Test-Time Compute Scaling Enhances Language Model Reasoning Benchmarks

Assignee Research

June 5, 2026

Abstract

This report synthesises findings from 1 peer-reviewed paper addressing the following research question: How does test-time compute scaling improve language model performance on reasoning benchmarks v7. 12 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: ∇-Reasoner: LLM Reasoning via Test-Time Gradient Descent in Latent Space. Research question: How does test-time compute scaling improve language model performance on reasoning benchmarks v7.

2 Methodology

Systematic literature search across multiple databases yielded 1 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.2/10.

3 Results

1 papers retrieved. 12 claims extracted; 0 independently verified. Quality review score: 3.2/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

5 Extracted Claims

Claim	Verified	Confidence
∇-Reasoner outperforms all test-time baselines and achieves performance on par with training-based methods (SFT and GRPO	×	0.09
∇-Reasoner is evaluated on four mathematical reasoning benchmarks: MATH-500, AIME24, AIME25, and AMC.	×	0.04
For BoN and SC, N = 8 is used to match Nmax = 8 used in ∇-Reasoner.	×	0.01
For TPO, Nsamples = 2 and Nrefine = 2 are set.	×	0.00
For ToT and RAP, default hyperparameters from Hao et al. (2024) are adopted.	×	0.00
∇-Reasoner uses reward models from the Skywork-V2 family: Skywork-V2-Qwen-4B for Qwen-based models and Skywork-V2-Llama-	×	0.02
∇-Reasoner is an iterative decoding algorithm driven by DTO.	×	0.07
∇-Reasoner applies gradient descent on the logits initialized from the base model to optimize a reward-informed loss to	×	0.11
∇-Reasoner combines the updated policy with rejection sampling, leading to high-reward responses.	×	0.06
∇-Reasoner resamples the first token of the generated sequence using the fine-tuned logits ez1.	×	0.02
If the resampled token differs from the original, the subsequent tokens are regenerated, and this candidate token is acc	×	0.03
∇-Reasoner scales inference-time reasoning by allocating additional computation to optimize the policy's outputs via ite	×	0.14

References

- http://arxiv.org/abs/2603.04948v1