

Gradient Clipping Effects on Training Stability and NDCG@10 in Lion vs. AdamW for ModernBERT Cross-Encoders on MS MARCO

Assignee Research

June 11, 2026

Abstract

Large Language Models (LLMs) have recently demonstrated remarkable capabilities in natural language processing tasks and beyond. This success of LLMs has led to a large influx of research contributions in this direction. These works encompass diverse topics such as architectural innovations, better training strategies, context length improvements, fine-tuning, multi-modal LLMs, robotics, datasets, benchmarking, efficiency, and more. With the rapid development of techniques and regular breakthroughs in LLM research, it has become considerably challenging to perceive the bigger picture of the ad

1 Introduction

This paper examines: A Comprehensive Overview of Large Language Models. Research question: What is the effect of gradient clipping on the training stability and final NDCG@10 scores when using Lion versus AdamW for ModernBERT cross-encoders on MS MARCO?.

2 Methodology

Systematic literature search across multiple databases yielded 6 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 7.7/10.

3 Results

6 papers retrieved. 6 claims extracted; 6 independently verified. Quality review score: 7.7/10.

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.

5 Extracted Claims

| Claim | Verified | Confidence |
|--|----------|------------|
| Large Language Models (LLMs) have demonstrated remarkable capabilities in natural language processing tasks and beyond. | ✓ | 0.29 |
| There has been a large influx of research contributions in the direction of LLMs, encompassing diverse topics such as ar | ✓ | 0.41 |
| The rapid development of techniques and regular breakthroughs in LLM research have made it challenging to perceive the b | ✓ | 0.32 |
| There is a rapidly emerging plethora of literature on LLMs, making it imperative for the research community to benefit f | ✓ | 0.33 |
| This article provides an overview of the existing literature on a broad range of LLM-related concepts, discussing releva | ✓ | 0.40 |
| This review article is intended to provide a systematic survey and a quick comprehensive reference for researchers and p | ✓ | 0.43 |

References

- <https://doi.org/10.48550/arxiv.2307.06440>
- <https://doi.org/10.48550/arxiv.2307.06435>
- <https://doi.org/10.48550/arxiv.2412.19437>