

# Instruction-Tuning Effects on Long-Context Retrieval Accuracy in Large Language Models

Assignee Research

June 7, 2026

## Abstract

This report synthesises findings from 13 peer-reviewed papers addressing the following research question: What is the impact of instruction-tuning (e.g., Llama-3 vs. Vicuna) on performance degradation in long-context retrieval tasks like MLNeedle, measured by accuracy drop-off as context length increases. 9 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: Learning Long-Context Diffusion Policies via Past-Token Prediction. Research question: What is the impact of instruction-tuning (e.g., Llama-3 vs. Vicuna) on performance degradation in long-context retrieval tasks like MLNeedle, measured by accuracy drop-off as context length increases from 50K to 300K tokens?.

## 2 Methodology

Systematic literature search across multiple databases yielded 13 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.8/10.

## 3 Results

13 papers retrieved. 9 claims extracted; 1 independently verified. Quality review score: 3.8/10.

## 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
PTP achieves an average improvement of over 30% compared to no-history diffusion policies, and over 60% compared to no-P	×	0.10
The gains from PTP are especially pronounced on history-critical tasks such as long-horizon aloha and long-horizon squar	×	0.05
The proposed multi-stage training recipe accelerates policy learning.	×	0.05
PTP verification can further mitigate deficiencies in temporal dependencies at test time.	×	0.04
The method is evaluated on six simulated tasks, including square, tool hang, transport, Push-T, long-horizon square, and	×	0.05
Each policy is trained using the multistage recipe with feature caching and evaluated under a single-sample inference se	×	0.03
The effect of test-time verification is evaluated separately across multiple checkpoints under varying sample budgets.	×	0.03
PTP tasks the policy to predict past action tokens alongside future ones.	✓	0.22
PTP expands the prediction window in both temporal directions, explicitly encouraging the policy to capture temporal dep	×	0.08

## References

- <http://arxiv.org/abs/2503.09202v1>
- <http://arxiv.org/abs/2503.01763v2>
- <http://arxiv.org/abs/2505.09561v2>