1897

VICTORIA UNIVERSITY OF
WELLINGTON
TE HERENGA WAKA

School of Engineering and Computer Science
Te Kura Mātai Pūkaha, Pūrorohiko

PO Box 600
Wellington
New Zealand

Tel: +64 4 463 5341
Internet: office@ecs.vuw.ac.nz

Implementation of a Privacy-Preserving Machine Learning Model Using Homomorphic Encryption

Thomas Green
Supervisor: Arman Khouzani

Submitted in partial fulfilment of the requirements for
Bachelor of Engineering Honours in Cybersecurity

Abstract

This project explores the feasibility of implementing a privacy-preserving machine learning (ML) model using fully homomorphic encryption (FHE). As data security becomes increasingly critical across business sectors, solutions that enable computation on encrypted data, without revealing raw inputs, are gaining importance. This project proposes FHE as a solution to this, as a cryptographic method that allows operations to be performed directly on encrypted data, offering strong privacy guarantees. This project will use open-source libraries to train and evaluate basic ML models entirely on encrypted inputs, assessing performance, accuracy, and usability compared to an unencrypted baseline. While a working prototype and user interface may be developed to demonstrate real-world applications, the primary goal is to evaluate whether FHE is a viable and practical solution for secure ML under current technological constraints.

1. Introduction

With growing concerns around data privacy and collecting sensitive information, the ability to process data securely without exposing its underlying form has become a key engineering challenge. As machine learning (ML) models become increasingly integrated into business sectors, their reliance on raw data raises significant privacy and security risks.

This project proposes developing a privacy-preserving ML system using Fully Homomorphic Encryption (FHE), a cryptographic method that enables computations to be performed directly on encrypted data [1]. The aim is to investigate whether an end-to-end solution, where a selected ML model is trained and used for inference without decrypting input data, is feasible and practical under current constraints.

The project will be evaluated by comparing the performance and accuracy of the encrypted model with an equivalent unencrypted version. Exploration into real-world use of FHE will be performed once requirements have been met, with a demonstration with a simple UI to show how this technology could be applied in practice.

2. The Problem

As machine learning (ML) becomes increasingly integrated into business operations and decision-making, concerns around privacy and protection are crucial to address [2]. ML models require vast amounts of data to function effectively, much of which may include sensitive or proprietary information such as personally identifiable information (PII), protected health information (PHI), intellectual property (IP), and confidential business data. Organisations may be hesitant to expose this data in plaintext form during training and inference, especially in sectors like healthcare, finance, and defence, where data breaches can lead to severe legal, financial, and reputational consequences.

From a privacy and rights perspective, the current landscape of AI development often lacks transparency and proper consent [3]. Large-scale AI systems are frequently trained on datasets scraped from public and privacy sources, sometimes without authorisation or regard for data ownership [4]. This raises serious ethical and legal concerns, particularly when copyrighted or sensitive content is used without proper permission or handled without appropriate security practices. In addition to legal and reputational risks, the lack of data privacy opens up new vectors for cyberattacks. Adversaries may target ML systems by injecting malicious training data, stealing sensitive information, or exploiting vulnerabilities during data transmission and storage. These risks highlight the urgent need for secure computation methods that eliminate the exposure of raw inputs throughout the ML lifecycle.

To maintain trust, comply with legal/regulatory requirements, and reduce the risk of exposure, it is essential to explore new approaches that enable secure computation of sensitive data. A system that enables ML workflows to preserve the confidentiality of input data during training and inference would represent a significant step forward in responsible AI development.

Investigating the feasibility, performance and practicality of such methods offers a pathway toward more secure, ethical, and deployable ML systems in real-world environments.

3. Proposed Solution

The proposed solution involves developing a privacy-preserving machine learning (ML) pipeline that uses Fully Homomorphic Encryption (FHE) to operate directly on encrypted data. FHE is an advanced cryptographic technique that enables computations to be performed on ciphertexts, producing encrypted results that, when decrypted, match the output of the same operations on the original plaintext [1]. In other words, it allows data to remain encrypted throughout the entire computational process, eliminating the need to expose raw inputs at any stage. This property makes FHE especially attractive for applications where data confidentiality and user privacy are important. FHE supports confidentiality, integrity, and privacy principles, offering a strong foundation for secure ML workflows. A selection of basic ML models, such as logistic regression, linear regression, and k-nearest neighbour, will be evaluated to determine which algorithms are most compatible with the computational constraints of FHE.

Despite its promise, FHE has yet to see widespread adoption in real-world ML systems due to its significant performance overhead, complex implementation, and limited library support. Trivial operations in plaintext become significantly more computationally expensive and restrictive when performed over encrypted data. To explore the trade-off between privacy and practicality, this project will also examine adjustable security parameters, such as key lengths or encryption schemes, to evaluate whether reducing secure practices would improve performance during prototyping. These controlled compromises will help assess the feasibility of integrating FHE into ML workflows under constrained conditions, offering insight into its potential for future real-world applications.

To implement the solution, an analysis of open-source libraries and repositories will be studied, and selection will be based on ease of integration and support of ML operations. A baseline ML model will first be implemented using plaintext data to establish benchmarks for performance and accuracy. The same model will then be recreated using encrypted data and FHE to perform training and inference. Publicly available or synthetic datasets will be used to test and evaluate the system. A simple user interface may also be developed to demonstrate how FHE-backed systems could be integrated into real-world software and made accessible to broader, possibly non-technical audiences.

Project Timeline

Task	Description	Estimated Duration
Background Research	Research FHE concepts, current libraries, and ML model constraints.	1 – 2 weeks
Project Design	Plan and Design Implementation. Select FHE library and candidate ML models. Justify choices.	1 – 2 weeks
FHE Setup	Implement encryption/decryption pipeline using chosen FHE libraries.	1 – 2 weeks
Implement Baseline ML Model	Implement a basic unencrypted ML model for comparison using selected ML candidates.	1 – 2 weeks

Encrypted ML Integration	Train the same model on encrypted data using FHE	1 - 2 weeks
Evaluation of Current FHE and ML Implementation	Compare encrypted vs. unencrypted models on accuracy, runtime, etc.	1 week
Explore Adjustments	Explore and contrast other ML models, key adjustments, etc.	1 – 2 weeks
Evaluate and Plan Real-World Implementation	Analyse produced FHE and ML implementation and decide on real-world UI implementation	1 week
UI Implementation	Build a simple UI	1 – 2 weeks
Full Integration	Connect UI implementation, encryption pipeline and ML model	1 – 2 weeks
Final Evaluation	Evaluate the entire system: performance, usability, feasibility, etc.	1 – 2 weeks
Report and Presentation	Write a final report and prepare a presentation/demo	1 – 3 weeks

4. Evaluating your Solution

The evaluation of this project will focus on both the technical performance and practical viability of the implemented system. Since the goal is to explore the feasibility of a privacy-preserving ML pipeline using encrypted data, the evaluation will assess how effectively the system preserves privacy while delivering meaningful and accurate model outputs. The evaluation will be structured around three key areas:

1. Correctness and Functionality: The primary objective is to ensure the encrypted ML model behaves as expected. Predictions on encrypted data should closely align with those of an equivalent plaintext model. This will involve unit testing of encryption routines, validating model outputs, and conducting consistency checks between encrypted and unencrypted workflows to confirm functional correctness.
2. Performance and Trade-Off Analysis: The encrypted model will be benchmarked against a baseline unencrypted version to evaluate runtime performance, memory usage, model accuracy, and encryption/decryption overhead. Based on these results, parameters such as key size, ciphertext modulus, or model architecture (e.g. linear regression vs. k-nearest-neighbour) may be adjusted to optimise performance while maintaining reasonable privacy and security standards. These trade-offs are an essential part of assessing FHE's practicality in real-world scenarios.
3. Usability and Real-World Demonstration: The system will also be evaluated in terms of usability and applicability. A basic user interface may be developed to simulate how the encrypted model could function in a real-world software environment. This will support evaluation of the system's accessibility, particularly for non-technical users, and serve as a communication tool to demonstrate the potential integration of privacy-preserving ML in business applications.

The evaluation process will be iterative, with findings from each phase informing potential changes in direction. For instance, if performance is insufficient, the project may pivot to a simpler ML model, or adjust encryption parameters to improve speed while balancing privacy. This flexible and exploratory approach ensures the final outcome is not only technically sound but also provides a realistic assessment of the viability of FHE-based ML in practical settings.

5. Resourcing and Ethics

5.1. Software, Datasets and Models

This project will involve a range of specialised software libraries and frameworks to support development. As the project is still in the early stages, specific tools, libraries, and frameworks will be selected based on factors such as ease of integration, performance, documentation, and community support. For FHE implementation, the project will explore modern FHE open-source Python and C++ libraries and repositories such as PySeal, TenSEAL, Helib. Other open-source FHE libraries or GitHub repositories may also be considered depending on implementation needs. For ML implementation, well-used and supported libraries such as scikit-learn, TensorFlow, PyTorch, or Keras will be explored, as each offers its own benefits regarding the simplicity and complexity of component and model tuning capabilities. Alternatively, some models can be implemented from scratch in Python to allow full control over encrypted operations and integration. For dataset selection, publicly available datasets may be used for testing and evaluation, depending on model requirements. Options such as Iris, Titanic, or MNIST datasets can be investigated, and other platforms such as Kaggle, OpenML, or UCI ML Repository offer a more simplistic, interpretable, and computationally cheap approach. Dataset selection will remain flexible, due to that not being the projects focus.

5.2. Budget

The proposed budget for this project is $200. This amount is allocated for potential software or tool licenses that may be required during the development. At this stage, the exact requirements are not fully determined, as I will be assessing various tools, libraries, and datasets as the project progresses. The budget will be used to cover the costs of any necessary software licenses or subscriptions, ensuring access to the tools required to complete the project successfully.

5.3. Ethics

This project does not directly involve any direct interaction with human participants or the use of personal, private, or sensitive real-world data. As such, formal human ethics approval is not required. However, the project operates within a domain that is inherently privacy-focused, as it explores methods to protect sensitive information during ML workflows. The project simulates scenarios where data confidentiality is critical. Ethical considerations include the responsible communication of the project's outcomes, ensuring that claims about privacy and security are realistic and do not overstate the guarantees provided by experimental implementations. The project will also explore the balance between practical usability and cryptographic security, and any trade-offs made will be clearly stated and contextualised to avoid misrepresentation and misinterpretation. No identifiable or sensitive information will be collected, stored, or processed throughout this project.

5.4. Safety

This project only involves the design and implementation of software systems and does not include any physical components, lab-based experiments, or interaction with hazardous materials or environments. The key safety considerations related to prolonged computer use include potential risks such as eye strain, back or neck discomfort, and repetitive strain injury. These will be mitigated by following standard workstation human factors, taking regular breaks, and using proper seating and posture while working. Electrical safety will also be observed by ensuring that any personal computing equipment used (e.g. laptop and peripherals) is in good working condition and used following university guidelines. All development will occur on personal or university-provided computing equipment in standard study or lab environments. A health and safety plan has been completed and submitted as required.

5.5. Intellectual Property

This project is being carried out under the Victoria University of Wellington Intellectual Property Policy, and the Student Intellectual Property Agreement has been signed in accordance with that policy. As the agreement outlines, all intellectual property (IP) developed as part of this project, such as original software, encryption workflows, and documentation, will be assigned to the University. While the project is academic and not intended for commercialisation, it may result in a proof-of-concept system demonstrating the practical application of privacy-preserving ML techniques. If the final implementation proves novel or commercially relevant, the University reserves the right to explore avenues for protection and commercialisation of the IP. No third-party IP agreements or industry partners are currently involved in this project. All tools will be open source, and any reuse will adhere to their respective licenses. Any public disclosure of results, code, or documentation will be subject to review and approval as required under the confidentiality terms of the Student IP Agreement.

References

[1] Wikipedia Contributors, “Homomorphic encryption,” *Wikipedia*, Nov. 03, 2019. https://en.wikipedia.org/wiki/Homomorphic_encryption

[2] E. Etheridge, “Cyber security measures: Secure your business with digital signatures,” *Dataguard.co.uk*, vol. 1, no. 1, Jun. 2024, Available: https://www.dataguard.co.uk/blog/what-is-a-digital-signature-and-how-it-can-secure-your-business/#which-industries-need-digital-signatures-most

[3] Dentons, “AI and intellectual property rights,” *Dentons.com*, 2025. https://www.dentons.com/en/insights/articles/2025/january/28/ai-and-intellectual-property-rights

[4] V. Berger, “Ex OpenAI Researcher: How ChatGPT’s Training Violated Copyright Law,” *Forbes*, Oct. 29, 2024. Available: https://www.forbes.com/sites/virginieberger/2024/10/29/ex-openai-researcher-how-chatgpts-training-violated-copyright-law/