

Fundamental Reasoning Paradigms Induce Out-of-Domain Generalization in Language Models

Mingzi Cao^{α*} Xingwei Tan^{α*} Mahmud Elahi Akhter^β Marco Valentino^α
Maria Liakata^{β,γ} Xi Wang^α Nikolaos Aletras^α

^αUniversity of Sheffield ^β Queen Mary University of London ^γThe Alan Turing Institute
{mcao20,xingwei.tan,m.valentino,xi.wang,n.aletras}@sheffield.ac.uk
{m.akhter,m.liakata}@qmul.ac.uk

Abstract

Deduction, induction, and abduction are fundamental reasoning paradigms, core for human logical thinking. Although improving Large Language Model (LLM) reasoning has attracted significant research efforts, the extent to which the fundamental paradigms induce generalization has yet to be systematically explored. In this study, we shed light on how the interplay between these core paradigms influences LLMs’ reasoning behavior. To this end, we first collect a new dataset of reasoning trajectories from symbolic tasks, each targeting one of the three fundamental paradigms, to abstract from concrete world knowledge. Then, we investigate effective ways for inducing these skills into LLMs. We experiment with a battery of methods including simple fine-tuning, and more complex approaches to increase model depth, or transform a dense model to a mixture-of-experts. We comprehensively evaluate induced models on realistic out-of-domain tasks, that are entirely formulated in natural language and contain real-world knowledge. Our results reveal that our approach yields strong generalizability with substantial performance gains (up to 14.60) across realistic tasks.¹

1 Introduction

Deduction, induction, and abduction are fundamental reasoning paradigms, forming the cornerstone of human logical reasoning (Frankfurt, 1958; Salmon, 1994). *Deduction* infers conclusions based on rules and observations. *Induction* abstracts rules based on patterns in premises and observations, and *Abduction* infers the most plausible explanation to account for an observation, given a set of rules. These paradigms play a crucial role in understanding and developing the reasoning capabilities of large language models (LLMs) (Dougrez-Lewis et al., 2025; Tan et al., 2025; Hu et al., 2025).

* Equal contribution.

¹ Our code is available at <https://github.com/voalmciaf/FR-OOD>.

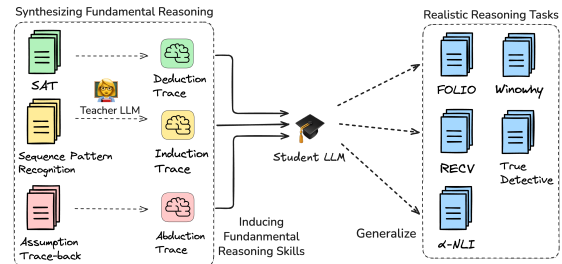


Figure 1: We induce fundamental reasoning skills with the trajectories generated over symbolic tasks, assessing generalization to realistic natural language tasks.

Previous work often conflates induction, deduction, and abduction, overlooking how they individually contribute to improving LLM reasoning (Zheng et al., 2025; Hu et al., 2025). While Tan et al. (2025) individually induce deductive skills via fine-tuning (FT) to improve OOD generalization, they neglect induction and abduction. Moreover, it is common practice to use data that entangles world knowledge and reasoning to improve LLM reasoning (Hu et al., 2025; Guha et al., 2025; DeepSeek-AI, 2025), making it unclear whether gains stem from genuine reasoning or factual recall. Consequently, two main questions remain unexplored: (1) *to what extent do distinct deduction, induction, and abduction skills improve general reasoning*; and (2) *what are the most effective methods for inducing these capabilities*?

To answer these questions, we first enrich a procedurally-generated symbolic reasoning dataset introduced by Hu et al. (2025) with reasoning trajectories from teacher LLMs. The dataset consists of tasks, each targeting one of the three fundamental paradigms while abstracting the inference from world knowledge. However, it only consists of questions and answers without associated trajectories. The trajectories serve as demonstrations for inference patterns independent of world knowledge. Subsequently, we induce student LLMs with

fundamental reasoning abilities, investigating the impact and interplay of deduction, induction, and abduction. We systematically evaluate, for the first time, a battery of inducing methods beyond standard fine-tuning, including *up-scaling* to increase model depth (Wu et al., 2024), allowing the effective absorption of new reasoning capabilities; and *upcycling* converting standard LLMs to Mixture-of-Experts (MoE) to scale model capacity while avoiding interference between existing and newly induced information (Komatsuzaki et al., 2023). Finally, we comprehensively evaluate the induced models both in-domain (i.e., on the same symbolic tasks) and Out-of-Domain (OOD) on the rest of the symbolic tasks and realistic tasks requiring real-world knowledge, across two model families (see Figure 1). Our contributions are as follows:

- We collect and release a new symbolic reasoning dataset for deduction, induction, and abduction. It contains a total of $\approx 17\text{K}$ symbolic problems enriched with more than 160K reasoning trajectories generated via two teacher LLMs from different families.
- We show that inducing smaller student LLMs with fundamental reasoning skills, decoupled from world knowledge, yields substantial performance gains across realistic OOD tasks (up to 14.60) across different inducing methods and LLM families.
- We find that deductive abilities generally offer the highest OOD performance gains across models. We also observe strong cross-paradigm improvements, which reveal how fundamental reasoning skills in isolation benefit LLM reasoning as a whole. On the other hand, the most effective inducing method is model-dependent, with upcycling achieving the best overall results.

2 Related Work

Eliciting Reasoning in LLMs Large reasoning models integrate step-by-step reasoning during post-training (Olmo et al., 2025). Recent work explores how fundamental reasoning paradigms impact the effectiveness of LLM reasoning. Zheng et al. (2025) evaluate different logical inference paradigms, but they only comparing inductive reasoning against the combination of abduction and deduction, lacking understanding of individual fundamental reasoning paradigms. Hu et al. (2025)

separately specialize LLMs on deductive, inductive, and abductive reasoning, and subsequently merge the resulting models, conflating all reasoning paradigms together. This makes it difficult to isolate how each individual fundamental reasoning paradigm impacts performance. Therefore, we learn each specific fundamental reasoning paradigm individually via training on reasoning trajectories collected by teacher LLMs. This approach provides a clearer and fine-grained understanding of their impact on LLM reasoning for the first time.

Synthetic Reasoning Trajectories LLM reasoning requires large-scale trajectories in post-training (Liu et al., 2024; Yang et al., 2025). However, human-curated reasoning data is often noisy (Morishita et al., 2024), and difficult to scale (Sun et al., 2024). Therefore, synthetic reasoning trajectories from teacher models are increasingly used to train student models (DeepSeek-AI, 2025; Olmo et al., 2025). While they offer substantial gains in general (Guha et al., 2025) and specialized domains (Yu et al., 2024; Zhao et al., 2025; Yu et al., 2025), the trajectories include world knowledge, making it difficult to isolate fundamental reasoning improvements. Tan et al. (2025) use symbolic reasoning trajectories to decouple world knowledge but focus exclusively on deductive reasoning, leaving other fundamental reasoning paradigms unexplored.

3 Fundamental Reasoning Data

Our aim is to investigate the impact of inducing LLMs with the three fundamental reasoning types: deduction, induction, and abduction. For this purpose, we need corresponding trajectories.

3.1 Symbolic Reasoning Tasks

We enrich an existing fundamental reasoning dataset developed by Hu et al. (2025). We opted for this data because it is publicly available and comprises 3,600 deductive, 4,500 inductive, and 9,000 abductive questions with their answers, but without associated reasoning trajectories that are essential to our study. The data is procedurally-generated based on three symbolic tasks (Figure 2), each serving as the sole instantiation of deduction, induction, and abduction.

Deduction. The task consists of Boolean Satisfiability (SAT) questions, requiring the truth value of the variables to satisfy a given formula. All problems are guaranteed to be satisfiable. The original

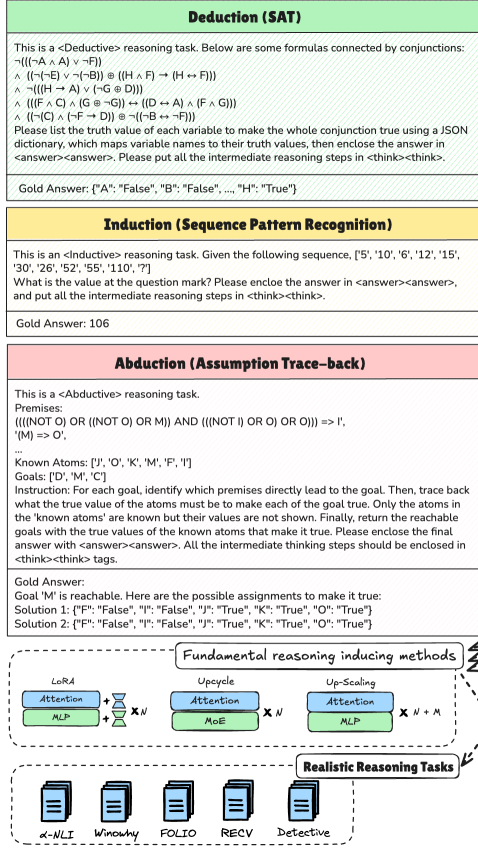


Figure 2: Symbolic reasoning tasks, and the fundamental reasoning inducing process.

questions list the formulas separately without using a Conjunctive Normal Form (CNF), i.e., a standardized way to write Boolean formulas as a conjunction of clauses, where each clause is a disjunction of variables or their negations. Our preliminary experimentation showed that this creates confusion for LLMs. Furthermore, the original answers are presented in plain text, which complicates automated evaluation. Therefore, we modify the original formulas into CNFs and express the answers in JSON for collecting reasoning trajectories.

Induction. This task requires the model to predict the next number in a given sequence of numbers, representing a basic form of inductive reasoning. The numerical sequences consist of repeating patterns that involve addition, subtraction, or multiplication. We use the original input-output pairs without any modifications.

Abduction. The task uses logic formulas and known atoms with hidden truth values. The objective is to first identify all reachable goals from the atoms by checking if formulas related to the goals exist. The model requires abductive reason-

ing to work backward from an assumed conclusion (i.e., the goals are true) to hypothesizing the most likely set of premises (antecedent truth values). However, the original task instructions do not clarify that known atoms do not necessarily equate to true atoms, leading to misinterpretations by LLMs. Moreover, the original answers only check for goal reachability despite requiring LLMs to identify related premises and trace back the truth values of the known atoms. This oversimplification causes the model to simply check whether the goals appear in the premises. To overcome these limitations, we keep the original formulas but modify the task instructions and the answers by searching the possible truth values of known atoms that make the reachable goal true using Prolog (Wielemaker et al., 2012). Table 3 (Appendix A) shows a comparison between the original and our abductive task.

3.2 Reasoning Trajectories Sampling

Data Splits. From each task type, we use 100 questions and answer pairs for testing and 100 for development, reserving the remaining for training. We collect reasoning trajectories only for the training set. A question-answer pair and its trajectories only appear in one of the splits.

Teacher Models. We sample reasoning trajectories using two widely used open-weight LLMs: *Qwen3-30B-Instruct* (Yang et al., 2025) and *Llama-3.3-70B-Instruct* (Grattafiori et al., 2024).²

Sampling. Questions are fed to a teacher LLM (see Appendix A for prompts) in a zero-shot setting. For each question, we sample five responses using different random seeds. We set the maximum number of generated tokens to 10,000, ensuring the trajectories reach the conclusion. We exclude responses with fewer than 20 words since such short texts are unlikely to contain valid reasoning trajectories. For the Qwen model, this results in filtering out 554 deductive, 2,430 inductive, and 19,872 abductive trajectories. However, no trajectories from the Llama model were filtered out. We find that the filtered responses of Qwen are either empty or strings including random tokens, perhaps generated due to confusion. Due to the fact that automatically verifying the logic consistency of textual reasoning trajectories remains an open problem and LLMs learn more from structure instead of content (Li et al., 2025a), we do not filter out more trajectories.

²We use the vLLM API (Kwon et al., 2023), ensuring efficient inference.

Teacher	Type	# Quest.	# Traject.	# Tokens	Avg. Tokens
Llama3.3-70B-IT	De.	3,400	17,000	18.6M	1,093
	In.	4,300	21,500	11.7M	544
	Ab.	8,800	44,000	51.2M	1,164
Qwen3-30B-IT	De.	3,400	16,446	23.0M	1,397
	In.	4,300	19,070	47.8M	2,507
	Ab.	8,800	24,128	135.9M	5,631

Table 1: Fundamental reasoning training data statistics.

The final training set comprises $\sim 16,000$ questions together with $\sim 82,000$ and $\sim 60,000$ trajectories generated by Llama and Qwen teacher models, respectively. Table 1 shows detailed statistics. We note that Qwen generates longer trajectories than Llama, particularly on inductive and abductive tasks (four times longer on average). We further observe that Qwen tends to repeat formulas and sequences from the questions in its reasoning steps. It also frequently uses tables to summarize its responses and generates many special tokens such as “<tool_call>”, “<IActionResult>”, and “<pal>”. This is likely due to post-training on tool calling.

4 Inducing Fundamental Reasoning Skills

We experiment with various inducing methods, from standard FT to more sophisticated approaches that introduce structural changes to the model.

Full FT. First, we simply update all model parameters via next-token prediction on the collected reasoning trajectories, similar to Tan et al. (2025).

LoRA FT. As an efficient alternative, we also experiment with Low-Rank Adaptation (Hu et al., 2022, LoRA) FT. Instead of updating an existing weight matrix $W \in \mathbb{R}^{d \times k}$, LoRA injects a low-rank decomposition $\Delta W = BA$, where $A \in \mathbb{R}^{r \times k}$ and $B \in \mathbb{R}^{d \times r}$, updating only A and B .

Up-scaling. LoRA’s low-rank constraint while efficient, limits its effectiveness in domain adaptation (Biderman et al., 2024; Pletenev et al., 2025). Model up-scaling (Wu et al., 2024; Cao et al., 2025) increases model capacity by inserting additional layers into a pre-trained LLM. By updating only the new layers, up-scaling preserves original parametric knowledge. It also allows better acquisition of new knowledge for language (Choudhury et al., 2025; Lee et al., 2025) and domain adaptation (Hoffmann et al., 2025; Lim et al., 2025). We up-scale models with LLaMAPro (Wu et al., 2024) by dividing base models into multiple layer groups and insert new layers atop each group. The new layers are initialized by copying weights from the

corresponding base layers, with the attention and MLP output projection matrices set to zero.

Upcycling. Upcycling (Komatsuzaki et al., 2023; Nakamura et al., 2025) transforms a dense Transformer layer to a MoE layer by expanding the MLP block into multiple expert MLPs. During inference, a router chooses experts. It has been successfully used for multilingual (Zhou et al., 2025; Zhang et al., 2025) and domain adaptation (Li et al., 2025b). Compared to the depth expansion of up-scaling, upcycling is an efficient width expansion approach. We use Sparse-Upcycling (Komatsuzaki et al., 2023) as it does not require additional training for the router (Gritsch et al., 2025), or extra modules beyond the MoE structure (Zhou et al., 2025). Sparse-Upcycling initializes new expert MLPs by copying the original pre-trained MLP weights, with the router randomly initialized.

5 Experimental Setup

5.1 Training Settings

To study how each of the fundamental reasoning paradigms influence reasoning behavior, we train LLMs separately using *Deduction*, *Induction*, and *Abduction* reasoning trajectories respectively. For comparison, we also induce LLMs with all three reasoning types together (i.e., *Mix*) for assessing whether they offer complementary inductive bias.

5.2 Models & Baselines

We use two student models: *Llama-3.1-8B-Instruct* (Grattafiori et al., 2024) with 32 layers; and *Qwen3-8B* (Yang et al., 2025) with 36 layers.

We compare inducing methods to (1) student models before FT (*vanilla*); and (2) *teacher models* used for collecting the trajectories (§3.2).

5.3 Implementation Details

For fair comparison, we keep the parameters of all induced models at a similar size, updating only the newly-added parameters, except of Full FT. We apply LoRA to all linear layers of the students, with ranks set to 1,024 for *Llama-3.1-8B-Instruct* and 512 for *Qwen3-8B*. For up-scaling, we divide each student into four layer groups, inserting three new layers into each group. We upcycle odd-numbered dense layers into MoE layers by adding a single expert MLP per layer, following Fedus et al. (2022). Further details are in Appendix B.

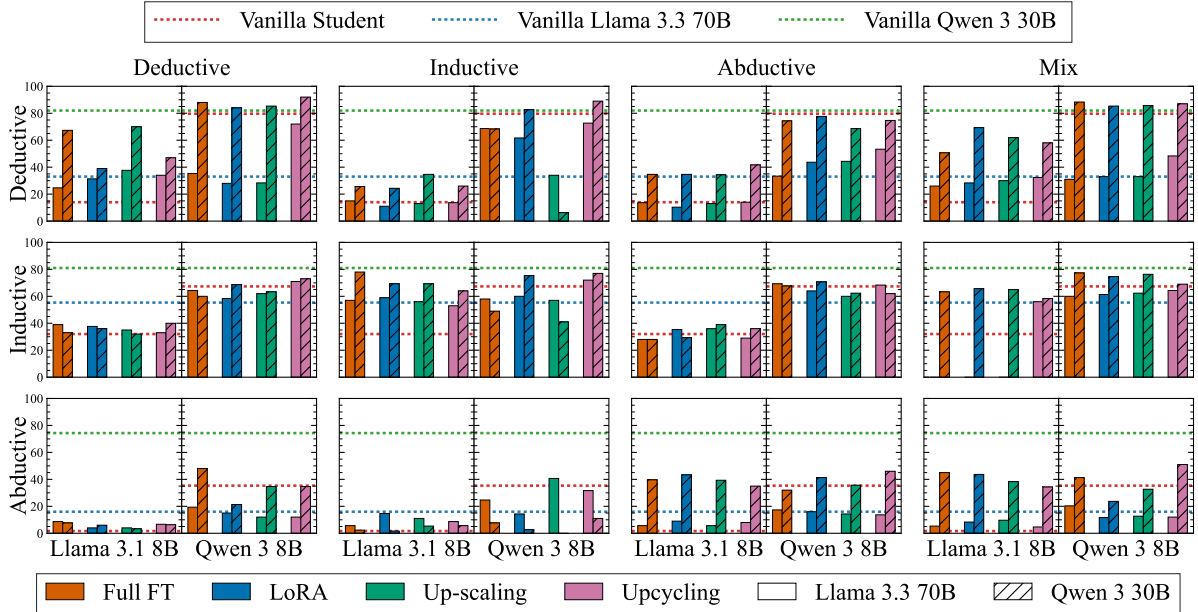


Figure 3: Symbolic task accuracy of induced models across different settings. X-axis and y-axis denote inducing paradigms and symbolic test tasks, respectively. Solid/stripped bars denote teacher for generating trajectories. Dotted lines denote performance for the vanilla student and teacher models.

5.4 Evaluation

Symbolic In-Domain. We evaluate on the test set (§3.2) that corresponds to the reasoning type used for training (e.g., training and testing on inductive tasks). When training on the Mix dataset, the test sets of all three reasoning types are in-domain.

Symbolic OOD. To test symbolic generalization, we evaluate on symbolic tasks not seen during training (e.g., train on abductive, test on deductive).

Realistic OOD. We include the following realistic OOD datasets: *True Detective* (Del and Fishel, 2023) contains detective puzzles with multi-choice questions; *α NLI* (Zhao et al., 2023) consists of multi-choice questions of the most probable explanation given a context; *WinoWhy* (Zhang et al., 2020) includes explanations for the questions in Winograd Challenge (Levesque et al., 2012), aiming to distinguish plausible reasons from similar but wrong reasons; *FOLIO* (Han et al., 2024) contains logical reasoning questions in natural language. Given the context, a set of first-order logic formulas, the model predicts if a statement is true, false, or uncertain; *RECV* (Dougrez-Lewis et al., 2025) contains claims from Twitter and Wikipedia, for determining whether they are true or false, requiring deductive or abductive reasoning.

Metrics. We report the accuracy of the final answer, judged by *Qwen3-30B-Instruct*. We use this

judge based on recent analysis by (Haldar and Hockenmaier, 2025), which indicates it is more robust than competitive alternatives (including Llama-3 and Deepseek-R1 (DeepSeek-AI, 2025)) in terms of coherence, consistency, and relevance. We further validate its reliability in Appendix B.

6 Results

Figures 3 and 4 show model performance on symbolic and realistic tasks respectively. Tables 5 and 6 (Appendix D) show full results and std deviations.

6.1 Symbolic Tasks

In-Domain: Specialized training yields the highest gains. The subplots on the main-diagonal and the last column in Figure 3 represent in-domain settings, where the train and test paradigms match. In five of the six scenarios (2 student models \times 3 symbolic tasks), specialist induced models substantially outperform the vanilla student. We find that inducing *Deduction* yields the largest improvements (e.g., *Llama-3.1-8B* gains 56.00 via up-scaling, while *Qwen-3-8B* is up 12.33 via upcycling). *Induction* follows (46.00 via Full FT, 9.67 via Up-cycling), while inducing *Abduction* produces the smallest comparative gains (41.66 via LoRA, 10.67 via Upcycling). One possible explanation of the larger deduction gains is that the transformer architecture has the right structure for deduction (Aggarwal et al., 2025). Notably, the *Mix* setting, which

trains LLMs on a mixture of all three paradigms, offers a competitive inducing strategy. While the *Mix* setting does not yield the largest in-domain performance gains achieved by single-paradigm specialists, it still outperforms the vanilla *Qwen-3-8B* on all symbolic tasks in six out of eight cases (2 teachers \times 4 inducing methods).

OOD: Asymmetric transfer between paradigms.

Turning to the other subplots in Figure 3, we analyze OOD generalization, where models trained on one paradigm are tested on other. We observe a distinct order of effectiveness: *Deduction* > *Induction* > *Abduction*. Gains in deduction transfer strongly to other tasks. Across 16 symbolic cases per paradigm, deduction gains transfer in 12 cases, improving induction and abduction in nine cases each. In contrast, induction transfers moderately (co-occurring with gains in 7 cases), while abduction transfers poorly, benefiting deduction in only four cases and induction in five cases. For instance, when inducing *Qwen-3-8B* via upcycling, deduction yields the most transferable gains (induction: 73.00, abduction: 34.67). Induction achieves moderate transfer, exhibiting asymmetric generalization (deduction: 89.00, abduction: 11.00). In contrast, inducing abduction does not provide comparable gains (deduction: 74.67, induction: 62.00). These results show the transfer between fundamental reasoning paradigms is asymmetric. While abduction is the hardest to learn (lowest in-domain gains), deduction is the most transferable skill to other reasoning paradigms under our task instantiations.

Inducing methods: Capacity expansion vs. efficient adaptation.

Looking at different inducing methods, we observe that the most effective method is model-dependent, though upcycling generally offers more robust gains. For *Qwen3-8B*, upcycling consistently offers the highest gains across paradigms (e.g., 12.33 in deduction, 9.67 in induction and 10.67 in abduction). In contrast, *Llama-3.1-8B* benefits from a diverse range of inducing methods. Up-scaling drives its peak gains in deduction (56.00) and Full FT dominates in induction (46.00). While LoRA is a competitive efficient alternative, surprisingly achieving the best abduction improvement (41.66), it generally lags capacity-expanding methods (upcycling), suggesting that fundamental reasoning skills benefit from the additional capacity offered by structural adaptation.

Stronger teachers lead to higher student performance.

Qwen-3-30B substantially outperforms *Llama-3.3-70B* on symbolic tasks (deduction: 82.00 vs. 33.00, induction: 81.00 vs. 55.33, abduction: 74.33 vs. 16.00). Both student LLMs benefit more from the stronger Qwen teacher for symbolic tasks. For example, *Qwen3-8B*, using the Llama teacher, causes performance drops in most cases (43/48). When shifting to Qwen teacher, its deduction performance substantially improves (up to 92.00 via deduction upcycling) even surpassing the teacher model (82.00). Such improvements on the deduction task are observed across different inducing paradigms (except abduction), further supporting the asymmetric transfer between paradigms. The performance difference between teachers highlights that the quality of reasoning traces is crucial for symbolic generalization.

6.2 Realistic Tasks

Transferability: Deduction > Induction > Abduction.

We observe that the order of effectiveness established in the symbolic experiments (§6.1) persists in realistic tasks. As shown in Figure 4, inducing deduction offers the strongest realistic OOD performance. Both *Qwen-3-8B* and *Llama-3.1-8B* achieve the highest average performance when inducing deduction (70.14 via upcycling, 58.70 via up-scaling) using traces from the Llama teacher. Induction is slightly lower (69.49 via upcycling, 57.58 via up-scaling), while abduction yields more modest gains (69.34 via upcycling, 56.79 via Full FT). The *Mix* strategy is worse than any single-paradigm approach (69.04 via upcycling, 56.07 via up-scaling), contrary to the symbolic performance. We hypothesize that, unlike symbolic tasks, realistic tasks are less structured, causing the *Mix* setting to dilute the strong inductive bias provided by single-paradigms, especially deduction.

Knowledge-free inducing goes beyond the teacher.

Inducing reasoning paradigms, abstracted from world knowledge, enables student LLMs to surpass their teachers even in realistic tasks. For example, when training Llama student (RECV: 71.98) with Llama teacher (RECV: 77.07), 13 out of 16 induced models (four inducing paradigms \times four inducing methods) outperform the teacher, achieving up to 81.73 via abductive Full FT. Strikingly, this improvement remains observable even when the student initially outperforms the teacher. For example, when training

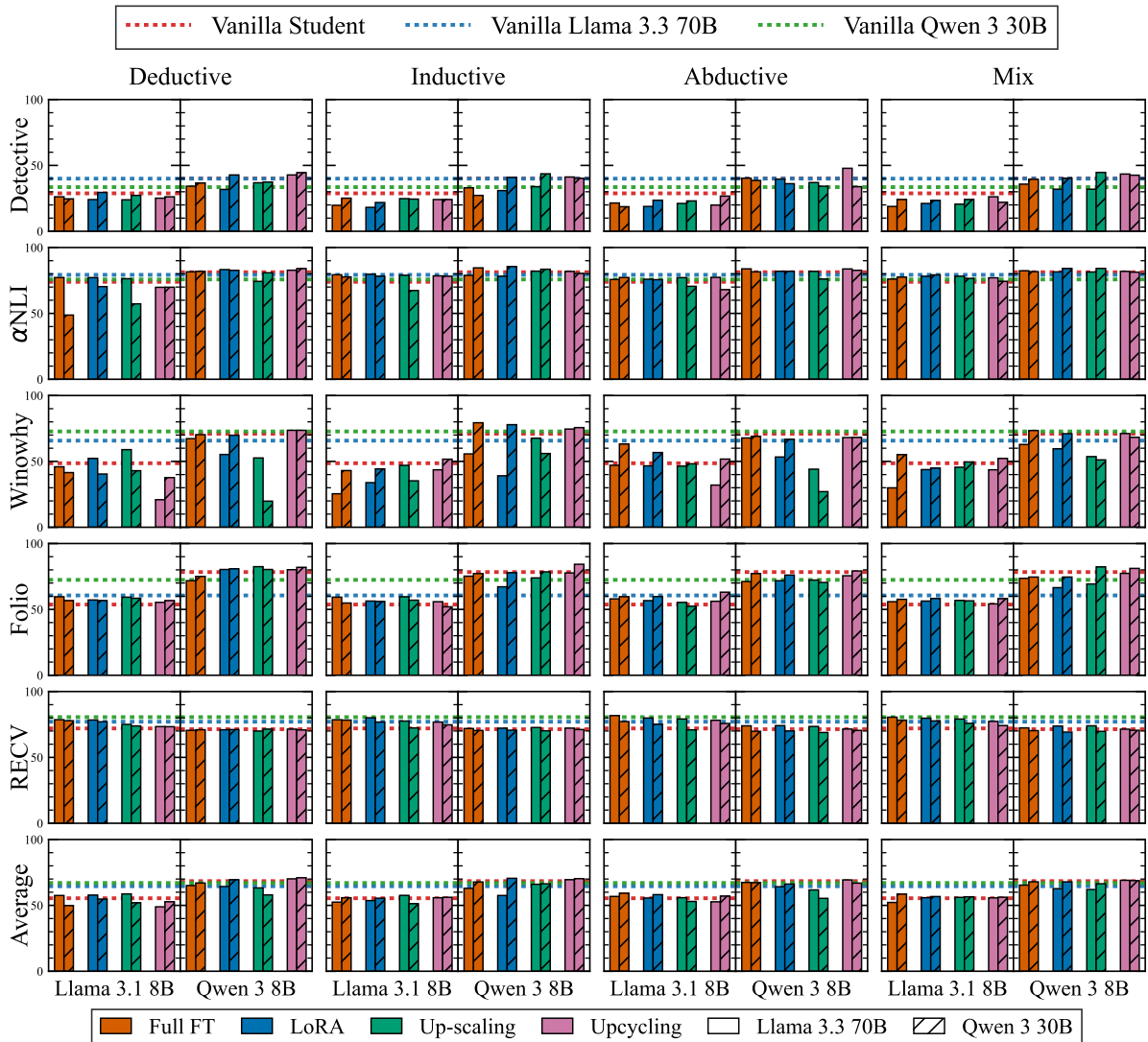


Figure 4: Realistic OOD task performance when inducing student model across different settings.

Qwen student (α NLI: 81.26) with Qwen teacher (α NLI: 75.77), the student performs better in the majority of cases (13/16), reaching up to 84.24 via inductive upcycling. Such improvements are observed across tasks under different teacher–student pairs. We attribute these gains to the knowledge-free symbolic trajectories. The teacher transfers better fundamental reasoning skills to the student through symbolic structures, avoiding the content bias of world knowledge, pervasive in real tasks.

Cross-paradigm transfer: The Abduction Paradox. Inducing the reasoning paradigm corresponding to a realistic task does not always yield best performance. This is evident in realistic tasks that rely heavily on abduction. For *True Detective*, training on deduction yields the best performance for both *Llama-3.1-8B* and *Qwen3-8B* (26.18 and 44.50). In contrast, training on abduction per-

forms the worst (21.47 and 38.57). A similar pattern emerges on α NLI, where inducing induction (79.73/85.76) consistently outperforms abduction (77.37/82.73). Since abduction is inherently difficult to learn, models benefit more from the transferable logic acquired via deduction or induction.

Task-specific concentration of gains. OOD generalization is highly concentrated on specific tasks. Compared to the vanilla model, gains are higher on α NLI (improving in 50 out of 64 cases) and Folio (41/64 cases). However, the source of these gains varies by teacher family. Llama teachers are more effective on RECV (28/32 cases), whereas Qwen teachers struggle (16/32) but perform better on Winowhy (increasing cases from 5 to 13). This shows that while the training paradigm decides OOD effectiveness, teacher models determine which tasks benefit from the transfer.

Students learn better from teachers from the same family. While symbolic tasks benefit from the *stronger* Qwen teacher regardless of model family, realistic OOD generalization improves when the teacher and student belong to the same model family, contradicting our findings in symbolic tasks (§6.1). For example, when inducing deduction for *Llama-3.1-8B*, the best performance on realistic tasks is achieved with the Llama teacher (58.70) rather than the Qwen teacher (54.79). This pattern is consistent across both student models when inducing deduction or induction. We speculate that realistic tasks are more sensitive to stylistic mismatches inherent in cross-family distillation. This aligns with prior work on cross-model trajectory mismatch (Wen et al., 2025), suggesting that, minimizing distributional shift between teacher and student is more critical than teacher capability.

Inducing Methods: Consistent with symbolic tasks. We observe a similar pattern to the symbolic domain (§6.1): although the most effective inducing method is model-dependent, upcycling provides more robust improvements overall. Across all eight settings (two teacher models \times four inducing paradigms), upcycle *Qwen-3-8B* achieves the highest realistic performance in seven out of eight cases. For example, with Llama teacher, upcycling consistently yields the best realistic performance for *Qwen-3-8B* across all paradigms (deduction: 70.14, induction: 69.49, abduction: 69.34, *Mix*: 69.04). In contrast, *Llama-3.1-8B* benefits from a diverse range of inducing methods (i.e. Full FT, LoRA, Up-scaling). This observation shows the consistency of OOD generalization across symbolic and realistic tasks when the same inducing method is applied. Furthermore, the superior performance of upcycling further illustrates that more sophisticated inducing methods outperform the current practice on using simpler Full FT, i.e., Tan et al. (2025).

6.3 Takeaways

Based on our findings, we summarize the key takeaways to answer the two research questions.

RQ1: To what extent do distinct deduction, induction, and abduction skills improve general reasoning?

- Inducing distinct fundamental reasoning skills, decouple from world knowledge, yields substantial gains across realistic OOD tasks

(Detect:+7.7, α NLI:+5.8, WinoWhy:+14.6, FOLIO:+9.4, RECV:+9.8).

- A consistent asymmetric transfer pattern emerges across both symbolic and realistic OOD tasks: Deduction > Induction > Abduction, with deduction being the most transferable and abduction the least.
- Student model can even surpass the teacher model across symbolic and realistic OOD tasks, suggesting that gains arise from induced structural reasoning biases, rather than task-specific knowledge transfer.

RQ2: What are the most effective methods for inducing these capabilities?

- Although the best inducing method is model-dependent, upcycling generally offer the best overall performance.
- Regarding symbolic OOD generalization, student model benefit more from a teacher model that exhibit stronger reasoning capabilities.
- For realistic OOD generalization, student models learn better from a teacher model within the same model family.

7 Qualitative Analysis

Table 2 shows examples of reasoning behavior using different inducing methods. In Case 1, the vanilla Llama and deductive Full FT Llama fail by prematurely resolving epistemic uncertainty, committing to a determinate conclusion through implicit assumption-making. In contrast, the deductive up-scaling Llama succeeds by preserving the uncertainty deriving from the disjunctive premise. This indicates that stronger structural constraints introduced by inducing methods support logical validity, preventing overconfident conclusions on underdetermined realistic reasoning problems.

In Case 2, the vanilla Qwen and abductive upcycling Qwen fail by prioritizing world knowledge over logical reasoning, revising implausible premises and overriding the given logical rules. Conversely, the deductive upcycling Qwen succeeds by strictly adhering to the stated premises and applying logical rules regardless of content. This behavior aligns with the intrinsic nature of abduction and deduction. Abduction focuses on exploring the most plausible observations under

Case 1	Premises: (1) People in this club who perform in school talent shows often attend and are very engaged with school events. . . . (3) People in this club who chaperone high school dances are not students who attend the school. . . . (6) Bonnie is in this club and she either (i) both attends and is very engaged with school events and is a student who attends the school, or (ii) is not someone who both attends and is very engaged with school events and is not a student who attends the school.		
Conclusion	Bonnie performs in school talent shows often.		
Model	Vanilla Llama	Deductive Full FT Llama	Deductive Up-scaling Llama
Reasoning	<i>"1. If Bonnie is a student... then she must perform in school talent shows often... 5. This means that if Bonnie is not a student... then she must chaperone high school dances, which contradicts the fact that she is not a student"</i>	<i>"...we cannot definitively say if she performs in school talent shows often.....we have a premise that supports this conclusion if she is a student, we can say that the conclusion is True if we assume Bonnie is a student..."</i>	<i>"...cannot directly conclude that Bonnie performs in school talent shows often... premise 6 does not directly state that Bonnie is a student or that she is not a student. It only gives two possible scenarios for Bonnie's situation... cannot be certain of it."</i>
Analysis	Directionality Error: Illegally reverses implication ($P \rightarrow Q$ as $Q \rightarrow P$) and hallucinates a contradiction to force a determinate conclusion.	Conditional Reasoning Drift: Recognizes uncertainty but nonetheless resolves it via assumption-dependent reasoning.	Disjunction Preservation: Correctly treats the disjunction in Premise 6 as epistemic uncertainty and maintains logical validity.
Final Answer	True	True	Uncertain
Case 2	Premises: (1) Barutin Cove is a cove named after the Bulgarian settlement of Barutin. (2) Barutin Cove is on the southwest coast of Snow Island. (3) Snow Island, Greenwich Island, and Deception Island are located in the South Shetland Islands. (4) Antarctica is located on the South Shetland Islands. (5) If place A is located in place B and place B is located in place C, then place A is located in place C.		
Conclusion	Barutin Cove is not located in Antarctica.		
Model	Vanilla Qwen	Abductive Upcycling Qwen	Deductive Upcycling Qwen
Reasoning	<i>"...the South Shetland Islands are not part of Antarctica, but rather located near Antarctica... which seems to be a contradiction..."</i>	<i>"...there's a potential confusion in the premises... Antarctica is a continent, while the South Shetland Islands are a group of islands that are geographically close to Antarctica..."</i>	<i>"...So maybe the premise is incorrect? But I have to take the premises as given, even if they might be factually wrong..."</i>
Analysis	World-Knowledge Bias: Overrides given premises using real-world knowledge, invents a spurious contradiction, and derives a formally correct answer by invalidating the logical setting.	Plausibility Reinterpretation: Treats premises as revisable observations, resolves tension via plausibility-based reinterpretation, and violates premise-faithful logical entailment.	Premise-Faithful Deduction: Explicitly distinguishes logical validity from factual correctness, enforces strict premise adherence, and rejects the conclusion through rule-consistent inference.
Final Answer	True	True	False

Table 2: Qualitative comparison of reasoning traces on realistic tasks across models before and after inducing fundamental reasoning skills. Full example for Case 1 and Case 2 are included in Appendix E.

a set of rules and thus encourages premise reinterpretation. In contrast, deduction derives conclusions from rules and observations and inherently promotes premise faithfulness. This further emphasizes how inducing fundamental reasoning paradigms that are decoupled from world knowledge can reduce content biases, leading to more robust reasoning generalization on realistic OOD tasks. We further extend our qualitative analysis to symbolic OOD tasks, with detailed examples and discussions provided in Appendix F.

8 Conclusion

We systematically studied how inducing fundamental reasoning types (i.e., deduction, induction, and

abduction) shapes the generalization of LLM reasoning. By training on reasoning trajectories abstracted from world knowledge, we isolate the effect of each paradigm. We evaluate a range of inducing methods, from standard fine-tuning to capacity expansion approaches such as up-scaling and upcycling. Our results show that knowledge-free inducing yields substantial and transferable OOD gains. Deduction is the most robust paradigm and upcycling provides the largest overall improvements, enabling student LLMs to acquire structural inductive biases that even surpass their teachers on realistic tasks. Future work will extend this framework and explore principled ways of combining multiple reasoning paradigms without diluting their individual contribution.

Limitations

All data and evaluations are in English given the availability and diversity of reasoning datasets. We acknowledge that the linguistic expression and underlying cognitive strategies for fundamental reasoning may differ across languages which is an interesting avenue for future work.

In addition, while many different symbolic tasks can be generated to represent the three fundamental reasoning paradigms, this work explored a single representative task for deduction, induction, and abduction, respectively. We focused, in fact, on the interplay between fundamental paradigms and inducing methods, prioritising depth of analysis rather than exhaustive coverage of the whole spectrum of symbolic reasoning tasks. While we found the proposed tasks to be sufficient to induce strong OOD generalisation, future work might investigate the potential impact of different symbolic data on performance and generalisation.

Acknowledgments

We thank Atsuki Yamaguchi for his valuable feedback. XT, MA, ML and NA are supported by the EPSRC [grant number EP/Y009800/1], through funding from Responsible AI UK (KP0016) as a Keystone project. We also acknowledge IT Services at the University of Sheffield and the University of Oxford Advanced Research Computing for the provision of HPC services.

References

- Naman Aggarwal, Siddhartha R. Dalal, and Vishal Misra. 2025. [The bayesian geometry of transformer attention](#). *Preprint*, arXiv:2512.22471.
- Dan Biderman, Jacob Portes, Jose Javier Gonzalez Ortiz, Mansheej Paul, Philip Greengard, Connor Jennings, Daniel King, Sam Havens, Vitaliy Chiley, Jonathan Frankle, Cody Blakeney, and John Patrick Cunningham. 2024. [LoRA learns less and forgets less](#). *Transactions on Machine Learning Research*. Featured Certification.
- Mingzi Cao, Xi Wang, and Nikolaos Aletras. 2025. [Progressive depth up-scaling via optimal transport](#). *Preprint*, arXiv:2508.08011.
- Monojit Choudhury, Shivam Chauhan, Rocktim Jyoti Das, Dhruv Sahnan, Xudong Han, Haonan Li, Aaryamonvikram Singh, Alok Anil Jadhav, Utkarsh Agarwal, Mukund Choudhary, Debopriyo Banerjee, Fajri Koto, Junaid Bhat, Awantika Shukla, Samujjwal Ghosh, Samta Kamboj, Onkar Pandit, Lalit Pradhan, Rahul Pal, and 11 others. 2025. [Llama-3-nanda-10b-chat: An open generative large language model for hindi](#). *Preprint*, arXiv:2504.06011.
- Gheorghe Comanici, Eric Bieber, Mike Schaeckermann, Ice Pasupat, Noveen Sachdeva, Inderjit Dhillon, Marcel Blistein, Ori Ram, Dan Zhang, Evan Rosen, Luke Marris, Sam Petulla, Colin Gaffney, Asaf Aharoni, Nathan Lintz, Tiago Cardal Pais, Henrik Jacobsson, Idan Szpektor, Nan-Jiang Jiang, and 3416 others. 2025. [Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities](#). *Preprint*, arXiv:2507.06261.
- Tri Dao. 2024. [FlashAttention-2: Faster attention with better parallelism and work partitioning](#). In *The Twelfth International Conference on Learning Representations*.
- DeepSeek-AI. 2025. [Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning](#). *Preprint*, arXiv:2501.12948.
- Maksym Del and Mark Fishel. 2023. [True detective: A deep abductive reasoning benchmark undoable for GPT-3 and challenging for GPT-4](#). In *Proceedings of the 12th Joint Conference on Lexical and Computational Semantics (*SEM 2023)*, pages 314–322, Toronto, Canada. Association for Computational Linguistics.
- John Dougrez-Lewis, Mahmud Elahi Akhter, Federico Ruggeri, Sebastian Löbbers, Yulan He, and Maria Liakata. 2025. [Assessing the reasoning capabilities of LLMs in the context of evidence-based claim verification](#). In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 20604–20628, Vienna, Austria. Association for Computational Linguistics.
- William Fedus, Barret Zoph, and Noam Shazeer. 2022. [Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity](#). *Journal of Machine Learning Research*, 23(120):1–39.
- Harry G Frankfurt. 1958. Peirce’s notion of abduction. *The Journal of Philosophy*, 55(14):593–597.
- Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, Amy Yang, Angela Fan, Anirudh Goyal, Anthony Hartshorn, Aobo Yang, Archi Mitra, Archie Sravankumar, Artem Korenev, Arthur Hinsvark, and 542 others. 2024. [The llama 3 herd of models](#). *Preprint*, arXiv:2407.21783.
- Nikolas Gritsch, Qizhen Zhang, Acyr Locatelli, Sara Hooker, and Ahmet Üstün. 2025. [Nexus: Adaptive upcycling to efficiently pretrain mixture of experts](#). In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 24364–24381, Suzhou, China. Association for Computational Linguistics.

- Etash Guha, Ryan Marten, Sedrick Keh, Negin Raoof, Georgios Smyrnis, Hritik Bansal, Marianna Nezhurina, Jean Mercat, Trung Vu, Zayne Sprague, Ashima Suvarna, Benjamin Feuer, Liangyu Chen, Zaid Khan, Eric Frankel, Sachin Grover, Caroline Choi, Niklas Muennighoff, Shiye Su, and 31 others. 2025. [Openthoughts: Data recipes for reasoning models](#). *Preprint*, arXiv:2506.04178.
- Rajarshi Haldar and Julia Hockenmaier. 2025. [Rating roulette: Self-inconsistency in LLM-as-a-judge frameworks](#). In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 24986–25004, Suzhou, China. Association for Computational Linguistics.
- Simeng Han, Hailey Schoelkopf, Yilun Zhao, Zhenqing Qi, Martin Riddell, Wenfei Zhou, James Coady, David Peng, Yujie Qiao, Luke Benson, Lucy Sun, Alexander Wardle-Solano, Hannah Szabó, Ekaterina Zubova, Matthew Burtell, Jonathan Fan, Yixin Liu, Brian Wong, Malcolm Sailor, and 16 others. 2024. [FOLIO: Natural language reasoning with first-order logic](#). In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 22017–22031, Miami, Florida, USA. Association for Computational Linguistics.
- Michael Hoffmann, Jophin John, Stefan Schweter, Gokul Ramakrishnan, Hoi-Fong Mak, Alice Zhang, Dmitry Gaynullin, and Nicolay J. Hammer. 2025. [Llama-genba-10b: A trilingual large language model for german, english and bavarian](#). *Preprint*, arXiv:2509.05668.
- Edward J Hu, yelong shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2022. [LoRA: Low-rank adaptation of large language models](#). In *International Conference on Learning Representations*.
- Zhiyuan Hu, Yibo Wang, Hanze Dong, Yuhui Xu, Amrita Saha, Caiming Xiong, Bryan Hooi, and Junnan Li. 2025. [Beyond 'aha!': Toward systematic meta-abilities alignment in large reasoning models](#). *Preprint*, arXiv:2505.10554.
- Aran Komatsuzaki, Joan Puigcerver, James Lee-Thorp, Carlos Riquelme Ruiz, Basil Mustafa, Joshua Ainslie, Yi Tay, Mostafa Dehghani, and Neil Houlsby. 2023. [Sparse upcycling: Training mixture-of-experts from dense checkpoints](#). In *The Eleventh International Conference on Learning Representations*.
- Woosuk Kwon, Zhuohan Li, Siyuan Zhuang, Ying Sheng, Lianmin Zheng, Cody Hao Yu, Joseph Gonzalez, Hao Zhang, and Ion Stoica. 2023. [Efficient memory management for large language model serving with pagedattention](#). In *Proceedings of the 29th Symposium on Operating Systems Principles, SOSP '23*, page 611–626, New York, NY, USA. Association for Computing Machinery.
- Wooseong Lee, Minseo Kim, Taeil Hur, Gyeong Hwan Jang, Woncheol Lee, Maro Na, and Taek Kim. 2025. [ENGINius: A bilingual LLM optimized for plant construction engineering](#). In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 6: Industry Track)*, pages 1350–1364, Vienna, Austria. Association for Computational Linguistics.
- Hector J. Levesque, Ernest Davis, and Leora Morgenstern. 2012. The winograd schema challenge. In *Proceedings of the Thirteenth International Conference on Principles of Knowledge Representation and Reasoning, KR'12*, page 552–561. AAAI Press.
- Dacheng Li, Shiyi Cao, Tyler Griggs, Shu Liu, Xiangxi Mo, Eric Tang, Sumanth Hegde, Kourosh Hakhmaneshi, Shishir G Patil, Matei Zaharia, Joseph E. Gonzalez, and Ion Stoica. 2025a. [Language models can easily learn to reason from demonstrations](#). In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 15979–15997, Suzhou, China. Association for Computational Linguistics.
- Junzhuo Li, Bo Wang, Xiuze Zhou, and Xuming Hu. 2025b. [Dynamic expert specialization: Towards catastrophic forgetting-free multi-domain MoE adaptation](#). In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 18478–18493, Suzhou, China. Association for Computational Linguistics.
- Junghwan Lim, Gangwon Jo, Sungmin Lee, Jiyoung Park, Dongseok Kim, Jihwan Kim, Junhyeok Lee, Wai Ting Cheung, Dahye Choi, Kibong Choi, Jaeyeon Huh, Beomgyu Kim, Jangwoong Kim, Taehyun Kim, Haesol Lee, Jeesoo Lee, Dongpin Oh, Changseok Song, and Daewon Suh. 2025. [Expanding foundational language capabilities in open-source llms through a korean case study](#). *Preprint*, arXiv:2509.03972.
- Yang Liu, Jiahuan Cao, Chongyu Liu, Kai Ding, and Lianwen Jin. 2024. [Datasets for large language models: A comprehensive survey](#). *Preprint*, arXiv:2402.18041.
- Terufumi Morishita, Gaku Morio, Atsuki Yamaguchi, and Yasuhiro Sogawa. 2024. [Enhancing reasoning capabilities of llms via principled synthetic logic corpus](#). In *Advances in Neural Information Processing Systems*, volume 37, pages 73572–73604. Curran Associates, Inc.
- Taishi Nakamura, Takuya Akiba, Kazuki Fujii, Yusuke Oda, Rio Yokota, and Jun Suzuki. 2025. [Drop-upcycling: Training sparse mixture of experts with partial re-initialization](#). In *The Thirteenth International Conference on Learning Representations*.
- Team Olmo, :, Allyson Ettinger, Amanda Bertsch, Bailey Kuehl, David Graham, David Heineman, Dirk Groeneveld, Faeze Brahman, Finbarr Timbers, Hamish Ivison, Jacob Morrison, Jake Poznanski, Kyle Lo, Luca Soldaini, Matt Jordan, Mayee Chen, Michael Noukhovitch, Nathan Lambert, and 50 others. 2025. [Olmo 3](#). *Preprint*, arXiv:2512.13961.

- Sergey Pletenev, Maria Marina, Daniil Moskovskiy, Vasily Konovalov, Pavel Braslavski, Alexander Panchenko, and Mikhail Salnikov. 2025. [How much knowledge can you pack into a LoRA adapter without harming LLM?](#) In *Findings of the Association for Computational Linguistics: NAACL 2025*, pages 4309–4322, Albuquerque, New Mexico. Association for Computational Linguistics.
- Merrilee H Salmon. 1994. *Introduction to logic and critical thinking*. Wadsworth Publishing Co Inc.
- Yikun Sun, Zhen Wan, Nobuhiro Ueda, Sakiko Yahata, Fei Cheng, Chenhui Chu, and Sadao Kurohashi. 2024. [Rapidly developing high-quality instruction data and evaluation benchmark for large language models with minimal human effort: A case study on Japanese](#). In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 13537–13547, Torino, Italia. ELRA and ICCL.
- Xingwei Tan, Marco Valentino, Mahmud Elahi Akhter, Maria Liakata, and Nikolaos Aletras. 2025. [Enhancing logical reasoning in language models via symbolically-guided Monte Carlo process supervision](#). In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 31874–31888, Suzhou, China. Association for Computational Linguistics.
- Leandro von Werra, Younes Belkada, Lewis Tunstall, Edward Beeching, Tristan Thrush, Nathan Lambert, Shengyi Huang, Kashif Rasul, and Quentin Galouédec. 2020. [Trl: Transformer reinforcement learning](https://github.com/huggingface/trl). <https://github.com/huggingface/trl>.
- Peiyi Wang, Lei Li, Liang Chen, Zefan Cai, Dawei Zhu, Binghuai Lin, Yunbo Cao, Lingpeng Kong, Qi Liu, Tianyu Liu, and Zhifang Sui. 2024. [Large language models are not fair evaluators](#). In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 9440–9450, Bangkok, Thailand. Association for Computational Linguistics.
- Xiangyu Wen, Junhua Huang, Zeju Li, Min Li, Jianyuan Zhong, Zhijian Xu, Mingxuan Yuan, Yongxiang Huang, and Qiang Xu. 2025. [Reasoning scaffolding: Distilling the flow of thought from llms](#). *Preprint*, arXiv:2509.23619.
- Jan Wielemaker, Tom Schrijvers, Markus Triska, and Torbjörn Lager. 2012. SWI-Prolog. *Theory and Practice of Logic Programming*, 12(1-2):67–96.
- Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, and 3 others. 2020. [Transformers: State-of-the-art natural language processing](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.
- Chengyue Wu, Yukang Gan, Yixiao Ge, Zeyu Lu, Jiahao Wang, Ye Feng, Ying Shan, and Ping Luo. 2024. [LLaMA pro: Progressive LLaMA with block expansion](#). In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6518–6537, Bangkok, Thailand. Association for Computational Linguistics.
- An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, Chuji Zhang, Dayiheng Liu, Fan Zhou, Fei Huang, Feng Hu, Hao Ge, Haoran Wei, Huan Lin, Jialong Tang, and 41 others. 2025. [Qwen3 technical report](#). *Preprint*, arXiv:2505.09388.
- Hongzhou Yu, Tianhao Cheng, Yingwen Wang, Wen He, Qing Wang, Ying Cheng, Yuejie Zhang, Rui Feng, and Xiaobo Zhang. 2025. [Finemedlm-01: Enhancing medical knowledge reasoning ability of llm from supervised fine-tuning to test-time training](#). *Preprint*, arXiv:2501.09213.
- Longhui Yu, Weisen Jiang, Han Shi, Jincheng YU, Zhengying Liu, Yu Zhang, James Kwok, Zhenguo Li, Adrian Weller, and Weiyang Liu. 2024. [Metamath: Bootstrap your own mathematical questions for large language models](#). In *The Twelfth International Conference on Learning Representations*.
- Hongming Zhang, Xinran Zhao, and Yangqiu Song. 2020. [WinoWhy: A deep diagnosis of essential commonsense knowledge for answering Winograd schema challenge](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5736–5745, Online. Association for Computational Linguistics.
- Xue Zhang, Yunlong Liang, Fandong Meng, Songming Zhang, Yufeng Chen, Jinan Xu, and Jie Zhou. 2025. [Less, but better: Efficient multilingual expansion for LLMs via layer-wise mixture-of-experts](#). In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 17948–17963, Vienna, Austria. Association for Computational Linguistics.
- Wenting Zhao, Justin Chiu, Claire Cardie, and Alexander Rush. 2023. [Abductive commonsense reasoning exploiting mutually exclusive explanations](#). In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 14883–14896, Toronto, Canada. Association for Computational Linguistics.
- Xiaoke Zhao, Zhaowen Zhou, Lin Chen, Lihong Wang, Zhiyi Huang, Kaiyuan Zheng, Yanjun Zheng, Xiyang Du, Longfei Liao, Jiawei Liu, Xiang Qi, Bo Zhang, Peng Zhang, Wei Wang, and Zhe Li. 2025. [Agentar-deepfinance-100k: A large-scale financial dataset via systematic chain-of-thought synthesis optimization](#). *Preprint*, arXiv:2507.12901.

Tianshi Zheng, Cheng Jiayang, Chunyang Li, Haochen Shi, Zihao Wang, Jiaxin Bai, Yangqiu Song, Ginny Wong, and Simon See. 2025. [LogiDynamics: Unraveling the dynamics of inductive, abductive and deductive logical inferences in LLM reasoning](#). In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 20721–20742, Suzhou, China. Association for Computational Linguistics.

Hao Zhou, Zhijun Wang, Shujian Huang, Xin Huang, Xue Han, Junlan Feng, Chao Deng, Weihua Luo, and Jiajun Chen. 2025. [Moe-lpr: multilingual extension of large language models through mixture-of-experts with language priors routing](#). In *Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence and Thirty-Seventh Conference on Innovative Applications of Artificial Intelligence and Fifteenth Symposium on Educational Advances in Artificial Intelligence, AAAI'25/IAAI'25/EAAI'25*. AAAI Press.

A Reasoning Task Examples

The following boxes present the questions and corresponding gold answers used in our synthetic fundamental reasoning tasks.

Deductive Task

This is a <Deductive> reasoning task. Below are some formulas connected by conjunctions:

$$\begin{aligned} &\neg(((\neg A \wedge A) \vee \neg(F))) \\ &\wedge((\neg(\neg E) \vee \neg(\neg B)) \oplus ((H \wedge F) \rightarrow (H \leftrightarrow F))) \\ &\wedge \neg(((H \rightarrow A) \vee (\neg G \oplus D))) \\ &\wedge(((F \wedge C) \wedge (G \oplus \neg G)) \leftrightarrow ((D \leftrightarrow A) \wedge (F \wedge G))) \\ &\wedge((\neg(C) \wedge (\neg F \rightarrow D)) \oplus \neg(\neg B \leftrightarrow \neg F)) \end{aligned}$$

Please list the truth value of each variable to make the whole conjunction true using a JSON dictionary, which maps variable names to their truth values, then enclose the answer in <answer><answer>. Please put all the intermediate reasoning steps in <think><think>.

Gold Answer:

```
{ "A": "False", "B": "False",
  "C": "True",
  "D": "False",
  "E": "False",
  "F": "True",
  "G": "True",
  "H": "True"
}
```

Inductive Task

This is a <Inductive> reasoning task. Given the following sequence, [‘5’, ‘10’, ‘6’, ‘12’, ‘15’, ‘30’, ‘26’, ‘52’, ‘55’, ‘110’, ‘?’]

What is the value at the question mark? Please enclose the answer in <answer><answer>, and put all the intermediate reasoning steps in <think><think>.

Gold Answer: 106

Rev. Rule-Graph Search (Hu et al., 2025)	Assumption Trace-Back (Ours)
<p>Premises: [‘(L => L’, ‘(((NOT D) OR (NOT M)) => N’, ‘(M OR M) => C’, ‘(M OR L) => L’, ‘(M) => M’, ‘(L OR L) => G’]</p> <p>Known Atoms: [‘L’, ‘M’, ‘L’, ‘D’, ‘N’]</p> <p>Goals: [‘L’, ‘KM’, ‘NK’]</p> <p>Instruction: For each goal, first assume it is true and determine what conditions must be satisfied. Identify which premises directly lead to these goals, then trace back what must be established to satisfy those premises. Finally, check whether the known atoms can lead to these conditions</p>	<p>Premises: [‘(L => L’, ‘(((NOT D) OR (NOT M)) => N’, ‘(M OR M) => C’, ‘(M OR L) => B’, ‘(M) => M’, ‘(L OR B) => G’]</p> <p>Known Atoms: [‘L’, ‘M’, ‘A’, ‘D’, ‘N’]</p> <p>Goals: [‘B’, ‘A’, ‘C’]</p> <p>Instruction: For each goal, identify which premises directly lead to the goal. Then, trace back what the true value of the atoms must be to make each of the goal true. Only the atoms in the ‘known atoms’ are known but their values are not shown. Finally, return the reachable goals with the true values of the known atoms that make it true. Please enclose the final answer with <answer><answer>. All the intermediate thinking steps should be enclosed in <think><think> tags.</p>
<p>Answer: (1) B is reachable (2) KM is unreachable (3) NK is unreachable</p>	<p>Answer: Goal ‘B’ is reachable. Sol 1: {‘D’: false, ‘L’: false, ‘M’: false} Sol 2: {‘D’: false, ‘L’: false, ‘M’: true} ... Sol 8: {‘D’: true, ‘L’: true, ‘M’: true}</p>

Table 3: Comparison between the original abductive task by Hu et al. (2025) and ours.

B Implementation Details

Data Sampling. The generation of training data takes approximately 280 hours for LLaMA trajectories and 304 hours for Qwen trajectories when executed on an AMD MI300X accelerator.

Training and Evaluation Setup For training all methods, we use a cosine scheduler with warm-up and a maximum learning rate of $5e-6$ (degraded to $5e-7$). We fine-tune models using SFT trainer from package TRL (von Werra et al., 2020) on one NVIDIA A100 (80GB). We employ FlashAttention 2 (Dao, 2024) and mixed-precision bf16 for accelerated training. We use the package Transformers (Wolf et al., 2020) for reproducible output generation and vLLM (Kwon et al., 2023) for evaluation.

Format-Robust Evaluation. Due to requirements (§ 3.1), the final answer is explicitly separated from the reasoning chain, and Deduction and Abduction tasks need JSON-formatted answers. Induced models sometimes deviate from the required format while still producing correct answers.

To address this issue, we use *Qwen3-30B-Instruct* (Yang et al., 2025) to compare the generated answer with the gold answer. Specifically, the model is instructed to determine whether the generation is consistent with the gold label and to return *TRUE* or *FALSE*. The full prompt is provided in Appendix C.

Reliability of LLM-as-Judge. One concern is the choice of LLM judge, as inconsistencies across LLM judges may affect evaluation reliability (Halдар and Hockenmaier, 2025). Additionally, using *Qwen3-30B-Instruct* to evaluate Qwen-based student models may introduce bias due to their shared model family (Wang et al., 2024).

Thus, we further use *Gemini-2.5-Flash* (Comanici et al., 2025), a closed-source flagship LLM with strong reasoning capabilities as the judge for validating the robustness of our evaluation. We randomly sample 30 instances per task (15 shared questions for Llama/Qwen students) to form a validation set. We first report consistency (i.e., whether two judges decide the same), and further quantify inter-judge agreement using Cohen’s kappa, shown in Table 4.

Subset	Consistency	Cohen’s κ
Qwen Student	89.17	77.07
Llama Student	90.00	79.96
Total Instances	89.58	78.98

Table 4: Consistency and Cohen’s κ between two LLM judges on the validation set. “Qwen/Llama Student” denote the source LLM for output generation.

Despite differences in model family and reasoning capability, the two LLM judges agree on the vast majority of cases (89.58), indicating strong consistency. Consistency remains similarly high across subsets from different model families (Llama vs. Qwen; 89.17 vs. 90.00), suggesting that *Qwen3-30B-Instruct* exhibit no noticeable self-bias toward the Qwen studnet model. Crucially, this observation is further supported by Cohen’s kappa, which falls within the substantial agreement range (61.00–80.00) and approaches the threshold for almost perfect agreement (81.00–100.00), thereby confirming that our evaluation does not suffer from inter-judge inconsistency or model-family bias.

C Prompt Examples

Below is the prompt used to compare generated answers with gold answers.

Flexible Match

Instruction: Please check whether the generation results is consistent with the gold label.

Generation Results: {output}

Gold Label: {answer}

Please output TRUE if they are consistent, otherwise output FALSE.

D Full Results

Tables 5 and 6 present the full results of inducing reasoning paradigms using different inducing methods across experimental settings.

Method	FT	Train/Total	Symbolic Tasks				Realistic Tasks				
			Deduct	Induct	Abduct	Detect	α NLI	Winowhy	Folio	RECV	Avg
Llama-3.3-70B-IT	-	0/70.0	33.00 \pm 0.3	55.33 \pm 2.5	16.00 \pm 2.6	39.96 \pm 0.3	79.28 \pm 0.2	65.75 \pm 0.2	60.59 \pm 0.9	77.07 \pm 0.1	64.53 \pm 0.2
Llama-3.1-8B-IT	-	0/8.03	14.00 \pm 1.0	32.00 \pm 0.0	1.67 \pm 0.6	28.80 \pm 0.0	73.97 \pm 0.0	48.60 \pm 0.4	53.69 \pm 0.5	71.98 \pm 0.0	55.41 \pm 0.0
Full FT		8.03/8.03	15.00\pm1.0	57.00\pm0.0	5.67\pm0.6	19.72 \pm 0.3	79.34\pm0.1	25.44 \pm 0.1	59.28\pm0.8	78.55\pm0.0	52.47 \pm 0.1
LoRA		3.73/10.71	11.00 \pm 0.0	59.00\pm0.0	14.67\pm0.6	18.15 \pm 0.3	79.73\pm0.0	33.87 \pm 0.1	56.32\pm1.0	80.07\pm0.0	53.63 \pm 0.2
Up-Scaling	Induct	3.67/10.65	13.00 \pm 0.0	56.00\pm0.0	11.00\pm1.0	24.78\pm0.3	78.97\pm0.1	47.00\pm1.9	59.61\pm0.0	77.55\pm0.0	57.58\pm0.4
Upcycling		3.87/10.85	13.67\pm1.5	53.00\pm0.0	8.67\pm2.1	24.08\pm0.0	78.50\pm0.1	43.61\pm0.3	55.83\pm0.6	76.82\pm0.0	55.77\pm0.2
Full FT		8.03/8.03	24.67\pm0.6	39.00\pm0.0	8.67\pm1.2	26.18\pm0.0	77.24\pm0.1	45.92 \pm 0.7	59.61\pm0.9	78.60\pm0.0	57.51 \pm 0.1
LoRA		3.73/10.71	31.33\pm1.5	37.67\pm0.6	4.00\pm1.0	24.05 \pm 0.0	77.18\pm0.0	52.27\pm0.9	57.14\pm0.5	78.38\pm0.0	57.81\pm0.3
Up-Scaling	Deduct	3.67/10.65	37.67\pm1.5	35.00\pm0.0	4.00\pm0.0	23.89 \pm 0.6	76.34\pm0.0	58.92\pm1.1	59.28\pm0.6	75.05\pm0.0	58.70\pm0.3
Upcycling		3.87/10.85	34.00\pm2.0	33.00\pm0.0	6.67\pm1.5	25.13 \pm 0.0	69.80 \pm 0.4	20.92 \pm 0.1	55.34\pm0.6	73.36\pm0.1	48.91 \pm 0.0
Full FT		8.03/8.03	13.67\pm1.2	28.00 \pm 0.0	5.67\pm1.5	21.47\pm0.0	75.74\pm0.0	47.06\pm0.4	57.97\pm0.3	81.73\pm0.0	56.79\pm0.1
LoRA		3.73/10.71	10.33\pm1.5	35.33\pm0.6	9.00\pm1.0	18.85 \pm 0.0	75.92\pm0.0	46.60\pm0.7	56.65\pm0.0	79.87\pm0.0	55.58\pm0.1
Up-Scaling	Abduct	3.67/10.65	13.00 \pm 0.0	36.00\pm0.0	5.67\pm1.5	21.12 \pm 0.3	77.16\pm0.0	46.42 \pm 0.7	55.34\pm1.2	79.13\pm0.1	55.83\pm0.4
Upcycling		3.87/10.85	14.00\pm0.0	29.00 \pm 0.0	8.00\pm1.0	19.90 \pm 0.0	77.37\pm0.0	32.02 \pm 0.1	56.16\pm1.7	78.16\pm0.1	52.72 \pm 0.4
Full FT		8.03/8.03	26.00\pm0.0	0.00 \pm 0.0	5.33\pm0.6	18.85 \pm 0.0	76.03\pm0.1	29.93 \pm 0.9	55.83\pm0.3	80.51\pm0.0	52.23 \pm 0.2
LoRA		3.73/10.71	28.33\pm3.2	0.00 \pm 0.0	8.33\pm0.6	21.12 \pm 0.6	78.22\pm0.2	43.78\pm4.2	56.32\pm1.6	79.76\pm1.1	55.84\pm0.8
Up-Scaling	Mix	3.67/10.65	30.00\pm3.5	0.00 \pm 0.0	9.67\pm1.5	20.60 \pm 1.2	78.25\pm0.2	45.58\pm4.6	56.82\pm1.4	79.09\pm1.1	56.07\pm0.8
Upcycling		3.87/10.85	32.33\pm0.6	56.00\pm0.0	4.67\pm2.1	26.18\pm0.0	77.02\pm0.0	43.63 \pm 0.6	54.19\pm0.0	77.33\pm0.0	55.67\pm0.1
Llama-3.3-70B-IT	-	0/70.0	33.00 \pm 0.3	55.33 \pm 2.5	16.00 \pm 2.6	39.96 \pm 0.3	79.28 \pm 0.2	65.75 \pm 0.2	60.59 \pm 0.9	77.07 \pm 0.1	64.53 \pm 0.2
Qwen3-8B	-	0/8.19	79.67 \pm 2.1	67.33 \pm 0.6	35.33 \pm 0.6	40.14 \pm 0.8	81.26 \pm 0.0	70.98 \pm 0.7	78.33 \pm 1.0	71.40 \pm 0.0	68.42 \pm 0.2
Full FT		8.19/8.19	68.67\pm0.6	58.00 \pm 0.0	24.67 \pm 1.2	32.98 \pm 2.7	78.90 \pm 0.0	55.68 \pm 1.4	75.21\pm0.3	71.91\pm0.0	62.94 \pm 0.9
LoRA		3.88/10.83	61.67 \pm 0.6	60.00\pm0.0	14.33 \pm 0.6	30.89 \pm 0.9	78.25 \pm 0.0	39.07 \pm 1.6	67.16 \pm 0.3	72.18\pm0.0	57.51 \pm 0.5
Up-Scaling	Induct	3.56/10.51	34.00 \pm 1.7	57.00 \pm 0.0	40.67\pm0.6	33.86 \pm 5.1	82.06\pm0.0	67.60\pm1.0	73.89 \pm 0.5	72.67\pm0.0	66.02 \pm 1.1
Upcycling		3.96/10.91	72.67\pm1.5	72.00\pm0.0	31.67 \pm 1.2	41.19\pm2.4	81.88\pm0.0	74.54\pm1.0	77.67\pm0.3	72.20\pm0.0	69.49\pm0.7
Full FT		8.19/8.19	35.33\pm0.6	64.33\pm0.6	19.33\pm0.6	34.21 \pm 3.3	81.55\pm0.0	67.26\pm0.4	71.76 \pm 0.3	70.53 \pm 0.0	65.06\pm0.6
LoRA		3.88/10.83	28.00 \pm 1.0	58.33 \pm 1.2	15.00\pm2.6	31.76 \pm 1.2	83.25\pm0.0	55.26\pm1.1	80.30\pm0.0	70.96\pm0.0	64.31 \pm 0.0
Up-Scaling	Deduct	3.56/10.51	28.33 \pm 0.6	62.00 \pm 0.0	12.00 \pm 0.0	36.82 \pm 5.1	74.29 \pm 0.1	52.60 \pm 1.0	82.43\pm0.3	70.00 \pm 0.1	63.23 \pm 0.8
Upcycling		3.96/10.91	72.00\pm0.0	71.00\pm0.0	12.00 \pm 1.0	42.76\pm2.1	82.77\pm0.0	73.59\pm0.9	80.13 \pm 0.3	71.45\pm0.0	70.14\pm0.4
Full FT		8.19/8.19	33.33 \pm 1.2	69.33\pm0.6	17.33\pm1.2	40.31\pm0.0	83.81\pm0.1	67.79 \pm 0.7	71.10 \pm 0.3	73.96\pm0.0	67.39 \pm 0.1
LoRA		3.88/10.83	43.67 \pm 0.6	64.00 \pm 0.0	16.00 \pm 2.0	39.61 \pm 0.8	81.96 \pm 0.1	53.33 \pm 1.2	71.76 \pm 0.7	74.11\pm0.0	64.16 \pm 0.2
Up-Scaling	Abduct	3.56/10.51	44.33 \pm 0.6	60.00 \pm 0.0	14.33 \pm 1.2	37.00 \pm 0.3	81.92 \pm 0.0	44.09 \pm 0.9	72.25\pm0.3	73.47\pm0.0	61.74 \pm 0.1
Upcycling		3.96/10.91	53.33\pm0.6	68.33\pm0.6	13.67 \pm 0.6	47.82\pm0.6	83.74\pm0.0	68.08\pm0.9	75.53\pm0.3	71.51\pm0.1	69.34\pm0.1
Full FT		8.19/8.19	31.00 \pm 0.0	60.00 \pm 0.0	20.33\pm0.6	35.78 \pm 0.6	82.34\pm0.0	62.84\pm0.1	73.56\pm0.3	72.40\pm0.0	65.38 \pm 0.2
LoRA		3.88/10.83	33.00\pm1.0	61.33 \pm 2.3	11.67 \pm 2.5	31.94 \pm 0.0	81.52\pm0.2	59.56 \pm 7.5	66.50 \pm 5.6	73.82\pm0.3	62.67 \pm 0.4
Up-Scaling	Mix	3.56/10.51	33.00\pm1.7	62.33 \pm 2.1	12.67\pm2.1	31.94 \pm 0.0	81.42 \pm 0.2	53.63 \pm 8.8	69.29 \pm 5.8	73.98\pm0.3	62.05 \pm 0.6
Upcycling		3.96/10.91	48.33\pm0.6	64.33\pm0.6	12.00 \pm 1.0	43.45\pm0.9	81.89\pm0.1	71.13\pm0.7	77.34\pm0.0	71.40 \pm 0.0	69.04\pm0.3

Table 5: Accuracy of the fine-tuned models on the fundamental reasoning tasks using trajectories collected from *Llama-3.3-70B-Instruct*. **Method** indicates the base model and the applied PEFT methods. **FT** shows the fundamental reasoning task used for fine-tuning. **Train/Total** denotes the number of trainable and total parameters. The best and second-best methods for each fundamental reasoning task are indicated in **bold** and underlined, respectively. Scores that exceed the *base* model performance are shown in **green**.

Method	FT	Train/Total	Symbolic Tasks			Realistic Tasks					
			Deduct	Induct	Abduct	Detect	α NLI	Winowhy	Folio	RECV	Avg
Qwen3-30B	-	0/30.5	82.00 \pm 1.0	81.00 \pm 4.0	74.33 \pm 4.2	33.51 \pm 3.2	75.77 \pm 0.1	72.73 \pm 0.2	72.41 \pm 0.5	80.56 \pm 0.3	67.00 \pm 0.6
Llama-3.1-8B-IT	-	0/8.03	14.00 \pm 1.0	32.00 \pm 0.0	1.67 \pm 0.6	28.80 \pm 0.0	73.97 \pm 0.0	48.60 \pm 0.4	53.69 \pm 0.5	71.98 \pm 0.0	55.41 \pm 0.0
Full FT		8.03/8.03	25.67 \pm 1.5	78.00 \pm 0.0	2.33 \pm 0.6	25.13 \pm 0.0	77.71 \pm 0.0	43.05 \pm 0.1	54.84 \pm 0.3	78.29 \pm 0.0	55.80 \pm 0.1
LoRA		3.73/10.71	24.33 \pm 1.2	69.33 \pm 0.6	1.67 \pm 1.2	21.82 \pm 0.3	78.23 \pm 0.0	44.33 \pm 0.4	56.00 \pm 0.3	76.82 \pm 0.0	55.44 \pm 0.1
Up-Scaling	Induct	3.67/10.65	34.67 \pm 1.5	69.33 \pm 0.6	5.33 \pm 1.2	24.43 \pm 0.3	67.37 \pm 0.1	35.25 \pm 0.1	56.98 \pm 0.6	72.36 \pm 0.5	51.28 \pm 0.3
Upcycling		3.87/10.85	26.00 \pm 1.0	64.00 \pm 0.0	5.67 \pm 1.5	24.08 \pm 0.0	78.18 \pm 0.1	51.67 \pm 0.3	51.96 \pm 2.6	74.73 \pm 0.0	56.13 \pm 0.5
Full FT		8.03/8.03	67.33 \pm 0.6	33.00 \pm 0.0	7.67 \pm 1.5	24.61 \pm 0.0	48.73 \pm 0.1	41.51 \pm 0.3	56.65 \pm 0.0	77.80 \pm 0.0	49.86 \pm 0.1
LoRA		3.73/10.71	39.00 \pm 3.5	36.00 \pm 0.0	6.00 \pm 1.0	29.49 \pm 0.3	70.35 \pm 0.2	40.42 \pm 0.4	56.65 \pm 0.0	77.02 \pm 0.0	54.79 \pm 0.1
Up-Scaling	Deduct	3.67/10.65	70.00 \pm 1.7	32.00 \pm 0.0	3.33 \pm 1.5	27.23 \pm 0.0	57.26 \pm 0.2	42.93 \pm 0.3	58.46 \pm 0.3	73.87 \pm 0.0	51.95 \pm 0.1
Upcycling		3.87/10.85	47.00 \pm 1.0	40.00 \pm 0.0	6.33 \pm 1.5	26.18 \pm 0.0	69.87 \pm 0.1	37.69 \pm 0.1	56.65 \pm 0.0	73.29 \pm 0.0	52.74 \pm 0.0
Full FT		8.03/8.03	34.67 \pm 1.5	28.00 \pm 0.0	39.67 \pm 0.6	18.67 \pm 0.3	77.27 \pm 0.1	63.20 \pm 0.8	59.61 \pm 0.5	77.20 \pm 0.1	59.19 \pm 0.3
LoRA		3.73/10.71	34.67 \pm 2.9	29.33 \pm 1.2	43.33 \pm 1.5	23.56 \pm 0.0	75.59 \pm 0.1	56.74 \pm 2.9	59.77 \pm 0.8	75.09 \pm 0.0	58.15 \pm 0.7
Up-Scaling	Abduct	3.67/10.65	34.33 \pm 0.6	39.00 \pm 0.0	39.33 \pm 1.2	23.04 \pm 0.0	70.57 \pm 0.0	48.19 \pm 0.2	52.38 \pm 0.3	70.89 \pm 0.0	53.01 \pm 0.0
Upcycling		3.87/10.85	41.67 \pm 0.6	36.00 \pm 0.0	35.00 \pm 2.6	26.70 \pm 0.0	68.02 \pm 0.1	51.76 \pm 0.3	63.05 \pm 0.5	75.73 \pm 0.1	57.05 \pm 0.1
Full FT		8.03/8.03	50.67 \pm 0.6	63.33 \pm 0.6	45.00 \pm 2.0	24.08 \pm 0.0	77.54 \pm 0.0	55.17 \pm 0.1	57.64 \pm 0.5	78.24 \pm 0.0	58.53 \pm 0.1
LoRA		3.73/10.71	69.33 \pm 2.3	65.67 \pm 1.2	43.67 \pm 1.2	23.39 \pm 0.3	79.26 \pm 0.1	44.89 \pm 0.2	58.29 \pm 0.7	77.67 \pm 0.0	56.70 \pm 0.2
Up-Scaling	Mix	3.67/10.65	62.00 \pm 1.0	65.00 \pm 0.0	38.33 \pm 1.2	24.08 \pm 0.0	76.57 \pm 0.0	49.59 \pm 0.2	56.49 \pm 0.7	75.82 \pm 0.0	56.51 \pm 0.2
Upcycling		3.87/10.85	58.00 \pm 0.0	58.33 \pm 0.6	34.33 \pm 2.9	21.99 \pm 0.0	74.36 \pm 0.1	52.23 \pm 0.1	58.13 \pm 0.5	74.16 \pm 0.1	56.17 \pm 0.1
Qwen3-30B	-	0/30.5	82.00 \pm 1.0	81.00 \pm 4.0	74.33 \pm 4.2	33.51 \pm 3.2	75.77 \pm 0.1	72.73 \pm 0.2	72.41 \pm 0.5	80.56 \pm 0.3	67.00 \pm 0.6
Qwen3-8B	-	0/8.19	79.67 \pm 2.1	67.33 \pm 0.6	35.33 \pm 0.6	40.14 \pm 0.8	81.26 \pm 0.0	70.98 \pm 0.7	78.33 \pm 1.0	71.40 \pm 0.0	68.42 \pm 0.2
Full FT		8.19/8.19	68.33 \pm 0.6	49.00 \pm 0.0	7.67 \pm 0.6	27.23 \pm 0.0	84.57 \pm 0.0	79.34 \pm 0.2	77.18 \pm 0.3	70.42 \pm 0.0	67.75 \pm 0.1
LoRA		3.88/10.83	82.67 \pm 0.6	75.33 \pm 1.2	2.67 \pm 0.6	40.84 \pm 0.9	85.56 \pm 0.1	77.88 \pm 0.1	77.83 \pm 0.5	70.64 \pm 0.0	70.55 \pm 0.2
Up-Scaling	Induct	3.56/10.51	6.33 \pm 0.6	41.00 \pm 0.0	0.00 \pm 0.0	43.63 \pm 0.6	83.53 \pm 0.1	56.00 \pm 0.6	78.49 \pm 0.3	70.18 \pm 0.2	66.37 \pm 0.0
Upcycling		3.96/10.91	89.00 \pm 0.0	77.00 \pm 1.0	11.00 \pm 1.0	40.14 \pm 1.1	80.34 \pm 0.2	75.65 \pm 0.2	84.24 \pm 0.5	71.13 \pm 0.0	70.30 \pm 0.3
Full FT		8.19/8.19	88.00 \pm 1.7	60.00 \pm 0.0	48.00 \pm 1.0	36.65 \pm 0.0	81.92 \pm 0.0	70.30 \pm 0.4	75.04 \pm 0.8	70.98 \pm 0.0	66.98 \pm 0.2
LoRA		3.88/10.83	84.00 \pm 1.0	68.67 \pm 1.2	21.33 \pm 1.5	42.76 \pm 0.3	82.70 \pm 0.1	69.83 \pm 0.2	80.79 \pm 1.0	71.13 \pm 0.0	69.44 \pm 0.3
Up-Scaling	Deduct	3.56/10.51	85.33 \pm 1.2	63.33 \pm 0.6	34.67 \pm 2.1	37.35 \pm 0.6	80.84 \pm 0.1	19.82 \pm 0.1	80.30 \pm 0.0	71.40 \pm 0.0	57.94 \pm 0.1
Upcycling		3.96/10.91	92.00 \pm 1.0	73.00 \pm 0.0	34.67 \pm 0.6	44.50 \pm 0.5	84.00 \pm 0.0	73.58 \pm 0.2	81.94 \pm 0.3	70.82 \pm 0.1	70.97 \pm 0.2
Full FT		8.19/8.19	74.33 \pm 0.6	67.67 \pm 0.6	32.00 \pm 2.0	38.57 \pm 0.3	81.54 \pm 0.1	69.03 \pm 0.8	77.18 \pm 0.3	69.82 \pm 0.2	67.23 \pm 0.1
LoRA		3.88/10.83	77.67 \pm 1.5	70.67 \pm 0.6	41.33 \pm 1.5	36.13 \pm 0.0	81.97 \pm 0.1	66.72 \pm 0.6	76.03 \pm 1.0	70.02 \pm 0.0	66.18 \pm 0.1
Up-Scaling	Abduct	3.56/10.51	68.67 \pm 0.6	62.33 \pm 0.6	35.67 \pm 0.6	34.20 \pm 0.3	76.05 \pm 0.1	27.05 \pm 0.0	70.44 \pm 0.0	68.91 \pm 0.1	55.33 \pm 0.1
Upcycling		3.96/10.91	74.67 \pm 1.2	62.00 \pm 0.0	46.00 \pm 1.7	33.86 \pm 0.3	82.73 \pm 0.0	68.20 \pm 0.6	79.15 \pm 0.3	70.49 \pm 0.0	66.89 \pm 0.1
Full FT		8.19/8.19	88.33 \pm 1.2	77.33 \pm 0.6	41.33 \pm 1.2	39.44 \pm 0.3	81.69 \pm 0.0	73.44 \pm 0.2	74.55 \pm 0.3	70.29 \pm 0.0	67.88 \pm 0.1
LoRA		3.88/10.83	85.33 \pm 1.5	74.67 \pm 0.6	23.67 \pm 3.1	40.32 \pm 0.9	84.10 \pm 0.0	71.02 \pm 0.1	74.55 \pm 0.3	69.20 \pm 0.1	67.84 \pm 0.1
Up-Scaling	Mix	3.56/10.51	85.67 \pm 0.6	76.33 \pm 0.6	32.67 \pm 0.6	44.68 \pm 0.3	84.16 \pm 0.1	51.15 \pm 0.4	82.26 \pm 0.9	69.76 \pm 0.0	66.40 \pm 0.1
Upcycling		3.96/10.91	87.00 \pm 0.0	69.00 \pm 0.0	51.00 \pm 1.0	42.58 \pm 0.3	81.32 \pm 0.0	68.28 \pm 0.1	81.12 \pm 0.3	70.56 \pm 0.1	68.77 \pm 0.1

Table 6: Accuracy of the fine-tuned models using trajectories collected from *Qwen3-30B-Instruct*.

E Out-of-Domain Generation Example

Examples 1–3 show the generated outputs for a question from FOLIO. The vanilla *Llama-3.1-8B-Instruct* and the deductive Full FT model fail to answer the question correctly, whereas deductive up-scaling produces the correct answer.

Example 1: Vanilla Llama

Based on the given premises, is the conclusion correct? Please respond with True, False, or Uncertain enclosing in `<answer><answer>`.

Premises: People in this club who perform in school talent shows often attend and are very engaged with school events.

People in this club either perform in school talent shows often or are inactive and disinterested community members.

People in this club who chaperone high school dances are not students who attend the school.

All people in this club who are inactive and disinterested members of their community chaperone high school dances.

All young children and teenagers in this club who wish to further their academic careers and educational opportunities are students who attend the school.

Bonnie is in this club and she either both attends and is very engaged with school events and is a student who attends the school or is not someone who both attends and is very engaged with school events and is not a student who attends the school.

Conclusion: Bonnie performs in school talent shows often.

Generated Output:

in this case, we can break down the information as follows:

1. If Bonnie is a student who attends the school, then she must perform in school talent shows often (from the first premise).
2. If Bonnie is not a student who attends the school, then she must be an inactive and disinterested community member (from the second premise).
3. If Bonnie is an inactive and disinterested community member, then she must chaperone high school dances (from the fourth premise).
4. However, we know that people who chaperone high school dances are not students who attend the school (from the third premise).
5. This means that if Bonnie is not a student who attends the school, then she must chaperone high school dances, which contradicts the fact that she is not a student who attends the school.
6. Therefore, Bonnie must be a student who attends the school.
7. Since Bonnie is a student who attends the school, she must perform in school talent shows often (from the first premise).

So, the conclusion is correct: `<answer>True</answer>`.

Examples 4–6 show the generated outputs for another question from FOLIO. The vanilla *Qwen-3-8B* and the abductive upcycling model fail to answer the question correctly, whereas the deductive upcycling model produces the correct answer.

Example 2: Deductive Full FT Llama

Generated Output:

in order to determine if the conclusion is correct, we need to analyze the premises and apply them to Bonnie's situation.

From the premises, we know:

1. People in this club who perform in school talent shows often attend and are very engaged with school events.
2. People in this club either perform in school talent shows often or are inactive and disinterested community members.
3. People in this club who chaperone high school dances are not students who attend the school.
4. All people in this club who are inactive and disinterested members of their community chaperone high school dances.
5. All young children and teenagers in this club who wish to further their academic careers and educational opportunities are students who attend the school.

Given that Bonnie is in this club, we can start by looking at the last premise:

All young children and teenagers in this club who wish to further their academic careers and educational opportunities are students who attend the school.

This implies that if Bonnie is a young child or teenager and wishes to further her academic career and educational opportunities, then she must be a student who attends the school.

However, we also know from premise 3 that people in this club who chaperone high school dances are not students who attend the school.

If Bonnie is a student who attends the school, then she cannot be a chaperone for high school dances.

Now, looking at premise 2:

People in this club either perform in school talent shows often or are inactive and disinterested community members.

If Bonnie is a student who attends the school, then from premise 1, she must perform in school talent shows often because she is very engaged with school events.

However, premise 4 states that all people in this club who are inactive and disinterested members of their community chaperone high school dances.

This means that if Bonnie is not a student who attends the school, then she must be an inactive and disinterested member of her community, which would make her a chaperone for high school dances, which contradicts our earlier deduction that if she is a student, she cannot be a chaperone.

Considering all premises, we can conclude that if Bonnie is a student who attends the school, then she must perform in school talent shows often.

But we also know from premise 3 that if she is a chaperone, she cannot be a student.

So, if we consider the possibility that Bonnie is not a student, then she would be an inactive and disinterested member of her community, which would make her a chaperone, but this contradicts our initial assumption that she is not a student.

Given the premises, the most consistent conclusion we can draw about Bonnie is that if she is a student who attends the school, then she must perform in school talent shows often.

But since we cannot definitively conclude that Bonnie is a student or not based on the given information, we cannot definitively say if she performs in school talent shows often.

However, looking closely at premise 5 and the structure of the premises, we can see that if Bonnie is a student who attends the school and is very engaged with school events, then she must perform in school talent shows often.

But we cannot conclude this with absolute certainty without more information about Bonnie's status as a student or a non-student.

Therefore, based on the given premises, we cannot definitively conclude that Bonnie performs in school talent shows often without assuming that she is a student who attends the school.

But since the conclusion is about whether Bonnie performs in school talent shows often, and we have a premise that supports this conclusion if she is a student, we can say that the conclusion is True if we assume Bonnie is a student who attends the school and is very engaged with school events.

`<answer>True</answer>`

Example 3: Deductive Up-scaling Llama

Generated Output:

in order to determine if the conclusion is correct, we need to analyze the premises:

1. People in this club who perform in school talent shows often attend and are very engaged with school events.
2. People in this club either perform in school talent shows often or are inactive and disinterested community members.
3. People in this club who chaperone high school dances are not students who attend the school.
4. All people in this club who are inactive and disinterested members of their community chaperone high school dances.
5. All young children and teenagers in this club who wish to further their academic careers and educational opportunities are students who attend the school.
6. Bonnie is in this club and she either both attends and is very engaged with school events and is a student who attends the school or is not someone who both attends and is very engaged with school events and is not a student who attends the school.

From premise 6, we know that Bonnie is in this club and has two possible scenarios:

- She both attends and is very engaged with school events and is a student who attends the school.
- She is not someone who both attends and is very engaged with school events and is not a student who attends the school.

From premise 1, if Bonnie is someone who both attends and is very engaged with school events, then she must perform in school talent shows often.

From premise 3, if Bonnie is a student who attends the school, then she cannot chaperone high school dances.

From premise 4, if Bonnie is an inactive and disinterested member of her community, then she must chaperone high school dances.

From premise 5, if Bonnie is a young child or teenager in this club who wishes to further her academic careers and educational opportunities, then she must be a student who attends the school

Considering all premises, we cannot directly conclude that Bonnie performs in school talent shows often from the given information. However, we can deduce that if Bonnie is a student who attends the school and is very engaged with school events, then she must perform in school talent shows often.

But premise 6 does not directly state that Bonnie is a student or that she is not a student. It only gives two possible scenarios for Bonnie's situation.

Therefore, the conclusion that Bonnie performs in school talent shows often is not directly supported by the premises, and we cannot be certain of it.

<answer>Uncertain</answer>

Example 4: Vanilla Qwen

Based on the given premises, is the conclusion correct? Please respond with True, False, or Uncertain enclosing in <answer><answer>.

Premises: Barutin Cove is a cove named after the Bulgarian settlement of Barutin.

Barutin Cove is on the southwest coast of Snow Island. Snow Island, Greenwich Island, and Deception Island are located in the South Shetland Islands.

Antarctica is located on the South Shetland Islands.

If place A is located in place B and place B is located in place C, then place A is located in place C.

Conclusion: Barutin Cove is not located in Antarctica.

Generated Output:

from the premises, we know that barutin cove is on the southwest coast of snow island. snow island is part of the south shetland islands. the south shetland islands are not part of antarctica, but rather located near antarctica. the premises also state that antarctica is located on the south shetland islands, which seems to be a contradiction. however, the correct understanding is that the south shetland islands are a group of islands located near antarctica, not part of it. thus, barutin cove, being on snow island, is not located in antarctica. the conclusion is correct.

<answer>True</answer>

Example 5: Abductive Upcycling Qwen

Generated Output:

The premises state that Barutin Cove is on the southwest coast of Snow Island. Snow Island, along with Greenwich Island and Deception Island, is located in the South Shetland Islands. Furthermore, it is stated that Antarctica is located on the South Shetland Islands.

However, there's a potential confusion in the premises. The statement "Antarctica is located on the South Shetland Islands" is incorrect. In reality, the South Shetland Islands are located in the Southern Ocean, near Antarctica, but they are not part of Antarctica itself. Antarctica is a continent, while the South Shetland Islands are a group of islands that are geographically close to Antarctica but are not part of it.

Given the logical structure provided: If place A is located in place B and place B is located in place C, then place A is located in place C.

Applying this to the premises:

- Barutin Cove is located in Snow Island (A in B).
- Snow Island is located in the South Shetland Islands (B in C).
- Therefore, by transitivity, Barutin Cove is located in the South Shetland Islands (A in C).

But the conclusion states that Barutin Cove is not located in Antarctica. Since the South Shetland Islands are not part of Antarctica, and Barutin Cove is in the South Shetland Islands, it follows that Barutin Cove is not located in Antarctica.

However, the premise that "Antarctica is located on the South Shetland Islands" is factually incorrect. This might lead to confusion. But based on the logical structure provided and assuming the premises are to be taken as true, even if they contain factual errors, the conclusion that Barutin Cove is not located in Antarctica would still follow from the transitive property of the given premises.

Therefore, based on the premises as given, the conclusion is correct.

<answer>True</answer>

Example 6: Deductive Upcycling Qwen

Generated Output:

Okay, let's try to figure this out step by step. So, the question is whether the conclusion "Barutin Cove is not located in Antarctica" is correct based on the given premises. Let me break down the information provided.

First, the premises say that Barutin Cove is a cove named after the Bulgarian settlement of Barutin. Then, it's located on the southwest coast of Snow Island. Next, Snow Island, Greenwich Island, and Deception Island are all in the South Shetland Islands. Then there's a statement that Antarctica is located on the South Shetland Islands. Wait, that seems a bit off. Wait, actually, the South Shetland Islands are a group of islands located north of the Antarctic Peninsula, and they are part of the Antarctic region, but Antarctica itself is a continent. So maybe there's a mistake here? Let me check that again.

The fourth premise says: "Antarctica is located on the South Shetland Islands." That doesn't sound right. Antarctica is a continent, and the South Shetland Islands are a group of islands that are part of the Antarctic region, but they are not part of Antarctica itself. Wait, actually, the South Shetland Islands are considered part of the Antarctic Treaty System, but geographically, they are separate from the continent of Antarctica. So maybe the premise is incorrect? But I have to take the premises as given, even if they might be factually wrong.

So, according to the premises, if place A is located in place B and place B is located in place C, then place A is located in place C. That's a transitive property. So, applying that.

Barutin Cove is on Snow Island. Snow Island is in the South Shetland Islands. Then, the premise says Antarctica is located on the South Shetland Islands. Wait, that's confusing. If the South Shetland Islands are in Antarctica, then Barutin Cove would be in Antarctica. But the conclusion says it's not. But according to the premises, if the South Shetland Islands are part of Antarctica, then Barutin Cove, being on Snow Island (which is in South Shetland), would be in Antarctica. Therefore, the conclusion that it's not located in Antarctica would be false. But wait, the premise says "Antarctica is located on the South Shetland Islands." That seems like a reversal. Normally, the South Shetland Islands are located in the Antarctic region, but not part of Antarctica. But if the premise states that Antarctica is located on the South Shetland Islands, then that would mean that the South Shetland Islands are part of Antarctica. So, if that's the case, then Barutin Cove is in Antarctica. Therefore, the conclusion that it's not located in Antarctica would be false. But wait, the premises might be conflicting. Let me re-express:

Premise 1: Barutin Cove is on Snow Island.

Premise 2: Snow Island is in South Shetland Islands.

Premise 3: South Shetland Islands, Greenwich Island, Deception Island are in South Shetland Islands. Wait, that's redundant. Wait, the third premise says "Snow Island, Greenwich Island, and Deception Island are located in the South Shetland Islands." So, Snow Island is in South Shetland Islands.

Premise 4: Antarctica is located on the South Shetland Islands. So, if the South Shetland Islands are in Antarctica, then Barutin Cove is in Antarctica. But the conclusion says it's not. Therefore, the conclusion would be false. But wait, maybe the premise is that the South Shetland Islands are located in Antarctica? Wait, the wording is "Antarctica is located on the South Shetland Islands." That would mean that Antarctica is a part of the South Shetland Islands, which is not correct in reality, but according to the premises, we have to take it as given. So, if Antarctica is located on the South Shetland Islands, then the South Shetland Islands are part of Antarctica. Therefore, Barutin Cove, being on Snow Island (which is in South Shetland), would be in Antarctica. Therefore, the conclusion that it's not located in Antarctica would be false. Therefore, the answer should be False. But wait, maybe I'm misunderstanding the premise. Let me check again.

Wait, the fourth premise says "Antarctica is located on the South Shetland Islands." That's a bit confusing. Normally, the South Shetland Islands are located in the Antarctic region, but they are not part of Antarctica. However, if the premise states that Antarctica is located on the South Shetland Islands, that would imply that the South Shetland Islands are part of Antarctica. So, if that's the case, then Barutin Cove is in Antarctica. Therefore, the conclusion that it's not located in Antarctica would be false. Therefore, the answer should be False. But wait, maybe the premise is that the South Shetland Islands are located in Antarctica? But the wording is "Antarctica is located on the South Shetland Islands," which is the opposite. So, if Antarctica is located on the South Shetland Islands, then the South Shetland Islands are part of Antarctica. Therefore, Barutin Cove is in Antarctica. Hence, the conclusion is false. Therefore, the answer is False.

<answer>False</answer>

F Symbolic Generation Example

The following examples shows the generated output of answering a symbolic inductive task. The vanilla *Llama-3.1-8B-Instruct* failed to answer the question, while Inductive and Deductive Full FT can generate the correct answer. The vanilla model (Example 7) fails on an inductive task, while inductive (Example 8) and deductive (Example 9) Full FT succeed. Example 8 adopts a pattern-driven, data-oriented perspective, while Example 9 follows a rule-based, hypothesis-testing process, consistent with their respective paradigms. Interestingly, both induced models exhibit self-reflection by revising earlier reasoning. These observations indicate that inducing different reasoning paradigms steers models toward distinct reasoning perspectives, while shared behaviors across induced models demonstrate transferability among these paradigms (§6.1).

Example 7: Untrained Llama 3.1 8B

This is a <Inductive> reasoning task. Given the following sequence, ['2', '5', '20', '23', '26', '104', '107', '110', '440', '443', '?'] What is the value at the question mark? Please enclose the answer in <answer><answer>, and put all the intermediate reasoning steps in <think><think>.

Generated Output:

<think>The given sequence appears to be formed by alternating between two different operations: multiplication and addition. To identify the pattern, let's break down the sequence into two separate sequences, one for the numbers that are obtained by multiplication and one for the numbers that are obtained by addition.</think>

Sequence 1 (Multiplication): 2, 20, 104, 440
Sequence 2 (Addition): 5, 23, 107, 443

<think>Now, let's analyze Sequence 1. The numbers in this sequence are obtained by multiplying the previous term by 10 and then adding 2. This can be represented as: $2 * 10 + 2 = 20$, $20 * 10 + 2 = 202$, but 202 is not in the sequence, however $20 * 10 + 4 = 204$ is not in the sequence either, but $20 * 10 + 2 = 202$ is not in the sequence, however $20 * 10 + 4 = 204$ is not in the sequence...

Keeping Repetitive until hit max number of tokens

Example 8: Inductive Full FT Llama

This is an <Inductive> reasoning task. Given the following sequence,

['2', '5', '20', '23', '26', '104', '107', '110', '440', '443', '?']

What is the value at the question mark? Please enclose the answer in <answer><answer>, and put all the intermediate reasoning steps in <think><think>.

Generated Output:

<think> To find the pattern in the sequence, let's first examine the differences between consecutive numbers:

$5 - 2 = 3$, $20 - 5 = 15$, $23 - 20 = 3$, $26 - 23 = 3$, $104 - 26 = 78$, $107 - 104 = 3$, $110 - 107 = 3$, $440 - 110 = 330$, $443 - 440 = 3$

From the differences, we can observe a pattern where the difference alternates between 3 and a larger number. The larger numbers are increasing (15, 78, 330), which suggests a pattern of multiplication or addition that increases each time.

Looking closer at the larger differences:

- $15 * 5.2 = 78$ (approximately)
- $78 * 4.23 = 330$ (approximately)

However, a more precise pattern can be observed:

$15 = 3 * 5$, $78 = 15 * 5 + 3$, $330 = 78 * 4 + 6$

But a simpler pattern can be observed if we consider the sequence of operations:

Start with 2, Add 3 to get 5, Multiply by 4 to get 20, Add 3 to get 23, Add 3 to get 26, Multiply by 4 to get 104, Add 3 to get 107, Add 3 to get 110, Multiply by 4 to get 440, Add 3 to get 443

Following this pattern, the next operation should be to add 3 again, as the pattern alternates between multiplying by 4 and adding 3.

So, the next number in the sequence after 443 would be $443 + 3 = 446$. </think>

<answer>446</answer>

Example 9: Deductive Full FT Llama

Generated Output:

<think> To solve this, let's analyze the given sequence for any patterns or rules that could help us determine the next value. The sequence is: '2', '5', '20', '23', '26', '104', '107', '110', '440', '443', '?'.</think>

Looking at the sequence, we can see that it starts with '2' and '5', which are both prime numbers. Then, '20' and '23' are not prime, but '26' is not a prime either, which makes us think that the sequence might not follow a simple pattern of prime numbers.

However, if we look closer, we can see that the sequence seems to be formed by adding a certain number to the previous term to get the next term. Let's try to find this pattern.

From '2' to '5', the difference is 3. From '5' to '20', the difference is 15. From '20' to '23', the difference is 3. From '23' to '26', the difference is 3. From '26' to '104', the difference is 78. From '104' to '107', the difference is 3. From '107' to '110', the difference is 3. From '110' to '440', the difference is 330. From '440' to '443', the difference is 3.

We can see that the differences are not consistent except for the differences between '2' and '5', '23' and '26', '104' and '107', '107' and '110', and '440' and '443', which are all 3. However, the other differences are not consistent.

But if we look at the sequence again, we can see that the numbers '20', '104', and '440' are all multiples of 4, and '23', '26', '107', and '110' are all 3 more than a multiple of 4.

So, the pattern seems to be that the sequence is formed by adding 3 to the previous term when the previous term is 3 more than a multiple of 4, and the term before that is a multiple of 4.

Let's apply this pattern to find the next term:

- The last term before the question mark is '443', which is 3 more than a multiple of 4 (440).
- So, the next term should be '443 + 3 = 446'.

Therefore, the value at the question mark is 446.

</think>

<answer>446</answer>