Privacy-Preserving Machine Learning: Protecting User Data in AI Systems

A Technical Report submitted to the Department of Computer Science

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Jaimin Thakkar

Spring, 2025

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Rosanne Vrugtman, Department of Computer Science

Privacy-Preserving Machine Learning: Protecting User Data in AI Systems

CS4991 Capstone Report, 2025

Jaimin Thakkar
Computer Science
The University of Virginia
School of Engineering and Applied Science
Charlottesville, Virginia USA
hcn6wd@virginia.edu

ABSTRACT

The reliance on machine learning systems has increased, raising concerns about the privacy and security of sensitive data. Privacy-preserving machine learning offers a solution by enabling model training without directly sharing user data. I have researched techniques such as differential privacy and secure multi-party computation, which allow AI models to be trained on decentralized data sources while maintaining confidentiality. The proposed solution includes incorporating these techniques into existing machine learning frameworks to increase data protection. Key findings show that these methods preserve privacy efficiently with minimal impact on model accuracy. Future work will focus on further improving these approaches for scalability and expanding their application to more diverse industries such as healthcare and finance.

1. INTRODUCTION

As machine learning models continue to be used in sectors such as healthcare, finance, and cybersecurity, concerns over data privacy have increased significantly. Traditional centralized learning methods require the collection of large datasets, which increases the risk of data breaches and unauthorized access (Shokri & Shmatikov, 2015).

Privacy-preserving machine learning (PPML) has emerged as a key approach to reducing these risks by enabling data analysis without exposing sensitive information. My proposal explores various PPML techniques, including differential privacy and secure multi-party computation, which allow organizations to train machine learning models while maintaining user confidentiality. By applying these approaches, companies can comply with data protection regulations such as HIPAA while still utilizing valuable insights from data.

2. RELATED WORKS

PPML **has gotten some recognition in recent years, with multiple studies proposing different techniques to safeguard user data. One widely studied approach is "differential privacy," which introduces noise into datasets to ensure individual data points cannot be differentiated (Dwork et al., 2006). Differential privacy has been adopted by major organizations, including Apple and Google, to protect user information while still allowing statistical insights to be drawn from data.**

Another promising approach is "secure multi-party computation" (SMPC), which enables multiple parties to collaboratively train machine learning models without revealing their respective datasets. A study by Bonawitz, et al. (2017) introduced a federated learning framework that allows decentralized model training while preserving privacy. This method has been used in applications such as Google's Gboard,

where user data remains on personal devices instead of being transmitted to centralized servers.

Additionally, "homomorphic encryption" which ensures privacy even during the training process (Gentry, 2009), has been explored to perform computations on encrypted data. Although expensive, recent advancements have made this method more feasible for real-world applications. My proposal builds on these existing works to evaluate the trade-offs between security, accuracy and efficiency in privacy-preserving machine learning.

3. PROPOSAL DESIGN

PPML aims to enable AI models to learn from decentralized datasets without compromising user privacy. This section details the design of a secure and efficient PPML framework using differential privacy, secure multi-party computation (SMPC), and homomorphic encryption.

3.1 System Architecture

The proposed framework consists of three key components:

- Data Owners: Institutions (e.g., hospitals, banks) with sensitive user data.
- Machine Learning Model: A central AI system that learns from distributed data without direct access to raw data.
- Privacy Mechanisms: Cryptographic methods to ensure secure computations while preventing data leakage.

The system will be implemented using a federated learning model, where user data remains on local devices. The model collects encrypted updates instead of raw data to ensure privacy.

3.2 Privacy-Preserving Techniques

To maintain confidentiality, the framework will incorporate:

- Differential Privacy: Adding controlled noise to data before model training to prevent individual user identification.
- Secure Multi-Party Computation (SMPC): Enabling multiple parties to jointly compute functions on encrypted data without revealing their inputs.
- Homomorphic Encryption: Allowing computations directly on encrypted data to enhance security.

A combination of these methods ensures balance between privacy, accuracy, and computational efficiency.

3.3 Implementation Strategy

- Data Processing: Pre-processing data to remove personally identifiable information (PII).
- Model Training: Using TensorFlow Privacy and PySyft to integrate privacy-preserving techniques into machine learning models.
- Performance Evaluation: Testing the model's accuracy and computational efficiency under various privacy constraints.

This approach will be validated through simulations using healthcare and financial datasets, ensuring its applicability to real-world privacy-sensitive scenarios.\

3.4 Deployment and Integration

Implementing PPML solutions in real-world applications requires careful planning to address various challenges:

- Infrastructure Compatibility: Ensuring that existing systems can support the computational demands of PPML techniques, such as secure multi-party computation and homomorphic encryption.
- User Adoption: Designing user-friendly interfaces and workflows to would be good for acceptance among stakeholders.

- Compliance and Regulation: Deployment strategies aligned with legal frameworks and industry standards to maintain compliance.

A structured integration plan is essential to seamlessly incorporate PPML into existing workflows, minimizing disruptions and maximizing efficiency.

4. ANTICIPATED RESULTS

The proposed PPML framework is expected to achieve high privacy protection while maintaining competitive model accuracy. Differential privacy will prevent individual data points from being re-identified, while SMPC and homomorphic encryption will ensure secure computations across multiple parties.

We anticipate a trade-off between privacy and model performance—higher privacy levels may slightly reduce accuracy due to added noise and encryption overhead. However, optimizing the balance between privacy strength and usability will enhance practical deployment.

Finally, this framework is expected to be scalable across industries such as healthcare, finance, and IoT applications which will provide organizations with a robust method for secure AI adoption while ensuring regulatory compliance (e.g., HIPAA, GDPR).

5. CONCLUSION

Privacy-preserving machine learning is becoming increasingly important as more industries rely on artificial intelligence to process sensitive information. Techniques like differential privacy, secure multi-party computation and homomorphic encryption provide practical solutions to protect data during model training and prediction. The proposed framework ensures that data privacy is maintained while allowing organizations to gain valuable insights from their data. This approach is particularly valuable in fields like healthcare and finance. Because in those fields, data breaches can have serious consequences. Ensuring privacy also helps organizations comply with strict data protection regulations, such as GDPR and HIPAA, while building user trust. Continued improvements in this area are essential to support ethical and responsible AI development.

PPML addresses critical challenges in data security and privacy. It makes it a key area of research and application. Implementing privacy-preserving techniques effectively requires finding a balance between maintaining data confidentiality and achieving acceptable model performance. While current methods are promising, there are still limitations in terms of scalability and computational efficiency. The proposed framework contributes to enhancing privacy protection without compromising usability. As more organizations adopt AI-based systems and the need for robust privacy-preserving techniques will only grow. By developing better methods for secure data analysis, this project offers valuable insights into creating safer and more trustworthy AI systems.

6. FUTURE WORK

The Framework for PPML presents a promising approach to enhancing data security. However, further work is needed to make these techniques more efficient and scalable. Improving the performance of methods like homomorphic encryption is crucial because it remains computationally expensive and difficult to apply to large datasets. Also, optimizing secure multi-party computation to reduce communication overhead could improve usability. Testing the framework on various datasets, especially real-world data from industries like healthcare and finance will provide valuable feedback on its effectiveness. Developing

more advanced algorithms that reduce computational costs while maintaining high privacy protection is a necessary step moving forward.

Another important area for future work is integrating privacy-preserving techniques with emerging technologies like blockchain. Combining blockchain's decentralized architecture with privacy-preserving machine learning can enhance data security and transparency. This integration could be particularly useful for applications where data integrity and trust are essential. Furthermore, exploring hybrid models that combine different privacy-preserving techniques could enhance overall efficiency and security. Finally, providing user-friendly tools and interfaces that make implementing these techniques easier for organizations will encourage wider adoption. Continued research and collaboration in this area are essential to make PPML more practical and accessible.

REFERENCES

Bonawitz, K., Eichner, H., Grieskamp, W., Huba, D., Ingerman, A., Ivanov, V., & Seth, K. (2017). Towards federated learning at scale: System design. *arXiv preprint arXiv:1902.01046.* https://arxiv.org/abs/1902.01046

Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography Conference* (pp. 265-284). Springer. https://link.springer.com/chapter/10.1007/11681878_14

Gentry, C. (2009). Fully homomorphic encryption using ideal lattices. In *Proceedings of the 41st annual ACM symposium on Theory of Computing* (pp. 169-178). https://dl.acm.org/doi/10.1145/1536414.1536440

Shokri, R., & Shmatikov, V. (2015). Privacy-preserving deep learning. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security* (pp. 1310-1321). https://dl.acm.org/doi/10.1145/2810103.2813687