PhD Research Proposal
Secure Privacy-Preserving Inference

Artem Grigor

April 5, 2024

Abstract

In the digital age, ensuring the privacy and security of machine learning (ML) systems is not merely a necessity but a fundamental prerequisite for the field's growth and societal acceptance. This research proposal is aimed at addressing the critical need for robust privacy-preserving mechanisms in ML, with a special emphasis on secure ML inference within adversarial environments. Our goal is to develop a practical Secure Privacy-Preserving Inference scheme that maintains data privacy and ensures the integrity of inference results.

At the core of our proposed methodology is the "Bringing ML to Data" strategy. This approach is designed to minimize the risks traditionally associated with data processing techniques. Our research will delve into verifiable inference optimization and the study of formal ML security mechanisms. This comprehensive approach is expected to significantly enhance the privacy and integrity of ML applications, thereby making them more applicable and efficient across various industries.

The primary outcome of our research will be a practical and scalable solution for conducting secure inference on private data. We anticipate that this solution will facilitate unprecedented access to data utility across multiple domains, driving transformative changes in data utilization practices. By carefully balancing the immense potential of ML with the critical demands for data security and user privacy, our work aims to encourage the adoption of more ethical and secure technological practices within the ML community. Ultimately, we envision our efforts contributing to a future where technology, guided by a deep respect for privacy and security, serves humanity more effectively and responsibly.

Contents

1 Introduction

In our current digital era, we have the unparalleled capability to generate, collect, and analyze vast quantities of data, unlocking significant benefits across multiple domains [32]. However, zettabytes of private data with the potential to transform lives remain largely untapped due to prevailing concerns over providing adequate privacy and security guarantees [18, 31][1].

Recent advancements, such as Differential Privacy [23], Federated Learning [35], Synthetic Data generation [47] have facilitated practical Machine Learning (ML) model training on distributed private data without explicit data sharing. Despite these advancements, a gap remains when the focus shifts to inference—an area of increasing interest [2].

This highlights a pressing void in the current landscape: the need for and absence of schemes **to perform efficient, privacy-preserving and tamper-resistant inference on private data within adversarial environments**. Throughout this proposal, we will refer to such schemes as **Secure Privacy-Preserving Inference**.

1.1 Security Model

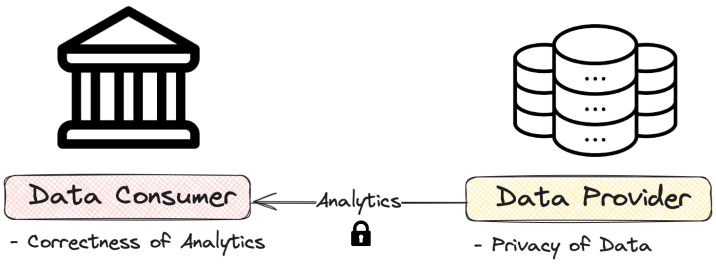

Figure 1: Setting: Two mutually distrusting entities interacting with each other.

The security model for Secure Privacy-Preserving Inference outlines the interactions and security requirements between two mutually distrusting entities:

1. **Data Provider** - An entity possessing sensitive, potentially third-party data, aims to protect the data's confidentiality while utilizing it in a privacy-preserving manner to extract benefits from the data.

   Example: Individuals could serve as Data Providers, with their smartphone texts and photos acting as sensitive data. Utilizing such data could, for instance, improve the accuracy of credit scores or reduce insurance premiums, showcasing the benefits of secure data usage [24, 15].

2. **Data Consumer** - An entity that analyzes the Data Provider's information without necessarily compromising its privacy. It is vital for the Data Consumer to ensure

[1] Data Privacy needs to be combined with Data Utility - https://www.prospectmagazine.co.uk/essays/52612/who-needs-digital-privacy

[2] AI: Nvidia Focused on Inference - https://medium.com/@mparekh/ai-nvidia-focused-on-inference-0ebf167238a0

the analytics results' accuracy and integrity are safeguarded from manipulations by the Data Provider.

> *Example:* Banks and Insurance Companies, acting as Data Consumers, could analyze clients' personal data to more accurately predict loan default risks [29] or health risks. This facilitates a more precise risk evaluation, illustrating the real-world benefits of accessing personal data.This facilitates a more precise risk evaluation, illustrating the real-world benefits of accessing personal data.

2 Background

Secure Privacy-Preserving Inference research has oscillated between two distinct paradigms: cryptography-focused and machine learning security-focused approaches. The former adapts cryptographic schemes to machine learning (ML) computations, often overlooking ML's unique security and privacy requirements [12, 45, 11]. The latter emphasizes empirical security enhancements while largely avoiding information security and cryptography principles, leading to repeated vulnerabilities discovered in ML models [37, 19, 34, 1, 2, 44, 41, 10]. This section aims to bridge insights from both domains, suggesting a route to a comprehensive approach of achieving practical, real-world Secure Privacy-Preserving Inference lies on the intersection of the two approaches.

2.1 Cryptography-focused Solutions

Cryptography-focused solutions for Secure Privacy-Preserving Inference vary in computational responsibility and data outsourcing. We categorize and examine the state of these solutions, highlighting potential security and efficiency issues.

2.1.1 Bringing Data to ML

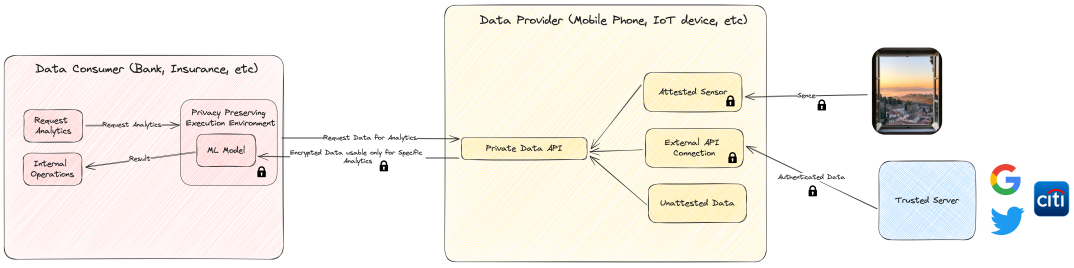

Figure 2: Bringing Data to the ML Model.

This method involves transferring data securely from the Data Provider to the Data Consumer, who then performs analytics. Techniques like Fully Homomorphic Encryption (FHE) [36, 7] and Trusted Execution Environments (TEE) [40, 43] aim to protect data during processing. However, challenges with FHE's practicality [38] and TEE's privacy guarantees [30, 46] suggest limitations, particularly in ensuring data deletion and preventing malicious data submission or model use. Moreover, the inability to provably delete data without continuous auditing [46] raises concerns about the Data Consumer

potentially retaining and later accessing the encrypted data [3]. There is also no safeguard against the Data Provider submitting incorrect or adversarial data as well as guarantee that the Data Consumer is not running a malicious models over the data. As all the Cryptography-focused solutions avoid this issue, we will cover this in the 2.2 section.

2.1.2 Continuous Interactions

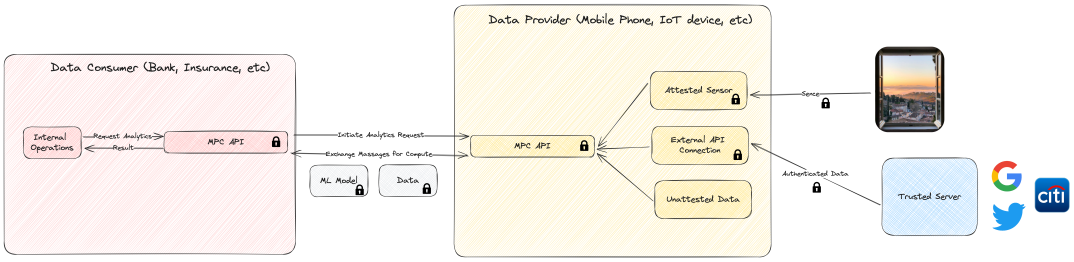

Figure 3: Ongoing Interaction between Data Provider and Data Consumer.

Featuring ongoing communication exchanges, this approach utilizes Secure Multi-Party Computation (SMPC) protocols [27] to balance computational loads. While SMPC reduces some inefficiencies [7], it introduces performance penalties and continuous communication challenges, highlighting the need for a balanced approach. Additionally, there is still a potential for Data Consumers to store interaction data for future decryption.

2.1.3 Bringing ML to Data

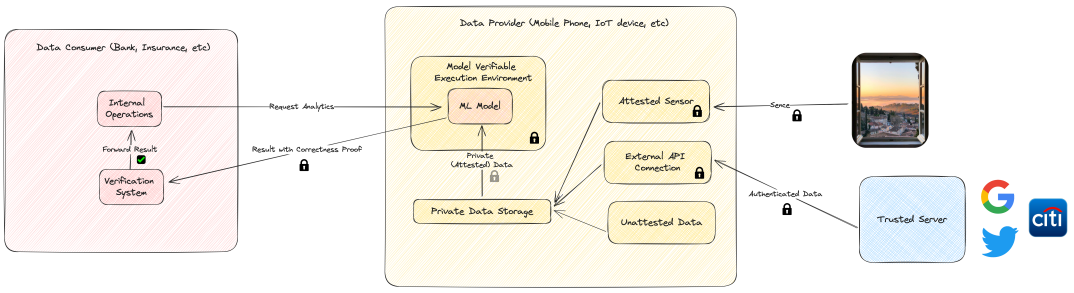

Figure 4: Bringing the ML Model to Data.

The "Bringing ML to Data" strategy involves the Data Consumer sending the ML model to the Data Provider, who then runs the model locally on their dataset and returns the results. This method inherently protects the Data Provider's sensitive information by eliminating the need for data transmission, thereby maintaining privacy by default. However, this approach requires the Data Consumer to trust the Data Provider's integrity in accurately executing the model and providing genuine results.

To address potential concerns of dishonesty and to assure the Data Consumer of the computation's integrity, verifiable computation schemes [48] are utilized. These schemes

[3] NSA stores encrypted data until it can be cracked - https://www.techdirt.com/2013/06/21/nsa-has-convinced-fisa-court-that-if-your-data-is-encrypted-you-might-be-terrorist-so-itll-hang-ont

allow the Data Provider to offer cryptographic proof that the model was executed correctly and that the results are authentic. This development, known as Verifiable Machine Learning [45], is gaining importance as ML models are increasingly applied in sensitive sectors like healthcare, finance, and legal decision-making [25, 42].

The primary method for verifying ML model outputs involves the use of zk-SNARKs (Zero-Knowledge Succinct Non-Interactive Argument of Knowledge) to construct computational circuits that mimic the architecture of the ML model [21]. These circuits provide a cryptographic means to attest to the accuracy of inferences, catering to the verification of the model's evaluation. Despite active research to scale verifiable computations to match the complexity and size of contemporary ML models [45], significant challenges remain. Existing methods struggle to efficiently handle large-scale models [39], which today exceed gigabytes in size, leading to substantial computational costs for Data Providers, those with constrained resources, such as IoT devices.

Despite these challenges, the approach of Bringing ML to Data, reinforced by verifiable computations, appears to us to be the most promising among existing methods. As the field of verifiable computations and non-interactive proofs attracts significant research and investment, highlighting the potential for future efficiency breakthroughs [45] [4]. Additionally, with edge devices becoming increasingly powerful [5], the capacity for conducting extensive computations locally is rapidly improving. This progress makes the "Bringing ML to Data" model not just feasible, but ideally suited for deployment in real-world scenarios.

2.2 Machine Learning Security-focused Solutions

While cryptography-focused solutions provide robust frameworks for Secure Privacy-Preserving Inference, they often treat Machine Learning (ML) models as black boxes, neglecting the considerable risks posed by malicious models and adversarial inputs. This oversight can lead to vulnerabilities that compromise the integrity of Secure Privacy-Preserving Inference systems. The following discussion highlights key areas of concern and potential mitigation strategies, emphasizing the importance of a more nuanced approach to ML model security.

2.2.1 Malicious Data Provider

Malicious Data Providers pose a significant threat by manipulating model outputs through the injection of specially crafted adversarial data. This technique can cause models to produce erroneous results, bypassing the protective measures of cryptography-based solutions like verifiable computations, which do not account for internal logic issues of ML models [17]. Addressing this challenge requires a robust ML security framework capable of detecting and neutralizing adversarial data threats. One potential countermeasure is the use of verified data, certified by third parties or sensor attestations, though this area remains underexplored in current research.

[4] VC firm opens a cryptography department - https://www.coindesk.com/tech/2023/08/10/vc-firm-a16z-wades-into-crypto-tech-research-with-zk-projects-jolt-and-lasso/

[5] Your Phone is more powerful than your PC - https://insights.samsung.com/2021/08/19/your-phone-is-now-more-powerful-than-your-pc-3/

Additionally, malicious Data Providers may seek to access or infer proprietary information embedded within the Data Consumer's model, aiming to reverse-engineer the model for competitive advantages [6] or to access private training data [33]. Solutions like Trusted Execution Environments (TEE) and Fully Homomorphic Encryption (FHE) have been proposed to shield the models' parameter, but vulnerabilities in machine learning still allow to perform parameter extraction [6]. This situation underscores the imperative for more sophisticated security measures that extend beyond TEE and FHE, advocating for ongoing research in this domain [5].

2.2.2 Malicious Data Consumer

On the flip side, Malicious Data Consumers aim to breach data privacy by exploiting models to extract Data Provider information. Direct methods include deploying models that simply output received data rather than performing intended computations. Mitigating such threats may require rigorous audits and transparency regarding model weights and architectures. Emerging research has revealed techniques for embedding covert triggers in model parameters or architectures, triggering unauthorized functions like data exfiltration upon detecting specific inputs [22, 9]. These revelations highlight the urgent need for enhanced security protocols to prevent such clandestine operations.

Moreover, excessive querying by Data Consumers can jeopardize data privacy by performing reconstruction attack [26]. Local Differential Privacy (DP)[28] has been touted as a solution for safeguarding individual data points within datasets , [16], but its applicability to inference processes is still unclear. Furthermore, the current conservatives of DP estimation techniques leads it to be too impractical to be used [14]. There is a critical need for research aimed at optimizing DP measures, possibly through dynamic adjustment of the Privacy Budget based on previous queries, to strike a better balance between privacy preservation and utility [8].

2.3 The Intersection

Addressing the security challenges posed by both malicious Data Providers and Consumers is crucial for the development of Secure Privacy-Preserving Inference systems. While cryptography-based solutions lay a strong foundation, the unique threats within the ML landscape demand specialized security measures. A comprehensive approach, incorporating robust ML security frameworks and ongoing research into advanced cryptographic and verification techniques, is essential for safeguarding against the sophisticated tactics employed by adversaries in the digital age.

3 Research Aims

The core objective of this research is to develop a practical scheme for Secure Privacy-Preserving Inference, leveraging the Cryptography-focused "Bringing ML to Data" strategy. This direction builds on our substantial expertise and the relative readiness of this approach for real-world applications. Our research will then broaden to address Machine

[6] Apple NeuroHash Model reverse-engineered - `https://www.schneier.com/blog/archives/2021/08/apples-neuralhash-algorithm-has-been-reverse-engineered.html`

Learning Security, safeguarding the system against ML-specific threats, and will explore alternative cryptographic solutions to enhance security and privacy further.

3.1 Primary Research: Cryptography Solution of Bringing ML to Data

- **Optimization Techniques for ML Verification:**
  1. Research optimization strategies for inference-focused ML models, expanding on our previous work in WNN verification [4].
  2. Develop efficient methodologies for converting standard ML models to formats optimized for inference.
  3. Examine model optimization techniques, including distillation and quantization, for compatibility with verifiable computation frameworks.
  4. Investigate inference optimization methods, such as employing hints, drawing from established verifiable computation practices.
- **Optimizing Verifiable Computations for ML Models:**
  1. Assess alternative verifiable computation systems suitable for performance requirements, especially on constrained devices like IoT [7].
  2. Improve the efficiency of converting models for use with selected verifiable computation frameworks.
  3. Explore the potential of integrating Trusted Execution Environments (TEEs) in hybrid security models.
  4. Consider proof generation delegation or assisted proving to alleviate computational burdens on Data Providers [13].

3.2 Supplementary Research Directions

- **Minimizing Data Leakage from Requests:**
  1. Evaluate the efficacy of Differential Privacy (DP) and Privacy Budgets in mitigating information leakage.
  2. Research dynamic Privacy Budget calculation methods for effective privacy preservation.
  3. Identify and counteract vulnerabilities in public model architectures that could lead to data leakage.
- **Enhancing Model Security Against Adversarial Data:**
  1. Formulate defensive strategies to protect models from manipulation by malicious Data Providers.

[7] Exploring GKR for ML Inference Verification optimization - `https://www.theblockbeats.info/en/flash/231429`

2. Develop provable security measures to defend against data-driven attacks.
3. Investigate techniques to reduce the success rate and impact of attacks.

- **Bridging Cryptography with ML Security:**
  1. Examine the security implications of integrating ML in cryptographic protocols, targeting the Universal Composability framework.
  2. Construct a security framework for the "Bringing ML to Data" strategy that includes ML-specific security considerations.
- **Exploring Authenticated Data and Sensor Authentication:**
  1. Analyze security challenges and attack vectors related to using authenticated data sources.
  2. Investigate the reliability of attested sensor data as a source for verifiable computations.
  3. Examine data manipulation risks within authenticated data scenarios.
- **Optimizing Alternative Cryptography Solutions:**
  1. Explore various schemes of Fully Homomorphic Encryption (FHE) for enhancing data privacy in ML computations.
  2. Identify optimization strategies for ML models that are compatible with FHE and other privacy-preserving computation environments.

Our research aims to tackle the complex challenges of Secure Privacy-Preserving Inference, striving for advancements that optimize efficiency, security, and privacy in adversarial contexts.

4 Significance and Impact

Our research is poised to make substantial contributions to the field of Machine Learning (ML) by addressing crucial privacy and security concerns:

- **Unlocking Personal Data:** By developing a privacy-centric ML framework, particularly through the "Bringing ML to Data" strategy, we aim to significantly reduce the risks currently associated with personal data analytics. This breakthrough has the potential to transform industries, enabling the creation of more personalized solutions and democratizing access to sophisticated data analysis tools. This would benefit not only large corporations but also empower smaller entities and individuals, making advanced analytics accessible to a broader audience.
- **Advancing Privacy-Preserving ML:** Adopting a security-first approach in ML research can drive widespread improvements within the domain, leading to a more secure privacy-preserving ML ecosystem. This effort is expected to bolster confidence in ML technologies, fostering their adoption across diverse sectors. By strengthening data protection and enhancing user privacy, we anticipate a significant boost in trust, further accelerating the integration of ML technologies into everyday applications.

- **Preparing ML for Critical Applications:** With ML models increasingly being applied in critical sectors such as healthcare, finance, and legislative decision-making, enhancing the security of ML systems is paramount [25, 42]. Our research is aimed at making ML systems robustly secure for mission-critical applications, thereby unlocking new horizons for digital automation, where automation and AI can play pivotal roles in society[8].

Overall, the research seeks to not only improve the security and privacy aspects of ML inference but also to inspire a shift towards more ethical and secure technological practices in the field of ML. Through our contributions, we aim to catalyze a transition to a digital era where technology serves humanity with utmost respect for privacy and security.

5 Methodology and Timeline

5.1 Overall Timeline

Our research is meticulously planned over a four-year period, each year dedicated to advancing specific area of Secure Privacy-Preserving Inference:

- **Year 1:** Focus on the Cryptography Solution of "Bringing ML to Data", enhancing verifiable computations and optimizing ML models for verification, laying the foundational work for subsequent years.
- **Year 2:** Expand on developing ML architectures optimized for verification and explore efficient model conversion techniques, building upon the optimizations introduced in Year 1.
- **Year 3:** Prioritize enhancing ML model security against adversarial threats and begin integrating cryptographic solutions with ML security measures, synthesizing the work from the first two years.
- **Year 4:** Finalize the integrated security framework for Secure Privacy-Preserving Inference, exploring broader applications and solidifying the project's contributions to ML and cryptography.

[8] Our proposal of AI-Powered Personal Delegates - `https://github.com/ConfidentiDemokratia`

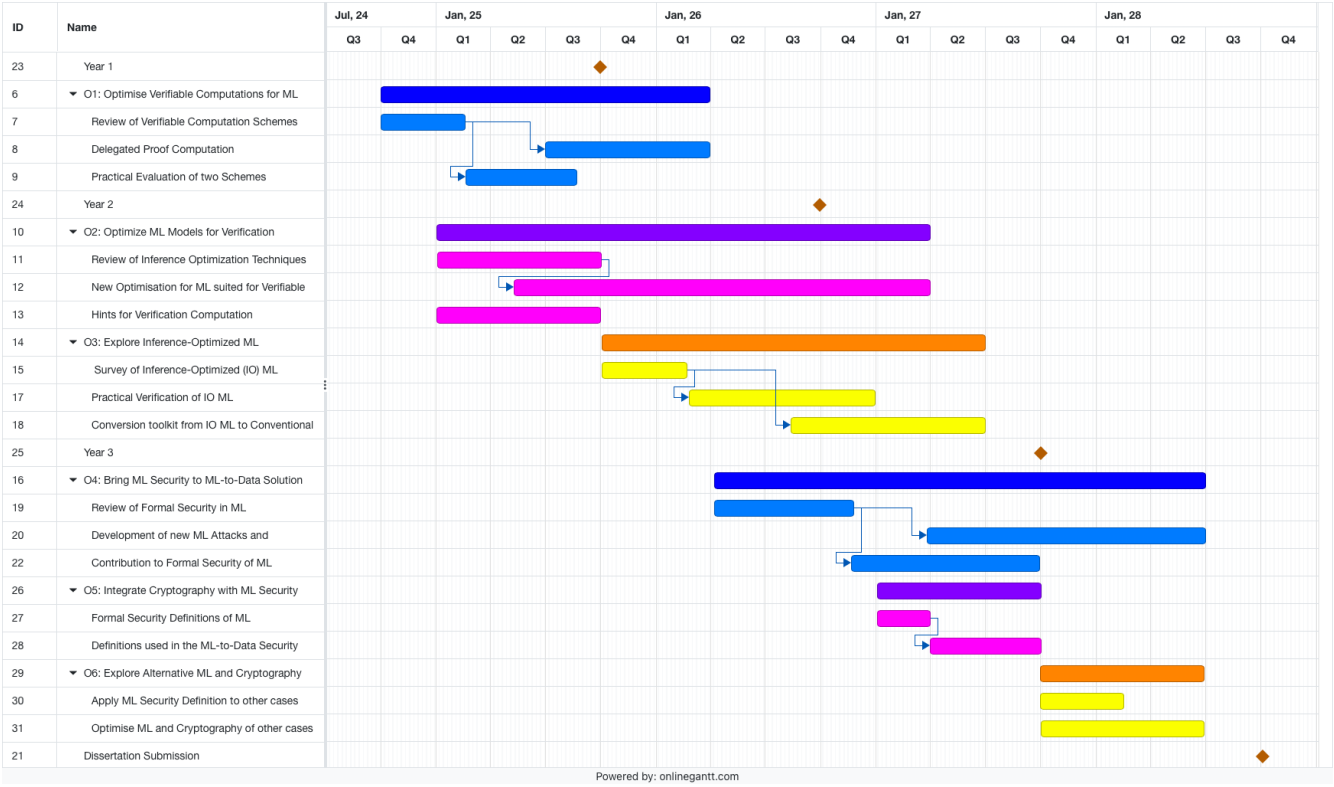

Figure 5: Gantt chart illustrating the research effort timeline.

5.2 Detailed Project Timeline

This timeline outlines a series of deliverables designed to systematically achieve our research objectives:

O1: Optimize Verifiable Computations for ML:

(a) Review verifiable computation systems for ML inference, focusing on compatibility with various computational capacities.

(b) Investigate delegated proof computation to alleviate computational demands on less powerful devices.

(c) Assess the practicality of selected verifiable computation schemes with ML models, emphasizing efficiency.

Timeline: Year 1, Q1 to Year 2, Q2

O2: Optimize ML Models for Verification:

(a) Document existing and novel inference optimization techniques, considering their compatibility with verifiable computation.

(b) Develop a toolkit for ML model optimization, facilitating efficient model conversion.

(c) Examine optimization hints to expedite verification, assessing their impact on security and scalability.

Timeline: Year 1, Q2 to Year 3, Q2

O3: Explore Inference-Optimized ML Architectures:

(a) Survey existing architectures for suitability with verifiable computations, identifying promising candidates.

(b) Validate chosen architectures through experimentation, aiming to develop a conversion toolkit.

Timeline: Year 2, Q1 to Year 3, Q3

O4: Enhance ML Security within the ML-to-Data Approach:

(a) Conduct a study on formal security in ML, focusing on adversarial defenses.

(b) Advance ML system security through innovative attack and defense mechanisms.

(c) Formulate new security definitions and reasoning methodologies for ML.

Timeline: Year 2, Q2 to Year 4, Q3

O5: Integrate Cryptography with ML Security:

(a) Develop formal security definitions for ML within cryptographic protocols, enhancing the "Bringing ML to Data" security framework.

Timeline: Year 3, Q2 to Year 3, Q4

O6: Explore Alternative ML and Cryptography Solutions:

(a) Apply established frameworks to new ML and cryptography intersections, assessing effectiveness.

Timeline: Year 4, Q1 to Year 4, Q3

O7: Dissemination of Findings:

(a) Engage in knowledge transfer through academic and industry publications and presentations.

(b) Establish forums, such as blog-posts, workshops, etc, for dialogue on ML Security and Privacy, enhancing public awareness.

Timeline: Year 1, Q3 to Year 4, Q4

6 Research Competence and Background

My academic and professional journey intersects Machine Learning (ML), Privacy-Preserving Technologies, and Security, underscoring my motivation and expertise for the proposed research. My fascination with ML began in high school, leading to a summer semester at Stanford in 2019, where I achieved the top grade. Concurrently, I pursued an interest in Cryptography and Privacy Enhancing Technologies (PET), completing Dan Boneh's course and starting as a Solution Architect at R3 working with Trusted Execution Environments (TEE) to devise innovative ML solutions in adversarial environments. This position enabled me to explore research challenges, such as data revocation problem, culminating in a TEE-powered Verifiable Delay Function (VDF) proposal detailed in a pre-print[9].

Later, as a Cryptography Research Engineer at Aragon, I deepened my expertise in PETs and Cryptography, particularly Zero-Knowledge Proofs (ZKP), a variant of Verifiable Computations, resulting in two co-authored publications[20, 3]. This role expanded my technical skills and aligned with my research interests at the ML and Verifiable Computations intersection, by contributing to the project exploring AI Agents for governance[10]. An Ethereum Foundation grant to investigate alternative neural network architectures for verifiable inference further solidified my commitment to this research domain[11]. My achievements in hackathons, deploying solutions for Privacy-Preserving Inference, and my ongoing Master's in Cybersecurity at UCL—where my thesis focuses on Verified On-Device Inference—demonstrate a strong theoretical and practical foundation paired with pation in the topic.

Moreover, my dedication to sharing knowledge through teaching Mathematics and delivering lectures on Advanced Cryptography, AI, and Blockchain[12] highlights my passion for education and mentorship. My commitment to research and aspiration to contribute to the academic and broader community showcase my readiness to explore and advance the fields of AI, security, and privacy.

7 Ethical Considerations and Data Privacy

The proposed research operates at the intersection of machine learning and privacy, inherently dealing with sensitive private data. Recognizing the potential for privacy invasion and misuse of data, the research will adhere to stringent ethical guidelines, ensuring that all data is anonymous and securely handled. The project will also undergo regular ethical reviews, aligning with the latest data protection regulations and ethical standards in AI research.

[9] VDF with Confidential Computing - `https://github.com/ElusAegis/Confidencial_Computing_VDF/blob/main/Confidencial_Computing_VDF_Revised.pdf`

[10] An Article of AI in Governance, with my contribution - `https://blog.aragon.org/ai-daos-the-future-of-daos-powered-by-artificial-intelligence/`

[11] WNN Inference Verification - `https://github.com/zkp-gravity/optimisation-research/blob/main/writeup.pdf`

[12] Our Public Activity List - `https://tinyurl.com/vpbznvcm`

8 Collaborations and Support Needs

The interdisciplinary nature of this research, weaving together the threads of machine learning, privacy, and cryptography, necessitates a collaborative approach drawing on a diverse range of expertise. Engaging with faculty members specializing in Information Security and Machine Learning will will enrich the research, offering nuanced perspectives and expert guidance. The ambitious scope of our work, from training intricate ML models to delving into cryptographic methodologies, places significant demands on computational resources. Therefore, access to advanced computing infrastructure, including GPUs and high-performance computing environments, is critical.

9 Conclusion

This research proposal sets forth a comprehensive plan to advance the field of Secure Privacy-Preserving Inference, operating at the intersection of machine learning, privacy, and cryptography. At its core, the proposal aims to develop robust, privacy-preserving ML technologies that can be deployed in sensitive and adversarial environments without compromising data security or model integrity.

We have outlined a series of research aims and methodologies designed to tackle the multifaceted challenges associated with privacy-preserving ML. Through innovative approaches such as "Bringing ML to Data" and the exploration of verifiable computations and inference-optimized ML architectures, we aim to enhance the security and efficiency of ML applications.

The interdisciplinary nature of this research underscores the necessity of collaboration across domains. Engaging with experts in Information Security, Machine Learning, and Cryptography will be pivotal in enriching our research perspective and methodology. Additionally, the success of this project is contingent upon access to substantial computing resources, highlighting the need for support in securing the necessary infrastructure.

Anticipated contributions of this research include advancing knowledge in the domain of privacy-preserving ML, providing scalable and secure solutions for ML applications, and fostering a greater understanding of the integration between ML and cryptography. Through these efforts, we aim to make a significant impact not only in the academic community but also in broader societal applications, promoting a future where technology serves humanity with respect for privacy and security.

In conclusion, we embark on this research journey with a deep sense of responsibility and an unwavering commitment to innovation, collaboration, and ethical excellence. Our ultimate goal is to contribute meaningful advancements that address critical challenges at the forefront of Secure Privacy Preserving ML, paving the way for secure and trustworthy digital technologies.

References

[1] Abay, N.C., Zhou, Y., Kantarcioglu, M., Thuraisingham, B., Sweeney, L.: Privacy preserving synthetic data release using deep learning. In: Berlingerio, M., Bonchi, F.,

Gärtner, T., Hurley, N., Ifrim, G. (eds.) Machine Learning and Knowledge Discovery in Databases. pp. 510–526. Springer International Publishing, Cham (2019)

[2] Abdalla, M., Abdalla, M., Hirst, G., Rudzicz, F.: Exploring the privacy-preserving properties of word embeddings: Algorithmic validation study. J Med Internet Res **22**(7), e18055 (Jul 2020)

[3] Artem, G., Iovino, V., Roșie, R.: Multi-input non-interactive functional encryption: Constructions and applications. In: El Hajji, S., Mesnager, S., Souidi, E.M. (eds.) Codes, Cryptology and Information Security. pp. 158–177. Springer Nature Switzerland, Cham (2023)

[4] Artem Grigor, G.W.: Optimization and scalability of weightless neural networks: A comprehensive study. `https://github.com/zkp-gravity/optimisation-research/blob/main/writeup.pdf` (September 2023), available online at `https://github.com/zkp-gravity/optimisation-research/blob/main/writeup.pdf`

[5] Atrey, A., Sinha, R., Mitra, S., Shenoy, P.J.: SODA: protecting proprietary information in on-device machine learning models. CoRR **abs/2312.15036** (2023). https://doi.org/10.48550/ARXIV.2312.15036, `https://doi.org/10.48550/arXiv.2312.15036`

[6] Atrey, A., Sinha, R., Sarkhel, S., Mitra, S., Arbour, D., Maharaj, A., Shenoy, P.: Towards preserving server-side privacy of on-device models. In: Companion Proceedings of the Web Conference 2022. p. 282–285. WWW '22, Association for Computing Machinery, New York, NY, USA (2022). https://doi.org/10.1145/3487553.3524257, `https://doi.org/10.1145/3487553.3524257`

[7] Azraoui, M., Bahram, M., Bozdemir, B., Canard, S., Ciceri, E., Ermis, O., Masalha, R., Mosconi, M., Önen, M., Paindavoine, M., Rozenberg, B., Vialla, B., Vicini, S.: SoK: Cryptography for Neural Networks, pp. 63–81. Springer International Publishing, Cham (2020), `https://doi.org/10.1007/978-3-030-42504-3_5`

[8] Bai, Y., Yang, G., Xiang, Y., Wang, X.: Generalized and multiple-queries-oriented privacy budget strategies in differential privacy via convergent series. Security and Communication Networks **2021**, 1–17 (12 2021). https://doi.org/10.1155/2021/5564176

[9] Bober-Irizar, M., Shumailov, I., Zhao, Y., Mullins, R., Papernot, N.: Architectural backdoors in neural networks (2022)

[10] Boenisch, F., Dziedzic, A., Schuster, R., Shamsabadi, A.S., Shumailov, I., Papernot, N.: Reconstructing individual data points in federated learning hardened with differential privacy and secure aggregation (2023)

[11] Boura, C., Gama, N., Georgieva, M., Jetchev, D.: Simulating homomorphic evaluation of deep learning predictions. Cryptology ePrint Archive, Paper 2019/591 (2019), `https://eprint.iacr.org/2019/591`, `https://eprint.iacr.org/2019/591`

[12] Cabrero-Holgueras, J., Pastrana, S.: Sok: Privacy-preserving computation techniques for deep learning. Proceedings on Privacy Enhancing Technologies **2021**, 139 – 162 (2021), https://api.semanticscholar.org/CorpusID:236213847

[13] Canard, S., Pointcheval, D., Sanders, O.: Efficient delegation of zero-knowledge proofs of knowledge in a pairing-friendly setting. In: Krawczyk, H. (ed.) Public-Key Cryptography – PKC 2014. pp. 167–184. Springer Berlin Heidelberg, Berlin, Heidelberg (2014)

[14] Cummings, R., Desfontaines, D., Evans, D., Geambasu, R., Huang, Y., Jagielski, M., Kairouz, P., Kamath, G., Oh, S., Ohrimenko, O., Papernot, N., Rogers, R., Shen, M., Song, S., Su, W., Terzis, A., Thakurta, A., Vassilvitskii, S., Wang, Y.X., Xiong, L., Yekhanin, S., Yu, D., Zhang, H., Zhang, W.: Advancing Differential Privacy: Where We Are Now and Future Directions for Real-World Deployment. Harvard Data Science Review **6**(1) (jan 16 2024), https://hdsr.mitpress.mit.edu/pub/sl9we8gh

[15] Djeundje, V.B., Crook, J., Calabrese, R., Hamid, M.: Enhancing credit scoring with alternative data. Expert Systems with Applications **163**, 113766 (2021). https://doi.org/https://doi.org/10.1016/j.eswa.2020.113766, https://www.sciencedirect.com/science/article/pii/S095741742030590X

[16] Ebadi, H., Sands, D., Schneider, G.: Differential privacy: Now it's getting personal. In: Proceedings of the 42nd Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages. p. 69–81. POPL '15, Association for Computing Machinery, New York, NY, USA (2015). https://doi.org/10.1145/2676726.2677005, https://doi.org/10.1145/2676726.2677005

[17] Fenaux, L., Kerschbaum, F.: Sok: Analyzing adversarial examples: A framework to study adversary knowledge (2024)

[18] García-Gasco Romero, M.: Personal Data: The New Black Gold, pp. 171–182. Springer International Publishing, Cham (2021), https://doi.org/10.1007/978-3-030-67973-6_12

[19] Glukhov, D., Shumailov, I., Gal, Y., Papernot, N., Papyan, V.: Llm censorship: A machine learning challenge or a computer security problem? (2023)

[20] Grigor, A., Iovino, V., Visconti, G.: The referendum problem in anonymous voting for decentralized autonomous organizations. Cryptology ePrint Archive, Paper 2023/741 (2023), https://eprint.iacr.org/2023/741, https://eprint.iacr.org/2023/741

[21] Groth, J.: On the size of pairing-based non-interactive arguments. In: Fischlin, M., Coron, J.S. (eds.) Advances in Cryptology – EUROCRYPT 2016. pp. 305–326. Springer Berlin Heidelberg, Berlin, Heidelberg (2016)

[22] Gu, T., Dolan-Gavitt, B., Garg, S.: Badnets: Identifying vulnerabilities in the machine learning model supply chain (2019)

[23] Ji, Z., Lipton, Z.C., Elkan, C.: Differential privacy and machine learning: a survey and review (2014)

[24] Lane, N.D., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T., Campbell, A.T.: A survey of mobile phone sensing. IEEE Communications Magazine **48**(9), 140–150 (2010). https://doi.org/10.1109/MCOM.2010.5560598

[25] Li, B., Qi, P., Liu, B., Di, S., Liu, J., Pei, J., Yi, J., Zhou, B.: Trustworthy ai: From principles to practices (2022)

[26] Liu, S., Wang, Z., Lei, Q.: Data reconstruction attacks and defenses: A systematic evaluation (2024)

[27] Long, Y., Gangwani, T., Mughees, H., Gunter, C.: Distributed and secure ml with self-tallying multi-party aggregation (2018)

[28] Mahawaga Arachchige, P.C., Bertok, P., Khalil, I., Liu, D., Camtepe, S., Atiquzzaman, M.: Local differential privacy for deep learning. IEEE Internet of Things Journal **7**(7), 5827–5842 (2020). https://doi.org/10.1109/JIOT.2019.2952146

[29] Meier, S., Sprenger, C.: Impatience and credit behavior: Evidence from a field experiment. Federal Reserve Bank of Boston, Working Papers (01 2007). https://doi.org/10.2139/ssrn.982398

[30] Muñoz, A., Ríos, R., Román, R., López, J.: A survey on the (in)security of trusted execution environments. Computers and Security **129**, 103180 (2023). https://doi.org/https://doi.org/10.1016/j.cose.2023.103180, `https://www.sciencedirect.com/science/article/pii/S0167404823000901`

[31] Niu, C., Wu, F., Tang, S., Ma, S., Chen, G.: Toward verifiable and privacy preserving machine learning prediction. IEEE Transactions on Dependable and Secure Computing **19**(3), 1703–1721 (2022). https://doi.org/10.1109/TDSC.2020.3035591

[32] Nuccio, M., Guerzoni, M.: Big data: Hell or heaven? digital platforms and market power in the data-driven economy. Competition & Change **23**(3), 312–328 (2019). https://doi.org/10.1177/1024529418816525, `https://doi.org/10.1177/1024529418816525`

[33] Oh, S.J., Augustin, M., Schiele, B., Fritz, M.: Towards reverse-engineering black-box neural networks (2018)

[34] Papernot, N., McDaniel, P., Sinha, A., Wellman, M.P.: Sok: Security and privacy in machine learning. In: 2018 IEEE European Symposium on Security and Privacy (EuroS&P). pp. 399–414 (2018). https://doi.org/10.1109/EuroSP.2018.00035

[35] Papernot, N., Song, S., Mironov, I., Raghunathan, A., Talwar, K., Úlfar Erlingsson: Scalable private learning with pate (2018)

[36] Podschwadt, R., Takabi, D., Hu, P., Rafiei, M.H., Cai, Z.: A survey of deep learning architectures for privacy-preserving machine learning with fully homomorphic encryption. IEEE Access **10**, 117477–117500 (2022). https://doi.org/10.1109/ACCESS.2022.3219049

[37] Psychoula, I., Chen, L., Chen, F.: Privacy modelling and management for assisted living within smart homes. In: 2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom). pp. 1–6 (2017). https://doi.org/10.1109/HealthCom.2017.8210782

[38] Stoian, A., Frery, J., Bredehoft, R., Montero, L., Kherfallah, C., Chevallier-Mames, B.: Deep neural networks for encrypted inference with tfhe. Cryptology ePrint Archive, Paper 2023/257 (2023), https://eprint.iacr.org/2023/257, https://eprint.iacr.org/2023/257

[39] Team, M.L.: The cost of inteligence: Proving machine learning inference with zero knoweldge. Online (2023), https://github.com/Modulus-Labs/Papers/blob/master/Cost_Of_Intelligence.pdf, accessed: 2024-04-02

[40] Truong, J.B., Gallagher, W., Guo, T., Walls, R.J.: Memory-efficient deep learning inference in trusted execution environments (2021)

[41] Vakili, T., Dalianis, H.: Using membership inference attacks to evaluate privacy-preserving language modeling fails for pseudonymizing data. In: Alumäe, T., Fishel, M. (eds.) Proceedings of the 24th Nordic Conference on Computational Linguistics (NoDaLiDa). pp. 318–323. University of Tartu Library, Tórshavn, Faroe Islands (May 2023), https://aclanthology.org/2023.nodalida-1.33

[42] Vice: A judge just used chatgpt to make a court decision. https://www.vice.com/en/article/k7bdmv/judge-used-chatgpt-to-make-court-decision (2023), accessed: 2023-09-29

[43] Wang, Q., Zhou, L., Bai, J., Koh, Y.S., Cui, S., Russello, G.: Ht2ml: An efficient hybrid framework for privacy-preserving machine learning using he and tee. Computers and Security **135**, 103509 (2023). https://doi.org/https://doi.org/10.1016/j.cose.2023.103509, https://www.sciencedirect.com/science/article/pii/S0167404823004194

[44] Wang, Z., Liu, K., Hu, J., Ren, J., Guo, H., Yuan, W.: Attrleaks on the edge: Exploiting information leakage from privacy-preserving co-inference. Chinese Journal of Electronics **32**(1), 1–12 (2023). https://doi.org/10.23919/cje.2022.00.031

[45] Xing, Z., Zhang, Z., Liu, J., Zhang, Z., Li, M., Zhu, L., Russello, G.: Zero-knowledge proof meets machine learning in verifiability: A survey (2023)

[46] Yang, C., Liu, Y., Zhao, F., Zhang, S.: Provable data deletion from efficient data integrity auditing and insertion in cloud storage. Computer Standards & Interfaces **82**, 103629 (2022). https://doi.org/https://doi.org/10.1016/j.csi.2022.103629, https://www.sciencedirect.com/science/article/pii/S0920548922000101

[47] Yoon, J., Jordon, J., van der Schaar, M.: PATE-GAN: Generating synthetic data with differential privacy guarantees. In: International Conference on Learning Representations (2019), https://openreview.net/forum?id=S1zk9iRqF7

[48] Yu, X., Yan, Z., Vasilakos, A.V.: A survey of verifiable computation. Mobile Networks and Applications **22**(3), 438–453 (06 2017), https://www.proquest.com/scholarly-journals/survey-verifiable-computation/docview/1908019401/se-2, copyright - Mobile Networks and Applications is a copyright of Springer, 2017; Last updated - 2023-11-30