# Maximizing Prefix-Confidence at Test-Time Efficiently Improves Mathematical Reasoning

ETH zürich

Matthias Otth, Jonas Hübotter, Ido Hakimi, Andreas Krause

Learning & Adaptive Systems

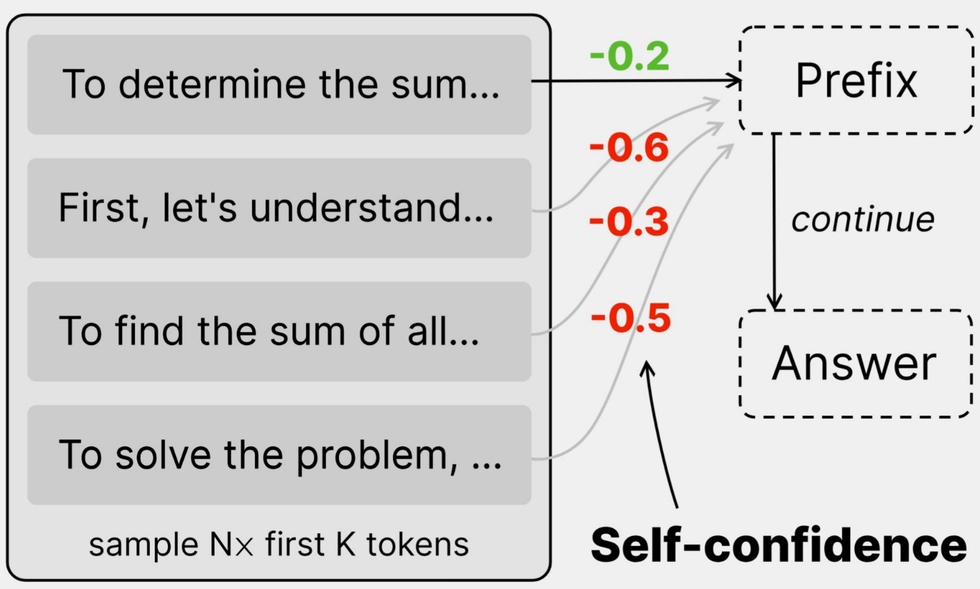
## Background

- **Goal:** Improve at reasoning by leveraging the models "confidence" about its attempts.
- **Problem:** BoN sampling on full attempts is expensive & unreliable due to length biases.

**Can LLMs reliably self-improve at test-time without relying on a verifier or reward?**

## Contributions

### Test-Time Prefix-Confidence Scaling




- We propose **Test-Time Prefix-Confidence** scaling, which samples N prefixes of length K, and then completes the prefix about which the model is most confident.
- Compared to majority voting and BoN ("confidence voting" on full attempts), prefix-confidence scaling achieves a better accuracy-compute trade-off.

**Confidence measures:**

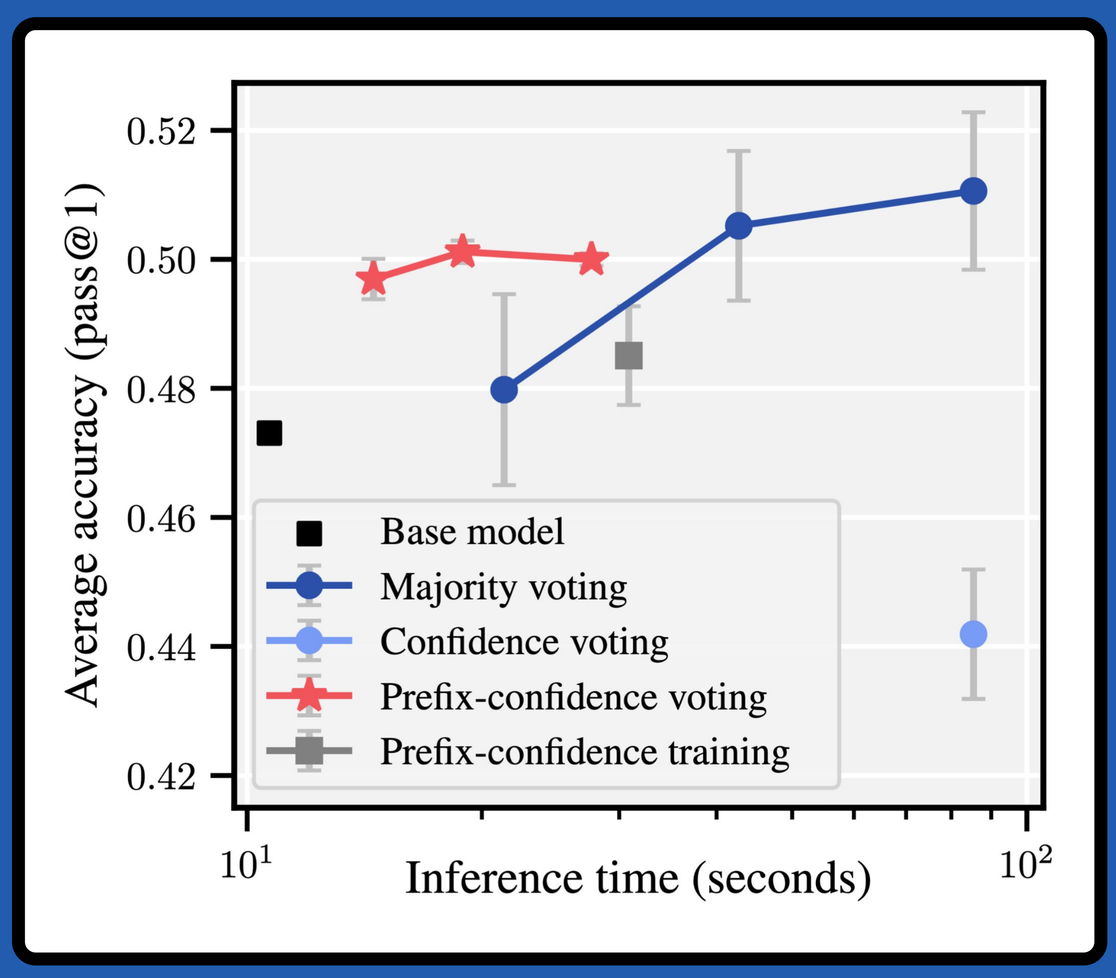
- Self-consistency / majority voting
  - Baseline requiring full attempts
- Self-certainty

$$\frac{1}{n}\sum_{i=1}^{n}\mathrm{KL}(\mathrm{Unif}\|\pi(y_i \mid x, y_{<i}))$$

- Self-confidence (performs best)

$$\log \pi(y \mid x) = \sum_{i=1}^{n} \log \pi(y_i \mid x, y_{<i})$$

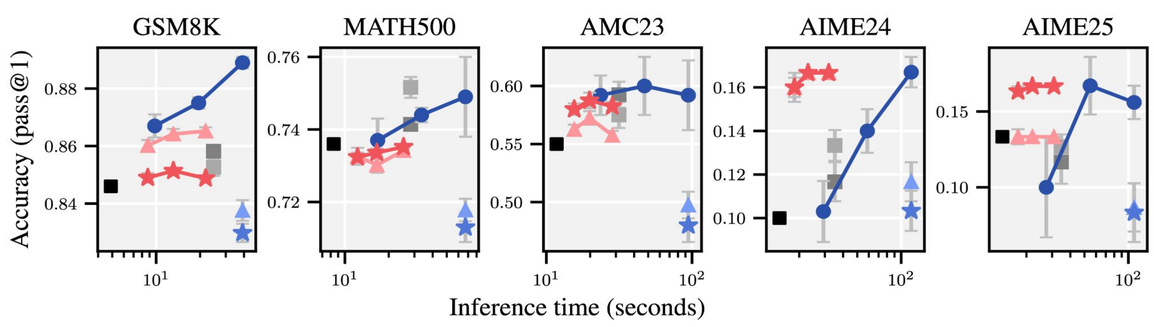
## LLMs improve at mathematical reasoning when continuing only their most confident prefixes.




## Details

**Model:** Qwen2.5-Math-1.5B-Instruct

**N:** {8,16,32}, **K:** 32 (seems relatively robust to these)




| | | GSM8K | MATH500 | AMC23 | AIME24 | AIME25 | **avg** | time (s) |
|---|---|---|---|---|---|---|---|---|
| Base | | 84.6 | 73.6 | 55.0 | 10.0 | 13.3 | 47.3 | 10.68 |
| Bo*N*@8 | self-confidence | 83.0±0.3 | 71.3±0.4 | 48.0±1.4 | 10.3±0.9 | 8.3±1.9 | 44.2±0.5 | 85.43 |
| | self-certainty | 83.8±0.4 | 71.8±0.3 | 49.8±1.1 | 11.7±0.9 | 8.7±1.6 | 45.2±0.4 | |
| Bo*N*@16 | self-confidence | 82.0±0.2 | 69.7±0.2 | 44.8±1.2 | 9.0±1.3 | 5.3±0.9 | 42.2±0.4 | 170.86 |
| | self-certainty | 83.0±0.2 | 70.2±0.4 | 47.5±1.4 | 9.7±1.0 | 6.3±1.8 | 43.3±0.5 | |
| Maj@2 | self-consistency | 86.7±0.4 | 73.7±0.6 | **59.2**±1.7 | 10.3±1.4 | 10.0±3.3 | 48.0±0.8 | 21.36 |
| Maj@4 | self-consistency | 87.5±0.2 | **74.4**±0.2 | **60.0**±2.5 | 14.0±1.0 | **16.7**±1.9 | **50.5**±0.7 | 42.72 |
| Maj@8 | self-consistency | **88.9**±0.2 | **74.9**±1.1 | **59.2**±3.0 | **16.7**±0.7 | **15.6**±1.1 | **51.1**±0.7 | 85.43 |
| PC@8 | self-confidence | 84.9±0.2 | 73.2±0.1 | 58.0±0.5 | 16.0±0.4 | 16.3±0.3 | 49.7±0.1 | 14.52 |
| | self-certainty | **86.0**±0.2 | 73.3±0.2 | 56.3±0.4 | **16.0**±0.7 | 13.3±0.5 | 49.0±0.2 | |
| PC@16 | self-confidence | 85.1±0.1 | 73.4±0.1 | **58.8**±0.7 | **16.7**±0.0 | **16.7**±0.0 | **50.1**±0.1 | 18.89 |
| | self-certainty | **86.4**±0.2 | 73.0±0.2 | 57.3±0.6 | **16.7**±0.0 | 13.3±0.0 | 49.3±0.1 | |
| PC training | NLL (4) | 85.8±0.2 | 74.1±0.2 | **59.3**±1.1 | 11.7±0.9 | 11.7±1.4 | 48.5±0.4 | 30.87 |
| | entropy (5) | 85.3±0.3 | **75.2**±0.3 | 57.5±1.1 | 13.3±0.7 | 12.7±0.8 | 48.8±0.3 | |