

Research Papers on 'AI'

[Google Scholar](#)

[Semantic Scholar](#)

[IEEE Xplore](#)

[PubMed](#)

Paper 1:

An Interactive Framework for Implementing Privacy-Preserving Federated Learning: Experiments on Large Language Models

Date: 2025-02-14

Time: 18:52:34

Authors:

Kasra Ahmadi, Rouzbeh Behnia, Reza Ebrahimi, Mehran Mozaffari Kermani,
Jeremiah Birrell, Jason Pacheco, Attila A Yavuz

Summary:

- Federated learning (FL) enhances privacy by keeping user data on local devices. However, emerging attacks have demonstrated that the updates shared by users during training can reveal significant information about their data. This has greatly thwarted the adoption of FL methods for training robust AI models in sensitive applications. Differential Privacy (DP) is considered the gold standard for safeguarding user data. However, DP guarantees are highly conservative, providing worst-case privacy guarantees. This can result in overestimating privacy needs, which may compromise the model's accuracy. Additionally, interpretations of these privacy guarantees have proven to be challenging in different contexts. This is further exacerbated when other factors, such as the number of training iterations, data distribution, and specific application requirements, can add further complexity to this problem. In this work, we proposed a framework that integrates a human entity as a privacy practitioner to determine an optimal trade-off between the model's privacy and utility. Our framework is the first to address the variable memory requirement of existing DP methods in FL settings, where resource-limited devices (e.g., cell phones) can participate. To support such settings, we adopt a recent DP method with fixed memory usage to ensure scalable private FL. We evaluated our proposed framework by fine-tuning a BERT-based LLM model using the GLUE dataset (a common approach in literature), leveraging the new accountant, and employing diverse data partitioning strategies to mimic real-world conditions. As a result,

we achieved stable memory usage, with an average accuracy reduction of 1.33% for $\epsilon = 10$ and 1.9% for $\epsilon = 6$, when compared to the state-of-the-art DP accountant which does not support fixed memory usage.

[Click here for more](#)

Paper 2:

Robustness tests for biomedical foundation models should tailor to specification

Date: 2025-02-14

Time: 18:52:10

Authors:

R. Patrick Xian, Noah R. Baker, Tom David, Qiming Cui, A. Jay Holmgren, Stefan Bauer, Madhumita Sushil, Reza Abbasi-Asl

Summary:

- Existing regulatory frameworks for biomedical AI include robustness as a key component but lack detailed implementational guidance. The recent rise of biomedical foundation models creates new hurdles in testing and certification given their broad capabilities and susceptibility to complex distribution shifts. To balance test feasibility and effectiveness, we suggest a priority-based, task-oriented approach to tailor robustness evaluation objectives to a predefined specification. We urge concrete policies to adopt a granular categorization of robustness concepts in the specification. Our approach promotes the standardization of risk assessment and monitoring, which guides technical developments and mitigation efforts.

[Click here for more](#)