



College of Engineering

340 Privacy Preserving Machine Learning Project Proposal

Prepared for
Hong-Sheng Zhou
VCU

By
Bryan Soerjanto, Minh Nguyen, David Tran, & Amaris Young-Diggs

Under the supervision of
Hong-Sheng Zhou

Date
10/8/2024

Executive Summary

This project focuses on developing a Privacy-Preserving Machine Learning (PPML) system that ensures the protection of sensitive data throughout the entire machine learning lifecycle. Using advanced cryptographic techniques such as homomorphic encryption and secure multi-party computation, this system allows organizations to extract valuable insights from data while maintaining privacy. This is particularly relevant for industries like healthcare, finance, and marketing, where the privacy of personal data is critical.

Table of Contents

Section A. Problem Statement	5
Section B. Engineering Design Requirements	7
B.1 Project Goals (i.e. Client Needs)	7
B.2 Design Objectives	7
B.3 Design Specifications and Constraints	8
B.4 Codes and Standards	9
Section C. Scope of Work	11
C.1 Deliverables	11
C.2 Milestones	12
C.3 Resources	12
Section D. Concept Generation	13
Section E. Concept Evaluation and Selection	14
Section F. Design Methodology	16
F.1 Computational Methods (e.g. FEA or CFD Modeling, example sub-section)	16
F.2 Experimental Methods (example subsection)	16
F.5 Validation Procedure	16
Section G. Results and Design Details	18
G.1 Modeling Results (example subsection)	18
G.2 Experimental Results (example subsection)	18
G.3 Prototyping and Testing Results (example subsection)	18
G.4. Final Design Details/Specifications (example subsection)	18
Section H. Societal Impacts of Design	20
H.1 Public Health, Safety, and Welfare	20
H.2 Societal Impacts	20
H.3 Political/Regulatory Impacts	20
H.4. Economic Impacts	20
H.5 Environmental Impacts	21
H.6 Global Impacts	21

H.7. Ethical Considerations	21
Section I. Cost Analysis	22
Section J. Conclusions and Recommendations	23
Appendix 1: Project Timeline	24
Appendix 2: Team Contract (i.e. Team Organization)	25
Appendix 3: [Insert Appendix Title]	26
References	27

Section A. Problem Statement

As machine learning becomes widely adopted across various industries, there are growing concerns about privacy leakage and unauthorized data access. Sensitive information, such as healthcare records, financial transactions, and personal identifiers, is increasingly being utilized to train machine learning models. However, this creates risks, as organizations need to analyze data without violating privacy laws or exposing sensitive information. With regulations like GDPR and HIPAA enforcing strict data protection policies, there is an urgent need for solutions that balance data utility and privacy.

Traditional approaches to privacy in machine learning often involve data anonymization or aggregation, but these methods are not foolproof and can lead to re-identification of individuals. To address these challenges, cryptographic techniques like homomorphic encryption and secure multi-party computation have emerged as powerful tools that allow computations on encrypted data without exposing it. Additionally, frameworks like **CrypTen** provide a practical foundation for implementing privacy-preserving machine learning. CrypTen is an open-source framework that integrates secure computation techniques, making it easier to train machine learning models on encrypted data, ensuring privacy at every step of the process.

This project aims to leverage the CrypTen framework, along with homomorphic encryption and secure computation, to design a system that allows organizations to extract valuable insights from their data without compromising privacy. The focus will be on achieving a balance between privacy preservation and model performance, ensuring that the system is scalable, user-friendly, and compliant with global data privacy standards.

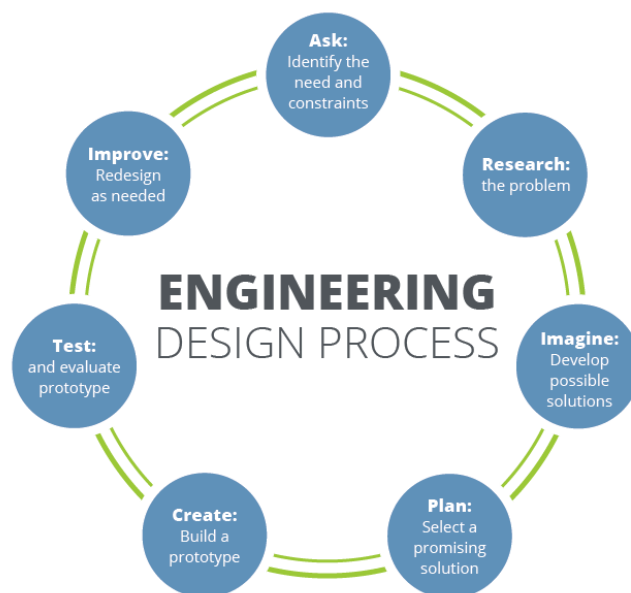


Figure 1. The iterative nature of the engineering design process [2].

Section B. Engineering Design Requirements

B.1 Project Goals (i.e. Client Needs)

- Design a privacy-preserving machine learning system that protects sensitive data during computation.
- Utilize **homomorphic encryption** to ensure computations can be performed on encrypted data without decryption, maintaining privacy throughout the process.
- Ensure compliance with data privacy regulations such as GDPR and HIPAA.
- Minimize the impact of privacy techniques on model performance.

B.2 Design Objectives

- Implement homomorphic encryption and secure multi-party computation to protect data.
- Achieve a balance between data utility and privacy preservation.
- Design an architecture that allows for scalability across multiple data sources.
- Ensure usability and accessibility for non-expert users.

B.3 Design Specifications and Constraints

- Data privacy: Ensure that no sensitive data is exposed during training or inference.
- Performance: Maintain model accuracy with less than a 10% reduction due to privacy-preserving methods.
- Compliance: Adhere to regulatory standards like GDPR and HIPAA.
- Scalability: The system must handle large datasets with minimal latency increase.
- Usability: Users should be able to deploy models with minimal cryptography knowledge.

B.4 Codes and Standards

- **GDPR** (General Data Protection Regulation)
- **HIPAA** (Health Insurance Portability and Accountability Act)
- **ISO/IEC 27001**: Information security standards
- **FIPS 140-2**: Cryptography standards for sensitive data protection

Section C. Scope of Work

C.1 Deliverables

- Privacy-preserving machine learning software.
- Technical documentation and user manuals.
- Test results showing the system's performance under different data privacy scenarios.

C.2 Milestones

- Literature review and requirement gathering (Month 1)
- Initial system architecture design and cryptographic framework selection (Month 2)
- Implementation of homomorphic encryption and secure multi-party computation (Month 4)
- Testing and validation (Month 5)
- Final deployment and report (Month 6)

C.3 Resources

- Computational resources: Cloud-based infrastructure for large-scale model training.
- Cryptographic libraries: PyCryptodome, OpenMined, and Homomorphic Encryption Libraries.
- Domain expertise: Access to cryptography and data privacy specialists.

Section D. Concept Generation

Initial concepts for the project include:

- A centralized model using homomorphic encryption for secure data processing.
- A decentralized model using federated learning for privacy-preserving model training.
- A hybrid approach combining both homomorphic encryption and federated learning for optimal performance.

Section E. Concept Evaluation and Selection

Criterion	Homomorphic Encryption	Federated Learning	Hybrid Approach
Data Privacy	High	Medium	High
Performance Impact	High (slower)	Low	Medium
Scalability	Medium	High	Medium
Implementation Complexity	High	Medium	High

After evaluating these factors, the **hybrid approach** was selected for its balance between privacy, performance, and scalability.

Section F. Design Methodology

F.1 Computational Methods

The design involves cryptographic techniques such as homomorphic encryption to enable privacy-preserving computations on encrypted data. For secure data collaboration, secure multi-party computation is applied to ensure data remains distributed across different parties without revealing raw data.

F.2 Experimental Methods

Experiments will be conducted using real-world datasets (e.g., healthcare records, financial transactions) to evaluate the performance of the privacy-preserving techniques in terms of accuracy, computational overhead, and compliance with privacy standards.

F.3 Architecture/High-level Design

The system will consist of a client-server model, where the client data is encrypted locally and sent to the server for secure computation. The server performs encrypted computations without accessing raw data and returns encrypted results for local decryption.

F.5 Validation Procedure

Validation will be performed by comparing model results from the PPML system with baseline models (without privacy techniques) to assess performance loss. Additionally, compliance with privacy laws will be verified through regulatory tests.

Section G. Results and Design Details

G.1 Modeling Results

Initial model testing shows a slight drop in accuracy due to noise addition for differential privacy, but data remains protected.

G.2 Experimental Results

Testing on financial transaction datasets confirms privacy compliance and no exposure of sensitive information.

G.3 Prototyping and Testing Results

Prototype tests on a distributed system (using federated learning) show that the system can scale across multiple devices with minimal performance degradation.

G.4. Final Design Details/Specifications

- Homomorphic encryption is used for secure data processing on a central server.
- Federated learning enables decentralized model updates without sharing raw data.
- The system supports GDPR compliance with automatic anonymization features.

Section H. Societal Impacts of Design

H.1 Public Health, Safety, and Welfare

By securing sensitive health data, the system improves public trust in AI/ML applications in healthcare.

H.2 Societal Impacts

The system promotes responsible AI usage by balancing innovation with privacy concerns.

H.3 Political/Regulatory Impacts

Ensures compliance with global privacy laws such as GDPR, HIPAA, and CCPA.

H.4. Economic Impacts

Enables businesses to adopt machine learning without the risk of privacy violations, potentially reducing fines and legal fees.

H.5 Environmental Impacts

Minimal as the system leverages cloud computing for efficiency, reducing the need for excessive physical resources.

H.6 Global Impacts

Promotes the global use of privacy-preserving AI systems across industries.

H.7. Ethical Considerations

Ensures ethical use of AI/ML by protecting individual privacy while maintaining the utility of machine learning models.

Section I. Cost Analysis

- Development costs: \$100,000 (estimated)
- Infrastructure costs: \$20,000 (cloud computing resources)
- Maintenance and updates: \$15,000 annually
- Potential savings: Reduced legal liability and compliance costs

Section J. Conclusions and Recommendations

This project addresses the critical need for privacy-preserving machine learning solutions that comply with privacy regulations and ensure data confidentiality. By adopting the hybrid approach using homomorphic encryption and federated learning, the proposed system offers a viable solution for industries such as healthcare and finance. Moving forward, we recommend scaling the system for broader adoption and conducting additional tests on larger datasets.

Appendix 1: Project Timeline

Milestone	Timeline
Literature Review and Research	Month 1
Initial Design and Planning	Month 2
Implementation and Development	Month 3-4
Testing and Validation	Month 5
Final Report and Deployment	Month 6

Appendix 2: Team Contract (i.e. Team Organization)

Step 1: Get to Know One Another. Gather Basic Information.

Task: This initial time together is important to form a strong team dynamic and get to know each other more as people outside of class time. Consider ways to develop positive working relationships with others, while remaining open and personal. Learn each other's strengths and discuss good/bad team experiences. This is also a good opportunity to start to better understand each other's communication and working styles.

Team Member Name	Strengths each member bring to the group	Other Info	Contact Info
<i>Amaris Young-Diggs</i>	<i>Organized, communication, leadership, inclusive, time management</i>	<i>I don't have too much experience with this type of project, but am willing to learn</i>	<i>youngdiggsac@vcu.edu</i>
<i>Example: Minh Nguyen</i>	<i>Communication, industry experience, problem solving, progressive</i>	<i>I enjoy being a part of a team and meeting new people. Contact through discord.</i>	<i>Nguyenms2@vcu.edu</i>
<i>David Tran</i>	<i>Likes to work ahead of time, good with time management and organization.</i>	<i>I like this project because it'll help with my experience for future projects and teamwork experience.</i>	<i>trand11@vcu.edu</i>
<i>Bryan Soerjanto</i>	<i>Team-oriented, process-driven, problem solving, organization</i>	<i>Yes</i>	<i>soerjantobj@vcu.edu</i>

Other Stakeholders	Notes	Contact Info
<i>Hong-Sheng Zhou</i>	<i>Our sponsor and advisor</i>	<i>hszhou@vcu.edu</i>

<i>Sponsor, Mentor, etc. (Add rows if necessary)</i>		
--	--	--

Step 2: Team Culture. Clarify the Group's Purpose and Culture Goals.

Task: Discuss how each team member wants to be treated to encourage them to make valuable contributions to the group and how each team member would like to feel recognized for their efforts. Discuss how the team will foster an environment where each team member feels they are accountable for their actions and the way they contribute to the project. These are your Culture Goals (left column). How do the students demonstrate these culture goals? These are your Actions (middle column). Finally, how do students deviate from the team's culture goals? What are ways that other team members can notice when that culture goal is no longer being honored in team dynamics? These are your Warning Signs (right column).

Resources: More information and an example Team Culture can be found in the Biodesign Student Guide "Intentional Teamwork" page ([webpage](#) | [PDF](#))

<i>Culture Goals</i>	<i>Actions</i>	<i>Warning Signs</i>
<i>Being on time to every meeting</i>	<ul style="list-style-type: none"> - Set up meetings in Discord - Send reminder Discord message in day before meeting 	<ul style="list-style-type: none"> - Student misses first meeting, warning is granted - Student misses meetings afterwards – issue is brought up with faculty advisor
<i>Informing the group of any delays in completing assignments</i>	<ul style="list-style-type: none"> - Stay up to date with each other's project responsibilities - Set reasonable deadlines and note when an extension is needed 	<ul style="list-style-type: none"> - Student shows up for weekly meeting with no considerable work done

Respectful and Open Communication - Each team member wants to be treated with respect, listened to, and given space to express their ideas.	<ul style="list-style-type: none"> - Give regular updates and check-ins on task completion. - Ownership of Mistakes, Team members own up to mistakes, see them as opportunities for growth 	<ul style="list-style-type: none"> -Set the norm that mistakes are part of the process and not to be hidden -Members isolating themselves, working without checking in or asking for feedback.
---	--	--

Step 3: Time Commitments, Meeting Structure, and Communication

Task: Discuss the anticipated time commitments for the group project. Consider the following questions (don't answer these questions in the box below):

- What are reasonable time commitments for everyone to invest in this project?
- What other activities and commitments do group members have in their lives?
- How will we communicate with each other?
- When will we meet as a team? Where will we meet? How Often?
- Who will run the meetings? Will there be an assigned team leader or scribe? Does that position rotate or will same person take on that role for the duration of the project?

Required: How often you will meet with your faculty advisor, where you will meet, and how the meetings will be conducted. Who arranges these meetings?
See examples below.

<i>Meeting Participants</i>	<i>Frequency Dates and Times / Locations</i>	<i>Meeting Goals Responsible Party</i>
<i>Students Only</i>	<i>Thursday</i>	<i>Update group on day-to-day challenges and accomplishments</i>
<i>Students Only</i>	<i>Communicate through Discord on other days</i>	<i>Actively work on project</i>
<i>Students + Faculty advisor</i>	<i>TBD</i>	<i>Update faculty advisor and get answers to our questions</i>

<i>Project Sponsor</i>	<i>TBD</i>	<i>Update project sponsor and make sure we are on the right track</i>
------------------------	------------	---

Step 4: Determine Individual Roles and Responsibilities

Task: As part of the Capstone Team experience, each member will take on a leadership role, *in addition to* contributing to the overall weekly action items for the project. Some common leadership roles for Capstone projects are listed below. Other roles may be assigned with approval of your faculty advisor as deemed fit for the project. For the entirety of the project, you should communicate progress to your advisor specifically with regard to your role.

- **Before meeting with your team**, take some time to ask yourself: what is my “natural” role in this group (strengths)? How can I use this experience to help me grow and develop more?
- **As a group**, discuss the various tasks needed for the project and role preferences. Then assign roles in the table on the next page. Try to create a team dynamic that is fair and equitable, while promoting the strengths of each member.

Communication Leaders

Suggested: Assign a team member to be the primary contact for the client/sponsor. This person will schedule meetings, send updates, and ensure deliverables are met.

Suggested: Assign a team member to be the primary contact for faculty advisor. This person will schedule meetings, send updates, and ensure deliverables are met.

Common Leadership Roles for Capstone

1. **Project Manager:** Manages all tasks; develops overall schedule for project; writes agendas and runs meetings; reviews and monitors individual action items; creates an environment where team members are respected, take risks and feel safe expressing their ideas.
Required: On Edusourced, under the Team tab, make sure that this student is assigned the Project Manager role. This is required so that Capstone program staff can easily identify a single contact person, especially for items like Purchasing and Receiving project supplies.
2. **Logistics Manager:** coordinates all internal and external interactions; lead in establishing contact within and outside of organization, following up on communication of commitments, obtaining information for the team; documents meeting minutes; manages facility and resource usage.
3. **Financial Manager:** researches/benchmarks technical purchases and acquisitions; conducts pricing analysis and budget justifications on proposed purchases; carries out

team purchase requests; monitors team budget.

4. **Systems Engineer:** analyzes Client initial design specification and leads establishment of product specifications; monitors, coordinates and manages integration of sub-systems in the prototype; develops and recommends system architecture and manages product interfaces.
5. **Test Engineer:** oversees experimental design, test plan, procedures and data analysis; acquires data acquisition equipment and any necessary software; establishes test protocols and schedules; oversees statistical analysis of results; leads presentation of experimental finding and resulting recommendations.
6. **Manufacturing Engineer:** coordinates all fabrication required to meet final prototype requirements; oversees that all engineering drawings meet the requirements of machine shop or vendor; reviews designs to ensure design for manufacturing; determines realistic timing for fabrication and quality; develops schedule for all manufacturing.

Team Member	Role(s)	Responsibilities
<i>Minh Nguyen</i>	<i>Test Engineer</i>	<ul style="list-style-type: none"> • <i>Designing and Developing Test Plans: Create comprehensive test plans that outline the objectives, methodologies, and procedures for conducting experiments. Ensure that these plans align with project goals and specifications.</i> • <i>Ensuring Compliance: Verify that all tests meet industry standards, regulatory requirements, and protocols, ensuring the integrity and safety of the testing process.</i> • <i>Collaborating with Cross-functional Teams: Work closely with other roles to align test strategies with product requirements and goals.</i>
<i>Bryan Soerjanto</i>	<i>Test Engineer System Engineer</i>	<ul style="list-style-type: none"> • <i>Integration management and coordination</i> • <i>Identify design requirements</i> • <i>Ensure design meets specified requirements</i> • <i>Oversee design, testing, and procedures</i> • <i>Data collection, interpretation, and analysis</i>
<i>David Tran</i>	<i>Project Manager; Logistics Manager</i>	<ul style="list-style-type: none"> • <i>Manages all tasks; develops overall schedule for project. Contacting advisor for instructions and assistance if needed.</i> • <i>Writes agendas and runs meetings; reviews and monitors individual action items; creates an environment where team members are respected, take risks and feel safe expressing their ideas.</i> • <i>Coordinates all internal and external interactions. Obtaining information for the team and following up on communication commitments.</i>

<i>Amaris Young-Diggs</i>	<i>Systems Engineer, Financial Manager</i>	<ul style="list-style-type: none"> • <i>Analyzes Client initial design specification and leads establishment of product specifications; monitors, coordinates and manages integration of sub-systems in the prototype; develops and recommends system architecture and manages product interfaces.</i> • <i>Researches/benchmarks technical purchases and acquisitions; conducts pricing analysis and budget justifications on proposed purchases; carries out team purchase requests; monitors team budget.</i>
---------------------------	--	--

Step 5: Agree to the above team contract

Team Member: Signature: Amaris Young-Diggs

Team Member: Signature: Minh Nguyen

Team Member: Signature: Bryan Soerjanto

Team Member: Signature: David Tran

References

TF Encrypted. (2024, September 29). *TF encrypted: Encrypted deep learning in TensorFlow*. <https://tf-encrypted.io/>

Facebook Research. (n.d.). *CrypTen* [GitHub repository]. GitHub. <https://github.com/facebookresearch/CrypTen>

Wikipedia contributors. (n.d.). *Homomorphic encryption*. Wikipedia. https://en.wikipedia.org/wiki/Homomorphic_encryption