Privacy Preserving Machine Learning

Team members: David Tran, Bryan Soerjanto, Amaris Young-Diggs, Minh Nguyen | **Faculty adviser:** Hong-Sheng Zhou, Ph.D. | **Sponsor: VCU College of Engineering** | **Mentor:** Hong-Sheng Zhou

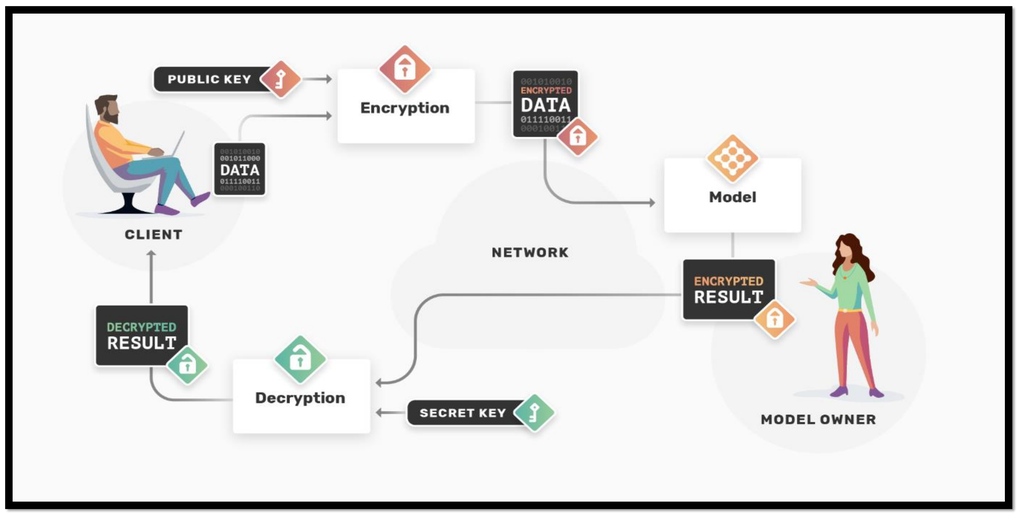
Executive Summary:
As organizations adopt machine learning, privacy concerns in fields like healthcare, finance, and marketing remain a challenge. This project develops a Privacy-Preserving Machine Learning (PPML) system using cryptographic techniques like homomorphic encryption and secure multi-party computation. By enabling secure data analysis without exposure, it ensures compliance and builds user trust.

Problem Statement:
How can organizations use machine learning while preserving data privacy throughout its lifecycle?

Key Constraints:
- The system must balance security, computational efficiency, and model accuracy.
- Cryptographic techniques should minimize overhead to remain practical for real-world applications.
- The solution must align with existing ML frameworks and comply with privacy regulations (e.g., GDPR, HIPAA).

Project Goals:
- Develop a Privacy-Preserving Machine Learning (PPML) system that enables secure data analysis without exposing sensitive information.
- Implement cryptographic techniques such as homomorphic encryption and secure multi-party computation to ensure data confidentiality.
- Evaluate the system's performance in real-world scenarios, particularly in industries like healthcare, finance, and marketing.
- Ensure compliance with data privacy regulations while maintaining high model accuracy and efficiency.

Privacy-Preserving Machine Learning Techniques:
- **Homomorphic Encryption (HE)**: Enables computations on encrypted data without decryption, ensuring privacy throughout ML workflows.
- **Federated Learning (FL)**: Distributes model training across decentralized data sources without sharing raw data.
- **Differential Privacy (DP)**: Adds controlled noise to data to prevent individual re-identification.
- **AES Encryption (Fernet)**: Secures trained models at rest, ensuring they are never stored in plaintext.
- **TenSEAL's CKKS Encryption**: Supports encrypted inference, enabling ML predictions without exposing input data.



Future Goals:
- Improve encryption efficiency using GPU Tensor Cores for batch operations and matrix multiplication.
- Optimize encrypted ML operations to reduce computational overhead and improve model scalability.
- Reduce accuracy loss in encrypted deep learning models
- Improve PPML interpretability.

Criterion	Homomorphic Encryption	Federated learning	Hybrid Approach
Data privacy	Higher	Medium	High
Performance Impact	High	Low	Medium
Scalability	Medium	High	Medium
Implementation Complexity	High	Medium	High

Conclusion: As Privacy-Preserving Machine Learning (PPML) becomes essential, organizations must integrate AI while safeguarding sensitive data. Full homomorphic encryption remains a long-term goal, but TenSEAL, built on Microsoft SEAL, provides a practical approach to partial homomorphic encryption (PHE). Combined with federated learning, differential privacy, and AES encryption, this system enables secure model training and inference without exposing raw data, ensuring confidentiality and compliance in privacy-critical domains.