

Knowledge Distillation and Dynamic Learning Rates for Stable Code Generation Models

Assignee Research

June 6, 2026

Abstract

This report synthesises findings from 15 peer-reviewed papers addressing the following research question: Does integrating knowledge distillation with dynamic learning rate schedules improve the stability of code generation models when evaluated on out-of-distribution LiveCodeBench problems. 4 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.3/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: A Closer Look at Deep Learning Heuristics: Learning rate restarts, Warmup and Distillation. Research question: Does integrating knowledge distillation with dynamic learning rate schedules improve the stability of code generation models when evaluated on out-of-distribution LiveCodeBench problems?.

2 Methodology

Systematic literature search across multiple databases yielded 15 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.3/10.

3 Results

15 papers retrieved. 4 claims extracted; 1 independently verified. Quality review score: 4.3/10.

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
Learning rate restarts, warmup, and distillation are heuristics used in deep learning.	✓	0.26
The paper presents benchmark tables comparing different deep learning heuristics.	×	0.08
The benchmark table includes columns labeled LB, LB + war, no wa mup, rmup, LB + FC freeze(no war mup).	×	0.01
The benchmark table includes columns labeled LB, LB, LB + war, no wa, + FC mup, rmup, freeze(no war mup).	×	0.01

References

- <http://arxiv.org/abs/2310.07831v2>
- <http://arxiv.org/abs/1810.13243v1>
- <http://arxiv.org/abs/2405.09820v1>