

# CausalMixFT and DP-LAMB Robustness Against Membership Inference on SuperGLUE Tasks

Assignee Research

June 9, 2026

## Abstract

This report synthesises findings from 13 peer-reviewed papers addressing the following research question: How does the robustness of CausalMixFT against membership inference attacks compare to DP-LAMB when evaluated on diverse NLP tasks within the SuperGLUE suite. 0 claims were extracted from source literature; 0 were 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: Towards Demystifying Membership Inference Attacks. Research question: How does the robustness of CausalMixFT against membership inference attacks compare to DP-LAMB when evaluated on diverse NLP tasks within the SuperGLUE suite?.

## 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. 0 claims extracted; 0 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.

## References

- <http://arxiv.org/abs/2509.13625v4>
- <http://arxiv.org/abs/2001.08855v1>
- <http://arxiv.org/abs/1807.09173v2>