PAPER RE07
Verifiable Compute to Trust AI in Regulated Workflows

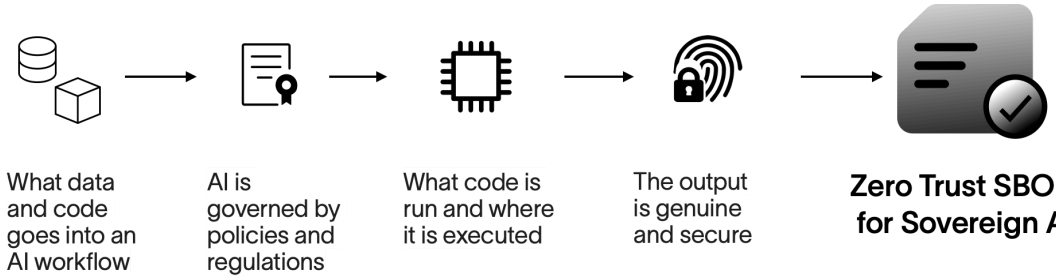
Ali Dootson, Chapel Hill, NC, USA
Jonathan Dotan, Los Angeles, CA, USA

Abstract
Verifiable Compute is a new approach to trust AI. Using hardware-based cryptography, tamper-proof records of data, code, and computations, developers can provide explainability and prove compliance with pre-specified governance controls. Using a digital notary within a secure enclave on both the CPU/GPU, developers create verifiable yet confidential audit records. Retroactively, these immutable proofs can be inspected for inference quality and regulatory compliance. By proving models meet any pre-determined governance controls, drug regulators could proactively review models before their fine-tuning by sponsors or CROs, aligning closer to SaMD device reviews. To demonstrate, an open-source pharmacovigilance model to detect safety signals using RWD has been verified, proving compliance with the EU AI Act and several more governance and security frameworks. This presentation aims to shape the model review process being pioneered by the FDA Emerging Drug Safety Technology Program, accelerating the market acceptance of AI within regulated drug development workflows.

Introduction
The opportunities created by AI elicit a simple yet fundamental question: How can we establish trust in AI data, models, and agents to keep pace with AI innovation? In this whitepaper, we'll demonstrate how a new root of trust for AI can begin right on the silicon for every AI computation. Using trusted execution environments available on next-generation CPUs and GPUs it is possible to create a novel digital notary that cryptographically signs at runtime the inputs, outputs, and work done on both a CPU and GPU to provide a certificate of authenticity of an AI process. We introduce this solution as a new paradigm for creating trust in AI, powered by a new class of computing: Verifiable Compute.

In the life sciences, where patient privacy and data integrity are paramount, Verifiable Compute can prove that their AI models are trained on authenticated datasets in secure environments. This is critical for applications like drug discovery, where the lineage and integrity of training data directly impact results. Verifiable Compute also facilitates compliance with strict regulatory requirements, providing audit trails that prove adherence to data governance and ethical standards.

Overview of Verifiable Computing



Why the AI Era Requires a New Trust Paradigm

The rapid advancements in artificial intelligence have revolutionized industries but introduced new challenges around trust. Today's AI systems operate on vast and dynamic datasets, often with limited oversight, making transparency and accountability essential. Trust in AI ensures stakeholders can collaborate effectively, accelerate innovation, and mitigate potential legal and ethical pitfalls.

The ability to reproduce the output of an AI system is one of the primary methods of establishing trust. However, verifying reproducibility isn't always feasible or practical due to the complexity, cost, and inherent limitations of a modern AI system. Non-deterministic models, distributed workflows, and the brittle nature of certain AI frameworks can make exact reproduction impossible. Moreover, in some cases, full transparency could lead to unintended consequences, such as exposing sensitive data or revealing proprietary methods.

The paradigm of Verifiable Compute shifts the foundation of trust to the silicon level, enabling cryptographically verifiable processes that persist across the AI lifecycle. With real-time certifications of AI data, computations, and governance, Verifiable Compute establishes a robust framework for managing AI systems securely and transparently.

Risks of AI Supply Chain Attacks

AI supply chains are uniquely vulnerable to security threats due to their complexity and reliance on multi-party collaboration. Conventional software systems already face widespread supply chain attacks, with 91% of organizations reporting incidents. For AI, these risks are magnified by the scale of data and models involved and the autonomy of agentic systems.

Supply chain attacks on AI can manifest through model poisoning, backdoors, and data exfiltration, potentially undermining decision-making processes or compromising sensitive information. Unlike traditional software-based systems, AI errors are often subtle and evolve over time, making them difficult to detect and mitigate.

The lack of transparency in AI supply chains further exacerbates these risks. Current systems often rely on unverifiable inputs and outputs, creating blind spots in governance and accountability. Verifiable Compute addresses these vulnerabilities by enabling end-to-end verification of AI assets, from model lineage to runtime processes. This ensures that every component in the supply chain is auditable and secure, laying the groundwork for a new standard of trust in AI systems.

TEE and SLSA Level 3

Trusted Execution Environments (TEE) anchor Verifiable Compute in hardware-based security, isolating and encrypting sensitive data and operations within a secure enclave. TEE ensures that computations are executed on authentic devices and protected from unauthorized access or tampering. This approach mitigates risks in environments where traditional security measures fall short, such as public clouds or multi-tenant systems.

The framework also incorporates the SLSA Security Level 3 standard, a rigorous security level designed for supply chain integrity. SLSA Level 3 confirms that all build artifacts are tamper-proof, cryptographically signed, and auditable. By aligning with SLSA, Verifiable Compute provides a trusted chain of custody for AI processes, establishing confidence in the inputs, environments, and outputs across the lifecycle.

The Notary System

The Notary system is a foundational component of Verifiable Compute, providing persistent cryptographic proofs for AI processes. This subsystem generates tamperproof certificates to verify the confidentiality, correctness, and governance of computations. These certificates are signed in real-time using keys generated within the TEE, ensuring that all artifacts are auditable indefinitely.

The Notary system extends beyond data and code to encompass the hardware environment, enabling comprehensive trust in AI systems. By integrating with open standards like Verifiable Credentials (VCs) and System Package Data Exchange (SPDX), it ensures compatibility and portability across various enterprise systems, cementing its role as the backbone of Verifiable Compute.

Five Example Proofs

Verifiable Compute delivers five persistent proofs that address critical aspects of trust in AI: confidentiality, environment, correctness, computation, and governance. These proofs ensure integrity at each stage of an AI process, enabling comprehensive auditability and accountability:

***1. Confidentiality**:* Verifiable Compute ensures that data and computations remain confidential, even in shared environments. Through TEE and secure GPU channels, sensitive data is encrypted in memory and accessible only to authorized processes. This guarantees that no information leaks occur, protecting both user privacy and proprietary models.

***2. Environment**:* The system provides a Proof of Environment that verifies the authenticity of the hardware and software stack. By attesting to the integrity of the CPU, GPU, and firmware, this proof establishes trust in the execution environment, ensuring that computations occur on genuine, unaltered devices.

***3. Correctness**:* Proof of Correctness ensures that computations were executed as intended without interference. The Notary system cryptographically validates the execution of specified programs, preventing tampering and providing assurance that inputs and outputs align with expectations.

***4. Computation**:* This proof captures and signs metrics such as GPU utilization, power usage, and memory consumption during execution. Proof of Computation enables stakeholders to verify that adequate resources were allocated and utilized, ensuring transparency and validation in performance claims.

***5. Governance**:* Verifiable Compute enforces runtime governance policies and generates cryptographically signed attestations of compliance. These governance proofs provide a layer of assurance that AI processes adhere to regulatory standards, business policies, and ethical guidelines, creating an audit trail for accountability.

Implementation

The implementation of Verifiable Compute seamlessly integrates into existing AI workflows and infrastructure, providing verifiable assurances at every step. This system addresses the full AI lifecycle, from training and inference to deployment and governance, ensuring that all processes are secure, accountable, and compliant. By leveraging a modular product suite, Verifiable Compute allows organizations to adapt the framework to their unique needs while maintaining cryptographic trust.

Workflows

Verifiable Compute transforms traditional AI workflows by embedding cryptographic proofs into every stage. These workflows, powered by Verifiable Compute, provide auditable evidence of data integrity, process correctness, and governance compliance, ensuring that AI systems are not only performant but also transparent and secure.

1. Verifiable Training: AI model training often involves large datasets and computational resources, making it a critical point of vulnerability. Verifiable Training ensures that data, algorithms, and environments are authenticated and unchanged throughout the training process. This creates a tamper-proof lineage for the resulting model, guaranteeing its integrity and provenance.

2. Verifiable Inference: During inference, Verifiable Compute provides assurances that models operate as intended and maintain confidentiality. Each inference session generates cryptographically signed proofs that verify the correctness of inputs, outputs, and runtime environments. This enables users to trust the results without exposing sensitive data.

3. Verifiable RAG: Retrieval-Augmented Generation (RAG) workflows benefit from Verifiable Compute's ability to authenticate the lineage of data references. By ensuring that embeddings and metadata are derived from trusted sources, Verifiable RAG empowers AI systems to provide transparent, traceable outputs with guaranteed data integrity.

4. Verifiable Benchmarks: Benchmarking AI models is essential for performance evaluation. Verifiable Benchmarks authenticate the validity of test datasets, models, and results, eliminating the need for re-execution. These proofs allow stakeholders to assess AI performance with confidence, even years after initial testing.

5. Verifiable Safeguards: Runtime safeguards, such as content moderation or privacy filters, are integral to responsible AI use. Verifiable Compute generates proofs that these safeguards are applied correctly, affirming that all policies and protections are in place during execution.

6. Verifiable Governance: By embedding governance checks into workflows, Verifiable Governance enables compliance with regulations and organizational policies. This includes verifying data usage, runtime conditions, and ethical AI practices, creating a transparent audit trail for all stakeholders.

7. Verifiable Builds: AI systems often rely on complex build processes. Verifiable Builds use cryptographic proofs to authenticate the integrity of software builds, validating that every component in the AI stack is secure and untampered. This prevents vulnerabilities in supply chains from compromising AI systems.

Real World Data

Verifiable Compute embeds cryptographic proofs into real world data and processes, enabling end-to-end verification of confidentiality, correctness, and governance. This is of particular importance when referencing the workflow of real world data becoming useful as real world evidence. The ability to verify the provenance, lineage and traceability of patient and then vendor data is paramount for the downstream applicability of the data. That proof can open doors to unprecedented regulated use of a multitude of data.

Summary

Verifiable Compute establishes a groundbreaking framework for trust in the AI era, addressing critical challenges of security, accountability, and compliance across diverse workflows. By leveraging advanced technologies from next-generation GPUs and CPUs, along with innovative new Notary systems, Verifiable Compute embeds cryptographic proofs into AI processes, enabling end-to-end verification of confidentiality, correctness, and governance.

By shifting trust to the silicon level and providing persistent proofs, this new paradigm redefines how stakeholders interact with and trust AI systems. Verifiable Compute is not just a technical solution – it's a critical enabler of Real World Data and AI's responsible and transformative potential.

Learn More

To explore how Verifiable Compute can transform AI workflows and ensure trust in every step of the process, please read the full whitepaper.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Author Name: Ali Dootson
Company: EQTY Life Sciences
Address: 136 E. Rosemary St. Suite 100 Chapel Hill, NC 27514
Work Phone: N/A
Email: alistair.dootson@eqtylab.io
Website: https://www.eqtylab.io/life-sciences

Co-author Name: Jonathan Dotan
Company: EQTY Life Sciences
Address: 136 E. Rosemary St. Suite 100 Chapel Hill, NC 27514
Work Phone: N/A
Email: jd@eqtylab.io
Website: https://www.eqtylab.io/life-sciences