LLM-Driven Temporal User Profiling Trade-offs in Recommendation Efficiency and Alignment

Assignee Research

May 30, 2026

Abstract

This report synthesises findings from 16 peer-reviewed papers addressing the following research question: What is the trade-off between inference efficiency (tokens/sec) and alignment performance when joint modeling user reviews in recommendation tasks using LLMs. Effectively modeling the dynamic nature of user preferences is crucial for enhancing recommendation accuracy and fostering transparency in recommender systems. Traditional user profiling often overlooks the distinction between transitory short-term interests and stable long-term. 8 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Effectiveness of LLMs in Temporal User Profiling for Recommendation. Research question: What is the trade-off between inference efficiency (tokens/sec) and alignment performance when joint modeling user reviews in recommendation tasks using LLMs?.

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 4.5/10.

3 Results

16 papers retrieved. 8 claims extracted; 1 independently verified. Quality review score: 4.5/10.

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
The full framework outperformed the Short-Term Only (ST) variant by +15.2% in Recall@20.	×	0.06
The full framework consistently outperformed both Short-Term Only (ST) and Long-Term Only (LT) variants individually.	×	0.11
Replacing MLP scoring with a dot product (DP variant) resulted in the steepest performance degradation among tested vari	×	0.02
The framework demonstrated notable performance gains in high-activity domains such as Movies&TV.	×	0.06
The framework demonstrated less pronounced benefits in sparser environments such as Video Games.	×	0.11
The framework generates two textual summaries per user: a short-term profile based on recent interactions and a long-ter	×	0.12
The interaction prediction function f in the proposed framework is defined as a multi-layer perceptron (MLP).	×	0.02
The framework leverages LLM-driven temporal user profiling to create intrinsically interpretable user profiles via natur	✓	0.23

References

- http://arxiv.org/abs/2601.18457v1
- http://arxiv.org/abs/2511.00176v1
- http://arxiv.org/abs/2402.11651v2