

GRPO Rollout Count Effects on LLM Alignment Stability in HELM Benchmarks

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 16 peer-reviewed papers addressing the following research question: What is the impact of varying the number of rollouts per prompt in GRPO-based RL on the alignment stability of LLMs, as measured by the HELM benchmark's fairness and robustness scores. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.7/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: PPO-BR: Dual-Signal Entropy-Reward Adaptation for Trust Region Policy Optimization. Research question: What is the impact of varying the number of rollouts per prompt in GRPO-based RL on the alignment stability of LLMs, as measured by the HELM benchmark's fairness and robustness scores?.

2 Methodology

Systematic literature search across multiple databases yielded 16 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.7/10.

3 Results

16 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 4.7/10.

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.

References

- <http://arxiv.org/abs/2505.17714v1>
- <http://arxiv.org/abs/2506.01937v2>
- <http://arxiv.org/abs/2202.03586v2>