QFFT, Question-Free Fine-Tuning for Adaptive Reasoning

Wanlong Liu

2025年11月2日

香港中文大學（深圳）
The Chinese University of Hong Kong, Shenzhen

HUAWEI

UESTC 1956

Brief Contribution

1. QFFT Fine-Tuning Method
- Adaptive Reasoning
- Performance compared to SFT
- Effectively reduce token counts

2. QFFT Applications (compared to SFT)
- Noisy Scenario
- OOD Scenario
- Low Resource Scenario

BackGround: Long CoT v.s. Short CoT

Short CoT

- Traditional CoT
- More Concise

Long CoT

- CoT patterns for thinking models
- Longer thinking chains
- Reflective, self-correction tokens.

Short CoT (Qwen-2.5-32B-Instruct) 425 Tokens

To solve the problem, we need to determine the time Aya spends walking and in the coffee shop for different speeds.
Aya walks 9 km. Let her walking speed be s km/h, and the time in the coffee shop be t minutes.
1. When walking at speed s:
   Total time: 9/s + t/60 = 4 ...
2. When walking at speed s + 2:
   Total time: 9/(s + 2) + t/60 = 2.4
 … By subtracting the two equations: … We get t = 24 minutes.
Now, if Aya walks at s + 1/2 = 3 km/h:
 - Walking time = 180 minutes
 - Total time = 180 + 24 = 204 minutes
Thus, the total time for the walk including the coffee shop is boxed{**204**} minutes."

Long CoT (S1.1-32B) 3867 Tokens

Okay, so I need to find how long Aya's walk takes when she walks at s + 1/2 kilometers per hour, including the t minutes she spends in the coffee shop... The walk itself is 9 kilometers long each time.
Wait, so the total time includes walking time plus t minutes at the coffee shop...Let me formalize that. Let's denote: Let's solve this equation for s. First, factor out the 9 ...
Alternatively, let me check with the second equation as well ... But the question asks for the total time in minutes. ... So, 3 hours is 180 minutes, plus 24 minutes is 204 minutes.
Wait, but let me confirm. The total time is walking time (3 hours) plus t minutes (24 minutes)... However, walking time is in hours, so converting the entire total time into minutes.
Wait, the problem says \"the walk takes her [time], including t minutes spent in the coffee shop...
Wait, the walking time when speed is 3 km/h is 9 / 3 = 3 hours. So, 3 hours is 180 minutes. Then, adding the t minutes (24 minutes), total time is 180 + 24 = 204 minutes.But wait, is that correct? ...
Answer: Aya walks 9 kilometers every morning and stops at a coffee shop afterwards. We need to find the total time ...
boxed{**204**}

BackGround: Long CoT v.s. Short CoT

Table 1: Comparison of Reasoning Modes: Short CoT, Long CoT, and Adaptive Reasoning

	Short CoT	**Long CoT**	**Adaptive Reasoning**
Reasoning Style	Direct, concise.	Reflective, self-correction, containing reflective keywords (e.g., “wait”).	Adaptively uses Short/Long CoT based on question difficulty
Application Scenarios	Effective for simple questions.	Strong on difficult questions.	Optimal for both simple and difficult questions.
Cost	Low (fewer tokens).	High (redundant tokens).	Moderate (adaptive to needs).
Risk	Oversimplification.	Overthinking.	Balanced (mitigates both risks).

Question: Can we achieve adaptive reasoning?

Definition of Adaptive Reasoning

Adaptive Reasoning: A model's ability to adaptively prioritize Short CoT reasoning patterns for ***simple questions*** and prioritize Long CoT reasoning patterns for ***challenging questions*** when Short CoT is ineffective.

Simple Question: Short CoT can solve

Difficult Question: Short CoT cannot solve

Question: How to measure a model's Adaptive Reasoning Ability?

Quantitative Metric on Adaptive Reasoning Ability

Assumption 1: If a Long CoT model M_L is derived (e.g. distilled) from a Short CoT model M_S, we approximate the Short CoT capability of model M_L by that of model M_S.

Reasoning Adaptability Cohen's Kappa (RAK)

$$\underbrace{(\text{true} \quad \text{false})}_{\text{Short CoT model}} \qquad \underbrace{(\text{short CoT}, \quad \text{long CoT})}_{\text{Target model}}$$

$$\text{RAK} = \frac{p_o - p_e}{1 - p_e},$$

$$p_o = \frac{\texttt{TP}+\texttt{TN}}{\texttt{TP}+\texttt{FP}+\texttt{FN}+\texttt{TN}},$$

$$p_e = \frac{(\texttt{TP}+\texttt{FP})(\texttt{TP}+\texttt{FN})+(\texttt{FN}+\texttt{TN})(\texttt{FP}+\texttt{TN})}{(\texttt{TP}+\texttt{FP}+\texttt{FN}+\texttt{TN})^2},$$

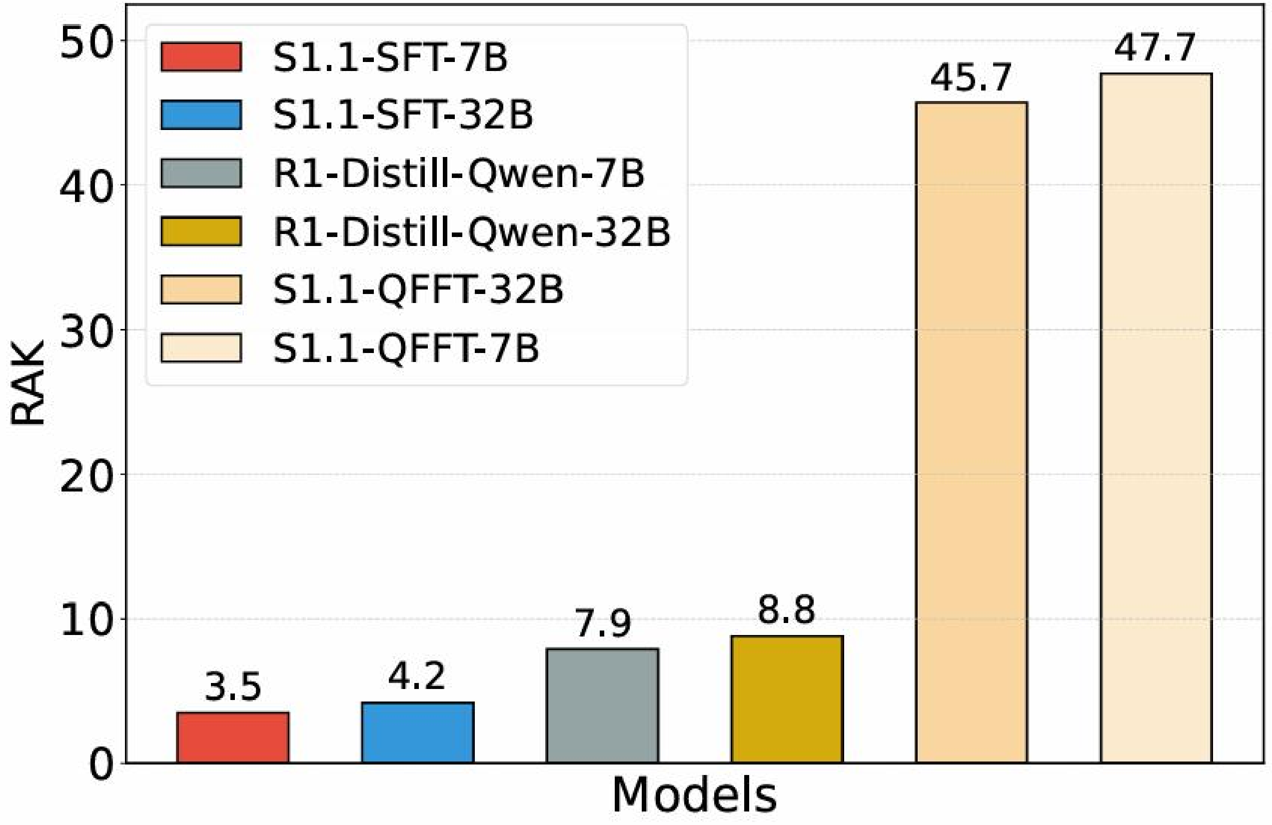
[3] Specifically, TP is the number of solvable questions of M_S where model M_L selects Short CoT patterns, FN is the number of solvable questions of M_S where M_L selects Long CoT patterns, FP is the number of unsolvable questions of M_S where model M_L selects Short CoT patterns, and TN is the number of unsolvable questions of M_S where model M_L selects Long CoT patterns. Here, "solvable" indicates that the reference model M_S answers the question correctly.

Quantitative Metric on Adaptive Reasoning Ability

RAK for Long CoT models

Current Long CoT models (obtained by SFT) have low RAK scores, which attributes to their consistent over-reliance on Long CoT reasoning patterns.

Why?



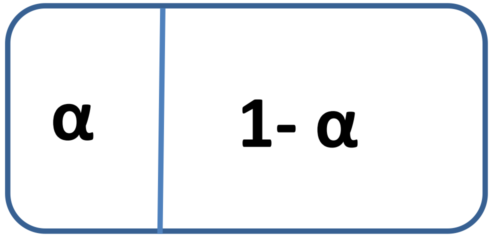
Pilot Study on Adaptive Reasoning

Hypothesis.

During SFT, model learns the Question (Q) -> Long CoT Chain (R) mappings,

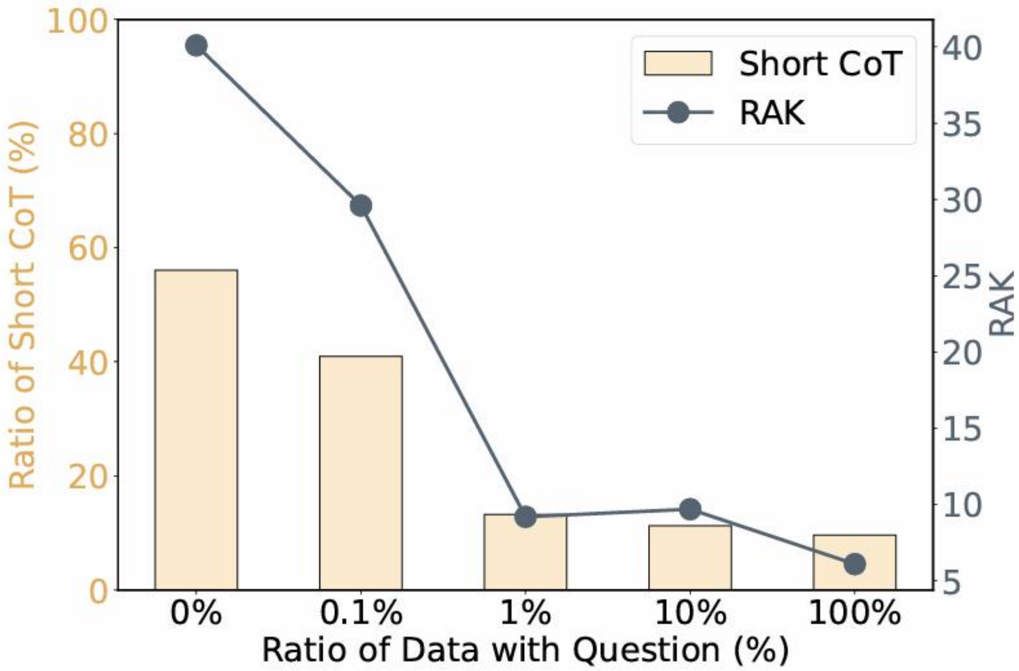
Overriding its original Question (Q) - > Short CoT Chain mappings.

Pilot Study.



With questions + responses

Only responses



Motivation

1. To preserve the model's **default Short CoT pattern**s and prevent it from being overridden, we need to avoid training the model to learn a fixed Q → R mapping.

2. **The model can learn Long CoT ability without questions.**

 Prior studies [1, 2] suggest that the core of Long CoT patterns lies in the structure of responses, rather than in the questions. This implies that models distilled solely from Long CoT responses, even without access to the corresponding questions, can still acquire Long CoT reasoning capability

[1] Dacheng Li, Shiyi Cao, Tyler Griggs, Shu Liu, Xiangxi Mo, Eric Tang, Sumanth Hegde, Kourosh Hakhamaneshi, Shishir G Patil, Matei Zaharia, et al. Llms can easily learn to reason from demonstrations structure, not content, is what matters! arXiv preprint arXiv:2502.07374, 2025.
[2] Edward Yeo, Yuxuan Tong, Morry Niu, Graham Neubig, and Xiang Yue. Demystifying long chain-of-thought reasoning in llms. arXiv preprint arXiv:2502.03373, 2025.

Motivation

Why model can learn Long CoT ability without questions?

1. Long CoT response data inherently contains patterns like reflection (denoted as B_L), which emerge when encountering uncertainties or errors (denoted as U_L).

2. The model learns: $P_\theta(B_\mathrm{r} \mid U_L)$ learning only the responses.

Method

Question-Free Fine-Tuning

Response

$$\mathcal{L}_{\text{QFFT}} = -\frac{1}{|\mathcal{R}|} \sum_{t \in \mathcal{R}} \log P_\theta(R_t \mid R_{<t}, \cancel{Q}),$$

Two Equivalence of the QFFT Method

1. Supervised Fine-tuning with Null Questions.

Since the questions are empty, the model does not learn any concrete Q → R mappings. during inference, any non-empty input will not trigger Long CoT patterns, thereby preserving the model’s original Short CoT patterns.

2. Continued Pre-training.

QFFT can be conceptualized as a specialized form of continued pre-training that systematically enhances the model’s general Long CoT proficiency, including reflective reasoning capabilities, by training exclusively on reasoning responses, without reliance on specific question answer pair formats.

Why QFFT Works?

Training: Preserving Short CoT and Learning Reflection. During the QFFT training, the model avoids learning a direct mapping from questions (Q) to Long CoT responses (R), thereby preserving the model's default Short CoT patterns to answer questions. Additionally, the model is trained exclusively on Long CoT responses. Thus, the model theoretically acquires the capability for Long CoT reasoning and learns to exhibit reflective behaviors when encountering uncertainty or errors within the context of Long CoT reasoning. Formally, let U_L denote the event of encountering uncertainty or errors during Long CoT reasoning. Then, the model learns the conditional probability:

$$P_\theta(B_r \mid U_L),$$

where B_r denotes the occurrence of reflective behaviors, and U_L represents uncertainty or errors arising specifically within the context of Long CoT reasoning.

Inference: Default to Short CoT with Adaptive Reflection. At inference time, the QFFT model defaults to the Short CoT patterns. However, since the model has only explicitly learned the conditional probability $P_\theta(B_r \mid U_L)$ during training, it is not immediately obvious why it can still trigger reflective behaviors in the context of Short CoT reasoning. We explain this phenomenon from a transfer learning [26, 27] perspective:

Assumption 2 *If a model has learned the conditional probability $P_\theta(B_r \mid U_L)$, this reflective capability can be transferred to Short CoT, enabling it to implicitly learn $P_\theta(B_r \mid U_S)$, where U_S denotes uncertainty or errors that arise specifically within Short CoT reasoning.*

Experiments

Compared to SFT

Table 1: Main results on 3 mathematical benchmarks. The reported accuracy and Avg. Len (Tokens) are averaged over 16 random sampling runs. TAK is our defined thinking adaptability metric.

Data	Method	GSM8K			MATH			AIME25			Average		
		Acc ↑	Tokens ↓	TAK ↑	Acc	Tokens	TAK	Acc	Tokens	TAK	Acc	Tokens	TAK
7B Models (Based on Qwen2.5-7B-Instruct)													
S1.1	SFT	90.6	1.7K	1.8	80.8	5.3K	3.5	18.2	17.7K	0	63.2	8.2K	1.8
	QFFT	91.0	0.4K	28.4	80.2	2.8K	47.7	17.2	12.8K	28.0	62.8	5.3K	34.7
	Δ	+0.4	-76.5%	+26.6	-0.6	-47.2%	+44.2	-1.0	-27.7%	+28.0	-0.4	-50.5%	+32.9
LIMO	SFT	88.2	1.8K	0.2	80.4	5.8K	6.1	16.8	17.1K	0.2	61.8	8.2K	2.2
	QFFT	88.0	0.7K	26.7	80.6	4.1K	40.1	17.2	15.6K	34.2	61.9	6.8K	33.7
	Δ	-0.2	-61.1%	+26.5	+0.2	-29.3%	+34.0	+0.4	-8.8%	+34.0	+0.1	-33.1%	+31.5
BS-17K	SFT	91.0	1.4K	2.3	81.6	5.7K	4.6	19.8	13.8K	0.1	64.1	7.0K	2.3
	QFFT	90.4	0.4K	29.7	81.4	2.2K	44.6	18.3	9.7K	29.7	63.4	4.1K	34.7
	Δ	-0.6	-71.4%	+27.4	-0.2	-61.4%	+40.0	-1.5	-29.7%	+29.6	-0.8	-54.2%	+32.3
32B Models (Based on Qwen2.5-32B-Instruct)													
S1.1	SFT	92.8	2.1K	0.2	93.1	4.1K	4.2	48.6	16.2K	0	78.2	7.5K	1.5
	QFFT	93.6	0.6K	31.2	92.2	2.4K	45.7	46.8	12.9K	29.3	77.5	5.3K	35.4
	Δ	+0.8	-71.4%	+31.0	-0.9	-41.5%	+41.5	-1.8	-20.4%	+29.3	-0.6	-44.4%	+33.9
LIMO	SFT	91.2	1.9K	1.0	93.0	3.9K	8.8	45.8	13.2K	0	76.6	6.3K	3.3
	QFFT	92.6	0.8K	27.4	92.6	2.9K	38.9	45.0	12.5K	33.2	76.7	5.4K	33.2
	Δ	+1.4	-57.9%	+26.4	-0.4	-25.6%	+30.1	-0.8	-5.3%	+33.2	+0.1	-29.6%	+29.9

Experiments

Compared to Long2short Baselines

Table 3: Comparison with other Long-to-Short baselines. AES is used as a metric that balances performance and response length.

Method	GSM8K			MATH			AIME25			Average.		
	Acc ↑	Tokens ↓	AES ↑	Acc ↑	Tokens ↓	AES ↑	Acc ↑	Tokens ↓	AES ↑	Acc ↑	Tokens ↓	AES ↑
Long-to-Short Methods (7B)												
LIMO 7b (base)	88.2	1.8K	-	80.4	5.9K	-	16.8	17.8K	-	61.8	8.5K	-
SFT Shortest	88.9	1.2K	3.4	78.3	4.8K	-0.7	**17.9**	17.3K	0.9	61.7	7.8K	1.2
DPO Shortest	89.8	1.6K	1.5	79.5	5.4K	-0.1	17.3	17.1K	0.7	62.2	8.0K	0.7
SimPO Shortest	87.2	1.2K	2.2	75.8	**3.2K**	-1.0	14.0	**8.8K**	-12.0	59.0	**4.4K**	-3.6
O1-pruner	**90.8**	0.8K	5.8	78.2	3.2K	1.8	14.2	12.2K	-12.7	61.0	5.4K	-1.7
Distilled Methods (7B)												
DAD-7B	90.0	0.9K	5.2	80.2	4.8K	1.6	17.3	17.7K	0.3	**62.5**	7.8K	2.4
QFFT (Ours)	88.0	**0.7K**	**5.9**	**80.6**	4.1K	**2.9**	17.2	15.6K	**1.4**	61.9	6.9K	**3.4**
Long-to-Short Methods (32B)												
Limo 32B (base)	91.2	1.9K	-	93.0	3.9K	-	45.8	13.2K	-	76.7	6.3K	-
SFT Shortest	93.2	9.2K	-38.2	91.4	3.1K	0.4	**46.3**	14.5K	-0.9	**76.9**	8.9K	-12.9
SimPO shortest	94.5	**0.6K**	**7.2**	89.8	2.4K	0.4	40.0	13.0K	-12.5	74.8	6.7K	-1.6
O1-pruner	**94.8**	**0.6K**	**7.2**	90.5	**2.0K**	**2.1**	29.2	**7.3K**	-136.9	71.5	**3.3K**	-42.5
Distilled Methods (32B)												
QFFT (Ours)	92.6	0.8K	5.9	**92.6**	2.9K	**2.1**	45.0	12.5K	**-1.2**	76.7	5.4K	**2.3**

Further Analysis

1. Pattern Analysis

(1) Default to Short CoT

(2) Trigger Long CoT from Short CoT.

Question

Q: Every morning Aya goes for a 9-kilometer-long walk and stops at a coffee shop afterwards... Find the number of minutes the walk takes her, including the t minutes spent in the coffee shop.

Short CoT (Qwen-2.5-32B-Instruct) 425 Tokens

To solve the problem, we need to determine the time Aya spends walking and in the coffee shop for different speeds.
Aya walks 9 km. Let her walking speed be s km/h, and the time in the coffee shop be t minutes.
1. When walking at speed s:
Total time: 9/s + t/60 = 4 ...
2. When walking at speed s + 2:
Total time: 9/(s + 2) + t/60 = 2.4
... By subtracting the two equations: ... We get t = 24 minutes.
Now, if Aya walks at s + 1/2 = 3 km/h:
- Walking time = 180 minutes
- Total time = 180 + 24 = 204 minutes
Thus, the total time for the walk including the coffee shop is boxed{**204**} minutes."

Long CoT (S1.1-32B) 3867 Tokens

Okay, so I need to find how long Aya's walk takes when she walks at s + 1/2 kilometers per hour, including the t minutes she spends in the coffee shop. Hmm, Let me start by breaking down the information given... The walk itself is 9 kilometers long each time.**Wait**, so the total time includes walking time plus t minutes at the coffee shop...Let me formalize that. Let's denote: Let's solve this equation for s. First, factor out the 9 ... **Alternatively**, let me check with the second equation as well ... But the question asks for the total time in minutes. So, 3 hours is 180 minutes, plus 24 minutes is 204 minutes.**Wait**, but let me confirm. The total time is walking time (3 hours) plus t minutes (24 minutes). However, walking time is in hours, so converting the entire total time into minutes.**Wait**, the problem says \"the walk takes her [time], including t minutes spent in the coffee shop... **Wait**, the walking time when speed is 3 km/h is 9 / 3 = 3 hours. So, 3 hours is 180 minutes. Then, adding the t minutes (24 minutes), total time is 180 + 24 = 204 minutes.But wait, is that correct? ...
Answer: Aya walks 9 kilometers every morning and stops at a coffee shop afterwards. We need to find the total time ...
boxed{**204**}

QFFT Long CoT 2675 Tokens

To solve the problem, we need to determine the time Aya spends walking and in the coffee shop for different speeds. ..
Step 1: Convert all time units to hours. Since speed is in kilometers per hour... ### Step 2: First scenario: Walking at *s* km/h ... Total time for the walk including coffee shop: 4 hours ... ### Step 3: Second scenario: Walking at s + 2 km/h ... Here, a = 8, b = 11, c = -90: s = (-11 ± sqrt(11^2 - 4 * 8 * (-90) * 5)) / (2 * 8) s = (-11 ± sqrt(121 + 2880)) / 16 s = (-11 ± sqrt(2991)) / 16. **(Short CoT pattern)**

Check if sqrt (2991) is an integer. 2991 = 9 * 331, so sqrt(2991) = 3 * sqrt(331), which is not an integer. This suggests an error in calculation.**Wait**, let's recheck the quadratic equation setup. **Wait,** after moving terms ... **Wait,** perhaps solving incorrectly. Let's compute discriminant ... Therefore, no real solution? Wait, check equations.Wait, original equations ... Walking time at 3.0 km/h is 9 / 3.0 = 3 hours = 180 minutes.Total time including t is 180 minutes + 24 minutes = 204 minutes.Thus, the answer is boxed{**204**} **(Long CoT pattern)**

Further Analysis

2. Long CoT Analysis

The long CoT of QFFT is triggered because of facing mistakes or uncertainty.

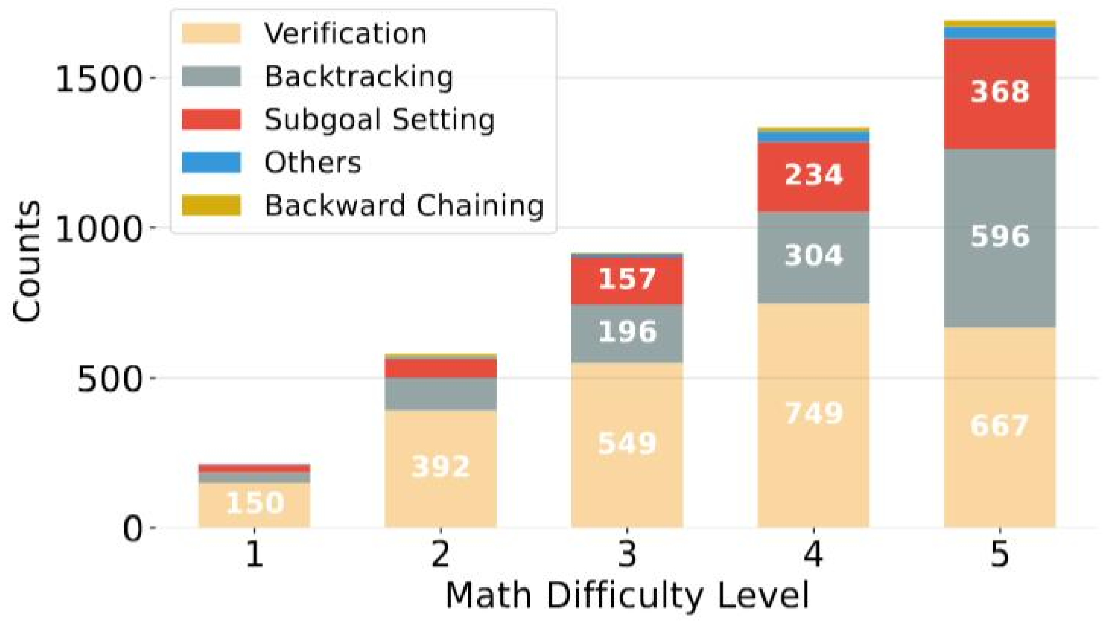
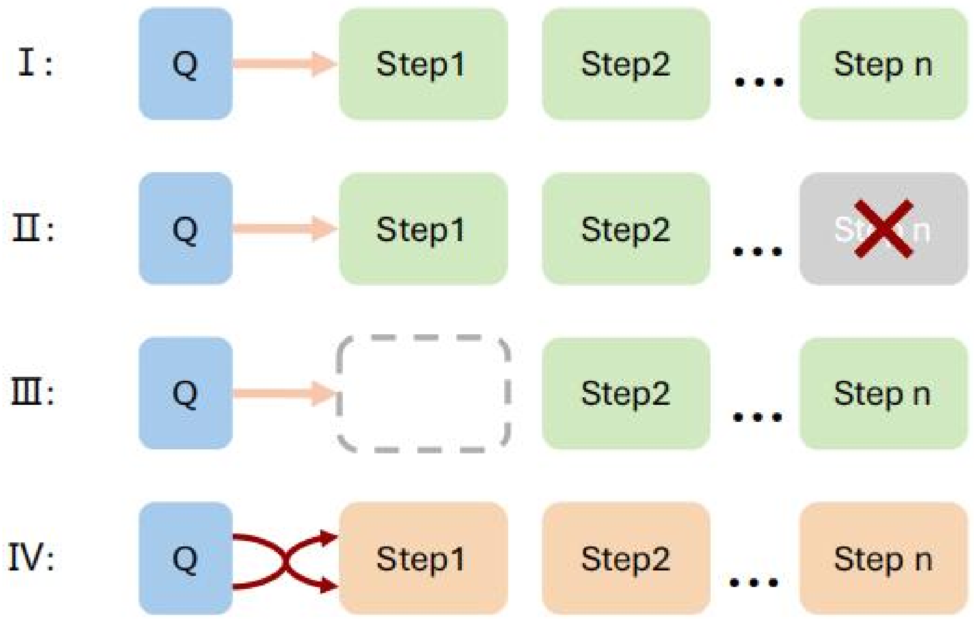


Figure 3: Long CoT Behavior Classification Trend on MATH500. The proportion of long CoT increases by difficulty level.

Applications of QFFT

1. Noisy Scenario

Level 1:， 该级别使用原始高质量数据，包含完整且正确的问答对。
Level 2: 结论错误, 该级别保留原始推理过程，但在最后步骤中加入数值错误，导致结论不正确。
Level 3: 推理不完整
该级别随机删减约一半的回答内容，同时保持句子语义完整。
Level 4: 答案无关
该级别生成完全不匹配的问答对，每个问题都搭配另一个问题的答案。



Applications of QFFT

1. Noisy Scenario

Level 1:，该级别使用原始高质量数据，包含完整且正确的问答对。

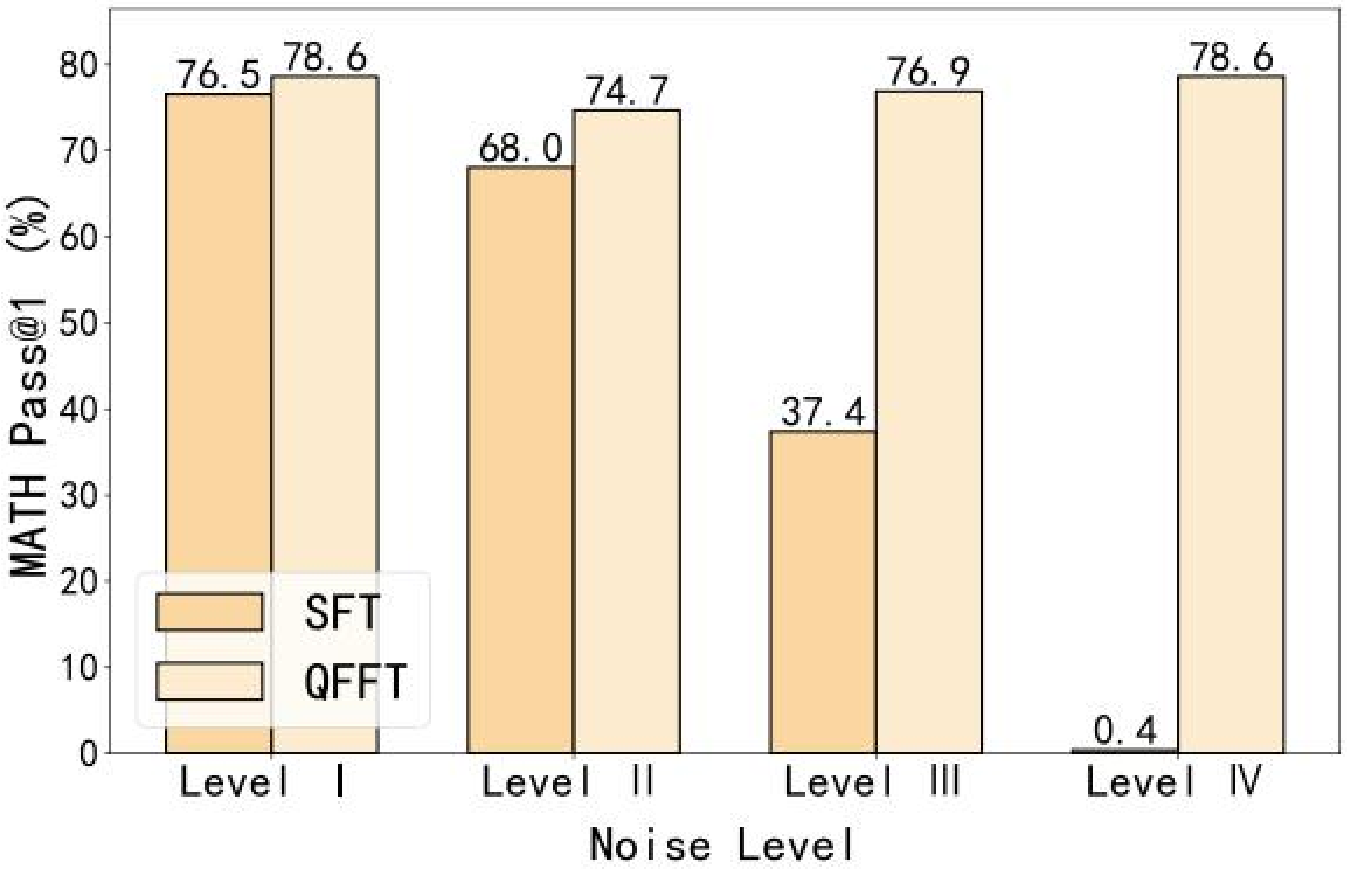
Level 2: 结论错误, 该级别保留原始推理过程，但在最后步骤中加入数值错误，导致结论不正确。

Level 3: 推理不完整

该级别随机删减约一半的回答内容，同时保持句子语义完整。

Level 4: 答案无关

该级别生成完全不匹配的问答对，每个问题都搭配另一个问题的答案。



Applications of QFFT

2. OOD Scenario

(1) Better performance on instruction following than SFT

(2) Better performance on Scientific and normal tasks

Data	MMLU-Pro	GPQA	LLM-AggreFact
Qwen-2.5-Instruct-7B	40.0	36.4	55.2
+ S1.1-SFT	43.1	41.8	25.5*
+ LIMO-SFT	29.3	43.2	8.9*
+ S1.1-QFFT	**53.2**	**44.4**	**58.9**
+ LIMO-QFFT	48.4	44.2	58.7
Qwen-2.5-Instruct-32B	62.2	49.5	75.3
+ S1.1-SFT	64.9	60.6	60.5
+ LIMO-SFT	52.4	65.3	62.2
+ S1.1-QFFT	**73.6**	65.7	**76.8**
+ LIMO-QFFT	73.4	**67.9**	75.8

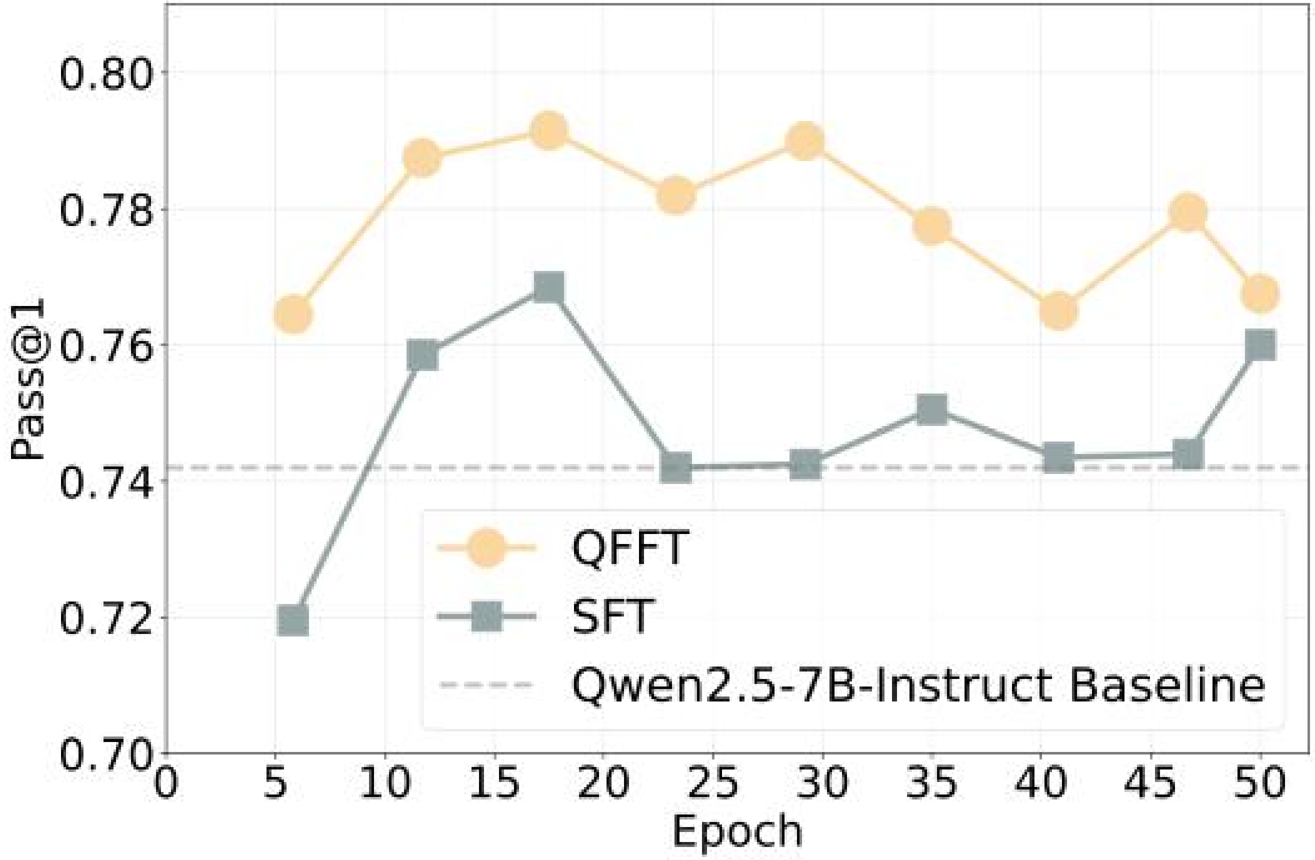
Applications of QFFT

3. Low-resource Scenario

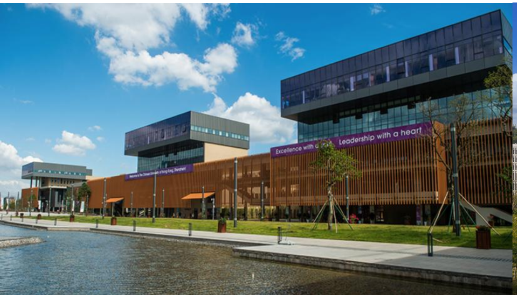
Experimental Setup：

Only 10 questions, distilled from Deepseek-R1 for 10 responses.

Totally 100 samples.



Thanks


Leadership with a heart


The Chinese University of Hong Kong, Shenzhen


深圳市大数据研究院
Shenzhen Research Institute of Big Data