ACL 2026 SAN DIEGO JULY 2–7

Guided By Gut: Efficient Test-Time Scaling with Reinforced Intrinsic Confidence

Amirhosein Ghasemabadi[1], Keith G. Mills[2], Baochun Li[3] and Di Niu[1]

[1]Dept. ECE, UAlberta [2]Division of CSE, LSU [3]Dept. ECE, UToronto

Paper

Project

LSU College of Engineering

The Edward S. Rogers Sr. Department of Electrical & Computer Engineering
UNIVERSITY OF TORONTO

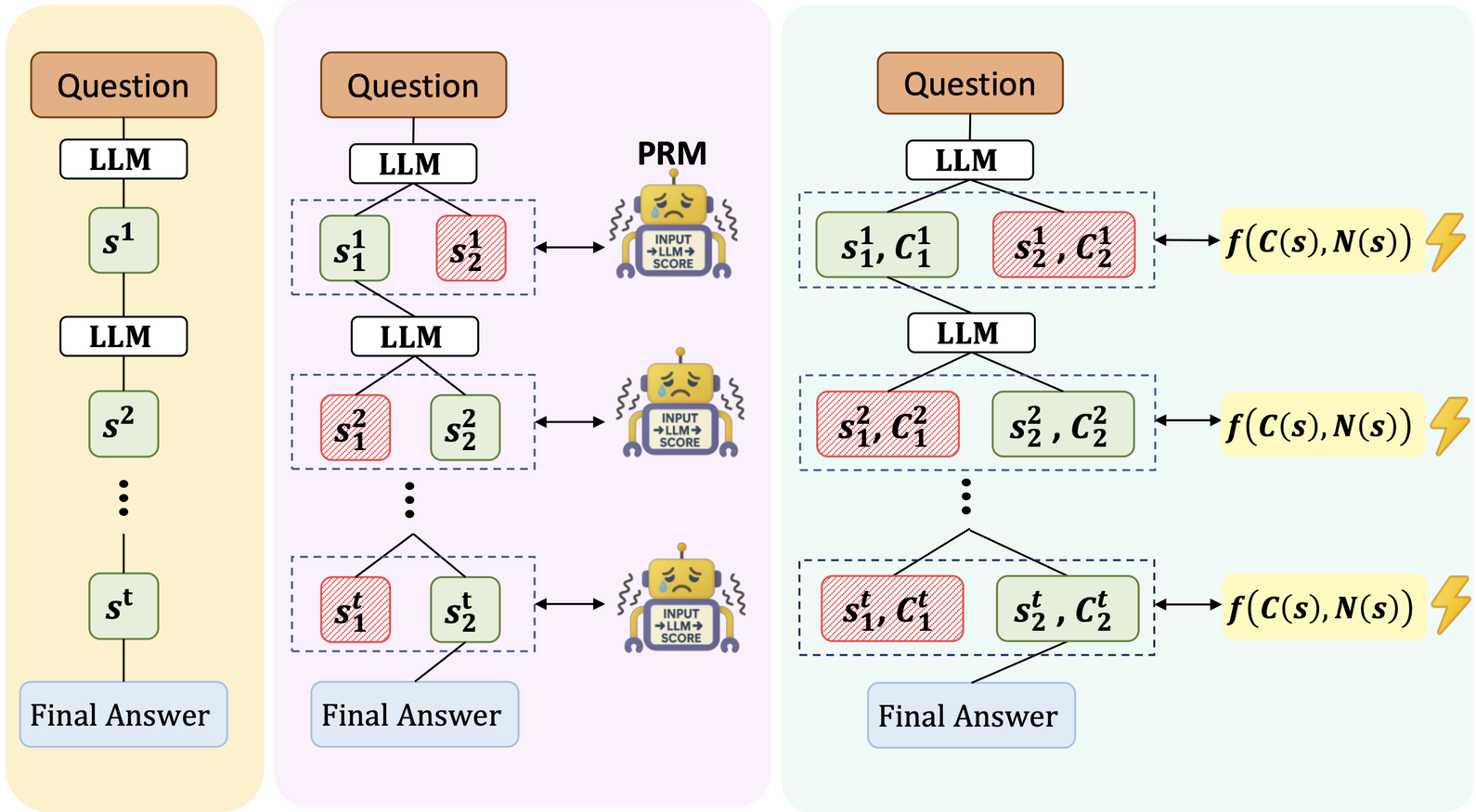
Large Language Model Reasoning

Chain-of-Thought (CoT)

- Go-to for larger (>100B param) models.

Best-of-N (BoN)

- Tree-based approach.
- Search + Rewards
- Helps smaller (~10B or <10B param) models reach the performance of larger ones.
- However, use of external verifiers, like Process Reward Models (PRM) impose compute cost.



LLM Internal Confidence Statistics

LLM internal statistics **already** provide information on whether an answer is good.

$$p(s_i^t) = \prod_i p(s_i^t | s_{<i}^t)$$

Next Token Prediction: Each token s_i^t in an answer is picked from a distribution, $p(s_i^t)$.

Higher $p(s_i^t)$ means the LLM is more **confident** that the token should be selected.

GG: Ditch PRMs for Confidence & Novelty

$$r_t = C(s^t) + \lambda_N N(s^t)$$

- $C(s^t)$ – *Confidence;* summation computed over logits of the tokens in response s^t.
- $N(s^t)$ – *Novelty;* Proportion of new tokens considered against existing context.
- Correlate **confidence** with **correctness** using GRPO RLHF calibration.

GG: Guided by Gut

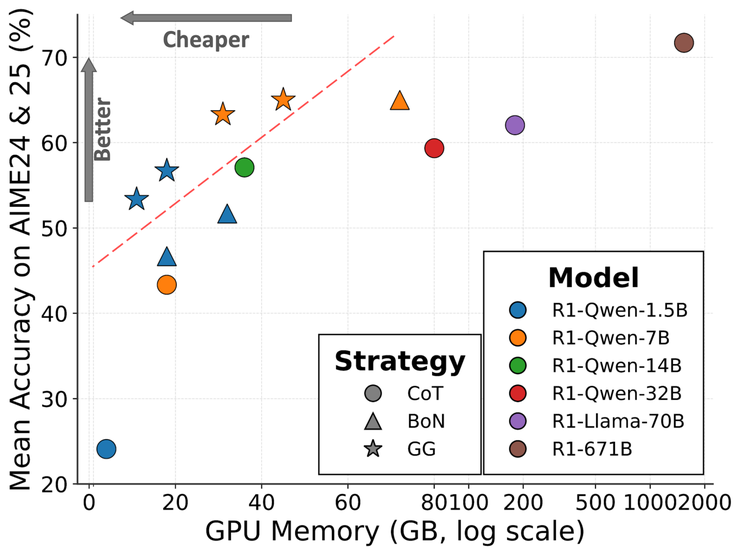
Our method replaces the costly PRM in tree-based Test-Time Scaling methods with internal confidence measures.

We test it on several reasoning benchmarks such as AIME24/25, MATH500 & AMC.

Model	TTS	N	Accuracy [%] ↑				KV (GB) ↓
			AIME24	AIME25	MATH500	AMC	
DS-R1-Qwen 1.5B	CoT	–	26.8	21.4	83.9	68.3	0.86
	BoN	32	52.2 [56.7]	34.4 [36.7]	91.7	90.0	13.7
	GG	32	58.9 [66.7]	37.5 [40.0]	91.9	92.5	6.9
	BoN	64	57.8 [66.7]	36.7 [36.7]	92.3	90.0	27.4
	GG	64	61.7 [66.7]	42.2 [46.7]	92.9	93.3	13.7
DS-R1-Qwen 7B	CoT	–	48.1	38.6	92.8	85.5	1.7
	BoN	32	72.2 [73.3]	51.1 [53.3]	96.1	92.5	27.2
	GG	32	71.7 [73.3]	52.5 [53.3]	96.3	94.1	13.7
	BoN	64	75.5 [76.7]	50.0 [53.3]	96.1	93.7	54.4
	GG	64	76.7 [76.7]	51.7 [53.3]	96.5	95.0	27.2

GG outperforms CoT/BoN on all of these metrics. Also, it significantly reduces the KV cache GB cost compared to BoN.

Acc. Increases & Memory Savings



Not only does GG outperform BoN at the same evaluation budget using fewer resources, it also enables smaller reasoning LLMs, like R1-Qwen-7B, to compute with much larger models like R1-671B at a fraction of the GPU memory cost.