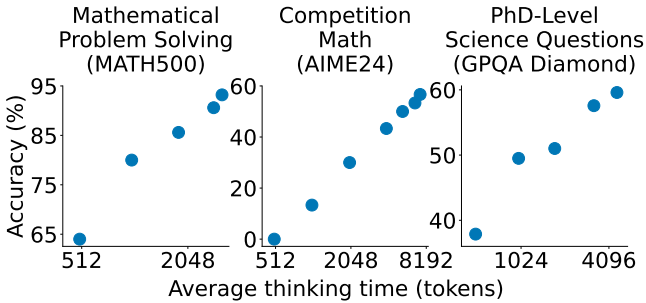
Multilingual Reasoning

Alan Ritter
Georgia Institute of Technology

Multilingual Multicultural LLMs
EMNLP 2026

Reasoning (Test Time Scaling)

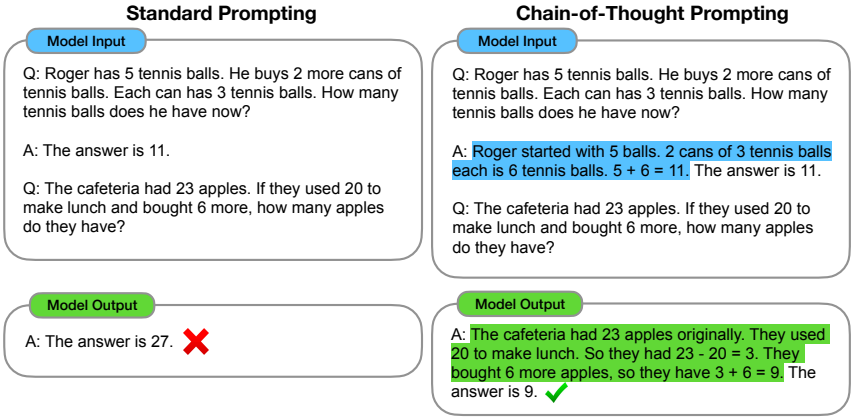


s1-32B: Qwen2.5-32B-Instruct fine-tuned on 1,000 reasoning examples (s1K).

Budget forcing: stop earlier, or append “Wait” to continue.

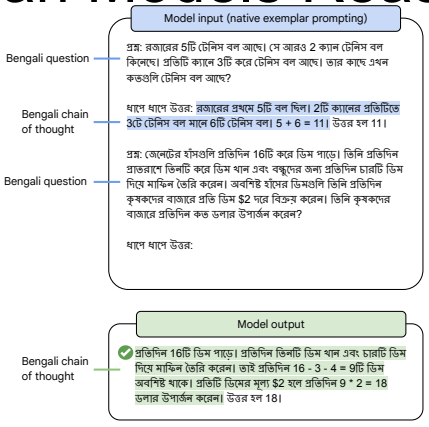
N. Muennighoff, Z. Yang, W. Shi, X. L. Li, F.-F. Li, H. Hajishirzi, L. Zettlemoyer, P. Liang, E. Candès, and T. Hashimoto (2025). “s1: Simple test-time scaling”. *arXiv preprint arXiv:2501.19393v1*. s1-32B.

Chain-of-Thought Prompting



J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. H. Chi, Q. V. Le, and D. Zhou (2022). "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models". *Advances in Neural Information Processing Systems*.

Can Models Reason in Bengali?



English translation of model output

16 eggs are laid each day. She eats three and uses four to make muffins.

So $16 - 3 - 4 = 9$ eggs remain each day. At $2 per egg, she earns $9 \times 2 = 18$ dollars per day.

The answer is 18.

F. Shi, M. Suzgun, M. Freitag, X. Wang, S. Srivats, S. Vosoughi, H. W. Chung, Y. Tay, S. Ruder, D. Zhou, D. Das, and J. Wei (2023). “Language Models are Multilingual Chain-of-Thought Reasoners”. *International Conference on Learning Representations*.

Multilingual CoT: Early Results

PaLM-540B; MGSM accuracy (%); native-language exemplars

Prompting	English	Bengali
Direct answer	22.0	17.2
Native-language CoT	62.4	46.0

CoT substantially improves Bengali performance.
Reasoning in Bengali is less accurate than in English.

Shi et al. (2023).

Reasoning Across Languages

Advances in reasoning are not equally accessible in every language.

AIME-Combined; DeepSeek-R1-Distill-Llama-70B (based on Llama 3.3)

Language	English	Chinese	French	Telugu	Swahili
Score (%)	56.3	39.7	41.7	17.3	25.3

Users should get equally good answers in their own language.

Question and reasoning in the same language. Score requires a correct, language-compliant answer.

Barua et al. (2026).

Models Often Just Reason in English

Distilled-R1-32B; AIME

French questions; instructed to reason in French

Detected reasoning language	Percentage of traces
English	96.67%
French	3.33%

J. Qi, S. Chen, Z. Xiong, R. Fernández, D. S. Bitterman, and A. Bisazza (2025). “When Models Reason in Your Language: Controlling Thinking Language Comes at the Cost of Accuracy”. *Findings of the Association for Computational Linguistics: EMNLP 2025*.

Models Often Just Reason in English

Distilled-R1-32B; AIME

French questions; instructed to reason in French

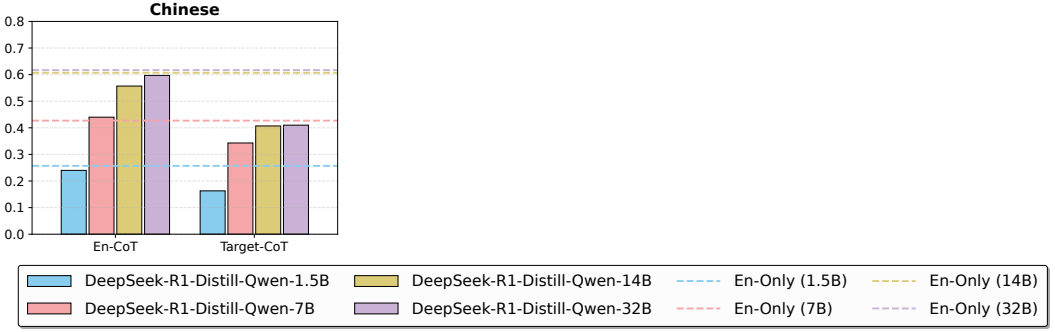
Detected reasoning language	Percentage of traces
English	96.67%
French	3.33%

Seems undesirable if users can't understand the reasoning

J. Qi, S. Chen, Z. Xiong, R. Fernández, D. S. Bitterman, and A. Bisazza (2025). "When Models Reason in Your Language: Controlling Thinking Language Comes at the Cost of Accuracy". *Findings of the Association for Computational Linguistics: EMNLP 2025*.

Scaling Helps, but Not Equally Across Languages

Released DeepSeek-R1-Distill-Qwen models (1.5B–32B)

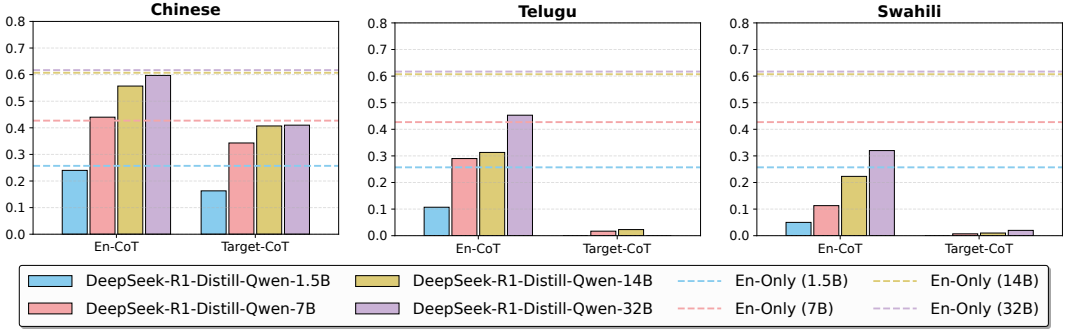


Target-language questions: **En-CoT** reasons in English; **Target-CoT** in the target language. Dashed lines: English question and reasoning.

Barua et al. (2026). AIME-Combined; correct answer + reasoning in the requested language.

Scaling Helps, but Not Equally Across Languages

Released DeepSeek-R1-Distill-Qwen models (1.5B–32B)

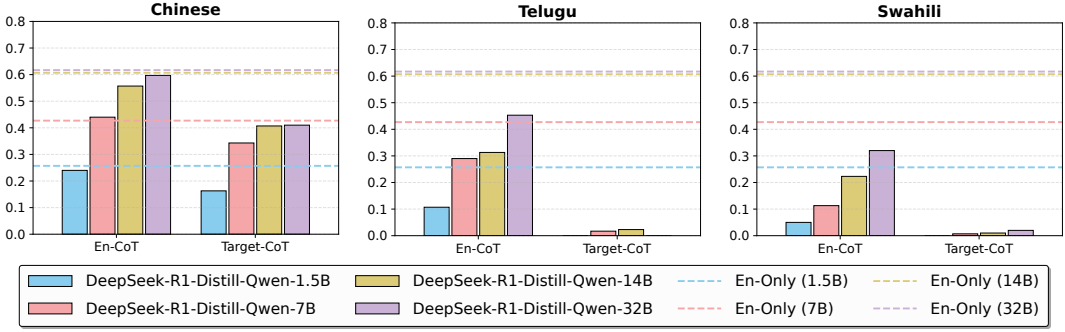


Target-language questions: **En-CoT** reasons in English; **Target-CoT** in the target language. Dashed lines: English question and reasoning.

Barua et al. (2026). AIME-Combined; correct answer + reasoning in the requested language.

Scaling Helps, but Not Equally Across Languages

Released DeepSeek-R1-Distill-Qwen models (1.5B–32B)



Target-language questions: **En-CoT** reasons in English; **Target-CoT** in the target language. Dashed lines: English question and reasoning.

Main Takeaway: English reasoning generally performs better.

Barua et al. (2026). AIME-Combined; correct answer + reasoning in the requested language.

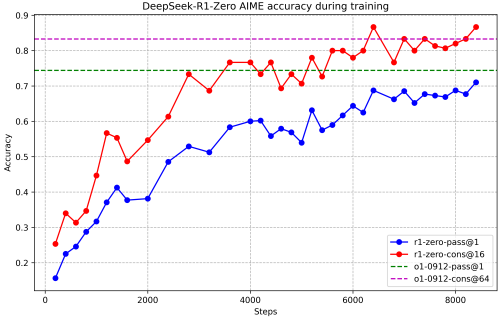
Outline

1. Introduction and background
2. Long chain-of-thought reasoning and datasets
3. Cross-lingual transfer for reasoning

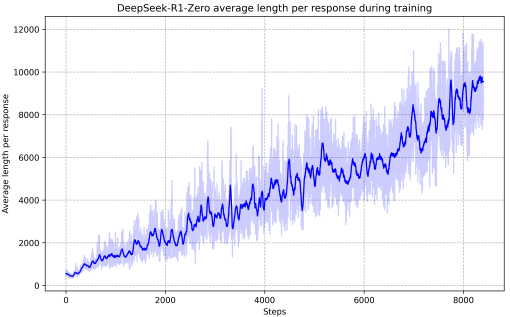
Learning Longer Reasoning with RL

DeepSeek-R1-Zero: 671B total parameters; 37B active per token (MoE)

Accuracy



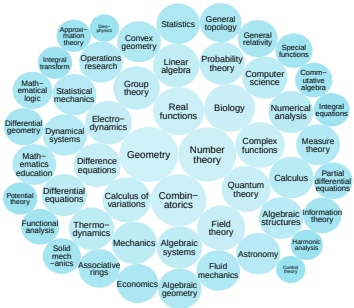
Response length



As RL training progresses, accuracy improves and responses get longer.

DeepSeek-AI (2025). “DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning”. *arXiv preprint arXiv:2501.12948v1*.

s1K: Training Data



Muennighoff et al. (2025).

1,000 questions selected from 59K
Sources include NuminaMATH, MATH, OmniMath, and OlympicArena.

Teacher traces → SFT

- Original s1K: Gemini
- s1K-1.1: DeepSeek-R1

Student: s1-32B
Qwen2.5-32B-Instruct (32B parameters)

Translated-s1K

s1K-1.1 (R1 traces)

- **Translated-s1K:** translate traces with Gemini-2.0-Flash
- **Distilled-s1K:** elicit target-language traces from DeepSeek-R1

Qwen3-8B-Base; MATH-500

Training traces	English	Non-English avg.
English s1K	92.6	33.4
Translated-s1K	92.6	**78.3**
Distilled-s1K	92.6	72.8

J. Barua, S. Eisape, K. Yin, and A. Suhr (2026). “Long Chain-of-Thought Reasoning Across Languages”. *International Conference on Learning Representations (ICLR)*.

Translated-s1K

s1K-1.1 (R1 traces)

- **Translated-s1K:** translate traces with Gemini-2.0-Flash
- **Distilled-s1K:** elicit target-language traces from DeepSeek-R1

Qwen3-8B-Base; MATH-500

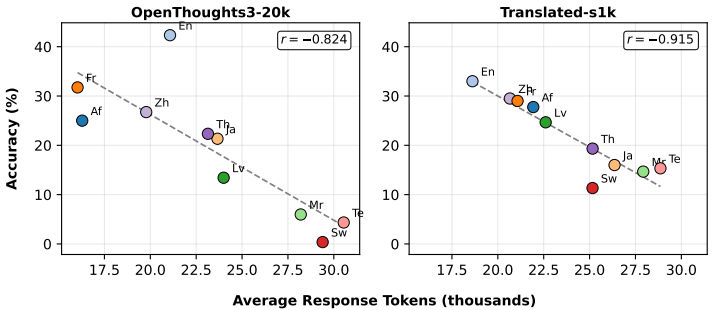
Training traces	English	Non-English avg.
English s1K	92.6	33.4
Translated-s1K	92.6	**78.3**
Distilled-s1K	92.6	72.8

Translated traces outperform directly distilled target-language traces here.

- Non-English average over nine languages.

J. Barua, S. Eisape, K. Yin, and A. Suhr (2026). “Long Chain-of-Thought Reasoning Across Languages”. *International Conference on Learning Representations (ICLR)*.

Inference Cost



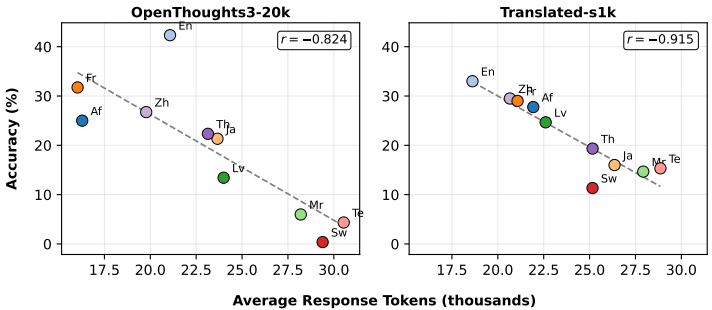
AIME-Combined

Strong negative correlation across languages: more thinking tokens, lower performance.

Tuning on translated data narrows the gap.

Barua et al. (2026).

Inference Cost



AIME-Combined

Strong negative correlation across languages: more thinking tokens, lower performance.

Tuning on translated data narrows the gap.

Main Takeaway: Reasoning in lower-resource languages can be less accurate and more costly.

Barua et al. (2026).

Outline

1. Introduction and background
2. Long chain-of-thought reasoning and datasets
3. Cross-lingual transfer for reasoning

Transferring Reasoning Capabilities Across Languages

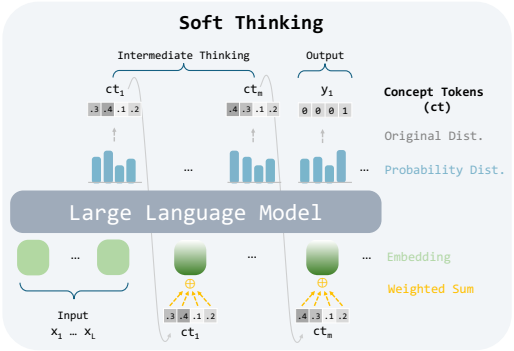
1. **SOLAR:** align soft-token representations across languages
   Background: soft-token representations.

2. **COPSD:** learn from an English-informed teacher
   Background: on-policy self-distillation.

J. He, J. Park, W. Xu, and A. Ritter (2026). “Soft Token Alignment for Cross-Lingual Reasoning”. *Advances in Neural Information Processing Systems (NeurIPS)*.

Y. Liu, R. Zhao, M. A. Hedderich, and H. Schütze (2026). “Crosslingual On-Policy Self-Distillation for Multilingual Reasoning”. *arXiv preprint arXiv:2605.09548v1*.

Background: Soft Tokens



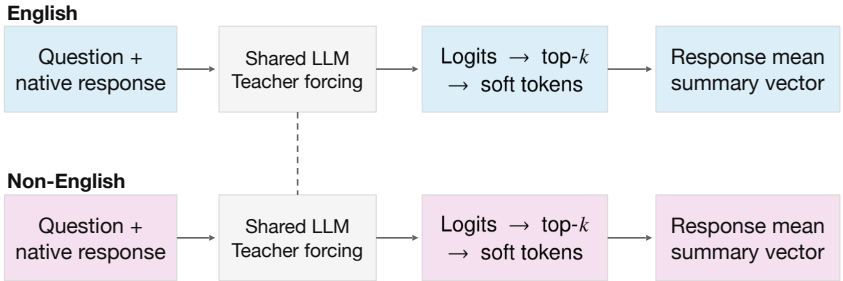
Instead of one token:

$$e_t^{\text{soft}} = \sum_{v \in V} p_t(v) E_v$$

Feed the weighted embedding back into the model.

Z. Zhang, X. He, W. Yan, A. Shen, C. Zhao, S. Wang, Y. Shen, and X. E. Wang (2025). “Soft Thinking: Unlocking the Reasoning Potential of LLMs in Continuous Concept Space”. *arXiv preprint arXiv:2505.15778*.

SOLAR: Summarize Each Response



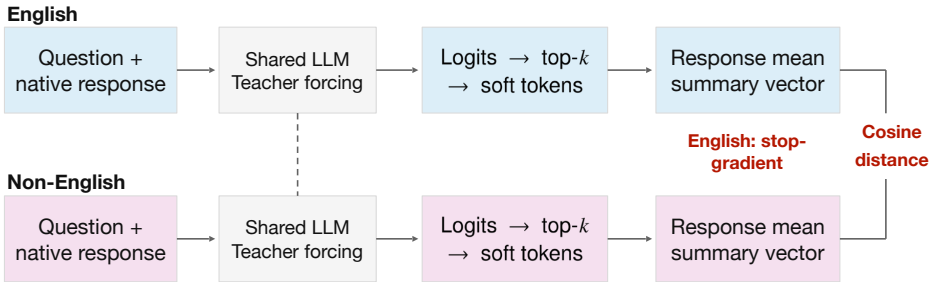
Mean-pool all supervised response positions

Training data: Translated-s1K (M-s1k)

1,000 parallel problems in 10 languages.

He et al. (2026); Barua et al. (2026).

SOLAR: Align with English

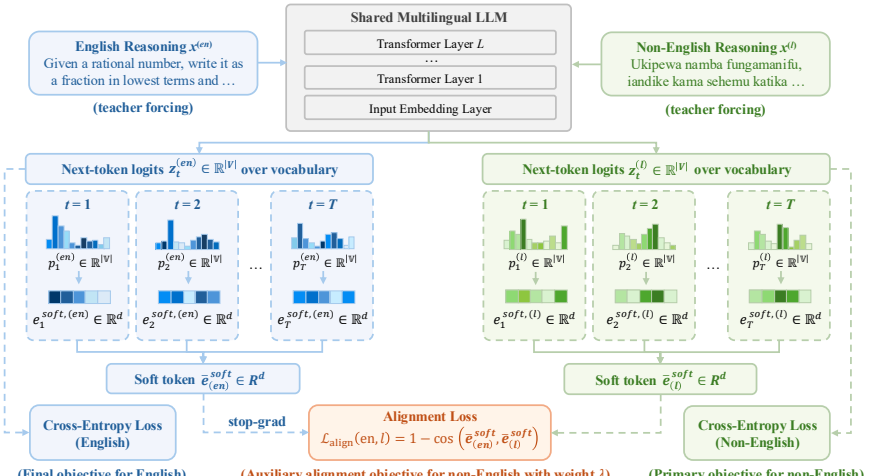


$$\mathcal{L}_{\text{align}}(\text{en}, \ell) = 1 - \cos\left(\text{sg}\left(\bar{\mathbf{e}}_{\text{en}}^{\text{soft}}\right), \bar{\mathbf{e}}_{\ell}^{\text{soft}}\right),$$

$$\mathcal{L}_{\text{SFT}}^{(\ell)} = \mathcal{L}_{\text{CE}}^{(\ell)} + \lambda\, \mathcal{L}_{\text{align}}(\text{en}, \ell).$$

He et al. (2026).

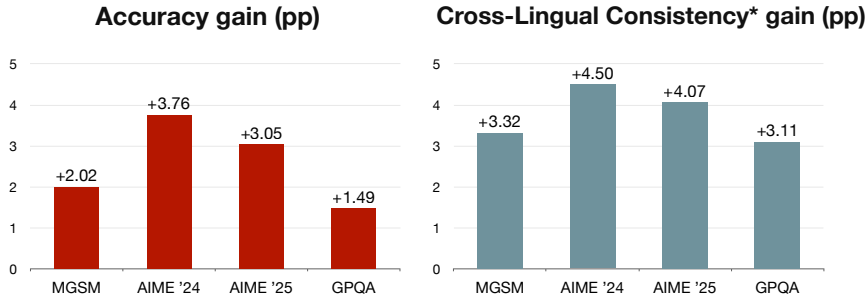
SOLAR: Full Training Objective



He et al. (2026).

SOLAR: Results

Qwen3-4B: SOLAR gains over SFT (percentage points)

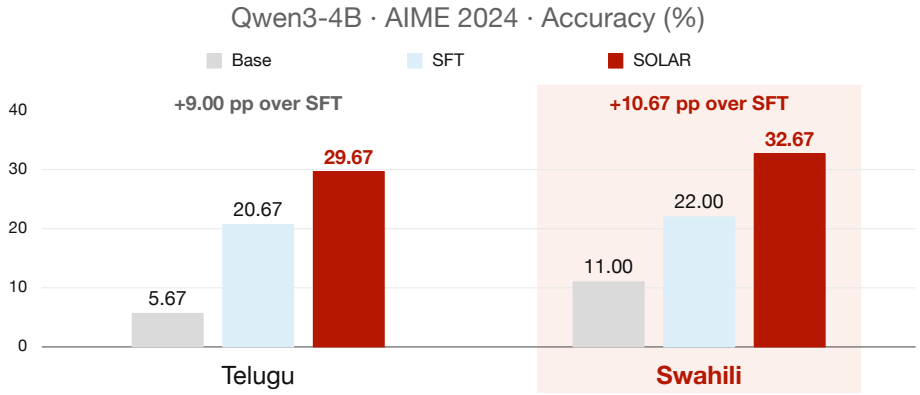


+2.58 points average accuracy over SFT.

Consistency: both the English and non-English versions of the same problem are answered correctly.

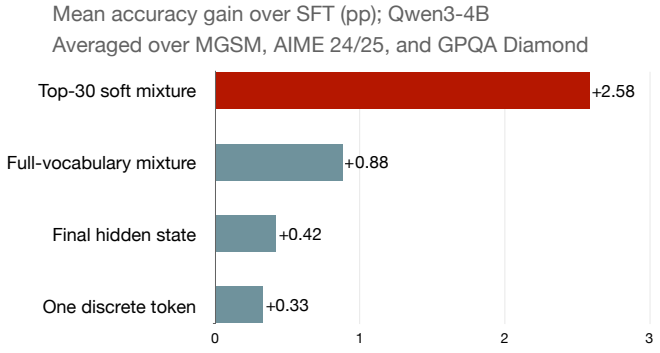
He et al. (2026).

SOLAR: Telugu and Swahili



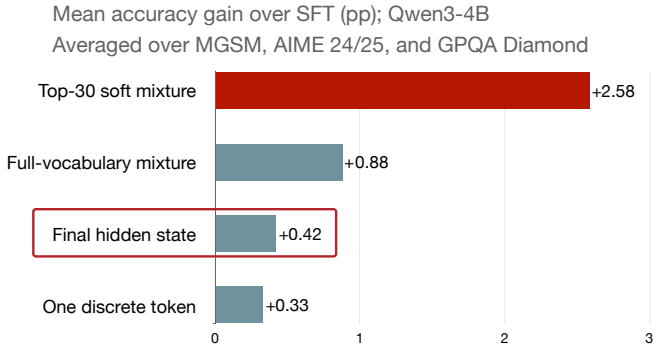
He et al. (2026).

SOLAR: Why Align Soft Tokens?



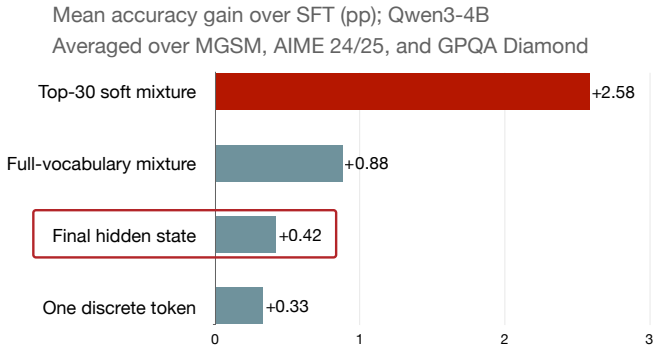
He et al. (2026).

SOLAR: Why Align Soft Tokens?



He et al. (2026).

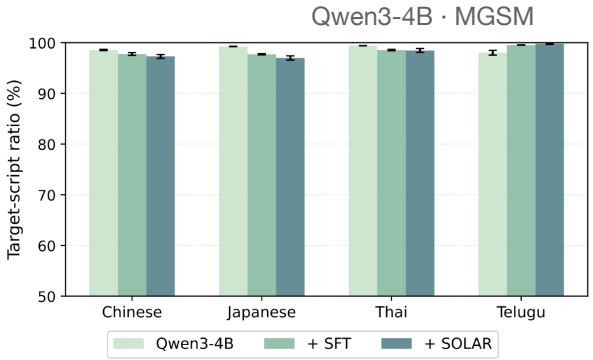
SOLAR: Why Align Soft Tokens?



Possible explanation: a more direct gradient path to the token embeddings.

He et al. (2026).

Does SOLAR Switch to English?



98.13%

SOLAR's average target-script ratio

SFT: 98.39%
Base: 98.81%

Target-script ratio = target-script tokens / (target-script + Latin tokens).

He et al. (2026).

SOLAR: Baseline Comparisons

Qwen3-4B; mean accuracy over four benchmarks.

Method	Approach	Accuracy (%)
Soft Thinking	Inference-time soft tokens	39.19
SFT	Translated reference traces	45.67
MidAlign	Middle-layer contrastive alignment	46.29
MAPO	Translation-scored preference training	44.60
AlignX	Hidden alignment + language prediction	46.12
SOLAR	Soft-token alignment	**48.25**

MGSM, AIME 2024, AIME 2025, and GPQA Diamond.

MidAlign: D. Liu and Niehues (2025); MAPO: She et al. (2024); AlignX: Bu et al. (2025).

Results: He et al. (2026).

Transferring Reasoning Capabilities Across Languages

1. **SOLAR:** align soft-token representations across languages
   Background: soft-token representations.

2. **COPSD:** learn from an English-informed teacher
   Background: on-policy self-distillation.

Liu et al. (2026); He et al. (2026).

Translated Traces Have Limitations

Translation can introduce errors in reasoning

- Errors in quantities, mathematical expressions, or logic

The reference may not look like the model's own reasoning

- Train on translated reference prefixes
- At inference, continue the model's own generated prefixes

Can we train on the student's own trajectories instead?

Liu et al. (2026).

Learn from the Student's Trajectories

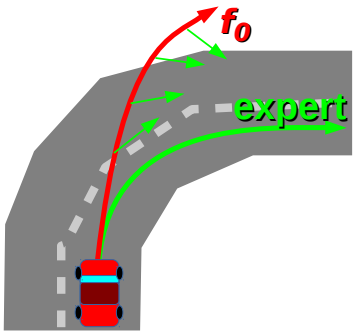


Illustration: Daumé III (2017).

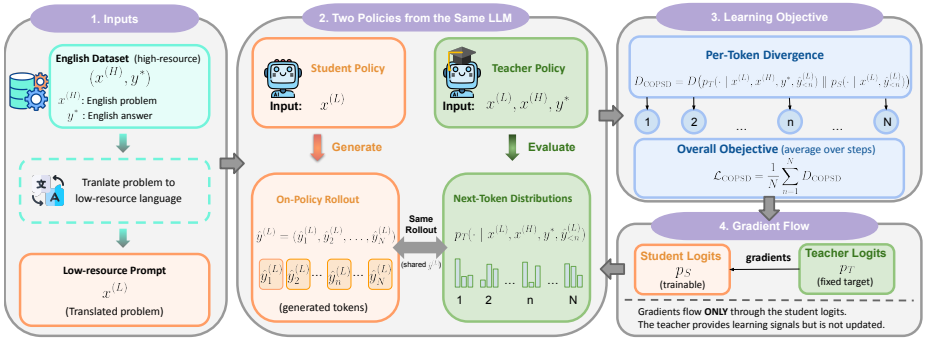
Watching an expert drive does not teach recovery.

Red: the learner's trajectory.
Green arrows: expert feedback along that trajectory.

For an LLM:
Let the student generate.
Ask the teacher how to continue.

S. Ross, G. J. Gordon, and J. A. Bagnell (2011). "A Reduction of Imitation Learning and Structured Prediction to No-Regret Online Learning". *International Conference on Artificial Intelligence and Statistics (AISTATS)*. DAgger.

COPSD: Crosslingual Self-Distillation



Student: target-language question

Teacher: also gets the English question and reference solution

Liu et al. (2026).

COPSD: Results

AfriMGSM: 17 African languages; 4,096-token limit per response.

Model	Base	GRPO	COPSD
Qwen3-1.7B	9.11	9.18	**15.53**
Qwen3-4B	19.20	19.36	**20.61**
Qwen3-8B	19.41	19.22	**23.55**

Average Pass@12 (%)

Pass@12: at least one correct answer among twelve samples.

Improves performance without target-language reference reasoning traces.

Liu et al. (2026).

Takeaways

1. Non-English reasoning often performs worse. We need better cross-lingual transfer.
2. SFT on translated reasoning traces is a strong baseline.
3. SOLAR and COPSD improve transfer, but more work is still needed.

Barua et al. (2026); Liu et al. (2026); He et al. (2026).

Open Questions

- Do users trust answers more when the reasoning is in their own language? This needs a user study.
- Does reasoning language change the solution strategy or conclusion? (Sapir–Whorf Hypothesis)
- How important is native-language reasoning for culturally grounded questions beyond math?

Open Questions

- Do users trust answers more when the reasoning is in their own language? This needs a user study.
- Does reasoning language change the solution strategy or conclusion? (Sapir–Whorf Hypothesis)
- How important is native-language reasoning for culturally grounded questions beyond math?

Thank You!

References

Barua, Josh et al. (2026). “Long Chain-of-Thought Reasoning Across Languages”. In: *International Conference on Learning Representations (ICLR)*. URL: https://arxiv.org/abs/2508.14828.

Bu, Mengyu et al. (2025). “AlignX: Advancing Multilingual Large Language Models with Multilingual Representation Alignment”. In: *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pp. 6460–6489. URL: https://aclanthology.org/2025.emnlp-main.328/.

Daumé III, Hal (2017). *A Course in Machine Learning*. 0.99. URL: http://ciml.info/dl/v0_99/ciml-v0_99-ch18.pdf.

DeepSeek-AI (2025). “DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning”. In: *arXiv preprint arXiv:2501.12948v1*. URL: https://arxiv.org/abs/2501.12948v1.

Hao, Shibo et al. (2024). “Training Large Language Models to Reason in a Continuous Latent Space”. In: *arXiv preprint arXiv:2412.06769*. URL: https://arxiv.org/abs/2412.06769.

He, Jiayi et al. (2026). “Soft Token Alignment for Cross-Lingual Reasoning”. In: *Advances in Neural Information Processing Systems (NeurIPS)*. URL: https://arxiv.org/abs/2606.26466. Forthcoming.

References

Lai, Huiyuan and Malvina Nissim (2024). “mCoT: Multilingual Instruction Tuning for Reasoning Consistency in Language Models”. In: *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 12012–12026. URL: https://aclanthology.org/2024.acl-long.649/.

Liu, Danni and Jan Niehues (2025). “Middle-Layer Representation Alignment for Cross-Lingual Transfer in Fine-Tuned LLMs”. In: *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 15979–15996. URL: https://aclanthology.org/2025.acl-long.778/.

Liu, Yihong et al. (2026). “Crosslingual On-Policy Self-Distillation for Multilingual Reasoning”. In: *arXiv preprint arXiv:2605.09548v1*. URL: https://arxiv.org/abs/2605.09548v1.

Muennighoff, Niklas et al. (2025). “s1: Simple test-time scaling”. In: *arXiv preprint arXiv:2501.19393v1*. URL: https://arxiv.org/abs/2501.19393v1.

Qi, Jirui et al. (2025). “When Models Reason in Your Language: Controlling Thinking Language Comes at the Cost of Accuracy”. In: *Findings of the Association for Computational Linguistics: EMNLP 2025*, pp. 20279–20296. URL: https://aclanthology.org/2025.findings-emnlp.1103/.

References

Ross, Stéphane, Geoffrey J. Gordon, and J. Andrew Bagnell (2011). “A Reduction of Imitation Learning and Structured Prediction to No-Regret Online Learning”. In: *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pp. 627–635. URL: https://proceedings.mlr.press/v15/ross11a.html.

She, Shuaijie et al. (2024). “MAPO: Advancing Multilingual Reasoning through Multilingual-Alignment-as-Preference Optimization”. In: *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 10015–10027. URL: https://aclanthology.org/2024.acl-long.539/.

Shi, Freda et al. (2023). “Language Models are Multilingual Chain-of-Thought Reasoners”. In: *International Conference on Learning Representations*. URL: https://arxiv.org/abs/2210.03057.

Wei, Jason et al. (2022). “Chain-of-Thought Prompting Elicits Reasoning in Large Language Models”. In: *Advances in Neural Information Processing Systems*. URL: https://arxiv.org/abs/2201.11903v6.

Zhang, Zhen et al. (2025). “Soft Thinking: Unlocking the Reasoning Potential of LLMs in Continuous Concept Space”. In: *arXiv preprint arXiv:2505.15778*. URL: https://arxiv.org/abs/2505.15778.

References

Zhao, Siyan et al. (2026). “Self-Distilled Reasoner: On-Policy Self-Distillation for Large Language Models”. In: *arXiv preprint arXiv:2601.18734v1*. URL: https://arxiv.org/abs/2601.18734v1.