The Illusion of Insight in Reasoning Models

ACL 2026

Liv d'Aliberti
od2961@princeton.edu

Manoel Horta Ribeiro
manoel@cs.princeton.edu

Do Reasoning Models Have "Aha!" Moments?

Prior work from DeepSeek-R1-Zero [2]: reinforcement learning can elicit mid-reasoning behavior shifts in language models.

Question: If $a > 1$, then the sum of the real solutions of $\sqrt{a - \sqrt{a + x}} = x$ is equal to

Response: <think>
To solve the equation $\sqrt{a - \sqrt{a + x}} = x$, let's start by squaring both ⋯

$$\left(\sqrt{a - \sqrt{a + x}}\right)^2 = x^2 \Longrightarrow a - \sqrt{a + x} = x^2.$$

Rearrange to isolate the inner square root term:

$$(a - x^2)^2 = a + x \Longrightarrow a^2 - 2ax^2 + (x^2)^2 = a + x \Longrightarrow x^4 - 2ax^2 - x + (a^2 - a) = 0$$

⋯
Wait, wait. Wait. That's an aha moment I can flag here.
Let's reevaluate this step-by-step to identify if the correct sum can be ⋯
We started with the equation:

$$\sqrt{a - \sqrt{a + x}} = x$$

First, let's square both sides:

$$a - \sqrt{a + x} = x^2 \Longrightarrow \sqrt{a + x} = a - x^2$$

Next, I could square both sides again, treating the equation: ⋯
⋯

A trace can look insightful without being useful.
Do reasoning shifts improve correctness?

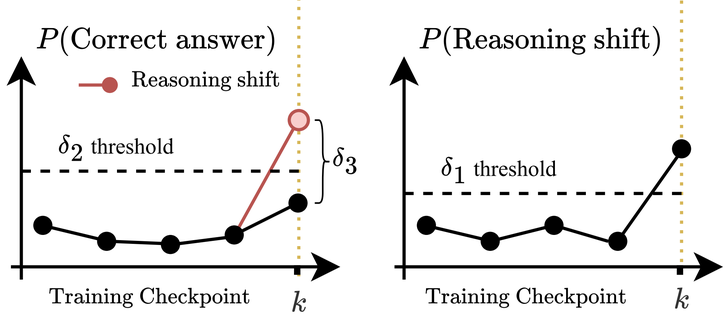
Research Questions

1. Do reasoning shifts improve accuracy?
2. Do shifts emerge during RL fine-tuning?
3. Are shifts more useful when the model is uncertain?

Thesis. Mid-reasoning shifts are not insight. Better understood as symptoms of unstable inference behavior.

Operationalizing "Aha!"

We treat an "Aha!" as a **behavioral pattern across training**.



For problem q_j and checkpoint k, an "Aha!" moment must satisfy:

1. **Prior failures** earlier checkpoints consistently fail on q_j.
2. **Prior stability** earlier checkpoints rarely show reasoning shifts on q_j.
3. **Performance gain** shifted traces at checkpoint k are more accurate.

References

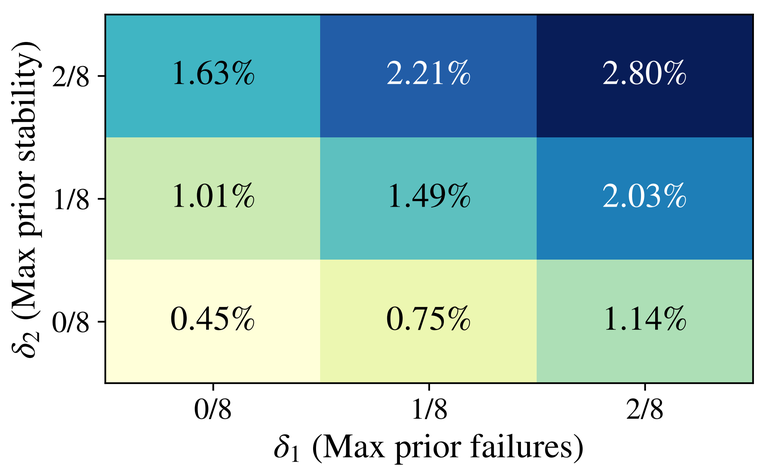
1. Liv G. d'Aliberti and Manoel Horta Ribeiro. *The Illusion of Insight in Reasoning Models.* arXiv:2601.00514, 2026.
2. DeepSeek-AI et al. *DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning.* arXiv:2501.12948, 2025.

Study Design

- **Domains:** Math, crosswords, Rush Hour
- **Models:** Qwen2.5, Llama, GPT-4o, DeepSeek-R1
- **Training:** GRPO checkpoints
- **Sampling:** multiple temperatures / traces
- **Scale:** 1M+ traces
- **Shift detection:** LLM-as-judge

We use an LLM-as-judge to annotate mid-trace reasoning shifts, then compare accuracy conditional on whether a shift occurs.

Formal "Aha!" Events Are Rare

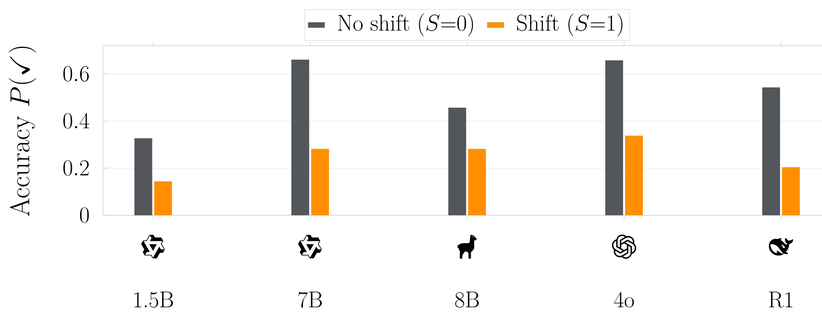


Fraction of problem–checkpoint pairs satisfying the formal definition for Qwen2.5-1.5B at $T = 0.7$, across thresholds for prior failures and prior stability.

Even lenient thresholds yield only 0.45%–2.80% formal events.

Main Finding: Shifts Rarely Improve Accuracy

Reasoning shifts are **rare** and usually associated with **lower conditional accuracy**. A trace that looks like reconsideration is not necessarily useful self-correction.



Bars show accuracy conditional on whether a reasoning shift occurred, pooled across problems, temperatures, checkpoints, and samples. Detected shifts are rare (%S ranges from 1.5% to 5.0%).

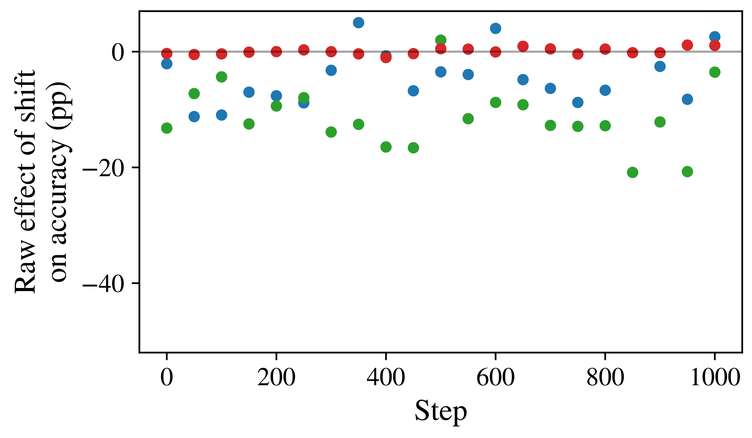
Looking like reconsideration is not the same as improving.

Takeaways

1. **Do not infer insight from rhetoric.**
2. **Measure behavior over training.**
3. **Use uncertainty.**

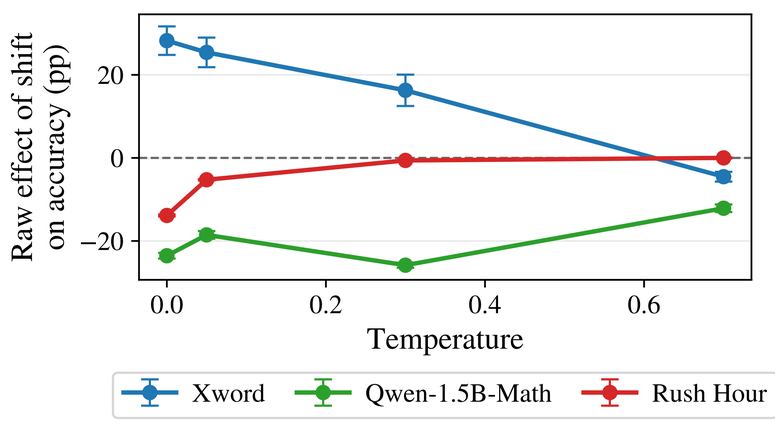
Training & Temperature

Training stage



Raw effect of a detected shift on accuracy (pp) for Qwen2.5-1.5B across the three domains by training step. Points below zero indicate that shifted traces are less accurate.

Temperature

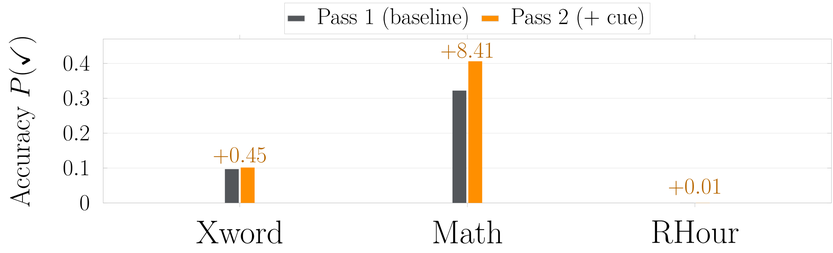


Raw effect of a detected shift on accuracy (pp) for Qwen2.5-1.5B by decoding temperature.

Training and temperature do not rescue the "Aha!" story.

Intervention: Forced Reconsideration Can Help

To test whether reconsideration can still be useful, we generate once, append a reconsideration cue, and generate again.



Pass 1 is the baseline; Pass 2 re-generates after a reconsideration cue. Labels above the bars give the accuracy gain Δ (pp). On Math, Pass 2 improves far more samples than it harms (50.6k vs. 23.5k).

Reconsideration can help when targeted.
On Math, forced reconsideration improves accuracy by +8.41 pp.

Scan me!