

# Latent Veracity Inference for Identifying Errors in Stepwise Reasoning

---

Minsu Kim\*, Jean-Pierre Falet\*, Oliver E. Richardson, Xiaoyin Chen, Moksh Jain, Sungjin Ahn, Sungsoo Ahn, Yoshua Bengio

\*Equal contribution

Published as a conference paper at ICLR 2026.

# Errors in Chain-of-Thought Reasoning

Chain-of-thought (CoT) reasoning improves LM performance.

---

# Errors in Chain-of-Thought Reasoning

Chain-of-thought (CoT) reasoning improves LM performance.

But individual reasoning steps can be incorrect.

---

[1] H. Lightman et al., "Let's Verify Step by Step," 2024.

[2] A. Jacovi et al., "A chain-of-thought is as strong as its weakest link: A benchmark for verifiers of reasoning chains," 2024.

[3] J. Huang et al., "Large language models can self-improve," 2023.

# Errors in Chain-of-Thought Reasoning

Chain-of-thought (CoT) reasoning improves LM performance.

But individual reasoning steps can be incorrect.

Existing approaches to identify step-level errors:

- Human-labeled step supervision: costly and hard to scale [1]
- Fact-verification: retrieval complexity and coverage gaps [2]
- Prompt-based techniques can be fragile [3]

---

[1] H. Lightman et al., "Let's Verify Step by Step," 2024.

[2] A. Jacovi et al., "A chain-of-thought is as strong as its weakest link: A benchmark for verifiers of reasoning chains," 2024.

[3] J. Huang et al., "Large language models can self-improve," 2023.

# Errors in Chain-of-Thought Reasoning

Chain-of-thought (CoT) reasoning improves LM performance.

But individual reasoning steps can be incorrect.

Existing approaches to identify step-level errors:

- Human-labeled step supervision: costly and hard to scale [1]
- Fact-verification: retrieval complexity and coverage gaps [2]
- Prompt-based techniques can be fragile [3]

## Main Contribution

Error identification as inference in a **latent-variable model of veracity**

---

[1] H. Lightman et al., "Let's Verify Step by Step," 2024.

[2] A. Jacovi et al., "A chain-of-thought is as strong as its weakest link: A benchmark for verifiers of reasoning chains," 2024.

[3] J. Huang et al., "Large language models can self-improve," 2023.

# Latent Variable Model of Veracity

Given context  $x$ , reasoning chain  $z$ , and answer  $y$

Augment each step  $z_i$  with a binary veracity variable  $V_{z_i} \in \{0, 1\}$

# Latent Variable Model of Veracity

Given context  $x$ , reasoning chain  $z$ , and answer  $y$

Augment each step  $z_i$  with a binary veracity variable  $V_{z_i} \in \{0, 1\}$

Similar to latent-variable models of reasoning, except we want to sample veracity assignments that help predict correct answer  $y^*$ :

$$\mathbb{P}(V_z=v \mid Y=y^*, x, z) = \frac{P_{\text{LM}}(v, y^* \mid x, z)}{\sum_{v'} P_{\text{LM}}(v', y^* \mid x, z)}$$

# Latent Variable Model of Veracity

Given context  $x$ , reasoning chain  $z$ , and answer  $y$

Augment each step  $z_i$  with a binary veracity variable  $V_{z_i} \in \{0, 1\}$

Similar to latent-variable models of reasoning, except we want to sample veracity assignments that help predict correct answer  $y^*$ :

$$\mathbb{P}(V_z=v \mid Y=y^*, x, z) = \frac{P_{\text{LM}}(v, y^* \mid x, z)}{\sum_{v'} P_{\text{LM}}(v', y^* \mid x, z)}$$

This is an intractable posterior  $\rightarrow$  not equivalent to  $P_{\text{LM}}(v \mid x, z, y^*)$

# Veracity Search and Amortized Veracity Inference

- **Reward:** proportional to the posterior over veracity vectors

$$R(v) := P_{\text{LM}}(v, y^* \mid x, z) \propto P(V_z=v \mid Y=y^*, x, z)$$

# Veracity Search and Amortized Veracity Inference

- **Reward:** proportional to the posterior over veracity vectors

$$R(v) := P_{\text{LM}}(v, y^* | x, z) \propto P(V_z=v | Y=y^*, x, z)$$

- **Search:** Single-bit Metropolis–Hastings with simulated annealing

# Veracity Search and Amortized Veracity Inference

- **Reward:** proportional to the posterior over veracity vectors

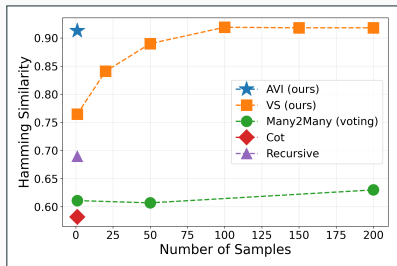
$$R(v) := P_{\text{LM}}(v, y^* | x, z) \propto P(V_z=v | Y=y^*, x, z)$$

- **Search:** Single-bit Metropolis–Hastings with simulated annealing
- **AVI:** Pseudo-labels from VS train a zero-shot verifier  $Q(V_z | x, z)$

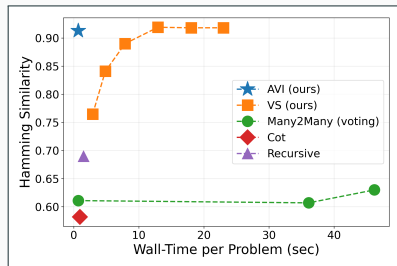
# VS outperforms baselines on Hamming (L1) distance $|v_z - v_{z^*}|$

| Dataset       | Method    | Qwen-4B      | Qwen-8B      | Llama-3B     | Llama-8B     |
|---------------|-----------|--------------|--------------|--------------|--------------|
| PRONTOQA      | Recursive | 0.691        | 0.667        | 0.538        | 0.471        |
|               | Many2Many | 0.590        | 0.683        | 0.506        | 0.530        |
|               | Voting    | 0.603        | 0.692        | 0.514        | 0.536        |
|               | CoT       | 0.591        | 0.384        | 0.459        | 0.515        |
|               | VS (ours) | <b>0.910</b> | <b>0.945</b> | <b>0.948</b> | <b>0.964</b> |
| GSM8K         | Recursive | 0.540        | 0.617        | 0.568        | 0.568        |
|               | Many2Many | 0.620        | 0.650        | 0.566        | 0.567        |
|               | Voting    | 0.623        | 0.654        | 0.566        | 0.553        |
|               | CoT       | 0.614        | 0.695        | 0.496        | 0.496        |
|               | VS (ours) | <b>0.711</b> | <b>0.751</b> | <b>0.614</b> | <b>0.646</b> |
| COMMONSENSEQA | Recursive | 0.607        | 0.509        | 0.505        | 0.506        |
|               | Many2Many | 0.517        | 0.534        | 0.504        | 0.503        |
|               | Voting    | 0.521        | 0.533        | 0.504        | 0.505        |
|               | CoT       | 0.695        | 0.590        | 0.507        | 0.535        |
|               | VS (ours) | <b>0.935</b> | <b>0.931</b> | <b>0.836</b> | <b>0.903</b> |

# AVI amortizes search cost



Samples vs. accuracy



Wall-clock time vs. accuracy

# AVI-guided self-correction improves reasoning performance

| Method          | Qwen 4B     |             |             | Qwen 8B     |             |             |
|-----------------|-------------|-------------|-------------|-------------|-------------|-------------|
|                 | 3-hop       | 4-hop       | 5-hop       | 3-hop       | 4-hop       | 5-hop       |
| No Correction   | 0.60        | 0.52        | 0.59        | 0.54        | 0.65        | 0.52        |
| Self Correction | 0.54        | 0.60        | 0.48        | 0.54        | 0.58        | 0.46        |
| AVI (ours)      | <b>0.68</b> | <b>0.72</b> | <b>0.77</b> | <b>0.87</b> | <b>0.85</b> | <b>0.81</b> |

**Latent-variable model of veracity:** disentangles reasoning step identity from correctness

**Latent-variable model of veracity:** disentangles reasoning step identity from correctness

**Veracity Search (VS):** MCMC-based error identification using the LM's own likelihood (no supervision at the level of reasoning)

**Latent-variable model of veracity:** disentangles reasoning step identity from correctness

**Veracity Search (VS):** MCMC-based error identification using the LM's own likelihood (no supervision at the level of reasoning)

**Amortized Veracity Inference (AVI):** zero-shot error detection, enabling self-correction and improved reasoning performance

# Conclusion

**Latent-variable model of veracity:** disentangles reasoning step identity from correctness

**Veracity Search (VS):** MCMC-based error identification using the LM's own likelihood (no supervision at the level of reasoning)

**Amortized Veracity Inference (AVI):** zero-shot error detection, enabling self-correction and improved reasoning performance

## Limitations:

- Analysis primarily on synthetic reasoning chains
- Error identification alone is insufficient for reasoning tasks