Technology Innovation Institute

Privacy and Security Foundations

for training scalable decentralized AI

Training more powerful AI and machine learning models requires access to large amounts of data. But much data is unavailable due to security, privacy, and compliance concerns. Federated learning (FL) shows promise for scaling decentralized AI training on private data but comes with some privacy vulnerabilities. Combining FL with Multi-party Computation (MPC) and Fully Homomorphic Encryption (FHE) can mitigate these vulnerabilities to scale AI training with strong privacy guarantees.

Authors:

Dr Najwa Aaraj, CEO, Technology Innovation Institute

Dr Victor Mateu, Chief Researcher of the Cryptography Research Center, Technology Innovation Institute

Dr Victor Sucasas, Senior Director Cryptography Engineer, Cryptography Research Center, Technology Innovation Institute

tii.ae

Introduction

The Data Dilemma for Scalable AI

There is a growing need to train AI without compromising privacy, compliance, trust, and accuracy. Access to increasingly larger volumes of data and tools for processing has been essential in driving advances in training artificial intelligence (AI). The Internet of Things (IoT) is also making it easier to aggregate a much larger variety and volume of data for physical AI. Public and private cloud services are frequently used to aggregate and process data from many parties to train AI models.

Yet some of the most valuable and useful data, particularly in compliance-heavy/privacy-sensitive industries (medicine, finance, enterprise secrets, smart cities), have been difficult to use due to privacy concerns. Emerging tools for protecting the security and privacy of data will be essential for unlocking access to more valuable and useful data sets that require privacy. This includes a range of data types relating to health, finance, and business secrets, which can contain personally identifiable information (PII).

Privacy-Enhancing Technologies (PETs) is an emerging field that broadly ensures the security and privacy of this data. Privacy-Preserving Machine Learning (PPML) is a subset of PETs that safeguards the security and confidentiality of this data while using AI models. PPML is crucial for leveraging recent progress and the excitement in using neural networks in generative AI (GenAI), and other neural network types such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNNs).

A variety of PPML techniques are being explored. Some, such as synthetic data and secure enclaves, are seeing some traction in the market but lack the utility and/or privacy guarantees. Federated learning (FL) shows tremendous promise for decentralizing AI training, but it can result in new privacy vulnerabilities.

Recent progress in bolstering FL with Multi-Party Computation (MPC) and Fully Homomorphic Encryption (FHE) shows promise for building the foundation for scalable and trustworthy AI. Several solutions developed by TII and others are building frameworks and platforms to support the widespread adoption of these new techniques. Ongoing improvements in these techniques will be essential for creating the privacy and security foundations for the next generation of AI.

From PETs to PPML

The term Privacy Enhancing Technologies (PETs) was first coined in 1995 by researchers from the Dutch Protection Authority to characterize technologies that enforce privacy measures. Early work focused on aspects such as data minimization, anonymization, secure communication channels, and identity management systems.

These efforts focused on protecting data at rest and in transit, but not on how it was processed. This spurred the growth of the confidential computing field, which focuses on private data processing. It's application to Machine Learning came to be known as Privacy-Preserving Machine Learning (PPML). This work involved finding new ways to repurpose confidential computing techniques to train AI models and run ML algorithms without exposing sensitive data.

Several private data processing techniques have been applied to PPML. Examples include Differential Privacy (DP), Trusted Execution Environments (TEE), synthetic data, and data anonymization techniques. These techniques mitigate privacy concerns by a) adding random noise to datasets; b) restricting data processing to secure enclaves; c) transforming raw data into random-looking data with similar statistical properties; and d) removing personally identifiable information (PII). However, these techniques face limitations owing to questionable privacy guarantees, limited utility or scalability issues. MPC and FHE go a step further by providing privacy without reducing utility.

The promise of Federated Learning (FL)

More recently, FL has emerged as a promising technology to address these limitations by decentralizing training without requiring data movement. It stands out compared to training on synthetic data, because FL is trained on actual data, which improves the utility and, consequently, the accuracy of the AI models.

In FL, models are trained across decentralized devices using local data, and only gradients/updates are shared. This keeps data local, reducing privacy risks. The devices could be mobile phones, IoT equipment, cars, or trusted staging servers that aggregate data from smaller devices or across an organization.

The process starts with the central server initializing the model and sending it to clients. Each client trains the model on its local data and sends the updated model parameters to the central server. The central server then aggregates the updates to improve the global model.

Google researchers introduced the core ideas behind FL in 2016 in a seminal paper that explained a practical method for iteratively learning from decentralized data while reducing the communication overhead.[1] The first implementation enabled Google to enhance the accuracy of the Google Android keyboard predictive text app without requiring the sharing of raw user data. Researchers subsequently explored how this foundation could improve models in healthcare, finance, and IoT use cases.

Since then, researchers have explored a variety of algorithms to address several challenges, such as working with non-independent and identically distributed (Non-IID) data sets, improving communication efficiency, and enhancing fairness. They have also built several FL frameworks, such as TensorFlow Federated, PySyft, and Flower, to streamline the development process.

[1] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, “Communication-Efficient Learning of Deep Networks from Decentralized Data,” Jan. 26, 2023, arXiv: arXiv:1602.05629. doi: 10.48550/arXiv.1602.05629.

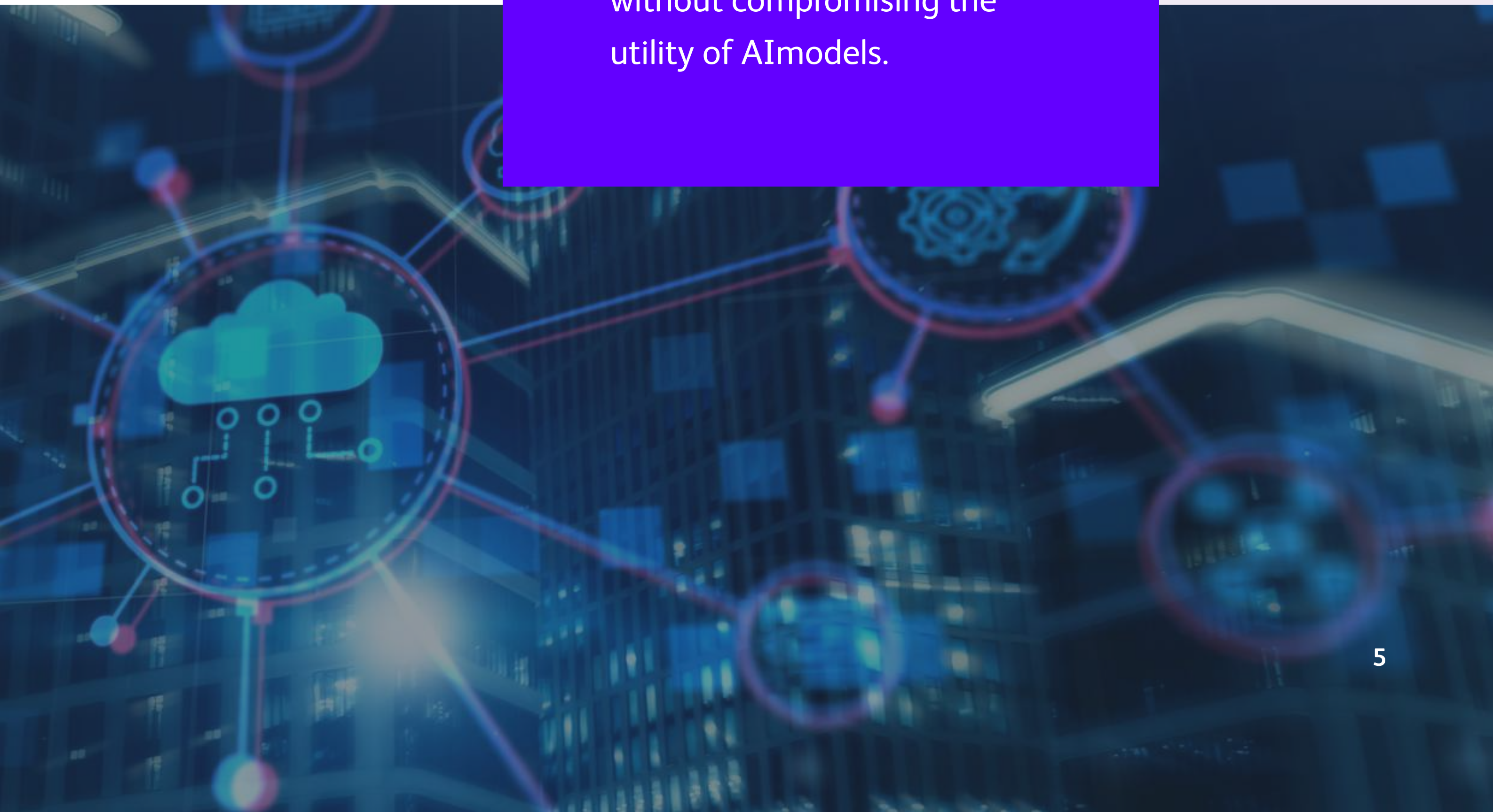
Mitigating inference attacks

In the early days, FL was thought to be secure against privacy attacks. However, researchers have discovered various inference attack vulnerabilities.

In a Membership Inference Attack (MIA), an adversary attempts to determine whether a specific data record was included in the training dataset of a machine learning model. The attacker doesn't necessarily want to reconstruct the entire training dataset, but rather to ascertain the presence or absence of a particular individual's data. In a model used to predict a medical condition, merely identifying that a patient's records were included could reveal sensitive information about that patient's medical status.

In a Source Inference Attack (SIA), a hacker takes control of the aggregating server to determine which specific client or data source contributed a particular data record to the global model's training. While a MIA tells you if a record was used, a SIA tells you who provided it.

MPC and FHE can improve the privacy guarantees of FL against these types of attacks without compromising the utility of AImodels.



Mitigating inference attacks

MPC could securely aggregate the client model updates without the server seeing individual updates, thus hindering server-side MIA based on raw updates. In MPC, multiple parties compute in a distributive fashion over secret-shared data. None of the parties can see any part of the data, not even a single item. The computation is fully decentralized but faces communication latency constraints.

FHE enables computation directly on encrypted data, which confounds the ability to conduct inference attacks. However, it is computationally expensive and increases communication overhead. There are many factors that contribute to the increased computational overhead including the impact of processing encrypted data that is much larger than raw data and the use of more complex encrypted processing operations that emulate the original ones.

Frameworks and Toolkits

Researchers are actively exploring ways that MPC and FHE could be used to bolster FL privacy guarantees or could be used on their own as part of PPML workflows.
For example, TII's PetalGuard[2] is a library that supports the use of MPC for federated learning. It provides a universal framework compatible with any model and scalable to any model size, including LLMs. It is currently available on AWS.

FL is only used for training ML models. Other approaches are required to protect the privacy of inference in ML systems using MPC, FHE, or TEEs. Inference is a separate problem that requires new tradeoffs and technical considerations, beyond the scope of this paper.

Examples of MPC frameworks that can be adapted to inference include TII's FANNG,[3] Meta's CrypTen and Microsoft's CryptoNets. These frameworks are good for smaller models but not large ones (e.g., billions of parameters).

TII's Versatile Encrypted Inference Library (VEIL) is a platform for inferencing using FHE. It's been tested in mobile skin health and spam detection applications. A compiler can transform some existing ML models into equivalents compatible with VEIL. Ongoing research and development are expanding the type and scale of ML applications it will support. Full details on VEIL are expected later in 2025.

2 [1] "Petal Guard | Home." Accessed: Jun. 10, 2025. [Online]. Available: https://petalguard.tii.ae/
3 N. Aaraj et al., "FANNG-MPC: Framework for Artificial Neural Networks and Generic MPC," 2023, 2023/1918. Accessed: Jun. 10, 2025. [Online]. Available: https://eprint.iacr.org/2023/1918

Research Directions and Open Problems

Both MPC and FHE techniques are being investigated for use on their own in the long run. However, they both incur significant overhead in computation and/or communication, which limits their practical adoption for large-scale deployments in the short run.

MPC limitations are often related to bandwidth and communication interactions, not just computational power. MPC tends to have high communication overhead due to the protocols' iterative nature. Newer designs can minimize this by using more computation for preprocessing. Outsourcing this to edge servers can help reduce limitations on individual devices.

FHE complexity remains a significant challenge despite considerable progress. Existing approaches require the use of complex encryption algorithms and large ciphertext sizes that can expand data several orders of magnitude during the encryption process. These issues can limit the performance or feasibility for many potential use cases. Hybrid FHE (combining conventional and homomorphic encryption, also known as transcription) is a potential solution for this bandwidth issue but comes at the cost of increasing computational complexity.

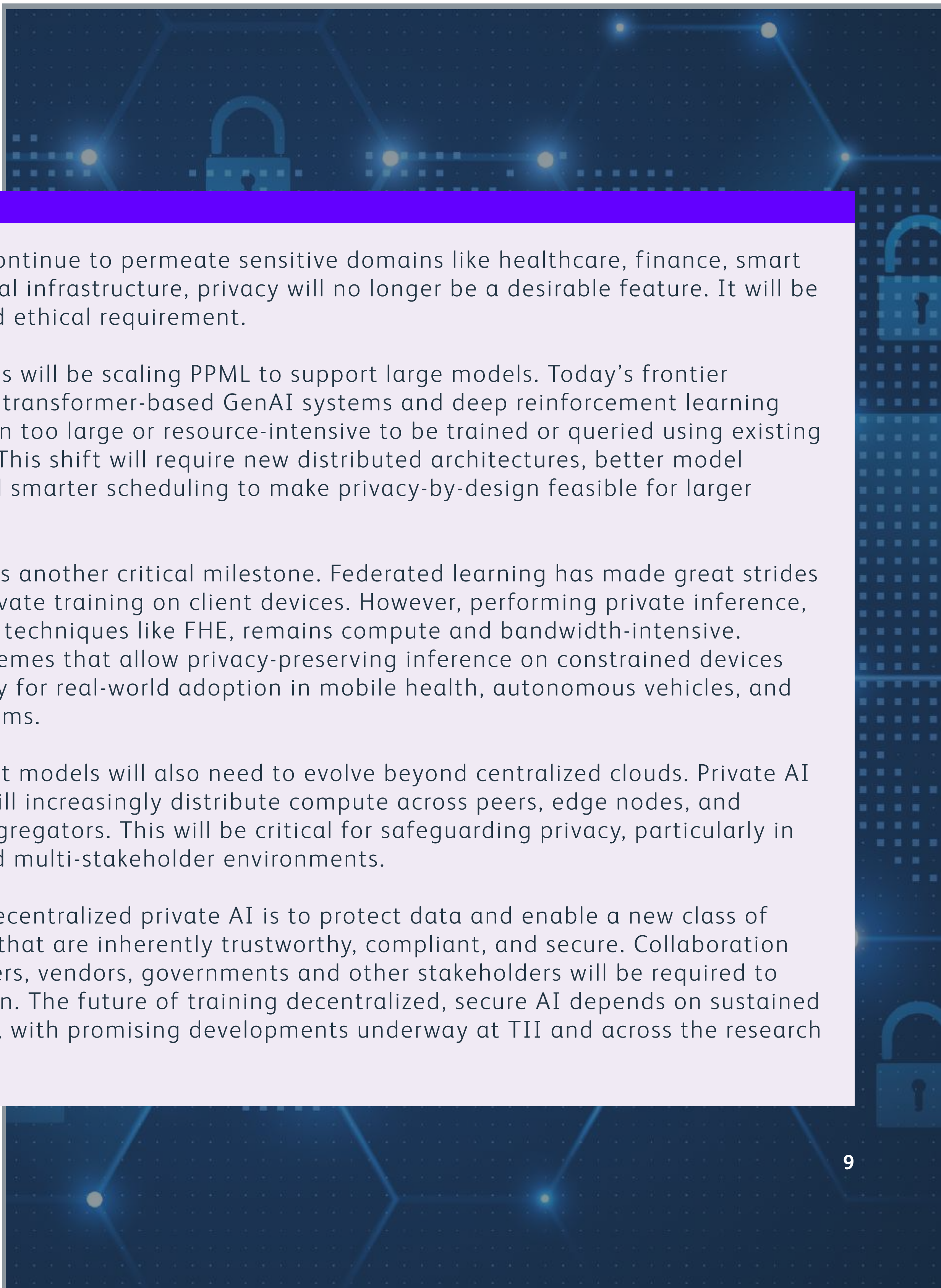
The FHE community is working on standardizing FHE and establishing a common implementation framework. This includes the development of tools for setting up FHE parameters,[4, 5, 6] the Homomorphic Encryption Standard, the creation of open FHE libraries,[8] and compilers for optimizing high-level programs into FHE implementations. FHE progress will also improve with hardware acceleration and new designs. Duality researchers have found that FPGA and GPU optimization techniques are already able to improve FHE performance by two orders of magnitude and predict that special-purpose application-specific integrated circuit (ASIC) chips could enhance performance by four orders of magnitude compared to multi-core CPUs. Efforts like OpenFHE will make it easier to write the algorithms once and then port them to new hardware as it becomes available.[9]

These kinds of improvements and efforts to combine MPC and/or FHE with FL will help foster the development of robust, scalable, and practical solutions that can enhance the security and privacy of the next generation of AI without compromising performance.

[4] Bergerat, L., Boudi, A., Bourgerie, Q. et al. (2022). Parameter optimization & larger precision for (T) FHE. Cryptology ePrint Archive. [5] Biasioli, B., Marcolla, C., Calderini, M., and Mono, J. (2023). Improving and automat- ing BFV parameters selection: an average-case approach. Cryptology ePrint Archive, Paper 2023/600. [6] Mono, J., Marcolla, C., Land, G. et al. (2023). Finding and evaluating parameters for BGV. International Conference on Cryptology in Africa - AFRICACRYPT 202. [7] Albrecht, M.R., Chase, M., Chen, H. et al. (2018). Homomorphic Encryption Security Standard. Technical Report. Toronto, Canada: HomomorphicEncryption.org. [8] "OpenFHE.org – OpenFHE – Open-Source Fully Homomorphic Encryption Library." Accessed: Jun. 09, 2025. [Online]. Available: https://openfhe.org/. [9] A. A. B. Rohloff David Bruce Cousins, Yuriy Polyakov, and Kurt, "Hardware Acceleration of Fully Homomorphic Encryption," Duality Technologies. Accessed: Jun. 09, 2025. [Online]. Available: https://dualitytech.com/blog/hardware-acceleration-of-fully-homomorphic-encryption-making-privacy-preserving-machine-learning-practical/

Conclusion

The Road Ahead for Private AI



As AI systems continue to permeate sensitive domains like healthcare, finance, smart cities, and critical infrastructure, privacy will no longer be a desirable feature. It will be a regulatory and ethical requirement.

A key part of this will be scaling PPML to support large models. Today's frontier models, such as transformer-based GenAI systems and deep reinforcement learning agents, are often too large or resource-intensive to be trained or queried using existing PPML schemes. This shift will require new distributed architectures, better model partitioning and smarter scheduling to make privacy-by-design feasible for larger models.

Edge inference is another critical milestone. Federated learning has made great strides in improving private training on client devices. However, performing private inference, especially using techniques like FHE, remains compute and bandwidth-intensive. Lightweight schemes that allow privacy-preserving inference on constrained devices will be necessary for real-world adoption in mobile health, autonomous vehicles, and embedded systems.

In addition, trust models will also need to evolve beyond centralized clouds. Private AI infrastructure will increasingly distribute compute across peers, edge nodes, and semi-trusted aggregators. This will be critical for safeguarding privacy, particularly in cross-border and multi-stakeholder environments.

The vision for decentralized private AI is to protect data and enable a new class of AI applications that are inherently trustworthy, compliant, and secure. Collaboration across researchers, vendors, governments and other stakeholders will be required to realize this vision. The future of training decentralized, secure AI depends on sustained PETs innovation, with promising developments underway at TII and across the research community.

TII Technology Innovation Institute

Innovation for a better world

tii.ae