First Step Advantage: Importance of Starting Right in Multi-Step Reasoning

Anonymous ACL submission

Abstract

Large Language Models (LLMs) can solve complex reasoning tasks by generating rationales for their predictions. Distilling these capabilities into a smaller, compact model can facilitate the creation of specialized, cost-effective models tailored for specific tasks. However, smaller models often face challenges in complex reasoning tasks and often deviate from the correct reasoning path. We show that LLMs can guide smaller models and bring them back to the correct reasoning path only if they intervene at the right time. We show that smaller models fail to reason primarily due to their difficulty in initiating the process, and that guiding them in the right direction can lead to a performance gain of over 100%. We explore different model sizes and evaluate the benefits of providing guidance to improve reasoning in smaller models.

1 Introduction

Over the years, Large Language Models (LLMs) have improved their reasoning skills by explaining their intermediate thoughts (Wei et al., 2022). This allows LLMs to transfer intermediate knowledge to student models [1] to improve their reasoning skills; often referred to as knowledge distillation (Yuan et al., 2023; Magister et al., 2023; Shridhar et al., 2023b; Hsieh et al., 2023). While training student models can certainly improve their reasoning skills, there are instances where teacher intervention remains essential to guide the student model when it encounters uncertainty or confusion. This is similar to the situation in a classroom, where a student can acquire knowledge independently by learning from textbooks, but often benefits from the guidance of a teacher (Wood et al., 1976; Van de Pol et al., 2015).

[1]We refer to smaller models as student models and larger models as teachers. The distinction is not based on the number of parameters, but rather on relative size, with smaller models often referred to as students.

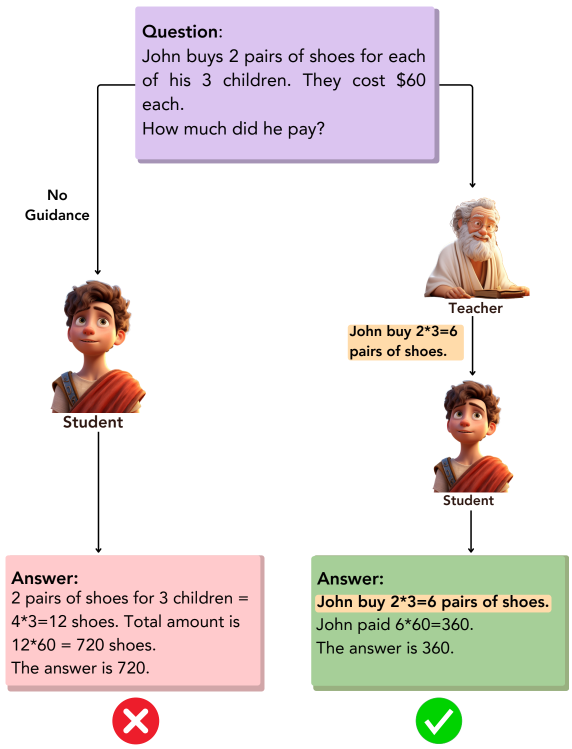

Figure 1: **The First Step Advantage**: Figure demonstrating the effect of first step guidance from a teacher on the student's overall performance (right) versus no guidance (left).

While teacher intervention can provide valuable assistance to students, understanding *when* and *how* to provide guidance plays a critical role. In general, more guidance typically leads to a consistent improvement in student performance (Saha et al., 2023), but the question arises: *should guidance at different stages be given equal importance?*

Our observations, as shown in Figure 1, suggest that for multi-step reasoning tasks, intervening at the first step yields the most significant benefits, with the effects diminishing at subsequent steps. As expected, expert teachers (models with superior performance) tend to provide more effective guidance, resulting in better student performance.

On a mathematical dataset of multi-step word problems, we demonstrate the effectiveness of *first step guidance* on various combinations of teacher-student pairs. A consistent improvement in student performance was observed with first step guidance. We show that a smaller student model (LLaMA 13B (Touvron et al., 2023)), when correctly guided at its first step, can achieve the same performance as a larger student model (LLaMA 70B) without any guidance. Furthermore, our results show a continuous and upward improvement of the student model's performance with expert guidance, with GPT-4 (OpenAI, 2023) as the teacher reaching a performance level similar to that of a human instructor.

2 Related Work

Previous work has shown that it is possible to elicit reasoning abilities from LLMs through in-context learning (Wei et al., 2022; Zhou et al., 2022). A key recipe in most methods is to spread the reasoning process over multiple tokens, rather than expecting them to provide an immediate response token. One way is to provide the model with the intermediate steps, or chain of thought (CoT), that leads to the final answer (Wei et al., 2022; Kojima et al., 2023; Yang et al., 2023; Wang et al., 2023). In parallel, in-context learning has been used to teach the model how to break a problem down into smaller, easier sub-problems, and then solve those sub-problems that eventually lead to the final answer (Shridhar et al., 2022; Zhou et al., 2023).

However, if the problem is misinterpreted, it can lead to a cascade of errors in subsequent steps. To counter this, several techniques have been proposed to intervene and correct intermediate steps (Welleck et al., 2022) or to provide feedback on their own generations, essentially "self-correcting" their own generations (Madaan et al., 2023; Shridhar et al., 2023a). It is important to note that complex reasoning and self-correcting capabilities only emerge in very large language models. For smaller models, CoT follows a rather flat scaling curve (Wei et al., 2022). Our work presents an effective way to transfer such reasoning capabilities into smaller models that does not require large scale pre-training, making it more accessible for researchers with limited compute.

While the LLM's ability to revise its own generations may prove helpful in many cases, it sometimes leads to worse outcomes upon refinement, requiring a "rolling-back" to the previous output (Shridhar et al., 2023a). The "rolling-back" dilemma can be avoided if we can know *when* to intervene. (Saha et al., 2023) presented an approach based on ToM (Kosinski, 2023; Kadavath et al., 2022), where a teacher model intervenes in a student model only for harder questions by creating an implicit mental model of the student's understanding. In contrast, our work shows that it is not necessary to intervene and help a student with the entire solution. Rather, just starting correctly has a significant impact on the student's performance and avoids the need for backtracking and correcting mistakes, saving time and effort.

3 Experimental Design

Dataset We examine our intervention on multi-step mathematical dataset : GSM8K (Cobbe et al., 2021). The dataset is a grade-school-level math word problem dataset with a training set of 7473 samples and a test set of 1319 samples, each requiring two to eight steps to solve.

Setup We used variants of LLaMA models (LLaMA 7B, LLaMA 13B and LLaMA 70B) (Touvron et al., 2023) and its variants (Mistral 7B (Jiang et al., 2023) and MetaMath 7B and 13B (Yu et al., 2023))as student models. On the other hand, we used LLaMA 13B and LLaMA 70B as teachers alongside ChatGPT (`gpt-3.5-turbo`) (Brown et al., 2020) and GPT-4 (`gpt-4`) (OpenAI, 2023). For pre-trained models as students in few-shot settings, 4-shot demonstrations were provided, which were chosen randomly from the train set. Fine-tuned students were trained on the training set with no modifications for 3 epochs. All models were evaluated in the greedy approach (temp=0, top p=1). Fine-tuning was performed on 1 node of 8 A100 GPUs. We report the accuracy (`maj@1`) on the test set.

4 Results and Discussion

Early intervention is key Figure 2 compares the effect of intervention by humans as teachers for LLaMA 7B student model on the GSM8K dataset. Intervention at the first step (in blue) proves to be most beneficial with maximum gains over baseline without any interventions (dotted line). The gains go down when the intervention is done at "step 2" (in orange) and intervening at later stages ("step 3" and beyond) leads to diminishing returns. This is because students may have already internalized

Student		Teacher					
Model	**Type**	**No**	**LLaMA 13B**	**LLaMA 70B**	**ChatGPT**	**GPT-4**	**Human**
LLaMA 7B	Pre-Trained	10.53	14.86	19.48	21.00	**23.27**	22.74
LLaMA 7B	Fine-Tuned	34.19	38.26	45.90	45.94	**47.61**	47.15
LLaMA 13B	Pre-Trained	24.70	-	26.39	33.24	**35.75**	33.75
LLaMA 13B	Fine-Tuned	46.24	-	55.34	59.28	60.50	**61.86**
LLaMA 70B	Pre-Trained	58.90	-	-	63.53	67.40	**67.55**
LLaMA 70B	Fine-Tuned	63.30	-	-	70.05	72.32	**74.14**
Mistral 7B	Pre-Trained	40.25	-	46.17	48.82	49.50	**50.34**
MetaMath 7B	Pre-Trained	62.69	-	61.48	66.94	**69.52**	65.95
MetaMath 13B	Pre-Trained	67.85	-	66.79	71.41	**75.51**	73.00

Table 1: Accuracy comparison for different configurations of the student and the teacher models. No refers to *no intervention* by any teacher. Best results are presented in **bold.**

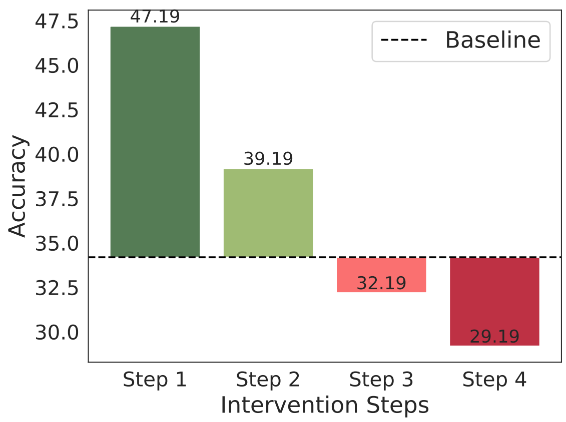

Figure 2: Accuracy of LLaMA 7B fine-tuned student model on GSM8K dataset with correct intervention at different steps by humans. The baseline is the represented by the dotted line with an accuracy of 34.19.

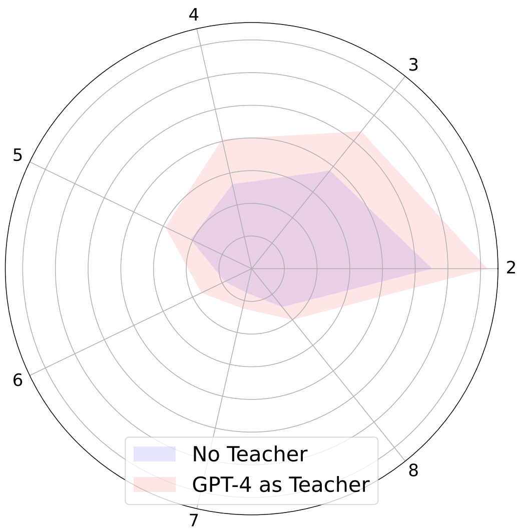

Figure 3: Accuracy comparison for No teacher vs GPT-4 as teacher for LLaMA 7B fine-tuned model across different steps for GSM8K dataset. 2-8 represents the number of steps needed to solve the problem.

incorrect concepts or approaches, making it harder to correct and replace them with the correct ones. This underscores the necessity of early and accurate guidance.

Starting right: first step to successful reasoning Table 1 demonstrates the usefulness of a teacher guiding a student with first-step guidance. LLaMA 7B pre-trained model's performance improves by 40% (10.5 → 14.8) when a slightly larger model (LLaMA 13B) acts as a teacher to provide first-step guidance. The performance increases monotonically with larger and expert models (10.5 → 19.5 with LLaMA 70B as teacher and to 21.0 with ChatGPT as teacher) and goes beyond 100% improvement with GPT-4 (10.5 → 23.2) or Human teachers (10.5 → 22.7). This demonstrates the importance of first-step guidance in reaching to the correct answer. Similar results are observed for LLaMA 13B as a student where the performance goes from 24.70 with no guidance to 35.75 with GPT-4 first step guidance. The trend follows similarly for even a relatively larger student model (LLaMA 70B) where the performance goes up from 58.9 without guidance to 67.4 with GPT-4 guidance.

Next, we fine-tune all the student models on the training data to see if first step guidance can be learned during training. Although fine-tuning improves the baseline performance significantly, the guidance still proved useful in boosting the performance for all models. LLaMA 7B fine-tuned model's performance improves by 40% (34.1 → 47.6) with expert guidance from GPT-4 as the teacher. A qualitative example is presented in Figure 4 where if LLaMA 7B model is guided to start correctly, it will to the correct answer. The trend is

Question Gene is sewing a quilt out of old souvenir t-shirts. He has one shirt from each vacation he has been on. Every shirt is its own quilt block. Each row is made of blocks from a different year of vacations. He goes on four vacations a year and has been vacationing since he was 23 years old. He is now 34. How many quilt blocks does he have in total?

No Guidance He has been on 4 vacations a year for 34 years so he has been on 4*34 = «4*34=136»136 vacations. Each vacation is a quilt block so he has 136 quilt blocks. Each quilt block is one shirt so he has 136 shirts. The answer is 136. ✗

GPT4 Guidance Gene has been on 34 - 23 = «34-23=11»11 years of vacations, so he has 11 rows of blocks. Each row is made of 4 blocks, so he has 11*4 = «11*4=44»44 blocks. The answer is 44. ✓

Figure 4: Example demonstrating the usefulness of first step guidance.

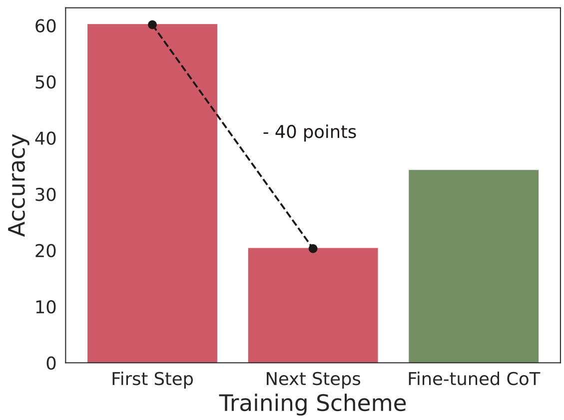

Figure 5: Comparison of curiculum style training for a 7B student model (first step training followed by next steps training) vs fine-tuning (Fine-tuned CoT) on GSM8K dataset.

similar for LLaMA 13B fine-tuned version with a gain of 30% with GPT-4 guidance (46.2 → 60.5) and LLaMA 70B goes up by 15% (63.3 → 72.3).

Finally, we test the applicability of first-step guidance across better student models: Mistral 7B and MetaMath 7B and 13B. Mistral 7B achieves a performance boost of 25% with GPT-4 as teacher (40.25 → 49.50) while MetaMath 7B and 13B gain more than 10% each (62.69 → 69.52 and 67.85 → 75.51 respectively). In all the cases above, it is worth noting that the guidance of GPT-4 is very close to human's guidance and in many cases surpasses it. This demonstrates the capabilities of GPT-4 as an alternative to teachers in educational domains.

Starting right helps even for longer reasoning chains Figure 3 demonstrates the performance of LLaMA 7B model with and without first step guidance from a teacher for different steps in GSM8K dataset. Across all steps (2 to 8), guidance improves the performance and suggests that starting with a solid foundation can help over longer context. However, the improvement is higher for problems with 2 to 5 steps compared to 6 to 8, suggesting that over a longer reasoning chain, the chances of making mistakes increase with the increase in the number of steps required to solve it. Nonetheless, starting right has a positive impact on longer reasoning chains too.

Can smaller models be aligned to start better? Each problem can be broken down into a first step and the next consecutive steps, where the first step can serve as a guidance for the consecutive steps. We train the student model to first learn the initial step and then fine-tune it further to learn the next steps required to solve the problem. This two-step training mechanism has similarity with curriculum learning (Platanios et al., 2019; Xu et al., 2020) where the simpler first step is learnt first, followed by the subsequent more difficult steps. Figure 5 shows a drop of 40 points once the next steps are learned and overall performance gets worse than learning all steps at once in a fine-tuning style. Since the first-step accuracy is close to 60%, only 3/5 samples get the correct guidance during the next steps training and we suspect this might be the reason for a lower overall performance.

5 Conclusion

Distilling reasoning capabilities in smaller models is a challenging task due to their limited abilities to learn complex reasoning strategies. To make these skills more accessible to smaller models, we present an effective way of first-step guidance, where LLMs can guide smaller models in the right direction to solve a reasoning task. On a multi-step reasoning dataset, we show the importance of starting right with a performance improvement of over 100%. Finally, our experiments reveal the quality of guidance which the monotonically increases with the size of the expertise of the teacher model.

6 Limitations

Our work has been tested on one multi-step mathematical reasoning dataset, and while the method can be extended to other reasoning datasets, we have not explicitly tested this in this work. LLMs in general are vulnerable to adversarial attacks and are often very sensitive to hyperparameter changes. We do not see any real-world application of our work.

References

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc.

Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, Christopher Hesse, and John Schulman. 2021. Training verifiers to solve math word problems.

Cheng-Yu Hsieh, Chun-Liang Li, Chih-Kuan Yeh, Hootan Nakhost, Yasuhisa Fujii, Alexander J. Ratner, Ranjay Krishna, Chen-Yu Lee, and Tomas Pfister. 2023. Distilling step-by-step! outperforming larger language models with less training data and smaller model sizes. *ArXiv*, abs/2305.02301.

Albert Qiaochu Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de Las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, L'elio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2023. Mistral 7b. *ArXiv*, abs/2310.06825.

Saurav Kadavath, Tom Conerly, Amanda Askell, Tom Henighan, Dawn Drain, Ethan Perez, Nicholas Schiefer, Zac Hatfield-Dodds, Nova DasSarma, Eli Tran-Johnson, Scott Johnston, Sheer El-Showk, Andy Jones, Nelson Elhage, Tristan Hume, Anna Chen, Yuntao Bai, Sam Bowman, Stanislav Fort, Deep Ganguli, Danny Hernandez, Josh Jacobson, Jackson Kernion, Shauna Kravec, Liane Lovitt, Kamal Ndousse, Catherine Olsson, Sam Ringer, Dario Amodei, Tom Brown, Jack Clark, Nicholas Joseph, Ben Mann, Sam McCandlish, Chris Olah, and Jared Kaplan. 2022. Language models (mostly) know what they know.

Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. 2023. Large language models are zero-shot reasoners.

Michal Kosinski. 2023. Theory of mind might have spontaneously emerged in large language models.

Aman Madaan, Niket Tandon, Prakhar Gupta, Skyler Hallinan, Luyu Gao, Sarah Wiegreffe, Uri Alon, Nouha Dziri, Shrimai Prabhumoye, Yiming Yang, Shashank Gupta, Bodhisattwa Prasad Majumder, Katherine Hermann, Sean Welleck, Amir Yazdanbakhsh, and Peter Clark. 2023. Self-refine: Iterative refinement with self-feedback.

Lucie Charlotte Magister, Jonathan Mallinson, Jakub Adamek, Eric Malmi, and Aliaksei Severyn. 2023. Teaching small language models to reason.

OpenAI. 2023. Gpt-4 technical report.

Emmanouil Antonios Platanios, Otilia Stretcu, Graham Neubig, Barnabas Poczos, and Tom Mitchell. 2019. Competence-based curriculum learning for neural machine translation. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 1162–1172, Minneapolis, Minnesota. Association for Computational Linguistics.

Swarnadeep Saha, Peter Hase, and Mohit Bansal. 2023. Can language models teach? teacher explanations improve student performance via personalization. In *Thirty-seventh Conference on Neural Information Processing Systems*.

Kumar Shridhar, Harsh Jhamtani, Hao Fang, Benjamin Van Durme, Jason Eisner, and Patrick Xia. 2023a. Screws: A modular framework for reasoning with revisions.

Kumar Shridhar, Jakub Macina, Mennatallah El-Assady, Tanmay Sinha, Manu Kapur, and Mrinmaya Sachan. 2022. Automatic generation of socratic subquestions for teaching math word problems.

Kumar Shridhar, Alessandro Stolfo, and Mrinmaya Sachan. 2023b. Distilling reasoning capabilities into smaller language models. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 7059–7073, Toronto, Canada. Association for Computational Linguistics.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, Daniel M. Bikel, Lukas Blecher, Cristian Cantón Ferrer, Moya Chen, Guillem Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, Brian Fuller, Cynthia Gao, Vedanuj Goswami, Naman Goyal, Anthony S. Hartshorn, Saghar Hosseini, Rui Hou, Hakan Inan, Marcin Kardas, Viktor

Kerkez, Madian Khabsa, Isabel M. Kloumann, A. V. Korenev, Punit Singh Koura, Marie-Anne Lachaux, Thibaut Lavril, Jenya Lee, Diana Liskovich, Yinghai Lu, Yuning Mao, Xavier Martinet, Todor Mihaylov, Pushkar Mishra, Igor Molybog, Yixin Nie, Andrew Poulton, Jeremy Reizenstein, Rashi Rungta, Kalyan Saladi, Alan Schelten, Ruan Silva, Eric Michael Smith, R. Subramanian, Xia Tan, Binh Tang, Ross Taylor, Adina Williams, Jian Xiang Kuan, Puxin Xu, Zhengxu Yan, Iliyan Zarov, Yuchen Zhang, Angela Fan, Melanie Kambadur, Sharan Narang, Aurelien Rodriguez, Robert Stojnic, Sergey Edunov, and Thomas Scialom. 2023. Llama 2: Open foundation and fine-tuned chat models. *ArXiv*, abs/2307.09288.

Janneke Van de Pol, Monique Volman, Frans Oort, and Jos Beishuizen. 2015. The effects of scaffolding in the classroom: support contingency and student independent working time in relation to student achievement, task effort and appreciation of support. *Instructional Science*, 43:615–641.

Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc Le, Ed Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. 2023. Self-consistency improves chain of thought reasoning in language models.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, brian ichter, Fei Xia, Ed H. Chi, Quoc V Le, and Denny Zhou. 2022. Chain of thought prompting elicits reasoning in large language models. In *Advances in Neural Information Processing Systems*.

Sean Welleck, Ximing Lu, Peter West, Faeze Brahman, Tianxiao Shen, Daniel Khashabi, and Yejin Choi. 2022. Generating sequences by learning to self-correct.

David Wood, Jerome S Bruner, and Gail Ross. 1976. The role of tutoring in problem solving. *Child Psychology & Psychiatry & Allied Disciplines*.

Benfeng Xu, Licheng Zhang, Zhendong Mao, Quan Wang, Hongtao Xie, and Yongdong Zhang. 2020. Curriculum learning for natural language understanding. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6095–6104, Online. Association for Computational Linguistics.

Chengrun Yang, Xuezhi Wang, Yifeng Lu, Hanxiao Liu, Quoc V. Le, Denny Zhou, and Xinyun Chen. 2023. Large language models as optimizers.

Long Long Yu, Weisen Jiang, Han Shi, Jincheng Yu, Zhengying Liu, Yu Zhang, James T. Kwok, Zheng Li, Adrian Weller, and Weiyang Liu. 2023. Metamath: Bootstrap your own mathematical questions for large language models. *ArXiv*, abs/2309.12284.

Zheng Yuan, Hongyi Yuan, Cheng Li, Guanting Dong, Chuanqi Tan, and Chang Zhou. 2023. Scaling relationship on learning mathematical reasoning with large language models. *ArXiv*, abs/2308.01825.

Denny Zhou, Nathanael Schärli, Le Hou, Jason Wei, Nathan Scales, Xuezhi Wang, Dale Schuurmans, Claire Cui, Olivier Bousquet, Quoc Le, and Ed Chi. 2023. Least-to-most prompting enables complex reasoning in large language models.

Hattie Zhou, Azade Nova, Hugo Larochelle, Aaron Courville, Behnam Neyshabur, and Hanie Sedghi. 2022. Teaching algorithmic reasoning via in-context learning.