

Executive Summary

Group Sequence Policy Optimization: A Breakthrough in Training Large Language Models

The Problem

Current reinforcement learning algorithms for training large language models, particularly state-of-the-art methods like GRPO, suffer from severe instability issues. When training gigantic models on complex tasks, these algorithms often experience catastrophic model collapse that's irreversible, hindering efforts to advance AI capabilities through continued training.

The Breakthrough

Researchers developed **Group Sequence Policy Optimization (GSPO)**, a revolutionary RL algorithm that fundamentally changes how language models are trained. Instead of applying importance corrections at the individual token level like previous methods, GSPO works at the sequence level—treating entire responses as the unit of optimization. This aligns perfectly with how rewards are actually given to language models.

How It Works

GSPO computes importance ratios based on the likelihood of complete response sequences rather than individual tokens, then applies clipping and optimization at the sequence level. This approach eliminates the high-variance noise that plagued previous algorithms. In experiments, GSPO demonstrated **superior training efficiency** over GRPO, achieving better performance with the same computational resources while maintaining remarkable stability throughout training.

Why This Matters

This breakthrough solves the fundamental instability that has limited the scaling of large language model training. It enables stable training of massive models, including complex Mixture-of-Experts (MoE) architectures that previously required complex workarounds.

The algorithm's stability and efficiency have already contributed to exceptional improvements in the latest Qwen3 models.

The Business Opportunity

GSPO provides a robust foundation for scaling AI capabilities, enabling companies to train larger, more capable models without fear of catastrophic collapse. This opens new possibilities for developing AI systems that can tackle increasingly sophisticated problems in mathematics, programming, and reasoning domains.