressures motivate a key research question: How can we endow small models with strong reasoning capabilities using limited data and compute budgets?

Motivated by these considerations, we investigate the potential of distilling reasoning capabilities into smaller, more deployable models through supervised fine-tuning, aiming to minimize inference costs at deployment. We utilize ~1,000 challenging competition math problems that demand strong reasoning but minimal reliance on external knowledge. While the domain is mathematical, our goal is not to model mathematical reasoning per se, but to evaluate generalized reasoning in a setting that is tightly scoped and easily measurable.

Our study examines how several core factors affect reasoning performance: student model size, training dataset scale, and teacher ability. We also investigate the extent to which reasoning capabilities distilled from a narrow domain generalize to out-of-domain tasks. To this end, we train and evaluate a range of teacher–student model pairings, spanning 1.5B to 32B parameters, and assess generalization on out-of-domain benchmarks.

Our contributions are as follows:

- We show that larger student models (e.g., 32B parameters) can develop strong reasoning skills and near state-of-the-art performance even from extremely small datasets (~1K examples) (Figure 1), while smaller students fail to benefit meaningfully.
- We demonstrate that teacher quality becomes increasingly important with student scale: large students show strong positive transfer from stronger teachers, whereas small students exhibit little sensitivity (Figure 2).
- We provide evidence that reasoning skills distilled from a narrow domain (competition math) can generalize to out-of-domain reasoning tasks (science reasoning).

Overall, our findings advance the understanding of how to efficiently equip smaller models with robust reasoning skills in data-constrained environments. Our models, data, and code are open-source at https://github.com/jonathanyin12/distilled-reasoning.

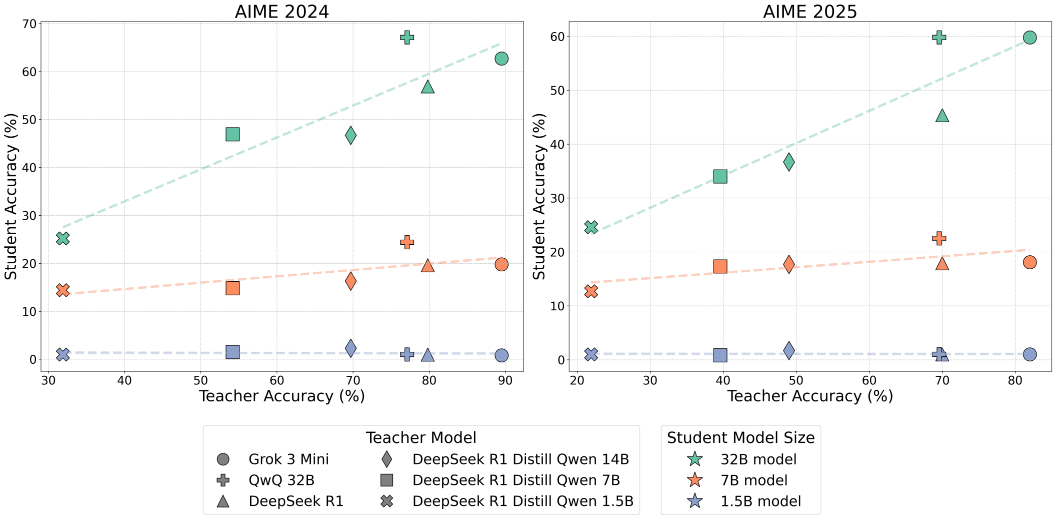

Figure 2: Pass@1 accuracy of student models versus their teacher model's accuracy on AIME 2024 (left) and AIME 2025 (right). Each marker represents a specific teacher–student pairing (marker shape denotes the teacher model; color denotes student scale: 1.5B in blue, 7B in orange, 32B in green). Larger students exhibit a much stronger positive correlation between teacher and student performance.

Teacher Model	AIME 2024	AIME 2025	GPQA Diamond
Grok 3 Mini (high)	**89.5**	**82.0**	**80.3**
QwQ-32B	77.1	69.6	63.6
DeepSeek-R1	79.8	70.0	71.5
DeepSeek-R1-Distill-Qwen-14B	69.7	49.0	59.1
DeepSeek-R1-Distill-Qwen-7B	54.2	39.6	50.9
DeepSeek-R1-Distill-Qwen-1.5B	31.9	21.9	34.3

Table 1: Various teacher models and their pass@1 accuracy (mean ± std) on AIME 2024, AIME 2025, and GPQA Diamond. We evaluated QwQ-32B, DeepSeek-R1-Distill-Qwen-7B, and DeepSeek-R1-Distill-Qwen-1.5B. Other results are from the respective reports [13] [1]

2 Methodology

2.1 Model Selection

We study the distillation of reasoning capabilities into the Qwen2.5 family of models [10], focusing on three student model sizes: 1.5B, 7B, and 32B parameters. These models were selected due to their open weights, solid general capabilities, and architectural consistency across scales, allowing for controlled comparisons. All student models are initialized from publicly available pretrained checkpoints without additional math-specific instruction tuning prior to our experiments.

We draw from a diverse pool of teacher models, chosen to span a wide range of performance on reasoning benchmarks as shown in Table 1. These include Grok 3 Mini (high) [13], QwQ-32B [11], DeepSeek-R1 [1], and DeepSeek-R1-Distill (1.5B, 7B, 14B) [1]. By pairing non-reasoning base models with reasoning-capable teachers, we isolate the effect of supervision quality on the student's ability to acquire structured reasoning behavior.

2.2 Reasoning Dataset

Our primary training dataset is composed of 919 problems from the American Invitational Mathematics Examination (AIME), spanning the years 1983–2023 [12]. Each problem is a self-contained, high-school-level math question that tests mathematical problem-solving with arithmetic, algebra, counting, geometry, number theory, probability, and other secondary school math topics. All AIME answers are integers ranging from 000 to 999, inclusive.

We chose AIME for four main reasons: (i) the problems are challenging and unambiguous, (ii) the dataset is small but well-curated, enabling rapid experimentation, (iii) the format (short-answer questions with numerical responses) allows for easy answer extraction and automated evaluation, and (iv) AIME 2024 and AIME 2025 are the standard benchmarks for the mathematical ability of reasoning models.

For each teacher model, we generate a reasoning trace and corresponding final answer for every problem in the training set. To ensure high-quality supervision, we regenerate responses until the teacher produces a correct final answer, preventing the student from learning from flawed or incorrect reasoning.

2.3 Training

We conduct full-parameter supervised fine-tuning on the student models using reasoning trajectories generated by teacher models. Minor hyperparameter adjustments, primarily learning rate and number of epochs, are made to accommodate differences in model scale.

Training times vary proportionally with model size: fine-tuning the Qwen2.5-1.5B model takes approximately 12 minutes, the Qwen2.5-7B model takes 40 minutes, and the Qwen2.5-32B model takes around 2 hrs. All experiments are conducted on 8 NVIDIA H200 GPUs using PyTorch Fully Sharded Data Parallel (FSDP).

2.4 Evaluation

We primarily evaluate model performance on the AIME 2024 and AIME 2025 benchmarks, which consist of 30 problems each, drawn from the 2024 and 2025 American Invitational Mathematics Examinations. Since our models do not support image inputs, we encode any figures present using the vector graphics language Asymptote and provide that in the prompt.

To assess out-of-domain generalization, we evaluate performance on the GPQA Diamond benchmark[8], a set of expert-authored questions in biology, physics, and chemistry.

All models are evaluated with a maximum generation length of 32,768 tokens. We use a decoding temperature of 0.6 and top-p sampling with a threshold of 0.95. For AIME24 and AIME25, we generate 16 completions per problem using 16 different random seeds to estimate Pass@1. For GPQA Diamond, we only generate 8 completions per problem using 8 different random seeds to estimate Pass@1. Evaluation is conducted using the LightEval framework [3].

3 Results

3.1 Performance Across Model Size and Data Scale

Table 2 reports Pass@1 accuracy on AIME 2024 and AIME 2025 for students at three different parameter scales (1.5B, 7B, and 32B), each distilled from various teachers using supervised fine-tuning (SFT) on datasets ranging from fewer than 1K to 1M examples. We benchmarked our distilled models against the corresponding Qwen2.5-Instruct base models, DeepSeek-R1-Distill models, OpenThinker models, BespokeStratos models, s1.1 series models, and LIMO.

The Qwen2.5-Instruct base models all exhibit very low accuracy. The models distilled from large-scale reasoning datasets—such as DeepSeek-R1-Distill and OpenThinker2—substantially outperform their base models across all scales (1.5B, 7B, 32B). This underscores that supervised fine-tuning on large, high-quality reasoning corpora (~1M examples) consistently yields large performance gains, regardless of model size.

However, when the training set size is severely restricted (~1K examples), a strong dependence on model scale emerges. At 1.5B, students fail to meaningfully outperform the base model, achieving Pass@1 scores of roughly 1–2%. At 7B, students improve to roughly 12–25%, while at 32B, students trained on only 1K examples achieve between 24% and 68% accuracy. These findings reveal that although dataset size drives absolute gains, student model size becomes the critical enabler of

Model	Training Samples	AIME 2024	AIME 2025
Based on: Qwen2.5-1.5B Instruct			
Qwen2.5 1.5B Instruct	N/A	2.7±2.4	1.5±2.0
DeepSeek-R1-Distill-Qwen-1.5B	800k	**31.9**±8.3	**21.9**±3.1
s1.1-1.5B	1k	1.0±1.6	2.3±1.9
Grok-3-Mini-Distill-1.5B	<1K	0.8±1.4	1.0±1.6
QwQ-32B-Distill-1.5B	<1K	1.0±1.9	1.0±1.6
R1-Distill-1.5B	<1K	1.0±1.6	1.0±1.9
R1-Qwen-14B-Distill-1.5B	<1K	2.3±2.6	1.7±2.4
R1-Qwen-7B-Distill-1.5B	<1K	1.5±2.6	0.8±1.9
R1-Qwen-1.5B-Distill-1.5B	<1K	1.0±1.6	1.0±1.6
Based on: Qwen2.5-7B Instruct			
Qwen2.5-7B-Instruct	N/A	11.7±4.1	8.1±4.4
DeepSeek-R1-Distill-Qwen-7B	800k	54.2±5.3	**39.6**±3.7
OpenThinker2-7B	1M	**57.9**±6.9	37.9±7.1
OpenThinker-7B	114K	29.8±4.6	25.8±4.3
BespokeStratos-7B	17K	19.2±6.3	19.4±3.6
s1.1-7B	1K	19.4±3.8	19.0±4.0
Grok-3-Mini-Distill-7B	<1K	19.8±5.2	18.1±4.1
QwQ-32B-Distill-7B	<1K	24.4±4.2	22.5±4.9
R1-Distill-7B	<1K	19.6±3.5	17.9±3.9
R1-Qwen-14B-Distill-7B	<1K	16.3±4.1	17.7±3.7
R1-Qwen-7B-Distill-7B	<1K	14.8±3.1	17.3±4.1
R1-Qwen-1.5B-Distill-7B	<1K	14.4±4.2	12.7±4.1
Based on: Qwen2.5-32B Instruct			
Qwen2.5-32B-Instruct	N/A	15.4±4.1	11.5±5.3
DeepSeek-R1-Distill-Qwen-32B	800k	70±3.9	53.1±5.2
OpenThinker2-32B	1M	**71.9**±4.9	**61.9**±6.3
OpenThinker-32B	114K	65.2±5.8	53.1±4.9
BespokeStratos-32B	17K	38.1±6.5	27.3±4.8
s1.1-32B	1K	60.4±6.0	44.2±5.1
LIMO	<1K	58.5±4.6	46.9±3.6
Grok-3-Mini-Distill-32B	<1K	62.7±5.4	59.8±5.3
QwQ-32B-Distill-32B	<1K	67.1±4.2	59.8±6.5
R1-Distill-32B	<1K	56.9±5.6	45.4±5.1
R1-Qwen-14B-Distill-32B	<1K	46.7±4.6	36.7±5.8
R1-Qwen-7B-Distill-32B	<1K	46.9±6.0	34.0±6.6
R1-Qwen-1.5B-Distill-32B	<1K	25.2±5.9	24.6±3.1

Table 2: Pass@1 accuracy (mean ± std) on AIME 2024 and AIME 2025 benchmarks for various reasoning-distilled models fine-tuned on Qwen2.5-1.5B, 7B, and 32B. Dataset sizes range from <1K to 1M samples (N/A indicates no fine-tuning).

extracting value from limited, high-quality supervision. Large models possess sufficient capacity to generalize useful reasoning patterns even from extremely small datasets.

Among all evaluated models, our distilled model QwQ-32B-Distill-32B achieves the second-highest accuracy overall, trailing only OpenThinker2-32B. Impressively, QwQ-32B-Distill-32B is trained on roughly 1,000 times fewer examples compared to models like OpenThinker2-32B and DeepSeek-R1-Distill-Qwen-32B, yet matches or exceeds their performance. This result highlights that for large enough students, data scale ceases to be the dominant limiting factor: careful selection of teacher quality and small, high-fidelity supervision can drive competitive or even state-of-the-art in-domain reasoning performance.

Model	Training Samples	Fine-Tuning Data Domain	GPQA Diamond
Based on: Qwen2.5-7B Instruct			
Qwen2.5-7B-Instruct	N/A	N/A	34.3±3.5
DeepSeek-R1-Distill-Qwen-7B	800k	Mixed	**50.9**±2.3
OpenThinker2-7B	1M	Mixed	44.9±1.9
s1.1-7B	1k	Mixed	39.5±2.5
QwQ-32B-Distill-7B	<1K	Math	40.0±2.8
R1-Distill-7B	<1K	Math	37.1±2.2
R1-Qwen-7B-Distill-7B	<1K	Math	36.4±2.8
R1-Qwen-1.5B-Distill-7B	<1K	Math	33.8±2.7
Based on: Qwen2.5-32B Instruct			
Qwen2.5-32B-Instruct	N/A	N/A	47.2±3.7
DeepSeek-R1-Distill-Qwen-32B	800k	Mixed	64.0±2.0
OpenThinker2-32B	1M	Mixed	63.7±1.4
s1.1-32B	1k	Mixed	62.1±1.9
LIMO	<1K	Math	**64.5**±2.8
QwQ-32B-Distill-32B	<1K	Math	62.4±1.3
R1-Distill-32B	<1K	Math	57.9±2.5
R1-Qwen-7B-Distill-32B	<1K	Math	51.6±1.6
R1-Qwen-1.5B-Distill-32B	<1K	Math	47.2±2.1

Table 3: Pass@1 accuracy (mean ± std) on GPQA Diamond for various reasoning-distilled models.

3.2 Dependence on Teacher Quality

Figure 2 examines how student performance depends on the underlying quality of the teacher model across both benchmarks. Each point corresponds to a particular teacher–student pairing, with marker shape encoding the teacher identity and color encoding the student scale.

For 1.5B students (blue), we observe a flat trend: despite teacher accuracies varying from 20% to 90%, student accuracies remain stuck around 0–2%. This plateau suggests a hard bottleneck in small models' capacity to absorb complex reasoning behaviors, irrespective of teacher quality.

At the 7B scale (orange), student accuracy exhibits a mild positive correlation with teacher strength: as teacher accuracy improves, student Pass@1 improves from ~12% to ~24%. This indicates that larger students benefit somewhat from better teachers but still face capacity limits under small-data supervision.

At 32B (teal), the dependence on teacher quality becomes markedly stronger. Gains in teacher performance translate almost linearly into gains in student performance, with student accuracies rising from ~25% to ~67% as teacher accuracy increases. This robust trend appears consistently across both AIME 2024 and 2025 benchmarks.

Collectively, these results highlight a scale-dependent teacher transfer effect: larger student models benefit dramatically and proportionally from stronger teachers, even with limited training data, while smaller students exhibit modest or no gains from stronger teachers.

3.3 Out-of-domain Generalization

To assess whether distilled reasoning capabilities transfer beyond the training domain, we evaluate on GPQA Diamond, an out-of-domain benchmark focused on scientific reasoning across biology, chemistry, and physics. Results are shown in Table 3.

Among 7B-scale models, those distilled on small math-only datasets yield modest improvements over the base model, with QwQ-32B-Distill-7B achieving the highest accuracy at 40.0%. However, these models are notably outperformed by models trained on large, mixed-domain datasets such as

DeepSeek-R1-Distill-Qwen-7B (50.9%), highlighting the benefits of training data diversity at this scale.

At the 32B scale, however, math-only distilled models perform surprisingly well. QwQ-32B-Distill-32B and LIMO, both trained on <1K math examples, achieve 62.4% and 64.5% respectively, matching the best mixed-domain models like DeepSeek-R1-Distill-Qwen-32B (64.0%) and OpenThinker2-32B (63.7%), which are trained on ~1000 times more data. This indicates that high-capacity models can extract domain-general reasoning capabilities from small, domain-specific supervision, enabling strong generalization to unfamiliar scientific tasks.

Finally, we again observe a dependence on teacher quality and student model size: students distilled from stronger teachers (e.g., QwQ-32B and R1) consistently outperform those distilled from weaker ones (e.g., R1-Distill-Qwen-7B and R1-Distill-Qwen-1.5B), even in out-of-domain settings, with the difference being more pronounced in larger student models. This mirrors the in-domain trend, suggesting that both in-domain and out-of-domain performance hinge on teacher strength and student model size when supervision is limited.

4 Discussion

Our study demonstrates that generalized reasoning capabilities can be effectively distilled into smaller models through supervised fine-tuning on small, high-quality datasets, provided the student model has sufficient scale. We find that even with as few as ~1K examples, 32B parameter student models can achieve near state-of-the-art in-domain performance when distilled from strong teachers. By contrast, smaller models (e.g., 1.5B) show limited ability to benefit meaningfully, highlighting a strong dependence on student capacity. Furthermore, our results show that teacher quality becomes increasingly important at larger student scales. Stronger teachers directly translate into stronger students, whereas small students remain relatively insensitive to teacher quality.

These findings suggest that student capacity fundamentally constrains the effectiveness of reasoning distillation in low-data regimes. In particular, while small students may require richer pretraining or more data to realize meaningful gains, large students can extract complex reasoning patterns even from limited supervision.

Moreover, the successful transfer of reasoning capabilities from a narrow mathematical domain to broader scientific reasoning tasks (i.e. GPQA) indicates that reasoning behaviors themselves may be domain-agnostic at sufficient model scale. This suggests a pathway for training deployable reasoning models: focus on a small, controlled domain to teach structure, then generalize to broader applications.

Limitations

Several limitations merit discussion. First, our training and evaluation focus on tasks with well-posed, deterministic answers. These benchmarks are valuable for evaluating structured reasoning, but they do not capture the full complexity of real-world reasoning tasks, which often require multi-step justification, formal proofs, or engagement with ambiguous and open-ended questions. For example, competition-style benchmarks only evaluate final numerical answers and omit proof-based reasoning, which is essential for many mathematical tasks. Similarly, GPQA targets factual scientific reasoning, but real scientific inquiry often involves synthesizing conflicting information or proposing hypotheses under uncertainty. Recent studies have highlighted that models achieving high scores on benchmarks like AIME often fail to produce coherent or logically valid reasoning when faced with proof-based tasks, such as those in the United States of America Mathematical Olympiad (USAMO) [7]. As a result, our findings may overestimate the models' ability to generalize to domains requiring more stringent, proof-level reasoning.

Second, our findings primarily apply to large student models (e.g., 32B), which are impractical for certain real-world deployment settings, such as local on-device applications. While we show that large models can extract substantial value from small, high-quality datasets, smaller models (e.g., 1.5B and 7B) show limited gains under the same distillation setup. This raises an important open challenge: how to effectively distill reasoning capabilities into truly lightweight models using minimal data. Achieving strong reasoning in this regime likely requires new strategies to close the capability gap observed in small models under constrained supervision.

5 Conclusion

We explore how to distill reasoning skills using limited but high-quality supervision. Our results show that student capacity is a key factor: larger models can acquire transferable reasoning abilities from just ~1K curated math problems, especially when paired with strong teachers. These findings suggest a practical path for training lightweight models that reason well, even in low-data settings, and highlight the importance of both scale and supervision quality in enabling generalizable reasoning.

References

[1] DeepSeek-AI. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning, 2025. URL https://arxiv.org/abs/2501.12948.

[2] Hugging Face. Open r1: A fully open reproduction of deepseek-r1, January 2025. URL https://github.com/huggingface/open-r1.

[3] Clémentine Fourrier, Nathan Habib, Hynek Kydlíček, Thomas Wolf, and Lewis Tunstall. Lighteval: A lightweight framework for llm evaluation, 2023. URL https://github.com/huggingface/lighteval.

[4] Bespoke Labs. Bespoke-stratos: The unreasonable effectiveness of reasoning distillation. www.bespokelabs.ai/blog/bespoke-stratos-the-unreasonable-effectiveness-of-reasoning-distillation, 2025. Accessed: 2025-01-22.

[5] Michael Luo, Sijun Tan, Justin Wong, Xiaoxiang Shi, William Y. Tang, Manan Roongta, Colin Cai, Jeffrey Luo, Li Erran Li, Raluca Ada Popa, and Ion Stoica. Deepscaler: Surpassing o1-preview with a 1.5b model by scaling rl. https://pretty-radio-b75.notion.site/DeepScaleR-Surpassing-O1-Preview-with-a-1-5B-Model-by-Scaling-RL-19681902c1468005bed8ca3030
2025. Notion Blog.

[6] Niklas Muennighoff, Zitong Yang, Weijia Shi, Xiang Lisa Li, Li Fei-Fei, Hannaneh Hajishirzi, Luke Zettlemoyer, Percy Liang, Emmanuel Candès, and Tatsunori Hashimoto. s1: Simple test-time scaling, 2025. URL https://arxiv.org/abs/2501.19393.

[7] Ivo Petrov, Jasper Dekoninck, Lyuben Baltadzhiev, Maria Drencheva, Kristian Minchev, Mislav Balunović, Nikola Jovanović, and Martin Vechev. Proof or bluff? evaluating llms on 2025 usa math olympiad, 2025. URL https://arxiv.org/abs/2503.21934.

[8] David Rein, Betty Li Hou, Asa Cooper Stickland, Jackson Petty, Richard Yuanzhe Pang, Julien Dirani, Julian Michael, and Samuel R. Bowman. Gpqa: A graduate-level google-proof qa benchmark, 2023. URL https://arxiv.org/abs/2311.12022.

[9] OpenThoughts Team. Open Thoughts. https://open-thoughts.ai, January 2025.

[10] Qwen Team. Qwen2.5: A party of foundation models, September 2024. URL https://qwenlm.github.io/blog/qwen2.5/.

[11] Qwen Team. Qwq-32b: Embracing the power of reinforcement learning, March 2025. URL https://qwenlm.github.io/blog/qwq-32b/.

[12] Hemish Veeraboina. Aime problem set 1983-2024, 2023. URL https://www.kaggle.com/datasets/hemishveeraboina/aime-problem-set-1983-2024.

[13] xAI. Grok 3 beta — the age of reasoning agents, 2025. URL https://x.ai/news/grok-3.

[14] Yixin Ye, Zhen Huang, Yang Xiao, Ethan Chern, Shijie Xia, and Pengfei Liu. Limo: Less is more for reasoning, 2025. URL https://arxiv.org/abs/2502.03387.