ZIRU LABS · CATEGORY MAP

The Trust Layer Category Map

A Reference Primer

Ziru Labs Research Publication · Version 1.0 · 2026 · Published under Ziru Labs corporate byline with contribution from Daniel Martin ·

Citable reference: Ziru Labs. *The Trust Layer Category Map: A Reference Primer*, v1.0. Published at zirulabs.com/research.

ABSTRACT

> This primer maps the operator landscape adjacent to the trust layer for AI category. It locates each surveyed operator on the eleven-layer AI infrastructure stack established in The AI Infrastructure Stack and the Trust Layer Position: A Reference Primer (Ziru Labs, 2026), surveys the categories of operators currently active at adjacent layers, identifies the standards bodies and sovereign programs codifying technical and policy requirements across the category, and documents the composition patterns through which the trust layer at Layer 3 connects to operators occupying adjacent layers in the trust-dependency ordering.

> The map is neutral with respect to specific commercial product comparisons. It is intended to support landscape navigation by readers including procurement teams evaluating AI infrastructure security architectures, partnership leads situating their organizations within the category, regulators identifying the operators implementing technical requirements under their frameworks, and academic researchers positioning new work against the established commercial landscape.

> The primer is a landscape companion to The Trust Layer for AI: A Reference Framework for the Category (Ziru Labs, 2026), which establishes the trust layer as a substrate category, and to the Stack Position primer noted above, which establishes its structural location. The accompanying Physics-Layer Threat Taxonomy for AI Infrastructure (Ziru Labs, 2026) catalogs the threat classes the substrate addresses. The Runtime Verification Gap in Federal AI Deployment: A Reference Primer (Ziru Labs, 2026) concretizes the substrate in the federal classified-deployment context. The Ziru Labs Capability Posture (Ziru Labs, 2026) documents Ziru Labs' specific posture within the category.

> Ziru Labs authored this map as one operator within the category it surveys. The map does not rank operators against one another and does not characterize any operator's offering as superior or inferior; that competitive neutrality is a deliberate and load-bearing property of a landscape reference. Where Ziru Labs is referenced, it is referenced as one operator within the category, and Section 7 and the companion Capability Posture state the company's own position directly.

1. Why a Category Map

A category in formation produces predictable confusion about who occupies which position in the landscape. Adjacent categories overlap conceptually; their operators are sometimes positioned by commercial messaging in ways that obscure structural distinctions; and practitioners evaluating the landscape need vocabulary that distinguishes layers, threat coverage, and composition patterns before evaluating any specific operator or product.

The trust layer for AI category is at this stage of formation as of the publication date of this primer. The category itself is established analytically in the companion Reference Framework. The category's structural position in the AI infrastructure stack is established in the companion Stack Position primer. The threat classes the category addresses are catalogued in the companion Threat Taxonomy. What remains, and what this primer provides, is the population of the landscape: which operators occupy adjacent layers, which operators occupy the trust-layer position itself, which standards bodies and sovereign programs are codifying the technical requirements within and around the category, and which composition patterns connect the trust layer to operators at adjacent positions.

This primer does not rank operators against one another. It does not characterize specific commercial offerings as superior or inferior to others. It locates each surveyed operator on the AI infrastructure stack, summarizes what the operator's offering addresses at category level, identifies the standards and frameworks the operator participates in, and articulates how the operator's category composes with the trust layer at Layer 3.

The map is structured to enable two reading patterns. A reader navigating to a specific layer (for example, evaluating confidential computing options at Layer 5) can read Section 4 directly. A reader navigating by composition pattern (for example, understanding how hardware roots of trust at Layer 4 compose with the trust layer at Layer 3) can read Section 8.

2. Scope and Method

The primer covers seven categories of operators and bodies: hardware roots of trust at Layer 4 (Section 4.1); confidential computing at Layer 5 (Section 4.2); AI security and governance platforms at Layer 11 (Section 4.3); hardware security specialists adjacent to but distinct from the eleven-layer stack (Section 4.4); cryptographic verifiable-inference and agent-action attestation operators producing zero-knowledge, proof-of-inference, and signed-receipt attestation systems (Section 4.5); standards bodies and consortia codifying technical requirements (Section 5); and sovereign and government programs codifying policy and procurement frameworks (Section 6). Layer 3 itself, the trust-layer-substrate position, is treated separately in Section 7.

In scope: organizations producing technical artifacts, commercial offerings, technical standards, or policy frameworks that materially affect the deployment, codification, or composition of AI infrastructure trust. The criterion for inclusion is structural relevance to the trust-layer category at Layer 3, not market size or commercial prominence.

Out of scope: pure model-layer AI safety research (covered well by other reference works including the OWASP Top 10 for Large Language Model Applications and the MITRE ATLAS framework), pure software security operators that do not touch hardware trust (distinct from the cryptographic verifiable-inference operators surveyed in Section 4.5, which are included because their verification claim is structurally adjacent to the trust layer), application-layer AI products that consume but do not produce trust infrastructure, and operators in categories so adjacent to the trust layer that their inclusion would dilute rather than sharpen the landscape.

The map's enumeration is current as of the publication date but is not exhaustive. Operators emerge, consolidate, and exit on a continuous basis. The planned quarterly State of Physics-Layer AI Trust report tracks landscape evolution between major-version updates of this primer.

The map is deliberately neutral. Where commercial messaging from a specific operator characterizes its offering's scope differently from the technical-layer placement documented here, the technical-layer placement governs for the purposes of this map. The map describes operators in terms of the structural layer at which they operate and the threat classes (per the Threat Taxonomy) they address; it does not endorse, evaluate, or rank any operator's specific implementation.

3. The Trust Layer Position Recapped

The trust layer for AI occupies Layer 3 in the eleven-layer AI infrastructure stack as developed in The AI Infrastructure Stack and the Trust Layer Position: A Reference Primer. The position is a trust-dependency position, not a physical-placement claim: the substrate is rooted in silicon architecture (Layer 2, at which ARM, x86, and RISC-V operate) and in compute silicon (Layer 4, at which AI accelerators including those from NVIDIA, AMD, Intel, Google, and AWS operate, and in which the silicon-resident hardware roots of trust reside), and it sits above them in the verification ordering because it consumes their execution and produces evidence about it. The layer's function is hardware-rooted runtime verification of AI computation properties, including weights used, inputs processed, constraints enforced, outputs produced, timing characteristics, and environmental state.

Four concepts that are commonly conflated with the trust layer for AI operate at distinct adjacent positions. Hardware roots of trust at Layer 4 are silicon features that validate firmware integrity at boot and at trusted measurement transitions; they do not extend to runtime evidence of AI inference operation. Confidential computing at Layer 5 establishes a software trust boundary around running workloads within accelerator dies; it does not extend below the trusted execution environment boundary to physical-layer, bus-level, or supply-chain threats. AI governance platforms at Layer 11 monitor, filter, and analyze AI behavior at software register; they cannot enforce against threats that operate below the substrate they observe. Cryptographic verifiable inference, surveyed in Section 4.5, produces software-generated proofs that a specific model produced a specific output; it establishes the correctness of the computation but is not rooted in the silicon that executed it and does not reach the physical-layer, bus-level, or supply-chain threats operating beneath the cryptographic abstraction.

The trust layer composes with each of these. The map's central sections survey the operators at each adjacent layer position; Section 8 details the composition patterns.

4. Operators at Adjacent Layers

4.1 Layer 4: Hardware Roots of Trust

The hardware root of trust position is occupied by silicon features and by silicon-feature vendors. The category provides cryptographic identity at the platform layer, measured-boot integrity attestation, and the foundational silicon-rooted assertions that subsequent attestation chains depend on. The category is mature, with codified standards (NIST SP 800-193, Trusted Computing Group TPM 2.0 Library Specification) and broad silicon-vendor adoption.

Trusted Platform Module vendors. The Trusted Platform Module (TPM) is the most widely deployed hardware root of trust, present in essentially all enterprise servers and most modern personal computing devices. The dominant TPM silicon vendors include Infineon Technologies, the largest TPM supplier by unit volume, along with STMicroelectronics, Nuvoton Technology, and Microchip Technology. TPM operates at Layer 4 as a silicon feature, providing cryptographic identity, sealed storage, platform configuration register measurement, and remote attestation capabilities through a standardized interface (Trusted Computing Group, current).

Apple Secure Enclave. Apple's Secure Enclave, present in Apple Silicon (M-series and A-series) processors and as a separate die in earlier Intel-based Macs, provides Apple-specific hardware root of trust functionality including biometric template storage, payment authentication, and platform integrity attestation (Apple Inc., 2024). The Secure Enclave is a closed-vendor implementation specific to the Apple platform stack.

Google Titan family. Google's Titan security chips operate as hardware roots of trust across the Google ecosystem. The Google Cloud Titan Security Module provides cloud-platform attestation at the data-center layer. The Pixel Titan M provides device-layer attestation in Pixel mobile devices. The OpenTitan project, hosted by the lowRISC consortium, provides an open-source silicon root-of-trust reference design that vendors and academic researchers can implement (OpenTitan Project, current).

Microsoft Pluton. The Microsoft Pluton processor, integrated into specific AMD Ryzen Pro processors, Intel Core processors in current generations, and Qualcomm Snapdragon platforms, provides hardware root of trust functionality with cryptographic identity managed through Microsoft cloud infrastructure (Microsoft Corporation, 2024). Pluton represents the most significant recent expansion of the hardware-root-of-trust category across the personal computing silicon ecosystem.

AWS Nitro. The Nitro System, deployed across AWS EC2 instances, provides hardware-rooted platform isolation and attestation for cloud workloads. The Nitro Security Chip operates as a hardware root of trust specific to the AWS infrastructure layer (Amazon Web Services, current). Nitro is a cloud-platform-specific implementation rather than a general-purpose silicon vendor offering.

ARM TrustZone. ARM TrustZone, present in the majority of ARM-based mobile and embedded silicon, provides hardware-enforced execution environment partitioning that supports root-of-trust functionality. TrustZone is a layered architecture spanning Layer 4 (root-of-trust feature) and Layer 5 (trusted-execution boundary); the architectural placement is mixed and depends on specific implementation choices by ARM licensees (ARM Holdings, current).

Hardware roots of trust at Layer 4 establish the boot-time and configuration-time integrity baseline that downstream attestation depends on. They do not produce runtime evidence of AI inference operation. The trust layer at Layer 3 composes with hardware roots of trust by extending the attestation envelope from trusted measurement transitions to continuous runtime operation.

4.2 Layer 5: Confidential Computing

The confidential computing category establishes a software trust boundary around running workloads, protecting workload data and computation from adversaries with privileged software access to the host operating system. The category is in mature commercial deployment with multiple silicon-vendor implementations, multiple cloud-

platform integrations, and a substantial pure-play operator ecosystem. The Confidential Computing Consortium under the Linux Foundation provides the cross-vendor technical specifications and reference architecture (Confidential Computing Consortium, current).

Silicon-vendor confidential computing. Four silicon-vendor confidential computing implementations are currently in commercial deployment. NVIDIA Confidential Computing extends the trusted execution environment boundary to GPU workloads, supporting confidential AI inference and training across H100, H200, and successor accelerator generations (NVIDIA Corporation, 2024). NVIDIA Confidential Computing is the most widely deployed AI-specific confidential computing implementation as of the publication date. Intel Trust Domain Extensions (TDX), introduced in Intel Xeon processors beginning with the Sapphire Rapids generation, provides VM-scoped confidential computing applicable across general-purpose data-center deployments, complementing the earlier enclave-scoped Intel Software Guard Extensions architecture for VM-level use cases (Intel Corporation, 2024). AMD Secure Encrypted Virtualization with Secure Nested Paging (SEV-SNP), introduced with AMD EPYC processors beginning with the Milan generation, provides confidential virtual-machine isolation with different architectural assumptions than Intel TDX (Advanced Micro Devices, 2024). ARM Confidential Compute Architecture (CCA), specified for ARMv9 platforms, provides confidential computing capabilities with realm-based partitioning; CCA deployment is more recent than the x86 implementations and is expanding across server-class ARM platforms (ARM Holdings, 2024).

Cloud-platform confidential computing. All major cloud platforms have integrated confidential computing offerings built on the underlying silicon-vendor implementations. Microsoft Azure Confidential Computing supports confidential virtual machines on AMD SEV-SNP and Intel TDX, confidential containers on Kubernetes, and confidential AI on NVIDIA Confidential Computing-enabled accelerators (Microsoft Corporation, 2024). Google Cloud Confidential Computing supports confidential virtual machines and Google Kubernetes Engine nodes on AMD SEV-SNP, with successor offerings on additional silicon (Google Cloud, current). AWS Nitro Enclaves provides isolated compute environments leveraging the Nitro System hardware root of trust (Amazon Web Services, current).

Pure-play confidential computing operators. A category of pure-play operators provides confidential computing orchestration, key management, and developer tooling that composes with the silicon-vendor implementations. Anjuna Security provides confidential computing deployment and orchestration, abstracting the underlying silicon-vendor differences and providing operational tooling for production deployment (Anjuna, current). Fortanix provides confidential computing combined with hardware security module functionality, addressing the key management surface that confidential computing deployments depend on (Fortanix, current). Edgeless Systems develops open-source tooling for confidential computing including Constellation (confidential Kubernetes), Contrast (confidential containers), and adjacent open-source projects (Edgeless Systems, current).

Open-source frameworks. The Open Enclave SDK, the Gramine library OS, and the Enarx project (originally developed by Profian and now stewarded under the Confidential Computing Consortium following Profian’s wind-down in early 2023) provide cross-vendor confidential computing development frameworks under Confidential Computing Consortium project hosting (Confidential Computing Consortium, current).

Confidential computing at Layer 5 protects running workloads within the accelerator die against software-layer adversaries. It does not extend to physical-layer threats outside the accelerator die boundary, to bus-level

interception of inference traffic between dies, to supply-chain compromise of the silicon being attested, or to runtime evidence of AI workload operation that operates below the trusted execution environment software trust boundary. The trust layer at Layer 3 composes with confidential computing by extending the attested surface below the trusted execution environment boundary into the physical-layer, bus-level, and supply-chain surfaces.

4.3 Layer 11: AI Security and Governance Platforms

The AI security and governance platform category operates at the software layer monitoring, filtering, and analyzing AI workload behavior. The category has emerged in commercial form across the past three years and is currently in active consolidation as larger security platforms acquire AI-specific specialists. Two industry-analyst category names are in common use: Gartner's AI Trust, Risk, and Security Management (AI TRiSM) and IDC's various AI security and AI governance category names. The category names are not formal standards but indicate the analytical groupings under which the operators below are typically discussed.

AI security specialists. Operators producing AI-specific security platforms include HiddenLayer, providing an AI security platform addressing adversarial machine learning detection, model integrity monitoring, and AI supply chain assessment (HiddenLayer, current); Robust Intelligence, acquired by Cisco Systems in 2024, providing AI risk management including pre-deployment model evaluation and runtime monitoring (Cisco Systems, 2024); Lakera, providing AI security including the Lakera Guard runtime protection product and the publicly available Gandalf prompt-injection evaluation environment that has produced widely cited research data (Lakera, current); Protect AI, acquired by Palo Alto Networks in 2025, providing AI model risk management, machine learning supply chain security, and AI red teaming (Palo Alto Networks, 2025); CalypsoAI, providing AI security including content moderation, policy enforcement, and AI usage monitoring across enterprise deployments (CalypsoAI, current); and Adversa AI, providing AI red teaming and AI security research with publication of cross-industry AI vulnerability research (Adversa, current).

AI governance platforms. Operators producing AI governance platforms include Credo AI, providing AI governance with model inventory, risk assessment, and regulatory compliance reporting across enterprise AI estates (Credo AI, current); Holistic AI, providing AI governance and audit with fairness assessment, regulatory compliance, and AI risk reporting (Holistic AI, current); Fiddler AI, providing AI observability including model monitoring, drift detection, and explainability (Fiddler, current); Cranium AI, providing AI security and governance with AI inventory mapping, risk monitoring, and supply chain assessment (Cranium, current); and IBM watsonx.governance, providing integrated AI governance within the IBM AI platform stack (IBM Corporation, current).

Hyperscaler-integrated offerings. Major cloud providers have integrated AI governance modules within their broader AI platforms, including Microsoft Azure AI Content Safety and Azure AI Governance, AWS Bedrock Guardrails, and Google Cloud Vertex AI safety filters and audit functionality.

AI security and governance platforms at Layer 11 operate at software register. They provide essential monitoring, observability, alerting, and policy-enforcement capabilities across deployed AI estates. They cannot enforce against threats that compromise the substrate they observe, because the platforms themselves run on that substrate. The trust layer at Layer 3 composes with AI governance platforms by providing hardware-rooted enforcement and evidence at the substrate level on which the platforms depend.

4.4 Adjacent Hardware Security Specialists

A set of operators outside the eleven-layer AI infrastructure stack proper but adjacent to several of its layers provides hardware security specialization that intersects with the trust layer category. The category includes hardware security IP licensors, hardware Trojan detection specialists, anti-tamper component vendors, and security-oriented open-architecture silicon vendors.

Hardware security IP. Hardware security intellectual property is licensed to silicon vendors and incorporated into shipping silicon. Major operators include Rambus (cryptographic IP, secure interfaces, anti-tamper technology), Synopsys DesignWare Security IP (root-of-trust IP, cryptographic accelerators), and Cadence Design Systems security IP. Hardware security IP licensors operate at the design and verification phases of silicon development, with their IP appearing in finished silicon shipped by their licensees (Rambus, current; Synopsys, current; Cadence, current).

Hardware Trojan and supply-chain integrity specialists. Specialists in hardware Trojan detection and supply-chain hardware integrity verification include Cycuity (formerly Tortuga Logic) and adjacent academic spinouts working on side-channel-based detection of hardware modification (Cycuity, current). The category is small but operationally significant for high-assurance deployments where supply chain integrity is a load-bearing concern. Academic foundations of the category include the hardware Trojan taxonomy literature (Tehranipoor and Koushanfar, 2010; Bhunia et al., 2014).

Anti-tamper component vendors. Component vendors providing tamper-detection and tamper-response integrated circuits include Analog Devices, which integrated the Maxim Integrated product family following its 2021 acquisition, and Microchip Technology. Their components appear in security-sensitive products including payment terminals, military hardware, and automotive security systems (Analog Devices, current; Microchip Technology, current).

Security-oriented open architecture. SiFive, the largest commercial RISC-V silicon vendor, provides the SiFive Shield platform integrating root-of-trust functionality with RISC-V cores (SiFive, current). OpenTitan, noted above under Layer 4, also belongs in this category as the open-source silicon root-of-trust reference design (OpenTitan Project, current).

These specialists provide components, IP, and reference designs that compose with the trust layer for AI category at multiple points. The composition patterns are detailed in Section 8.

4.5 Cryptographic Verifiable Inference and Agent-Action Attestation

A category of operators produces cryptographic proofs of AI computation at the software layer, establishing that a specific model produced a specific output through zero-knowledge and succinct-proof techniques rather than through hardware-rooted attestation. The category is surveyed here despite operating in software because the verification claim it makes, that the computation a model performed can be demonstrated rather than asserted, is structurally adjacent to the trust layer at Layer 3 and is commonly conflated with it. The distinction this map draws is one of method and rooting rather than of intent: both categories pursue verifiable AI computation, through different mechanisms at different layers.

Zero-knowledge machine learning (zkML) frameworks. zkML proving frameworks compile model inference into zero-knowledge circuits, allowing a prover to demonstrate correct execution of a committed model without revealing the model weights or the input. EZKL is a widely used proving framework in this category (EZKL, current). The underlying circuit constructions and proof systems applied to machine-learning workloads are surveyed in the academic literature catalogued in Section 11.

Proof-of-inference operators. Operators building production verifiable-inference systems include Lagrange Labs, whose DeepProve system targets zero-knowledge proofs over neural-network inference at production scale (Lagrange Labs, current); Inference Labs, which develops a proof-of-inference protocol for demonstrating the soundness of model outputs in outsourced and decentralized compute settings (Inference Labs, current); and Nexus, which develops verifiable-computation infrastructure applied to AI workloads (Nexus, current). Much of the category's current commercial activity is oriented toward decentralized-compute and outsourced-inference contexts, in which the consumer of an inference cannot trust the machine that produced it.

Agent-action attestation and trust-receipt operators. A related set of operators produces software-layer attestation of AI agent actions rather than proofs of model inference. These systems generate cryptographically signed, tamper-evident records of what an agent did, the handoffs and approvals along a workflow, and the identity under which each action was taken, so that an agent's activity can be demonstrated rather than asserted. Zerker Labs, an SRI International spinout, produces Treeship, an open-source local-first trust layer that creates and verifies signed, hash-chained receipts for agent actions, handoffs, and approvals, verifiable offline (Zerker Labs, 2026). HUMAN Security produces AgenticTrust and the associated open-source HUMAN Verified AI Agent, which signs agent requests using HTTP Message Signatures and public-key pairs (HUMAN Security, 2025). Keycard produces agent identity and access infrastructure that issues dynamic, identity-bound, task-scoped credentials and maintains cryptographically linked audit trails from user to agent to resource (Keycard, 2025). The category is oriented toward the emerging agentic-economy context, in which the consumer of an agent action cannot otherwise verify which agent, under which authority, performed it.

Method and rooting distinction. Cryptographic verifiable inference establishes the correctness of a computation in software: it proves that a committed model, run on a given input, produced a given output. Agent-action attestation establishes, also in software, that a given action occurred and is attributable to a given agent identity. Neither binds its proof or receipt to the physical silicon that executed the underlying computation, and neither extends to the physical-layer, bus-level, or supply-chain threats catalogued in the Physics-Layer Threat Taxonomy Attack Surfaces 1 and 2, which operate beneath the cryptographic abstraction. Proof-generation overhead at frontier-model scale also remains an active research constraint for the zero-knowledge operators, per the literature in Section 11. The trust layer at Layer 3 addresses the complementary surface: it roots verification in attested silicon and covers the physical surface a software proof or receipt abstracts away. The composition pattern is detailed in Section 8.

5. Standards Bodies and Consortia

The standards-and-consortia landscape codifying technical and policy requirements relevant to the trust layer category is broad. The operators below are those whose work materially affects the category as of the publication

date.

International standards bodies. ISO/IEC JTC 1 (Information Technology Joint Technical Committee) hosts two subcommittees centrally relevant to the category. SC 42 (Artificial Intelligence) produces AI-specific standards including ISO/IEC 42001:2023 (AI management systems), ISO/IEC 23894:2023 (AI risk management), ISO/IEC 5338:2023 (AI lifecycle processes), and the AI trustworthiness work-in-progress family. SC 27 (Information Security, Cybersecurity, and Privacy Protection) produces general information security standards including the ISO/IEC 27000 family and the in-progress ISO/IEC 27090 (AI-specific cybersecurity guidance, at FDIS stage as of early 2026).

Regional standards bodies. CEN-CENELEC JTC 21 (Artificial Intelligence) is developing the harmonised standards required by the EU AI Act, including the technical standards under Article 40 that establish presumption of conformity. JTC 21's work is structurally important to the category because the harmonised standards adopted will be the technical reference under which EU AI Act conformity assessment proceeds (CEN-CENELEC, current).

IEEE Standards Association. IEEE produces standards across AI, hardware security, and cryptography. The IEEE International Symposium on Hardware Oriented Security and Trust (HOST) is the primary academic forum for hardware security research. IEEE working groups on AI ethics, AI system development, and adjacent topics produce normative material that intersects with the category (IEEE Standards Association, current).

National standards bodies. The U.S. National Institute of Standards and Technology is the most prominent national standards body in the category, with the AI Risk Management Framework (NIST AI 100-1, 2023), the AI RMF Playbook (NIST, 2024), the Cybersecurity Framework 2.0 (NIST CSWP 29, 2024), and the FIPS, SP 800-series, and SP 1800-series publications that establish technical requirements for federal and federally aligned deployments. NIST hosts the U.S. AI Safety Institute within the broader institute structure.

Industry consortia. Three industry consortia produce technical specifications centrally relevant to the trust layer category. The Confidential Computing Consortium (Linux Foundation) produces cross-vendor confidential computing specifications, reference architectures, and project hosting (Confidential Computing Consortium, current). The Trusted Computing Group produces specifications including the TPM 2.0 Library Specification, the DICE Architecture, and adjacent trusted-computing specifications (Trusted Computing Group, current). The Cloud Security Alliance produces cloud security guidance including the Cloud Controls Matrix that intersects with the trust layer at the cloud-deployment surface (Cloud Security Alliance, current).

AI-specific industry frameworks. The OWASP Foundation publishes the OWASP Top 10 for Large Language Model Applications, providing the canonical application-layer AI security framework (OWASP Foundation, 2025). The MITRE Corporation publishes the MITRE ATT&CK Enterprise framework and the MITRE ATLAS (Adversarial Threat Landscape for Artificial Intelligence Systems) framework, providing the canonical adversary-tactics taxonomies for general information security and AI-specific contexts respectively (MITRE Corporation, 2024). The Coalition for Content Provenance and Authenticity publishes content authenticity standards relevant to the provenance application surface (C2PA, 2024).

These standards bodies and consortia codify the technical and procedural requirements through which the trust layer category will be operationalized in commercial, regulatory, and sovereign deployment contexts. The

category's reference implementation will be established through the standards-codification window currently open across these bodies, with primary codification activity expected to complete across the eighteen to thirty-six months from the publication date.

6. Sovereign and Government Programs

Sovereign and government programs codifying policy and procurement frameworks adjacent to the trust layer category are surveyed at category level below. Visibility into specific program scope is limited by classification posture in many jurisdictions; the survey captures publicly disclosed activity.

United States. The Defense Advanced Research Projects Agency hosts multiple programs relevant to the category. The Trust in Integrated Circuits program addressed supply-chain integrity for high-assurance silicon. The Supply Chain Hardware Integrity for Electronics Defense (SHIELD) program developed authentication mechanisms for silicon supply chain. The Automatic Implementation of Secure Silicon (AISS) program developed automated security verification for silicon designs (DARPA, program documentation). The Intelligence Advanced Research Projects Activity hosts the TrojAI program addressing AI Trojan detection (IARPA, program documentation). The U.S. Department of Defense Chief Digital and Artificial Intelligence Office coordinates departmental AI deployment including the operational guidance under which federal classified AI deployment proceeds (DoD CDAO, current). The U.S. AI Safety Institute, hosted within NIST, is the U.S. government's primary AI safety research entity.

United Kingdom. The UK AI Security Institute (renamed in early 2025 from the previously named UK AI Safety Institute) is the U.K. government's primary AI safety and security research entity (UK AI Security Institute, current). The National Cyber Security Centre provides the U.K.'s cybersecurity guidance baseline including for AI deployment (NCSC, current).

European Union. The European AI Office, established within Directorate-General for Communications Networks, Content and Technology, is the European Commission's primary AI governance entity and coordinates EU AI Act implementation (European Commission, 2024). The European Union Agency for Cybersecurity provides cross-member-state cybersecurity guidance including for AI systems (ENISA, current).

Allied jurisdictions. France's National Cybersecurity Agency (ANSSI), Germany's Federal Office for Information Security (BSI), Australia's Australian Signals Directorate, and Canada's Canadian Centre for Cyber Security provide national cybersecurity guidance baselines that intersect with the trust layer category through their respective frameworks (ANSSI, current; BSI, current; ASD, current; CCCS, current). Each operates under its respective sovereign AI policy framework as those frameworks codify.

Asia-Pacific. Singapore's Infocomm Media Development Authority has produced the Model AI Governance Framework that has been widely referenced internationally (IMDA, 2024). Japan has established the Japan AI Safety Institute as a counterpart to the U.S. and U.K. institutes. The Republic of Korea has established analogous structures under its AI Basic Act framework (Republic of Korea Government, 2024).

The sovereign and government program landscape determines the procurement frameworks under which trust-layer infrastructure is acquired by government customers, the technical standards under which classified AI

deployment proceeds, and the policy frameworks within which allied cooperation on AI infrastructure trust is structured. The category's deployment in sovereign and government contexts depends materially on the codification activity within these programs.

7. The Layer 3 Position: Operator Landscape

The Layer 3 position in the AI infrastructure stack, the trust-layer-substrate position, is sparsely populated as of the publication date. The category is in formation per the analysis in the companion Reference Framework, and the structural characteristics distinguishing the layer from adjacent layers were established in the Stack Position primer.

Ziru Labs operates at Layer 3 within the trust layer for AI category, with the operational posture documented in the companion Ziru Labs Capability Posture (Ziru Labs, 2026). Ziru Labs is the publisher of the analytical foundation papers that establish the category, the threat taxonomy the category addresses, the stack position the category occupies, and the present landscape map.

Other operators occupying Layer 3 positions may exist as of the publication date. The map's visibility into Layer 3 activity beyond Ziru Labs is limited by several structural factors: deep-tech operators at this layer often operate with cleared-handling discipline that restricts public disclosure of their work; the category is in formation and operators may not yet be positioned publicly at the Layer 3 level; and the standards codification activity that will eventually surface Layer 3 operators is still underway across multiple bodies. The map will be updated as visibility into Layer 3 operator activity improves.

The observation that Layer 3 is sparsely populated is structural rather than competitive. Substrate-category positions are typically sparsely populated at category formation; the historical analogs in the Reference Framework (public-key cryptography in the 1970s and 1980s, TCP/IP in the 1970s and early 1980s, the GPS civilian signal in the 1980s, the ARM instruction set architecture in the early 1990s) each occupied their respective Layer 3 positions with one or a small number of operators during the first decade of category formation. The trust layer for AI is at this same stage.

8. Composition Patterns

The trust layer at Layer 3 composes additively with operators at adjacent layers. The composition patterns determine how trust-layer deployment combines with existing operator deployments to produce defense-in-depth trust postures. Six composition patterns are documented below.

Composition with hardware roots of trust at Layer 4. Hardware roots of trust establish cryptographic identity and measured-boot integrity at trusted measurement transitions. The trust layer extends the attestation envelope from these discrete measurement transitions to continuous runtime evidence of AI workload operation. The composition pattern is: the hardware root of trust establishes the boot-time and configuration-time integrity baseline; the trust layer produces runtime evidence of operation against that baseline; the combination produces continuous trust posture from boot through operational lifecycle. Specific operators at Layer 4 including the TPM

ecosystem, the Apple Secure Enclave, the Google Titan family, the Microsoft Pluton processor, and the AWS Nitro Security Chip each compose with the trust layer through this pattern.

Composition with confidential computing at Layer 5. Confidential computing protects running workloads within the accelerator die against software-layer adversaries with privileged host access. The trust layer extends the attested surface below the trusted execution environment boundary to physical-layer, bus-level, and supply-chain threats that the TEE boundary does not cover. The composition pattern is: confidential computing protects within the die; the trust layer attests outside the die; the combination produces complete attested coverage from outside-the-chassis to inside-the-die. NVIDIA Confidential Computing, Intel TDX, AMD SEV-SNP, ARM CCA, and the pure-play operators built on these architectures each compose with the trust layer through this pattern.

Composition with AI security and governance platforms at Layer 11. AI governance platforms operate at software register, providing monitoring, observability, alerting, and policy-enforcement capabilities across deployed AI estates. The trust layer provides hardware-rooted enforcement and evidence at the substrate level that governance platforms cannot produce from their own software layer. The composition pattern is: governance platforms provide the breadth of monitoring and policy-management coverage; the trust layer provides the depth of hardware-rooted enforcement at specific verification surfaces; the combination produces governance posture that survives software-layer compromise. HiddenLayer, Robust Intelligence, Lakera, Protect AI, CalypsoAI, Credo AI, Holistic AI, Fiddler, Cranium, IBM watsonx.governance, and adjacent governance operators each compose with the trust layer through this pattern.

Composition with hardware security specialists. Hardware security IP licensors, hardware Trojan detection specialists, anti-tamper component vendors, and security-oriented open-architecture silicon vendors each provide capabilities that compose with the trust layer at specific points in the deployment lifecycle. Hardware security IP appears in finished silicon under which the trust layer operates. Hardware Trojan detection provides supply-chain integrity verification supporting the trust assumptions the trust layer depends on. Anti-tamper components provide chassis-level integrity that supports the trust layer’s physical-layer threat coverage per Threat Taxonomy Attack Surface 1. Open-architecture security silicon provides reference designs that intersect with the trust layer at the silicon-design phase.

Composition with cryptographic verifiable inference. Cryptographic verifiable-inference systems prove, in software, that a committed model produced a given output. The trust layer roots that claim in the silicon that executed it and extends coverage to the physical-layer, bus-level, and supply-chain surfaces the cryptographic abstraction does not reach. The composition pattern is: the cryptographic proof establishes the correctness of the computation; the trust layer establishes that the computation ran on attested hardware within an attested physical envelope, and can bind the software proof to that hardware-rooted evidence; the combination produces a verifiable-inference claim that is both mathematically sound and physically rooted. The zkML, proof-of-inference, and agent-action attestation operators surveyed in Section 4.5, including EZKL, Lagrange Labs, Inference Labs, Nexus, Zerker Labs, HUMAN Security, and Keycard, compose with the trust layer through this pattern.

Composition with standards bodies and sovereign programs. Standards-body and sovereign-program activity produces the technical specifications, procurement frameworks, and policy environments within which the trust layer is deployed. The composition pattern is bidirectional: standards codification establishes the requirements the

trust layer is designed to satisfy; trust-layer deployment validates and refines those requirements. The trust layer composes with EU AI Act conformity assessment under Articles 40 and 43, with FY2026 NDAA Section 1513 framework development, with ISO/IEC 27090 AI security guidance, with NATO STANAG AI trust frameworks, with NIST AI RMF and Cybersecurity Framework implementations, and with allied sovereign frameworks across Five Eyes, Gulf Cooperation Council, and adjacent member-state contexts.

The composition pattern across all six is consistent: the trust layer addresses the structural property each adjacent category does not address from its own layer position. The categories together produce defense-in-depth trust posture; no single category, including the trust layer for AI, is sufficient on its own.

9. What This Map Is Not

The map is explicit about what it does not provide.

It is not a competitive ranking. No operator is characterized as superior or inferior to any other. Operators are situated by structural layer position and by threat-class coverage; commercial evaluation requires deployment-specific analysis that the map intentionally does not provide.

It is not a procurement guide. Procurement teams evaluating specific deployments require analysis of specific operator implementations against specific deployment requirements. The map provides vocabulary for that analysis; it does not substitute for it.

It is not exhaustive. The category landscape contains more operators than are surveyed. Inclusion criteria favor structural relevance over commercial prominence; some commercially significant operators are not separately enumerated because their structural role is well-represented by other surveyed operators. The map will be updated as the category landscape evolves.

It is not a claim about Ziru Labs' relative positioning. The Layer 3 observation that Ziru Labs operates at Layer 3 is structural. The map does not make competitive claims about Ziru Labs versus other operators at any layer, including any other operators that may be positioned at Layer 3. The accompanying Ziru Labs Capability Posture (Ziru Labs, 2026) documents Ziru Labs' specific coverage and composition posture.

It is not a substitute for threat modeling. Specific deployments require specific threat modeling that combines this map with the Physics-Layer Threat Taxonomy and deployment-specific factors. The map provides landscape vocabulary; the Threat Taxonomy provides threat-class vocabulary; deployment-specific threat modeling combines both.

10. Operator Profile: Ziru Labs

Ziru Labs authored this primer. The company operates at Layer 3 within the trust layer for AI category. The company's specific coverage, composition posture, deployment surface, and operational characteristics are documented in the companion Ziru Labs Capability Posture (Ziru Labs, 2026).

This primer, the accompanying Trust Layer for AI: A Reference Framework for the Category (Ziru Labs, 2026), the AI Infrastructure Stack and the Trust Layer Position: A Reference Primer (Ziru Labs, 2026), the Physics-Layer Threat Taxonomy for AI Infrastructure (Ziru Labs, 2026), the Runtime Verification Gap in Federal AI Deployment: A Reference Primer (Ziru Labs, 2026), and the Ziru Labs Capability Posture (Ziru Labs, 2026) together establish the analytical foundation for the trust layer for AI category. The planned quarterly State of Physics-Layer AI Trust report tracks the category's evolution between major-version updates of these foundational documents. Continued research is published at zirulabs.com/research.

Counterparties evaluating engagement with Ziru Labs route through the Engage pathway at zirulabs.com.

11. Bibliography

References are organized into five categories: academic literature (A), standards and regulatory documents (B), industry and vendor technical documentation (C), historical and analytical reporting (D), and framework integration references (E).

A. ACADEMIC LITERATURE

Bhunia, S., Hsiao, M.S., Banga, M., and Narasimhan, S. (2014). "Hardware Trojan Attacks: Threat Analysis and Countermeasures." *Proceedings of the IEEE*, 102(8), 1229-1247.

Costan, V. and Devadas, S. (2016). "Intel SGX Explained." IACR Cryptology ePrint Archive, Report 2016/086.

Russinovich, M., et al. (2021). "Toward Confidential Cloud Computing." *Communications of the ACM*, 64(6), 54-61.

Tehranipoor, M. and Koushanfar, F. (2010). "A Survey of Hardware Trojan Taxonomy and Detection." *IEEE Design & Test of Computers*, 27(1), 10-25.

Wang, Y. (2025). "Zero-Knowledge Proof Based Verifiable Inference of Models." arXiv:2511.19902.

"Lightweight Cryptographic Proofs of Inference" (2026). IACR Cryptology ePrint Archive, Report 2026/541; to appear, IEEE Symposium on Secure and Trustworthy Machine Learning (SaTML).

B. STANDARDS AND REGULATORY DOCUMENTS

European Commission (2024). "Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)." Official Journal of the European Union.

International Organization for Standardization / International Electrotechnical Commission (2023). "ISO/IEC 42001:2023, Information technology, Artificial intelligence, Management system."

International Organization for Standardization / International Electrotechnical Commission (2023). "ISO/IEC 23894:2023, Information technology, Artificial intelligence, Guidance on risk management."

International Organization for Standardization / International Electrotechnical Commission (2023). "ISO/IEC 5338:2023, Information technology, Artificial intelligence, AI system life cycle processes."

International Organization for Standardization / International Electrotechnical Commission (2026, Final Draft International Standard). "ISO/IEC 27090, Cybersecurity, Artificial intelligence, Guidance for addressing security threats to artificial intelligence systems."

National Institute of Standards and Technology (2018). "NIST SP 800-193: Platform Firmware Resiliency Guidelines."

National Institute of Standards and Technology (2023). "AI Risk Management Framework (AI RMF 1.0)," NIST AI 100-1.

National Institute of Standards and Technology (2024). "AI RMF Playbook."

National Institute of Standards and Technology (2024). "The NIST Cybersecurity Framework (CSF) 2.0," NIST CSWP 29.

Republic of Korea Government (2024). “AI Basic Act and accompanying AI governance framework.”

Singapore Infocomm Media Development Authority (2024). “Model AI Governance Framework.”

Trusted Computing Group (current). “TPM 2.0 Library Specification” and “DICE Architecture.”

C. INDUSTRY AND VENDOR TECHNICAL DOCUMENTATION

Advanced Micro Devices (2024). “AMD SEV-SNP: Strengthening VM Isolation” and accompanying technical documentation.

Amazon Web Services (current). “AWS Nitro System Technical Documentation” and “AWS Nitro Enclaves Documentation.”

Analog Devices (current). Tamper detection integrated circuit product documentation.

Anjuna (current). Confidential computing deployment platform documentation.

Apple Inc. (2024). “Apple Platform Security Guide.”

ARM Holdings (current). “ARM TrustZone Technology Overview” and “ARM Confidential Compute Architecture Specification.”

Cadence Design Systems (current). Security IP product documentation.

CalypsoAI (current). Platform technical documentation.

Cisco Systems (2024). “Cisco Acquisition of Robust Intelligence” announcement and integration documentation.

Cloud Security Alliance (current). “Cloud Controls Matrix” and adjacent guidance.

Coalition for Content Provenance and Authenticity (2024). “C2PA Technical Specification.”

Confidential Computing Consortium (current). Technical documentation and member working group output, including Open Enclave SDK and Gramine library OS project documentation.

Cranium (current). Platform technical documentation.

Credo AI (current). Platform technical documentation.

Cycuity (current). Hardware security verification platform documentation.

Edgeless Systems (current). Constellation, Contrast, and adjacent project documentation.

EZKL (current). Zero-knowledge machine-learning proving-framework documentation.

Fiddler (current). AI observability platform documentation.

Fortanix (current). Confidential computing and hardware security module platform documentation.

HiddenLayer (current). AI security platform documentation.

Holistic AI (current). Platform technical documentation.

HUMAN Security (2025). “AgenticTrust” platform documentation and the open-source “HUMAN Verified AI Agent” project (HTTP Message Signatures, RFC 9421).

IBM Corporation (current). “watsonx.governance” platform documentation.

Inference Labs (current). “Proof of Inference” protocol documentation.

Intel Corporation (2024). “Intel Trust Domain Extensions (TDX) Whitepaper” and accompanying technical documentation.

Keycard (2025). Agent identity and access platform documentation (dynamic identity-bound, task-scoped credentials; agent-action audit trails).

Lagrange Labs (current). “DeepProve” zero-knowledge inference documentation.

Lakera (current). “Lakera Guard” platform documentation and “Gandalf” prompt-injection evaluation environment.

Microchip Technology (current). Anti-tamper product documentation.

Microsoft Corporation (2024). “Microsoft Pluton Security Processor” documentation; “Azure Confidential Computing” documentation.

MITRE Corporation (2024). “ATT&CK Framework, Enterprise Matrix,” Version 14, and “ATLAS: Adversarial Threat Landscape for Artificial-Intelligence Systems.”

Nexus (current). Verifiable-computation infrastructure and “Verifiable AI” documentation.

NVIDIA Corporation (2024). “NVIDIA Confidential Computing Deployment Guide” and “NVIDIA H100 Tensor Core GPU Architecture” technical whitepapers.

OpenTitan Project (current). Open-source silicon root-of-trust reference design documentation, hosted by lowRISC consortium.

OWASP Foundation (2025). “OWASP Top 10 for Large Language Model Applications v2.0.”

Palo Alto Networks (2025). “Palo Alto Networks Acquisition of Protect AI” announcement and integration documentation.

Rambus (current). Security IP product documentation.

SiFive (current). “SiFive Shield” platform documentation.

Synopsys (current). DesignWare Security IP product documentation.

Zerker Labs (2026). “Treeship” open-source trust-receipt project documentation and the Zerker synthetic-media analysis platform; SRI International spinout building on the SRI OLIVE platform.

D. HISTORICAL AND ANALYTICAL REPORTING

Ziru Labs (2026). *The Trust Layer for AI: A Reference Framework for the Category, v1.0.* Published at zirulabs.com/research.

Ziru Labs (2026). *The AI Infrastructure Stack and the Trust Layer Position: A Reference Primer, v1.0.* Published at zirulabs.com/research.

Ziru Labs (2026). *The Physics-Layer Threat Taxonomy for AI Infrastructure: A Reference Framework, v1.0.* Published at zirulabs.com/research.

Ziru Labs (2026). *The Runtime Verification Gap in Federal AI Deployment: A Reference Primer, v1.0.* Published at zirulabs.com/research.

Ziru Labs (2026). *Ziru Labs Capability Posture: A Reference Primer, v1.0.* Published at zirulabs.com/research.

E. FRAMEWORK INTEGRATION REFERENCES

Defense Advanced Research Projects Agency. Program documentation for Trust in Integrated Circuits, Supply Chain Hardware Integrity for Electronics Defense, and Automatic Implementation of Secure Silicon.

Intelligence Advanced Research Projects Activity. TrojAI program documentation.

U.S. Department of Defense Chief Digital and Artificial Intelligence Office (current). Departmental AI deployment guidance.

UK AI Security Institute (current). Institute charter and research program documentation.

UK National Cyber Security Centre (current). “Government Security Classifications Policy” and AI deployment guidance.

European Union Agency for Cybersecurity (current). AI threat landscape and cybersecurity guidance.

France National Cybersecurity Agency (ANSSI) (current). National cybersecurity guidance baseline.

Germany Federal Office for Information Security (BSI) (current). National cybersecurity guidance baseline.

Australia Australian Signals Directorate (current). National cybersecurity guidance baseline.

Canada Canadian Centre for Cyber Security (current). National cybersecurity guidance baseline.

Acknowledgments

This primer draws on the technical documentation, public regulatory and standards-body output, and the publicly disclosed activity of the operators surveyed. Specific acknowledgments to the silicon-vendor confidential computing technical communities at NVIDIA, Intel, AMD, and ARM; to the Confidential Computing Consortium and Trusted Computing Group technical working groups; to the AI security and AI governance research communities producing the operational platforms surveyed in Section 4.3; to the standards-body working groups at NIST, ISO/IEC JTC 1/SC 42, ISO/IEC JTC 1/SC 27, CEN-CENELEC JTC 21, and adjacent bodies; to the

sovereign-program research communities at DARPA, IARPA, and allied program offices; and to the founding team's prior working experience across U.S. federal classified AI, signals intelligence, cryptographic warfare, financial services, and AI infrastructure. Initial draft review provided by the Ziru Labs founding team. Any errors in characterization are solely Ziru Labs' responsibility.

Citation

Ziru Labs. *The Trust Layer Category Map: A Reference Primer, v1.0.* Published at zirulabs.com/research.

