



# Reasoning Scaffolding: Distilling the Flow of Thought from LLM

**Authors:** Xiangyu Wen, Junhua Huang, Zeju Li, Min Li, Jianyuan Zhong, Zhijian Xu, Mingxuan Yuan,  
Yongxiang Huang, Qiang Xu

**Presenter:** Xiangyu Wen



香港中文大學  
The Chinese University of Hong Kong



# Motivational Example

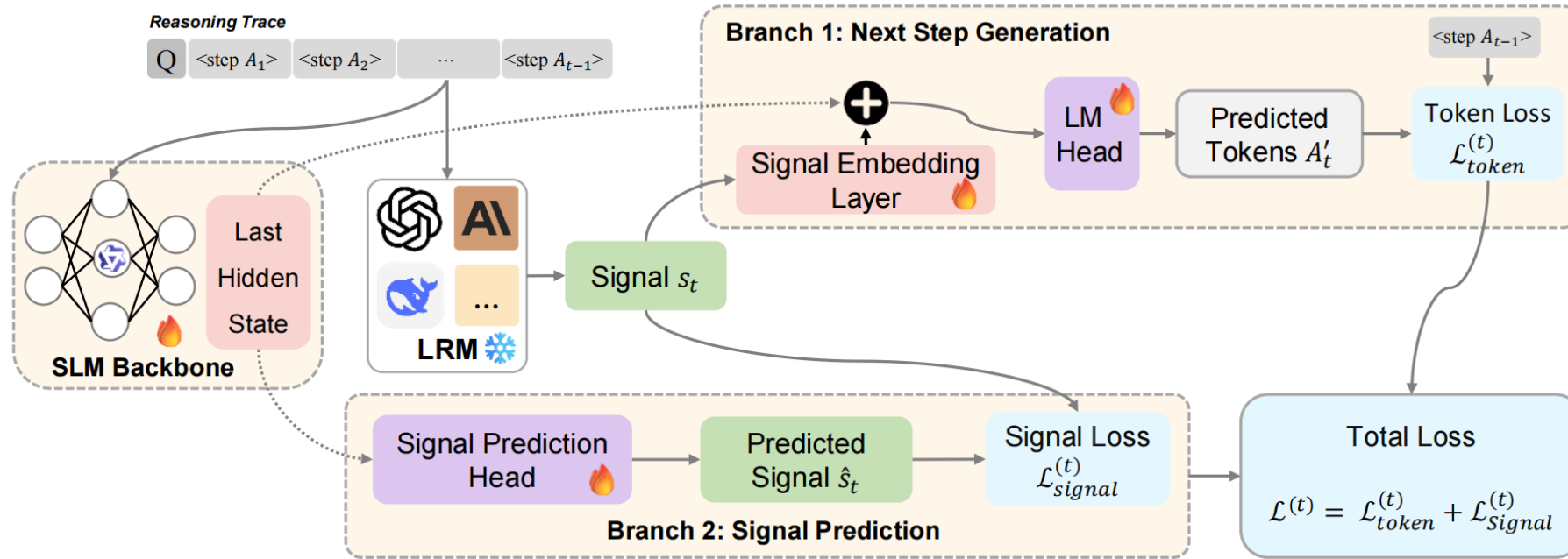
From the perspective of the reasoning logic, we may often observe there are many obvious semantic signals/hesitation words like ‘so’, ‘but’, and ‘then’, within the reasoning trace. These signals indicate when and where to do reflection and continual reasoning, so they build up the reasoning structure of a model.

Step Content	Signals	All Semantic Signals
... <b>Additionally</b> , she bakes muffins for her friends every day with four eggs. That means she uses four eggs for muffins each day.\nAfter these uses, she sells the remainder at the farmers' market.\nl need to find out how many eggs are left to sell, and then multiply by the price per egg to find her daily earnings.\n\n	Addition and Elaboration	Addition and Elaboration
... <b>So</b> , step by step:\n1. Total eggs laid per day: 16\n2. Eggs eaten for breakfast: 3\n3. Eggs used for muffins: 4\n\n	Conclusion and Summary	Examples and Illustration Personal Opinion and Recall
... <b>Then</b> , the eggs left for sale: total eggs minus eggs used = $16 - 7 = 9$ eggs.\nNow, she sells these 9 eggs at \$2 per egg.\n\n	Conclusion and Summary	Contrast and Concession Reasoning and Analysis
... <b>But</b> let me calculate again.\nTotal eggs: 16\nMinus breakfast: $16 - 3 = 13$ eggs left.\n\n	Contrast and Concession	Conclusion and Summary
... <b>Perhaps</b> the question is about whether she sells all or part, but it says remainder.\nAnother thought: does she sell the eggs daily, meaning she might have eggs from previous days?\n\n	Personal Opinion and Recall	Response Generation
... <b>Therefore</b> , the final answer must be 18.\n\n	Conclusion and Summary	
... Summary of above reasoning content and generate responses	Response Generation	

Figure 1: Examples from GSM8K illustrating clear logical transition signals that guide reasoning direction.



# Method Design



- Stage 1: Train the proposer model with two losses (token prediction + signal prediction)
  - Make the model sensitive to external semantic signals
  - Ensure the model can be explicitly guided by these signals during generation
- Stage 2: Train the semantic signal predictor using the annotated reasoning data with semantic signal labels
  - Learn to predict the next semantic signal from context
  - Provide control signals to guide the proposer at inference time

# Data Creation



Table 7: All keywords in the long-think trace and their corresponding semantic signal names

Semantic Signals	Keywords						
<b>Contrast and Concession</b>	but still	however although	on the other hand whereas	otherwise	nevertheless	nonetheless	in contrast
<b>Addition and Elaboration</b>	also	moreover	additionally	furthermore	in addition		
<b>Examples and Illustration</b>	for example	for instance					
<b>Personal Opinion and Recall</b>	i think i recall	i believe i remember	i guess	in my opinion	maybe	it seems	perhaps
<b>Reasoning and Analysis</b>	first i'll/need	actually let's see	in fact wait	let me ok	anyway well	by the way now	of course
<b>Conclusion and Summary</b>	so similarly	then unless	after all as a result	obviously therefore	clearly thus	indeed to conclude	meanwhile in conclusion

Prompts for Semantic Signal Checking

You are a judge who evaluates the semantic logic transition between a new sentence/paragraph (the "text") and its preceding context. Select the transition type strictly from the following list:

1. Contrast and Concession: The new sentence or paragraph presents a contrast or concession in relation to the previous context.
2. Addition and Elaboration: The new sentence or paragraph supplements or expands upon the previous context.
3. Examples and Illustration: The new sentence or paragraph provides examples or explanations related to the previous context.
4. Personal Opinion and Recall: The new sentence or paragraph primarily expresses a personal viewpoint or belief.
5. Reasoning and Analysis: The new sentence or paragraph serves as a starting point for reasoning or doing analysis based on the current information.
6. Conclusion and Summary: The new sentence or paragraph concludes from the previous context and present an intermediate result. outcome.

## Instructions: Given the previous context and the new sentence, analyze the semantic logic transition from the context to the text. Your answer should contain only the transition type, formatted as: `\boxed{logic transition name}`

## Context:  
{context}

## Text to be judged:  
{text}

## Your Response:

- Step 1 – Initial segmentation
  - Split long teacher reasoning traces into steps using double newlines (each segment = one step).
- Step 2 – Keyword-based initial labeling
  - Use a predefined keyword-to-signal map to assign each step an initial semantic signal if it starts with a keyword.
- Step 3 – LLM-based verification & correction
  - Feed context + step + initial label into a strong LLM to check consistency; if mismatched, let the LLM relabel with the correct semantic signal.
  - For steps without any keyword, let the LLM directly assign the signal.
- Step 4 – Build golden-labeled dataset
  - Collect all LLM-verified labels as golden semantic signals for training the proposer and the signal predictor.

# Experimental Results



Table 6: All keywords in the long-think trace and their corresponding semantic signal names

Semantic Signals	Keywords						
<b>Contrast and Concession</b>	but	however	on the other hand	otherwise	nevertheless	nonetheless	in contrast
<b>Addition and Elaboration</b>	still	although	whereas				
<b>Examples and Illustration</b>	also	moreover	additionally	furthermore	in addition		
<b>Personal Opinion and Recall</b>	for example	for instance					
<b>Reasoning and Analysis</b>	i think	i believe	i guess	in my opinion	maybe	it seems	perhaps
<b>Conclusion and Summary</b>	i recall	i remember					
	first	actually	in fact	let me	anyway	by the way	of course
	i'll/need	let's see	wait	ok	well	now	
	so	then	after all	obviously	clearly	indeed	meanwhile
	similarly	unless	as a result	therefore	thus	to conclude	in conclusion

Table 2: Signal prediction accuracy.

Model Size	StrategyQA	CommonsenseQA	TruthfulQA	GSM8K	MATH-500	Averaged
Qwen 0.5B	0.748	0.739	0.732	0.729	0.737	0.737
+ Adaptive	0.783	0.777	0.791	0.778	0.784	0.783
Qwen 7B	0.791	0.788	0.802	0.801	0.796	0.796
+ Adaptive	0.839	0.841	0.837	0.835	0.837	0.838
Qwen 14B	0.841	0.843	0.841	0.829	0.836	0.838
+ Adaptive	0.857	0.849	0.855	0.858	0.849	0.854

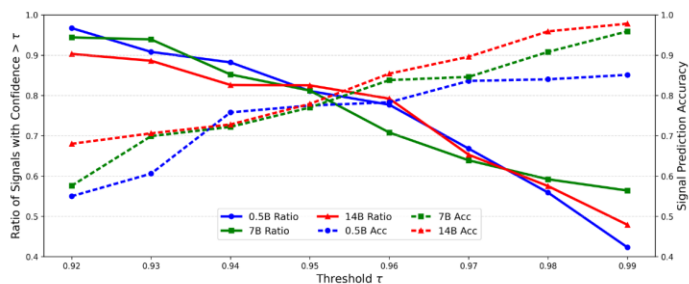


Figure 5: The effects of threshold  $\tau$  on the ratio of predicted signals with log probability  $> \tau$  and the signal prediction accuracy.

Table 1: Main results (Pass@1)

Methods	Models	StrategyQA	CommonsenseQA	TruthfulQA	GSM8K	MATH-500
<b>Original</b>	Qwen2.5-0.5B-Ins	0.543	0.475	0.268	0.379	0.335
	Qwen2.5-7B-Ins	0.726	0.785	0.706	0.875	0.738
	Qwen2.5-14B-Ins	0.755	0.785	0.750	0.921	0.764
	Llama3.1-8B-Ins	0.709	0.646	0.642	0.823	0.514
<b>CoT SFT</b>	Qwen2.5-0.5B-Ins	0.569	0.457	0.682	0.394	0.351
	Qwen2.5-7B-Ins	0.751	0.808	0.752	0.887	0.851
	Qwen2.5-14B-Ins	0.760	0.810	0.831	0.928	0.882
<b>Long Thinking SFT<sup>a</sup></b>	Qwen2.5-0.5B-Ins	0.571	0.463	0.670	0.412	0.388
	Qwen2.5-7B-Ins	0.759	0.817	0.771	0.862	0.879
	Qwen2.5-14B-Ins	0.768	0.845	0.812	0.931	0.901
<b>Long Thinking Distill<sup>b</sup></b>	Qwen2.5-7B-Ins	0.680	0.670	0.544	0.867	0.883
	Qwen2.5-14B-Ins	0.811	0.805	0.763	0.936	0.904
<b>GPT-oss-120B</b>	Few-Shot	0.783	0.825	0.862	0.768	0.872
<b>Deepseek-R1</b>	Few-Shot	<b>0.863</b>	<b>0.895</b>	<b>0.874</b>	<b>0.961</b>	<b>0.965</b>
<b>Ours (Teacher=GPT-oss)</b>	Qwen2.5-7B-Ins	0.748	0.822	0.868	-	0.850
	Llama3.1-8B-Ins	0.783	0.841	0.843	0.864	0.844
<b>Ours (Teacher=DS-R1)</b>	Qwen2.5-0.5B-Ins	0.659	0.601	0.861	0.488	0.417
	Qwen2.5-7B-Ins	0.832	0.866	0.879	0.899	0.922
	Qwen2.5-14B-Ins	<b>0.858</b>	<b>0.887</b>	<b>0.917</b>	<b>0.942</b>	<b>0.928</b>

<sup>a</sup> Official distilled model released by Deepseek.

<sup>b</sup> Our SFT model trained on long-thinking data distilled and generated by Deepseek-R1.

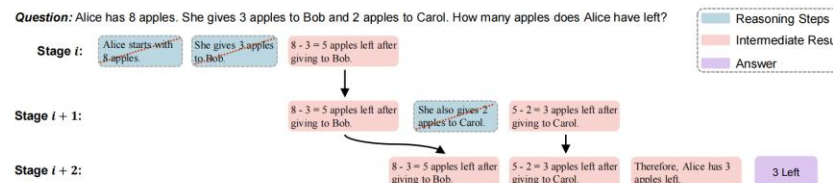


Figure 3: Token efficiency: By pruning reasoning steps before those labeled 'Conclusion and Summary', only intermediate results are retained.

Table 4: Token consumption analysis under different strategies.

Methods	StrategyQA	CommonsenseQA	TruthfulQA	GSM8K	MATH-500
Original	224	217	282	304	604
CoT SFT	438	392	509	343	715
Thinking Distill	597	613	705	771	2,616
Thinking SFT	1,878	1,959	1,877	1,996	5,921
<b>Ours + All Signals</b>	1,524	1,638	1,550	1,659	4,755
- Remove Reasoning Steps	746	823	773	845	2,628

**Thank you!**

---