POLITECNICO DI TORINO
Repository ISTITUZIONALE

Auditable Execution of Machine Learning Algorithms using Blockchain and Trusted Execution Environments

Original
Auditable Execution of Machine Learning Algorithms using Blockchain and Trusted Execution Environments / Romani, N., Gatteschi, V., De Caria, R., Zunino, L., Villa, V., Domaneschi, M., Marengo, G.. - (In corso di stampa). (8th Distributed Ledger Technology Workshop (DLT 2026) Pula (Italy) 4–6 June 2026).

Availability:
This version is available at: 11583/3015538 since: 2026-09-15T18:08:10Z

Publisher:
CEUR-WS.org

Published
DOI:

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

20 November 2026

Auditable Execution of Machine Learning Algorithms using Blockchain and Trusted Execution Environments

Noemi Romani[1,*], Valentina Gatteschi[1], Riccardo De Caria[2], Leonardo Zunino[1], Valentina Villa[1], Marco Domaneschi[1] and Giorgio Marengo[1]

[1]*Politecnico di Torino, Corso Duca degli Abruzzi, 24, Turin, 10129, Italy*
[2]*Università degli Studi di Torino, Via Verdi, 8, Turin, 10124, Italy*

Abstract

The increasing adoption of machine learning algorithms in data-driven decision-making systems raises critical challenges related to transparency, accountability, and trust. While blockchain technologies enable immutable records and non-repudiable audit trails, their inherent transparency may conflict with the confidentiality requirements of many real-world applications, where algorithms operate on sensitive datasets or proprietary models. In such contexts, ensuring verifiable and auditable computations without disclosing confidential information becomes a key challenge. This paper proposes a framework for the auditable execution of machine learning algorithms by integrating blockchain technology with Trusted Execution Environments (TEE). The proposed approach enables the certification of algorithm execution while preserving the confidentiality of both data and models. Input datasets are stored in decentralized storage systems, while the algorithm is executed inside a hardware-protected enclave capable of producing remote attestation proofs. These attestations certify that a specific algorithm has been executed on declared inputs within a genuine confidential-computing environment, without exposing the underlying data or internal computation. Cryptographic references to the input data, algorithm execution, and generated outputs are recorded on-chain through smart contracts, creating an immutable and publicly verifiable audit trail. Rather than relying on computationally expensive zero-knowledge proofs to verify the entire machine learning computation, the framework selectively applies ZKPs to certify properties of the decision-making process derived from the confidential model output. In this way, the framework establishes a verifiable computation pipeline that links datasets, algorithm execution, and outputs while maintaining data confidentiality. The proposed system also supports periodic data ingestion and automated algorithm execution, enabling auditors and stakeholders to independently verify the integrity of the analysis when required. The main contributions of this work are threefold. First, we introduce a blockchain-enabled architecture for auditable off-chain machine learning execution that reconciles the need for transparency with strict privacy requirements. Second, we introduce a decision-level verification mechanism in which Zero-Knowledge Proofs selectively certify properties of the decision-making process, avoiding the computational cost of proving the entire machine learning inference. Third, we illustrate the applicability of the framework in an infrastructure monitoring scenario where machine learning algorithms process sensor data to support decisions. The results highlight how blockchain and confidential computing technologies can jointly enable privacy-preserving and auditable AI-driven systems.

Keywords

Blockchain, Trusted Execution Environment, Machine Learning, Auditability

1. Introduction

Machine learning (ML) algorithms are increasingly used to analyze large volumes of data and support automated or semi-automated decision-making. In ML-based systems, where computations are typically executed off-chain on proprietary infrastructures and rely on sensitive datasets and confidential models, it becomes particularly critical to ensure the auditability of ML algorithms. On the one hand, it is challenging to guarantee the transparency and privacy of both the data and the model; on the other hand, it is hard to certify the correctness of the off-chain computations. Enabling ML pipelines to

DLT2026: 8th Distributed Ledger Technology Workshop, June, 4-6 2026 - Pula, Italy
*Corresponding author.
noemi.romani@polito.it (N. Romani); valentina.gatteschi@polito.it (V. Gatteschi); riccardo.decaria@unito.it (R. D. Caria); leonardo.zunino@polito.it (L. Zunino); valentina.villa@polito.it (V. Villa); marco.domaneschi@polito.it (M. Domaneschi); giorgio.marengo@studenti.polito.it (G. Marengo)
0009-0004-3725-8002 (N. Romani); 0000-0001-6075-6430 (V. Gatteschi); 0000-0002-6460-7616 (R. D. Caria); 0009-0007-5876-1151 (L. Zunino); 0000-0003-3903-8922 (V. Villa); 0000-0002-6077-8338 (M. Domaneschi)
© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

operate in systems that are both auditable and confidentiality-preserving could increase the adoption of such algorithms in particularly sensitive domains such as healthcare and cyber-physical systems [1]. Furthermore, certifying the execution of the model and the actions taken according to the output registered on-chain could minimize the additional risks associated to ML pipelines, including data poisoning and adversarial manipulation, and increase the trust in the model behavior and in the decision-making process [2].

To address these challenges, different approaches for verifiable and privacy-preserving machine learning have been proposed: blockchain and Trusted Execution Environment (TEE), among others. Distributed Ledger Technology (DLT) represents a major shift in data management, enabling information to be shared and maintained across decentralized networks rather than stored in centralized repositories. By distributing data across multiple nodes, DLT ensures transparency, immutability, and consensus-driven validation without relying on trusted intermediaries. Blockchain, as a prominent implementation of DLT, leverages cryptographic mechanisms to guarantee that once data are recorded, they remain tamper-resistant and chronologically ordered, allowing all participants to independently verify the integrity of the system. TEE enables the execution of programs within hardware-protected enclaves, ensuring confidentiality and integrity of the computation through remote attestation mechanisms [3].

Besides blockchain and TEE, recent advances have introduced additional technologies aimed at improving privacy-preserving machine learning, including Federated Learning (FL) and Zero-Knowledge Proofs (ZKPs). The main contribution of FL is enabling collaborative model training without directly sharing raw data. Data are kept on local devices, and only model updates are transmitted to a centralized server, thus preserving data privacy and ensuring data security. As for ZKPs, this technology allows a prover to demonstrate the correctness of a computation without revealing the underlying data or internal states, providing strong cryptographic guarantees [4, 5], thus guaranteeing both data privacy and data integrity.

In this work, building on the complementary properties of the different technologies discussed above, we propose a conceptual architecture that integrates blockchain, TEE, and ZKPs, rather than considering them as mutually exclusive alternatives. In the proposed framework, TEE provides confidential and integrity-preserving execution of the machine learning algorithm, while blockchain creates an immutable audit trail by recording cryptographic commitments associated with the relevant components of the ML pipeline. Unlike existing approaches that primarily focus on verifying the correctness of the entire machine learning computation, our framework adopts a decision-level verification strategy. Rather than proving the complete inference process, ZKPs are selectively applied to certify properties of the decision derived from the confidential model output, such as consistency with the recorded commitment or compliance with predefined decision rules.

The remainder of this paper is organized as follows: Section 2 presents relevant research in the context of this paper. Section 3 analyzes the design space of possible solutions, highlighting the trade-offs between TEE and ZKPs. Section 4 presents the proposed system architecture, describing its main components and data flow. Finally, Section 5 concludes the paper and outlines future research directions.

2. Related Works

2.1. Tecnological Primer

In the following, the contributes given by the different technological approaches to enable ML pipelines auditability are described. We consider solutions identified in the literature trying to address:

- verifiability of off-chain computations;
- data and model confidentiality.

Blockchain provides immutable and publicly verifiable records, making it particularly suitable for applications requiring trust, traceability, and accountability [6, 7]. However, in ML-based systems operating on sensitive datasets and proprietary models, blockchain alone does not guarantee the correctness of off-chain computations, creating a fundamental tension between transparency and confidentiality.

This limitation is widely recognized in the literature, where blockchain is considered an audit layer rather than a complete solution for secure and privacy-preserving computation [1, 4]. Furthermore, ML pipelines introduce additional risks, including data poisoning, adversarial manipulation, and lack of trust in training data and model behavior [2], especially in sensitive domains such as healthcare and cyber-physical systems [1].

Federated Learning (FL) enables collaborative model training without sharing raw data, but still introduces communication overhead and remains vulnerable to poisoning and inference attacks [8, 2].

TEEs are hardware-based technologies that enable the execution of programs within isolated and protected regions of a processor, commonly referred to as enclaves [3, 9]. These environments ensure that both code and data remain confidential and cannot be accessed or tampered with, even by privileged system components such as the operating system or hypervisor. The integrity of execution in TEE-based systems is guaranteed through remote attestation. During this process, the hardware generates a cryptographic measurement of the executed program and produces a signed attestation report, which can be verified by external parties to provide evidence that a specific binary has been executed within a genuine confidential-computing environment [9]. TEE-based approaches are particularly well suited for machine learning workloads, as they allow the execution of complex algorithms with near-native performance while preserving data and model confidentiality [10]. However, their trust model relies on the correctness of the hardware and its manufacturer, and several works have highlighted potential vulnerabilities, including side-channel attacks [3, 11]. Moreover, while attestation certifies that a program has been executed, it does not provide a cryptographic guarantee that the output is correctly derived from the input according to the intended computation semantics [12].

ZKPs are cryptographic protocols that allow one party (the prover) to demonstrate the correctness of a statement to another party (the verifier) without revealing any additional information beyond the validity of the statement itself. ZKP-based approaches provide strong security guarantees, as correctness is enforced through mathematical soundness rather than trust in external entities [13]. In the context of verifiable computation, in particular, this mechanism establishes a strong cryptographic binding between input data, computation, and output. Given a tuple (D, M, O) representing data, model, and output, a valid proof ensures that $O = M(D)$ holds according to the formalized computation, preventing any tampering or forgery without invalidating the proof. As a result, ZKP-based approaches provide end-to-end verifiability guarantees, removing the need to trust the computing entity. While both approaches (TEE and ZKP) aim to enhance trust in off-chain computation, they exhibit significantly different trade-offs in terms of trust assumptions, computational overhead, and applicability to real-world ML workloads [14]. TEE-based solutions rely on hardware trust assumptions and provide efficient execution with low overhead, but do not ensure a cryptographic binding between input and output. In contrast, ZKP-based approaches provide strong mathematical guarantees by binding input, computation, and output, at the cost of significantly higher computational overhead [5] which limits their applicability in practical scenarios involving large datasets and floating-point operations. This trade-off highlights a fundamental tension between efficiency and verifiability. TEE provides practical execution guarantees but requires trust in hardware, while ZKP eliminates trust assumptions but introduces scalability limitations. As a result, treating these approaches as mutually exclusive alternatives leads to solutions that are either impractical or rely on strong assumptions.

2.2. Analysis of the State of the Art

Existing works can be broadly grouped according to the role assigned to TEE and ZKP within the computation pipeline. For instance, [15] proposes a blockchain consensus algorithm specifically designed for data sharing in IoT environments. It leverages TEE to prevent fork attacks by ensuring that a node can never create two different events with the same "self-parent". Additional works apply TEE to cross-domain authentication [16] and trustworthy IoT data collection [17], where blockchain mainly provides immutable storage of authenticated information. An additional work proposed in [18] offers a high-performance rollup solution designed to improve blockchain scalability. Unlike ZK-rollups, which use complex zero-knowledge proofs, this framework uses heterogeneous TEEs to execute transactions

off-chain. The TEEs sign the execution results, and the blockchain only needs to verify these hardware-based signatures, significantly reducing on-chain gas costs. In all these works, TEE is primarily used to establish trust in execution or data acquisition, while blockchain acts as an immutable audit layer.

Unlike the previously described works, numerous studies focus exclusively on the integration of blockchain and ZKPs, often avoiding the use of TEEs due to concerns regarding hardware trust assumptions and potential side-channel attacks. The solution presented in [19] integrates machine learning and blockchain for intelligent energy trading, using Zero-Knowledge Machine Learning (ZKML) to protect user energy consumption data while making computations mathematically verifiable on-chain. Other approaches employ zk-SNARKs or NIZKs to provide authenticated data feeds [20] or anonymous device interaction [21], demonstrating how blockchain can verify cryptographic proofs without relying on trusted hardware.

Finally, the work presented in [22] proposes a lightweight framework for healthcare IoT, where zk-SNARKs are used to ensure the correctness of local training and model aggregation across a ring topology. The authors state that, with respect to TEE-based methods for securing federated learning which require specialized hardware, the zk-SNARK approach presented in this work provides a non-interactive, lightweight alternative for constrained devices. These approaches remove hardware trust assumptions by relying on mathematical proofs, but often incur a substantially higher computational cost when applied to complex ML workloads.

Differently from the already described works that integrate the blockchain technology with TEE or ZKP, based on the current state of art, some works have investigated the integration of TEEs with ZKPs and blockchain in a unique framework. Despite sharing the common objective of combining blockchain, TEE and ZKP, existing hybrid approaches primarily employ ZKPs to verify the correctness of the computation itself, or of aggregation rules in federated settings. For instance [23] defines a framework for medical AI privacy, where the three technologies work together to enable multi-institutional collaboration while maintaining strict data governance. The author mentions Intel SGX/TDX or NVIDIA H100 as TEE where the aggregator operates in order to isolate the computation of global model updates and to ensure that the system operator cannot view raw patient gradients or updates in plaintext. The TEE generates a succinct ZK proof (using systems like Halo2 or Nova) to prove that the global update was computed using exactly the legitimate, committed inputs and followed the prescribed aggregation rules. This provides public verifiability without revealing the sensitive underlying updates to the verifiers. A permissioned consortium blockchain acts as the decentralized verifier and immutable audit log. Smart contracts verify the ZKPs and record minimal metadata, such as hashes of commitments and policy parameters, to provide a permanent audit trail for regulators and hospitals. In [24] the authors propose an architecture designed for verifiable generative AI inference on the blockchain, where TEEs (specifically NVIDIA H100 Confidential Computing enclaves) serve as high-performance optimistic co-processors that perform complex model inference at near-native speeds and generate hardware attestations for real-time applications. ZKP acts as a defense layer against TEE side-channel vulnerabilities using stochastic spot-checks, randomly selecting a small subset of transactions (typically <1 %) via a Verifiable Random Function (VRF) to be proven correct through a mathematical ZKP, thereby deterring malicious nodes from exploiting hardware flaws. The blockchain manages the Model Registry, executes the VRF for spot-checks, and enforces economic penalties (slashing) for fraud.

Other hybrid approaches apply TEE and ZKP to secure healthcare data governance [25] or privacy-preserving decentralized storage [26], highlighting the complementary use of confidential execution and cryptographic verification in domain-specific scenarios.

Building on these observations, we identify a different design point. Instead of verifying the entire ML computation, we investigate whether verifying properties of the final decision can provide sufficient guarantees for many application scenarios while significantly reducing computational overhead. Compared to the existing hybrid approaches discussed above, our framework provides the following contributions:

- Focus on decision verifiability: while other works, such as [18, 23], use ZKPs to prove that the output correctly results from the execution of the ML model or from a federated aggregation

process, our work adopts a different perspective. We argue that, in many practical applications, the property requiring public verification is not the entire inference itself, but the correctness of the decision derived from the confidential output. Consequently, we propose a framework that uses ZKPs selectively to certify consistency, ensuring that the computed output matches the value recorded on-chain, and to verify that a decision (e.g., a classification or an alert) was correctly derived from the output based on predefined rules.
- Strategic separation of concerns: this work proposes a conceptual architecture that integrates blockchain, TEE, and ZKPs as complementary layers rather than mutually exclusive technologies. Each layer is assigned a specific role, which is better explained in Section 4. The proposed framework also relies on a decentralized data storage using IPFS.
- Mitigating TEE-Specific semantic gaps by coupling remote attestation with decision-level ZKP verification: a significant insight of this work is that while a TEE's remote attestation certifies that a specific binary was executed in a genuine enclave, it does not provide a formal cryptographic guarantee that the output was correctly derived from the input according to the intended semantics. The added ZKP layer closes this semantic gap by providing a mathematical proof that the subsequent decision-making process adhered to the intended rules.

Overall, even if each of the mentioned technologies addresses a specific requirement, none of them, when used in isolation, fully satisfies the need for auditable and privacy-preserving machine learning. Indeed, recent research highlights that no single technique is sufficient on its own, and that hybrid approaches combining multiple privacy-preserving mechanisms are increasingly necessary to achieve end-to-end secure and trustworthy ML pipelines [8]. The combination of the complementary properties of the key technologies underpinning this work (blockchain, TEE and ZKP) offers promising potential to achieve both practical efficiency and strong verifiability guarantees. The following Section 3 analyzes the design space of possible solutions, focusing on the trade-offs between TEE and ZKPs, and motivating a hybrid approach that leverages their complementary strengths.

3. Design Space for Auditable Machine Learning

The technologies introduced in the previous section provide complementary capabilities for addressing the challenges of trust, confidentiality, and verifiability in data-driven systems. However, their practical adoption requires a clear understanding of how they can be effectively combined within a machine learning pipeline. A key design choice in auditable machine learning systems concerns the level at which verifiability is enforced. Existing approaches typically focus on *execution verifiability*, aiming to guarantee that a computation has been correctly performed. In this setting, the objective is to prove that the output O is the result of applying a model M to input data D, i.e., $O = M(D)$. In this setting, the objective is to provide a mathematical proof that the model output is correctly derived from the declared input according to the intended computation.

While ZKP-based approaches provide strong guarantees for this purpose, their application to real-world machine learning workloads is still limited by the computational complexity of representing models as verifiable circuits and by the overhead of proof generation. Conversely, TEE-based solutions provide practical execution guarantees, but rely on hardware trust assumptions and do not offer a formal proof of correctness. For complex machine learning models, generating proofs for the entire inference may require encoding the complete computation as a verifiable arithmetic circuit, significantly increasing proof-generation costs.

An alternative perspective shifts the focus from execution to *decision verifiability*. In many applications, the final outcome of interest is not the raw output of a model, but a decision derived from it, such as a classification, an alert, or a threshold-based evaluation. In this context, verifying that the decision has been correctly derived from the output can be sufficient to ensure trust in the system, without requiring full verification of the underlying computation. This observation motivates a shift from execution verifiability to decision verifiability, where the focus is placed on the correctness of the business-relevant outcome rather than on the complete inference process.

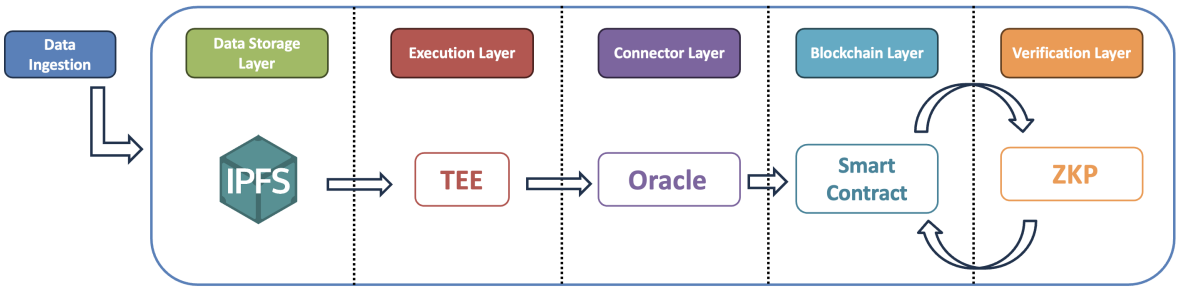


Figure 1: System components and data flow

Based on these considerations and on the comparison between the trade-offs characterizing TEE-based and ZKP-based systems described in the previous section, rather than selecting a single technology, this work explores a hybrid design in which TEE and ZKP are used in a complementary manner. In this approach, the execution of the machine learning algorithm is performed within a TEE, ensuring confidentiality and practical efficiency. ZKPs are then applied selectively to verify specific properties of the decision-making process, rather than the entire execution.

In particular, ZKPs can be used to certify invariants related to the decision-making process, such as:

- consistency between the computed output and the value registered on-chain;
- correctness of threshold-based decisions;
- compliance with predefined decision rules.

Unlike full execution proofs, these properties can be represented through compact verification circuits, for example encoding threshold comparisons or consistency checks between confidential outputs and their on-chain commitments. This verification approach, which is limited to specific output properties, reduces the computational burden associated with ZKPs, while improving the overall trustworthiness of the system. By combining TEE-based execution with ZKP-based validation of critical properties, the system reduces reliance on hardware trust assumptions without incurring the full cost of end-to-end cryptographic verification. This design enables a balance between efficiency, confidentiality, and verifiability, and forms the foundation for the system architecture described in the next section.

Unlike previous works that focus on certifying the execution of algorithms, this work focuses on verifying the correctness of decisions derived from their outputs.

4. System Components and Data Flow

Based on the design considerations discussed in the previous section, this work proposes a hybrid architecture that combines blockchain, TEE, and ZKP to enable auditable and privacy-preserving machine learning pipelines. The architecture, which combines off-chain processing with immutable on-chain recording, is designed to separate concerns between execution, verification, and auditability, assigning each technology a specific role.

In order to better understand the main components of the system and the data flow across the pipeline, it could be worth considering the Figure 1.

4.1. System Components

The architecture consists of the following main components:

- **Data Storage Layer**: Input datasets and intermediate artifacts are stored in a decentralized storage system (e.g., IPFS), where each file is identified by a unique content identifier (CID).

IPFS (InterPlanetary File System) is a peer-to-peer distributed storage system that ensures data integrity, availability, and content-addressability through immutable CIDs. It is used for a decentralized storage of raw data and intermediate processing results. This approach would enable all stakeholders to reference the exact datasets used in each analysis without relying on centralized repositories or trusted custodians.

- **Execution Layer (TEE)**: The machine learning model is executed inside a Trusted Execution Environment. This ensures confidentiality of both input data and model, while producing a verifiable attestation of the execution.
- **Connector Layer (Oracle)**: The oracle acts as a communication layer between off-chain computation and the blockchain. The oracle is not assumed to be trusted for correctness. Its role is limited to relaying TEE attestations, cryptographic commitments, and ZKP proofs to the blockchain. Integrity is ensured through remote attestation, hash commitments, and proof verification, while the oracle may only affect message availability or latency. It retrieves the CIDs associated with the input data, the attestation report, and the confidential output previously uploaded to IPFS, computes or retrieves the corresponding cryptographic commitments, and builds and broadcasts the transaction submitted to the smart contract.
- **Blockchain Layer (Decision and Auditing)**: The output of the machine learning model is stored to a Smart contract and processed to derive a decision (e.g., classification, threshold evaluation, anomaly detection). Smart contracts are also used to store cryptographic commitments (hashes) of data, outputs, and proofs, enabling public auditability and traceability.
- **Verification Layer (ZKP)**: ZKPs are generated to verify selected properties of the decision, ensuring that it has been correctly derived from the model output. Rather than proving the complete ML inference, the proof certifies business-relevant properties, such as threshold compliance or consistency between the confidential output and its on-chain commitment.

4.2. Data Flow

The system operates according to the following pipeline:

1. **Data Ingestion**: Input data D is collected and stored in a decentralized storage system, generating a content identifier (CID) that uniquely represents the dataset.
2. **Model Execution**: The machine learning model M is executed within the TEE using input data D, producing an output O. The TEE generates a remote attestation report certifying that the execution has been performed in a trusted environment.
3. **Output Registration**: The output O is stored off-chain and its hash is computed. The hash, together with the attestation report and references to input data, is submitted to the blockchain.
4. **Decision Derivation**: A decision function $f(O)$ is applied to the output, producing a final decision (e.g., alert, classification, threshold-based result).
5. **ZKP Generation**: A Zero-Knowledge Proof π is generated to certify that the decision $f(O)$ satisfies predefined conditions, without revealing sensitive information about O or the model. The private witness of the proof is the confidential model output O, while the public inputs include the on-chain commitment of the output and the decision to be verified. For example, a proof can certify that the confidential output satisfies a predefined threshold and therefore justifies the generated alert, without revealing the actual value of O.
6. **On-chain Verification**: The proof π, together with the relevant hashes, is stored on-chain. Smart contracts can verify the proof and ensure consistency between the recorded data and the decision.

In the proposed architecture, TEE and ZKP are not used as alternative solutions, but as complementary mechanisms addressing different aspects of the problem. TEE ensures that the machine learning model is executed in a confidential and integrity-preserving environment, while ZKPs provide verifiable guarantees on selected properties of the decision-making process.

This separation allows the expensive cryptographic verification to focus only on the business-relevant properties of the pipeline, while leaving the computationally intensive ML inference inside the TEE.

Importantly, ZKPs are not used to verify the entire machine learning computation, which would be computationally infeasible in many real-world scenarios. Instead, they are applied selectively to validate decision-level properties, enabling a practical trade-off between performance and verifiability.

This work represents a first step as it proposes a conceptual architecture and does not include a full implementation, which is left for future work.

5. Conclusion and Future Work

This paper addressed the challenge of enabling auditability in machine learning pipelines operating in decentralized and privacy-sensitive environments. While blockchain provides immutable and publicly verifiable records, it does not guarantee the correctness of off-chain computations. To overcome this limitation, we proposed a hybrid architecture that combines blockchain, TEE, and ZKPs.

The proposed system separates execution, verification, and auditability concerns. TEE ensures confidential and integrity-preserving execution of machine learning algorithms, while blockchain provides an immutable audit trail of the computation process. ZKPs are used to verify selected properties of the decision-making process derived from the model output, rather than the entire computation. This design enables a practical trade-off between efficiency and verifiability, without requiring full cryptographic verification of the ML inference, making the approach suitable for real-world applications.

Unlike existing works that focus on certifying the execution of algorithms, this work emphasizes decision-level verification, demonstrating that validating the correctness of decisions derived from model outputs can be sufficient to ensure trust in many application scenarios. This shift in perspective allows reducing computational complexity while maintaining meaningful guarantees on system behavior. The proposed approach is primarily conceptual, with an architecture designed to be practically feasible. A full implementation and experimental evaluation remain part of future work. In its current form, the integration of ZKP focuses on selected aspects of the decision-making process rather than covering the entire machine learning computation. Rather than replacing TEE-based execution with full cryptographic verification, the proposed framework leverages the complementary strengths of TEE, blockchain, and selective ZKP verification to achieve a practical balance between trust, confidentiality, and auditability.

Future implementations will also include a formal threat model covering components such as the oracle and the trust assumptions underlying the TEE.

Future work will focus on the implementation and evaluation of the proposed architecture in a real-world scenario. This includes the development of a prototype system, the integration of concrete zk-SNARK schemes (e.g., Groth16 or PLONK) for decision-level verification, and the assessment of performance and scalability trade-offs. Additionally, further research will investigate the applicability of the approach to different machine learning models and application domains, as well as the evolution of hybrid solutions combining TEE and ZKP as cryptographic technologies continue to mature. Overall, the proposed framework shows that decision-level verification can represent a practical alternative to full execution verification for many ML-driven applications, enabling auditable, confidentiality-preserving, and computationally efficient machine learning pipelines.

References

[1] R. Shinde, S. Patil, K. Kotecha, V. Potdar, G. Selvachandran, A. Abraham, Securing ai-based healthcare systems using blockchain technology: A state-of-the-art systematic literature review and future research directions, Transactions on Emerging Telecommunications Technologies 35 (2024) e4884.

[2] B. Liu, M. Ding, S. Shaham, W. Rahayu, F. Farokhi, Z. Lin, When machine learning meets privacy: A survey and outlook, ACM Computing Surveys (CSUR) 54 (2021) 1–36.

[3] M. Sabt, M. Achemlal, A. Bouabdallah, Trusted execution environment: What it is, and what it is not, in: 2015 IEEE Trustcom/BigDataSE/Ispa, volume 1, IEEE, 2015, pp. 57–64.
[4] G. Almashaqbeh, R. Solomon, Sok: Privacy-preserving computing in the blockchain era, in: 2022 IEEE 7th European Symposium on Security and Privacy (EuroS&P), IEEE, 2022, pp. 124–139.
[5] X. Sun, F. Yu, P. Zhang, Z. Sun, W. Xie, X. Peng, A survey on zero-knowledge proof in blockchain, IEEE Network 35 (2021) 198–205. doi:10.1109/MNET.011.2000473.
[6] K. Christidis, M. Devetsikiotis, Blockchains and smart contracts for the internet of things, IEEE access 4 (2016) 2292–2303.
[7] F. Casino, T. K. Dasaklis, C. Patsakis, A systematic literature review of blockchain-based applications: Current status, classification and open issues, Telematics and informatics 36 (2019) 55–81.
[8] J. Chen, H. Yan, Z. Liu, M. Zhang, H. Xiong, S. Yu, When federated learning meets privacy-preserving computation, ACM Computing Surveys 56 (2024) 1–36.
[9] V. Costan, S. Devadas, Intel sgx explained, Cryptology ePrint Archive (2016).
[10] R. Kunkel, D. L. Quoc, F. Gregor, S. Arnautov, P. Bhatotia, C. Fetzer, Tensorscone: A secure tensorflow framework using intel sgx, arXiv preprint arXiv:1902.04413 (2019).
[11] J. V. Bulck, M. Minkin, O. Weisse, D. Genkin, B. Kasikci, F. Piessens, M. Silberstein, T. F. Wenisch, Y. Yarom, R. Strackx, Foreshadow: Extracting the keys to the intel sgx kingdom with transient out-of-order execution, in: USENIX Security Symposium, 2018. URL: https://api.semanticscholar.org/CorpusID:52036294.
[12] F. Mo, Z. Tarkhani, H. Haddadi, Machine learning with confidential computing: A systematization of knowledge, ACM Comput. Surv. 56 (2024). URL: https://doi.org/10.1145/3670007. doi:10.1145/3670007.
[13] Z. Peng, T. Wang, C. Zhao, G. Liao, Z. Lin, Y. Liu, B. Cao, L. Shi, Q. Yang, S. Zhang, A survey of zero-knowledge proof based verifiable machine learning, arXiv preprint arXiv:2502.18535 (2025).
[14] G. Marengo, Design and Implementation of a Transparent and Verifiable Structural Health Monitoring Pipeline Based on Decentralized Technology, Master's thesis, Politecnico di Torino, Torino, Italia, 2026. URL: https://webthesis.biblio.polito.it/id/eprint/39277.
[15] X. Fu, H. Wang, P. Shi, X. Zhang, Teegraph: A blockchain consensus algorithm based on tee and dag for data sharing in iot, Journal of Systems Architecture 122 (2022) 102344.
[16] W. Mao, P. Jiang, L. Zhu, Btaa: Blockchain and tee-assisted authentication for iot systems, IEEE Internet of Things Journal 10 (2023) 12603–12615.
[17] C. Liu, H. Guo, M. Xu, S. Wang, D. Yu, J. Yu, X. Cheng, Extending on-chain trust to off-chain – trustworthy blockchain data collection using trusted execution environment (tee), IEEE Transactions on Computers 71 (2022) 3268–3280.
[18] X. Wen, Q. Feng, H. Lyu, J. Niu, Y. Zhang, C. Feng, Teerollup: Efficient rollup design using heterogeneous tee, IEEE Transactions on Computers (2025).
[19] C. Fan, A. Sohrabbeig, P. Musilek, Zero-knowledge machine learning models for blockchain peer-to-peer energy trading, Internet of Things 32 (2025) 101638.
[20] Z. Wan, Y. Zhou, K. Ren, zk-authfeed: Protecting data feed to smart contracts with authenticated zero knowledge proof, IEEE Transactions on Dependable and Secure Computing 20 (2023) 1335–1347.
[21] H. Ma, Y. Liu, Y. Liu, F. Feng, Z. Liu, Dpczk: Enhancing device privacy through certificate-free encryption and zero-knowledge proof in multidomain iot environments, IEEE Internet of Things Journal (2025).
[22] H. Byeon, A. Chaudhary, J. V. N. Ramesh, D. N. Reddy, B. V. Nandakishore, K. B. V. B. Rao, F. Dahan, A. Ostonokulov, M. Soni, Trusted aggregation for decentralized federated learning in healthcare consumer electronics using zero-knowledge proofs, IEEE Transactions on Consumer Electronics (2025).
[23] S. Sharma, G. Petrovic, S. Kaushik, zkfl-health: Blockchain-enabled zero-knowledge federated learning for medical ai privacy (2025).
[24] A. Chan, A. Ding, F. Chen, A. Wu, B. Zhang, A. Tian, Optimistic tee-rollups: A hybrid architecture

for scalable and verifiable generative ai inference on blockchain, arXiv preprint (2025).
[25] M. Ezz, A. S. Alaerjan, A. M. Mostafa, Ethical ai in healthcare: Integrating zero-knowledge proofs and smart contracts for transparent data governance, Bioengineering 12 (2025) 1236.
[26] J. John, Zero-knowledge proofs for privacy-preserving access in blockchain storage systems (2025).