Differentially Private LoRA Fine-Tuning and Cross-Domain Transfer Accuracy on GLUE and SuperGLUE

Assignee Research

June 8, 2026

Abstract

This report synthesises findings from 12 peer-reviewed papers addressing the following research question: How does differentially private LoRA fine-tuning affect cross-domain transfer accuracy on GLUE and SuperGLUE benchmarks compared to non-private baselines. 11 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 2.4/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Large Scale Private Learning via Low-rank Reparametrization. Research question: How does differentially private LoRA fine-tuning affect cross-domain transfer accuracy on GLUE and SuperGLUE benchmarks compared to non-private baselines?.

2 Methodology

Systematic literature search across multiple databases yielded 12 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 2.4/10.

3 Results

12 papers retrieved. 11 claims extracted; 0 independently verified. Quality review score: 2.4/10.

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
RGP achieves a validation accuracy of 97.1% on SVHN and 91.2% on CIFAR10 without privacy constraints.	×	0.04
RGP achieves a validation accuracy of 94.2% on SVHN and 63.4% on CIFAR10 with ϵ = 8.	×	0.02
RGP achieves a validation accuracy of 87.3% on SVHN and 44.0% on CIFAR10 with ϵ = 2.	×	0.05
RGP achieves a validation accuracy of 89.7% on SVHN and 53.3% on CIFAR10 with ϵ = 4.	×	0.02
RGP achieves a validation accuracy of 92.3% on SVHN and 59.6% on CIFAR10 with ϵ = 6.	×	0.02
RGP achieves an average validation accuracy of 88.1% on MNLI, QQP, QNLI, and SST-2 without privacy constraints.	×	0.05
RGP achieves an average validation accuracy of 83.9% on MNLI, QQP, QNLI, and SST-2 with privacy constraints.	×	0.06
The computational cost of RGP is O(md2 + Krd2 + Kr2d).	×	0.02
The memory cost of RGP is O(mrd).	×	0.09
The computational cost of DP-SGD is O(md2).	×	0.05
The memory cost of DP-SGD is O(md2).	×	0.08

References

- http://arxiv.org/abs/2411.14961v3
- http://arxiv.org/abs/2110.06500v2

- http://arxiv.org/abs/2106.09352v4