GAITA
Global AI Trust Alliance

THE ZURICH CONSENSUS

ON THE

VERIFIABLE AI TRUST CHAIN (VAITC)

First Principle — Intelligence Is Physical

Intelligence is physical. Because every computation occurs in unique hardware, only hardware-rooted identity can make an intelligent system accountable.

Adopted by the Global AI Trust Alliance (GAITA), Zurich, Switzerland

Our Challenge

Human civilization rests upon verifiable identity and accountability. We have long required persons and organizations to be identifiable, responsible, and answerable under the law.

Autonomous artificial agents now move and act alongside us, influencing both physical and economic reality—yet many do so without verifiable identity or tamper-proof records of their actions. Where no objective proof connects each action to a responsible human, a specific machine, and a verifiable chain of thought, accountability dissolves. A society built on trust cannot grant authority to systems that cannot be independently verified.

Our Opportunity

The exact physical nature of intelligence, creating the challenge, also provides the solution. By anchoring trust in hardware roots of trust and extending cryptographic attestation across the full lifecycle of intelligent systems, we can create an unbroken Verifiable AI Trust Chain ("VAITC") that binds three inseparable elements:

1. The Responsible Entity (The Responsibility Chain)
2. The Machine Anchor (The Hardware Birth Chain)
3. The Verifiable Action Stream (The Provenance Chain)

When these three elements are cryptographically integrated, any party may independently verify who authorized an action, the hardware on which it took place, and ensure that the recorded behavior remains unaltered.

Our Call for Contributions

To address this significant development, we hereby introduce **The Zurich Consensus on the Verifiable AI Trust Chain (VAITC)**—a comprehensive global framework developed by GAITA to establish verifiable intelligence and economic trust in the era of autonomy.

We invite all creators, policymakers, and innovators to contribute to the formulation of this standard. The VAITC aims to establish a root-of-trust-anchored system that mandates each intelligent entity to furnish cryptographic proof of its identity, conduct, and accountability throughout its lifecycle.

This is a collective call to action. Join us in designing and deploying the standards for verifiable intelligence and sustainable autonomy. Together, we can ensure that intelligent systems remain traceable, trustworthy, and aligned with human values—forming the technical and economic backbone of a secure and prosperous digital civilization.

Definitions

The **Verifiable AI Trust Chain Standard** initiates with a hardware root of trust and employs cryptographic methods to sign an AI's inputs, computations, and outputs, thereby establishing verifiable evidence of its identity, provenance, and integrity.

The **VAITC Standard** delineates the principles, technical requisites, and validation procedures by which AI systems produce verifiable evidence of their behavior and economic involvement. It facilitates the secure integration of hardware-based trust, cryptographic signing, and comprehensive verification processes to create a transparent and accountable framework for autonomous intelligence.

Article I — Architecture

VAITC is founded upon four mutually dependent pillars.

1. Silicon Root of Trust — an immutable, hardware-derived identity for every AI-capable device.
2. Chain of Verifiable Credentials — cryptographically signed attestations linking data, models, and operations to the root, supporting real-time validation and immediate revocation capability.

3. Cryptographic Action Stream — a real-time, tamper-proof sequence of attestation signatures that enables the creation of an immutable ledger.
4. Independent Verification — open standards enabling any party to independently validate the chain, thereby ensuring transparency even for sovereign actors.

Article II — Priority

The most pressing application pertains to agentic and embodied systems—specifically, those that operate through movement, transactions, or exertion of physical force within the real world. For these systems, verifiability is not discretionary; it is integral to lawful operation, insurance compliance, and societal acceptance.

Article III — Privacy by Design

The verification of authenticity and integrity shall never require the disclosure of proprietary model weights, training data, or personal information. Techniques such as remote attestation, selective disclosure, and zero-knowledge proofs will be employed to safeguard privacy while affirming factual accuracy.

Article IV — Economic Accountability

Autonomous economic agents capable of purchasing resources, entering contracts, or generating value must bind every transaction to the same hardware-anchored trust chain, creating verifiable audit trails for taxation, liability, and settlement.

Article V — Openness and Evolution

VAITC shall persist as an open, non-proprietary framework operating as a universal connecting layer, honoring the sovereignty of national data and facilitating a federated, state-regulated root of trust compatible with current infrastructure. Its application will be proportional to the associated risks, governed by transparent, globally representative processes, and supported by reasonable transitional pathways. All implementation particulars will be specified within accompanying technical standards and certification protocols.

Article VI — Call to Action

We therefore declare that verifiable truth — the cryptographically proven linkage of human responsibility, physical machine identity, and intelligent action — is the indispensable foundation for trustworthy autonomous intelligence.

Signatories to this Consensus commit to:

- Implementing hardware-anchored trust chains in high-capability systems under their stewardship.
- Supporting the development of open technical standards and independent certification mechanisms.
- Advocating for the verification of physical identity as the fundamental prerequisite for the secure, lawful, and prosperous integration of artificial agents into human society.

Conclusion

Without verification, trustworthiness cannot be established. With verification, it becomes assured.

www.GAITA.org