ZIRU LABS · REFERENCE FRAMEWORK

The Trust Layer for AI

A Reference Framework for the Category

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 for AI: A Reference Framework for the Category*, v1.0. Published at zirulabs.com/research.

ABSTRACT

> This paper establishes a reference framework for the trust layer for AI as a category of foundational substrate in the artificial intelligence infrastructure stack. The trust layer is the hardware-rooted substrate at which properties of AI computation become cryptographically verifiable. It occupies a structural position in the AI infrastructure stack analogous to the positions public-key cryptography, TCP/IP, the GPS civilian signal, and the ARM instruction set architecture occupy in their respective stacks.
>
> Foundational substrate categories are distinguishable from commoditized product categories by analytical criteria that the technology and economics literatures have developed across the past four decades. This paper names the trust layer for AI as a category, situates it in the AI infrastructure stack, applies the diagnostic criteria for foundational substrate categories to it, surveys the application domains it supports, and identifies the adjacent categories with which it composes additively.
>
> The framework is intended as a living reference. The accompanying Physics-Layer Threat Taxonomy for AI Infrastructure (Ziru Labs, 2026) catalogs the threats the substrate addresses, the AI Infrastructure Stack and the Trust Layer Position: A Reference Primer (Ziru Labs, 2026) concretizes the substrate's position in the AI infrastructure stack, the Trust Layer Category Map: A Reference Primer (Ziru Labs, 2026) maps the operator landscape adjacent to the category, the Runtime Verification Gap in Federal AI Deployment: A Reference Primer (Ziru Labs, 2026) concretizes the substrate in the federal classified-deployment context, and the Ziru Labs Capability Posture: A Reference Primer (Ziru Labs, 2026) documents Ziru Labs' specific posture within the category. The planned quarterly State of Physics-Layer AI Trust report tracks the category's evolution between major-version updates of this framework.
>
> Ziru Labs authored this framework because the company is building an operational response to the category, and the work of building it is what surfaced the category as a distinct object of analysis. The framework is offered as a reference for the ecosystem rather than as a description of any single product: it does not depend on Ziru Labs' specific architecture, and it is written to remain useful to silicon vendors, platform vendors, standards bodies, and other operators who reach the same layer by different routes. Section 7 states the company's position within the category directly.

1. Why the Trust Layer for AI Is a Category

1.1 The Distinction Between Product Categories and Substrate Categories

The technology and economics literatures distinguish two kinds of category in technology infrastructure. A **product category** consists of products that compete within an existing market on price, feature set, and distribution. A **substrate category** consists of a single foundational layer that enables applications above it across multiple markets, where the substrate's value derives from being structurally required for the applications rather than from competing within an established market (David and Greenstein, 1990; Bresnahan and Trajtenberg, 1995; Perez, 2002).

Substrate categories share five characteristics that distinguish them from product categories. They establish a new capability rather than improve an existing one. They occupy their layer by structural necessity rather than by competitive selection. They generate value through application enablement rather than through direct market capture. They produce positive externalities such that value compounds with adoption rather than saturates. They are codified by standards bodies and regulatory frameworks rather than displaced by competitive substitutes (Arthur, 1989; David, 1985).

The value capture profile of a substrate category differs from that of a product category. Product categories generate value approximately proportional to their direct market size. Substrate categories generate enabled economic value at ratios that the general-purpose-technology literature characterizes as spanning orders of magnitude above direct market revenue (Lipsey, Carlaw, and Bekar, 2005), distributed across the applications that depend on the substrate. The four historical analogs developed in Section 2 each exhibit this profile. The enabled-value ratios cited for those analogs are order-of-magnitude estimates offered to illustrate the distribution pattern, not derived market-sizing figures; they are intended as analytical analogy rather than as a forecast of the trust layer's realized value.

The analytical mistake of confusing a substrate category for a product category produces systematic errors in market sizing, pricing, partnership architecture, and strategic positioning. Foundational substrate categories analyzed under the product-category frame produce TAM estimates that systematically underestimate realized economic value, because the bottom-up product-category method cannot capture markets that exist only because the substrate exists (Bresnahan and Trajtenberg, 1995; Helpman, 1998).

1.2 The Trust Layer for AI in the Infrastructure Stack

The trust layer for AI occupies the substrate position in the artificial intelligence infrastructure stack at which properties of AI computation become cryptographically verifiable from hardware up. Every property of AI computation that software-layer mechanisms currently assert (weights used, inputs processed, constraints enforced, outputs produced, timing characteristics, environmental state) can be rooted in the trust layer and converted from assertion to evidence.

The placement is structural and is best read as a position in the trust-dependency ordering of the AI infrastructure stack rather than a claim about physical placement on a die or board. In that dependency ordering, in which each layer depends on the layers below it for its trust properties, the trust layer is rooted in silicon architecture (the

instruction-set-architecture layer at which ARM, x86, and RISC-V operate) and in compute silicon (the AI accelerator layer at which NVIDIA, AMD, Intel, and emerging silicon vendors operate, and in which the silicon-resident hardware roots of trust reside), and sits above them in the ordering because it consumes the execution those layers perform and produces evidence about it. Software layers above compute silicon, including AI runtimes, AI governance platforms, model frameworks, and applications, depend on the substrate for the verifiability properties they currently assert at software register. A hardware-rooted verification substrate cannot sit physically beneath the silicon whose computation it attests; its position in the stack is its position in the verification-dependency chain.

The substrate must be hardware-rooted because software cannot attest the silicon on which it runs. A software attestation mechanism is part of the substrate being attested; the integrity of the attestation evidence depends on the integrity of the substrate it is generated on. This is a structural property of attestation architecture documented across the trusted computing literature (Trusted Computing Group, 2003; Sailer et al., 2004; NIST, 2018) and is not a limitation of any specific software-attestation implementation. The runtime-verification surface specifically, addressing AI computation properties at the moment of inference rather than at boot or at configuration, requires the substrate to operate continuously at the hardware layer rather than only at measured transitions.

The trust layer for AI is distinguishable from confidential computing, from hardware roots of trust as conventionally defined, and from AI governance platforms. Each of these adjacent categories addresses a property of AI infrastructure security or trust that the trust layer composes with rather than substitutes for. Section 6 of this paper develops the composition relationships explicitly.

1.3 What This Paper Establishes

This paper makes three analytical claims and provides reference material in support of each.

The first claim is that the trust layer for AI is a substrate category in the sense distinguished by the technology and economics literatures. Section 3 applies the diagnostic criteria.

The second claim is that the structural position the trust layer for AI occupies in the AI infrastructure stack is analytically parallel to the positions four historical substrate categories occupied in their respective stacks: public-key cryptography in distributed-communication security, TCP/IP in inter-network communication, the GPS civilian signal in position-dependent applications, and the ARM instruction set architecture in energy-constrained computing. Section 2 develops the historical analogs.

The third claim is that the trust layer category supports application surfaces across security, verification, provenance, compliance, identity, and economic agency. Section 4 surveys the application surfaces.

Sections 5 and 6 situate the trust layer for AI in the standards landscape and in relation to adjacent categories. Section 7 identifies Ziru Labs as the author of the framework and one operator within the category. Section 8 provides the bibliography.

1.4 Relationship to Adjacent Frameworks

The trust layer for AI category composes with several adjacent security and AI governance frameworks already in operational use across the security community. The category is designed to interoperate with these frameworks

rather than substitute for them. Specific composition relationships:

The MITRE ATT&CK Enterprise Framework (MITRE Corporation, 2024) provides the canonical taxonomy of adversarial tactics and techniques in conventional information security. The trust-layer category interoperates with ATT&CK by providing the hardware-layer substrate extension that ATT&CK's enterprise scope does not currently address. A practitioner familiar with ATT&CK can use the trust layer for AI as the AI-infrastructure-specific extension that closes the hardware-layer gap.

The NIST Cybersecurity Framework 2.0 (NIST CSWP 29, 2024) provides the canonical functional framework for cybersecurity programs in U.S. organizations. The trust-layer category populates the Identify function (ID.AM, ID.RA) for AI infrastructure programs and signals where Protect (PR.DS, PR.AC) and Detect (DE.CM, DE.AE) obligations follow.

The NIST AI Risk Management Framework (NIST AI 100-1, 2023) and its accompanying Playbook (NIST, 2024) provide the canonical AI risk framework. The trust-layer category supplies the hardware-layer substrate that supports the AI RMF's GOVERN and MAP functions for AI infrastructure programs.

The OWASP Top 10 for Large Language Model Applications v2.0 (OWASP Foundation, 2025) catalogs application-layer LLM risks. The trust-layer category addresses the hardware-layer concerns that OWASP's application-layer scope does not currently address.

ISO/IEC 42001:2023 (AI management systems) and ISO/IEC 27001:2022 (information security management systems) provide management-system frameworks that the trust-layer category populates at AI infrastructure scope. ISO/IEC 15408:2022 (Common Criteria evaluation framework) is the canonical evaluation framework for sovereign and regulated-industry deployments where physics-layer concerns are evaluated.

FIPS 140-3 (NIST, 2019) and NIST SP 800-53 Rev. 5 (NIST, 2020) provide the cryptographic-module and security-control frameworks under which AI infrastructure components are evaluated for U.S. federal deployment. The trust-layer category is designed to compose with the security-control structure that NIST SP 800-53 establishes.

The composition pattern across all of these frameworks is consistent: the trust-layer category supplies the AI-infrastructure-specific hardware-layer substrate that the adjacent framework's scope does not currently address. The category strengthens rather than displaces the existing reference architecture.

2. Historical Foundations: Substrate Layers in Technology Stacks

Four foundational substrate categories from the past five decades provide the analytical reference class for the trust layer for AI. Each of the four occupied a single foundational layer in its respective technology stack, defined a capability that applications above the layer required structurally, and produced enabled economic value at ratios substantially exceeding direct-market revenue at the layer itself. (The companion Stack Position primer draws on a deliberately different and smaller reference set of three single-source layer-position cases to make a distinct argument about layer consolidation; the two analog sets serve different claims and are not intended to match.)

The pattern is consistent across the four cases. Modest direct-market revenue at inception, decade-scale commercial inflection, applications proliferating above the substrate at each subsequent decade, and enabled economic value distributed across the substrate-dependent application surfaces. The analytical reference class is the foundational-substrate category, not the product-category benchmarks that the substrates themselves are sometimes compared against in commercial-analysis contexts.

2.1 Public-Key Cryptography (1976 to Present)

The Diffie-Hellman key exchange (Diffie and Hellman, 1976) and the RSA algorithm (Rivest, Shamir, and Adleman, 1978) established the asymmetric-cryptography substrate that enables trust between parties without prior shared secret. The substrate occupies a single layer in the distributed-communication stack: above the mathematical primitives at which integer factorization and discrete logarithm hardness assumptions operate, and below the protocols (TLS, SSH, S/MIME, PGP, Bitcoin, every blockchain protocol, Tor, Signal, WireGuard, every modern secure-messaging platform) that depend on it.

Direct market revenue at the public-key cryptography layer itself is modest. The combined revenue of the dedicated public-key infrastructure industry (certificate authorities, hardware security modules, PKI software, identity management built on public-key primitives) is a few billion dollars annually. Enabled economic value across the applications that depend on the substrate is measured in trillions per year cumulatively, including digital commerce, secure messaging, cryptocurrency, identity verification, and the foundational integrity of nearly every contemporary digital transaction. The ratio of enabled value to direct market is on the order of one thousand to ten thousand.

The ratios cited across the four analogs in this section are not computed on a common basis and are not directly comparable across cases. Public-key cryptography is the canonical reference case for the substrate-category pattern because the four-decade horizon now allows the value distribution to be fully observed. The applications could not have been anticipated at the substrate’s inception in 1976. The category became indispensable because every application above it required it structurally and no software-layer substitute proved adequate.

2.2 The Internet Protocol Suite (1974 to Present)

The Cerf and Kahn architecture (Cerf and Kahn, 1974) and its codification in RFC 791 and RFC 793 (Postel, 1981) established the inter-network communication substrate that enables addressing and packet routing across heterogeneous physical networks. The substrate occupies a single layer in the networking stack: above the physical and data-link layers at which Ethernet, Wi-Fi, optical transmission, cellular, and satellite operate, and below the application protocols (HTTP, SMTP, FTP, SSH, every transport protocol above TCP and UDP, every web application, every cloud service, every API) that depend on it.

The internet protocol suite was a sufficiently unlikely outcome in the 1970s that the OSI Reference Model (ISO/IEC, 1984) was expected to be the substrate at this layer through most of the early commercial-deployment period. The TCP/IP substrate displaced OSI through a combination of operational adoption, simplicity of implementation, and the strategic decision to make the substrate openly specified rather than proprietary (Abbate, 1999). The historical contingency matters analytically: substrate categories do not always emerge through technical superiority alone, and the codification window in which the category’s reference implementation is established is structurally important.

Direct market revenue at the internet protocol layer itself is modest. Networking equipment vendors, internet service providers, and the protocol-implementation industry generate aggregate revenue in the tens of billions annually. Enabled economic value across the applications that depend on the substrate is measured in the tens of trillions annually, including approximately the entirety of the modern digital economy. The most valuable technology companies on earth at the time of this writing all rest on infrastructure for which TCP/IP is structurally required. The ratio of enabled value to direct market is on the order of two hundred to three hundred.

2.3 The Global Positioning System Civilian Signal (1983 to Present)

The Global Positioning System civilian signal, opened to civilian use by presidential directive in 1983 and made unrestricted by Selective Availability removal in 2000 (Federal Aviation Administration, 2000), established the position-verification substrate that enables location-dependent applications. The substrate occupies a single layer in the position-dependent application stack: above radio signal propagation, satellite transmission, and atomic time standards at which the underlying physical phenomena operate, and below the mapping, navigation, logistics, ride-sharing, food delivery, precision agriculture, autonomous vehicles, and emergency response applications that depend on verified position.

Direct market revenue at the GPS receiver and location-services layer, including GPS receivers, signal augmentation systems, and the chipset and services market built directly on the signal, is on the order of tens to over one hundred billion dollars annually, while the satellite constellation itself is publicly funded rather than commercially monetized. Enabled economic value across the applications that depend on the substrate is measured in trillions annually, including the ride-sharing economy, modern logistics and supply chain, precision agriculture, the emerging autonomous vehicle deployment, and the location-dependent services that have restructured retail, real estate, and emergency response (National Research Council, 1995; O’Connor et al., 2019). The ratio of enabled value to direct market is on the order of ten to twenty.

GPS is the canonical reference case for the substrate-category pattern in which the substrate itself is publicly funded but the enabled applications are predominantly commercial. The substrate category’s value capture distributes asymmetrically across the application surfaces it enables, with the publicly-funded substrate operator receiving none of the application-surface value capture by design.

2.4 The ARM Instruction Set Architecture (1991 to Present)

The ARM instruction set architecture, established as a licensing-model substrate by ARM Holdings in 1991 (Stokes, 2007; Garnsey and Heffernan, 2005), established the energy-constrained computation substrate that enables mobile, embedded, and increasingly general-purpose computing in operational envelopes for which the x86 instruction set architecture is unsuitable. The substrate occupies a single layer in the mobile and embedded computing stack: above silicon transistor implementation at which TSMC and Samsung Foundry operate, and below the operating systems, drivers, and applications that depend on the instruction set architecture for their compilation target.

Direct market revenue at the ARM ISA layer itself, measured by ARM Holdings revenue, is approximately three to four billion dollars annually at the time of this writing. Enabled economic value across the applications that depend on the substrate, measured by the aggregate output of devices built on ARM-licensed designs, is measured in the trillions annually, including approximately the entirety of the mobile computing industry, large portions of

the embedded computing industry, the Apple Silicon design family, the AWS Graviton family, and an accelerating share of general-purpose data-center computing. The ratio of enabled value to direct market is on the order of one hundred to three hundred.

ARM Holdings demonstrates the substrate-category value-capture pattern in a public-equity context. ARM Holdings is valued, at the time of this writing, in a market capitalization range above one hundred billion dollars on annual revenue of approximately three to four billion, a multiple that reflects substrate-category economics rather than product-category economics (ARM Holdings, 2023). The valuation pattern is comparable to other public substrate-category operators including VeriSign (in its historical PKI-substrate role) and Qualcomm (in its CDMA and 5G substrate roles).

2.5 The Pattern

The four historical analogs share a structural pattern that the technology and economics literatures have characterized as the general-purpose-technology pattern (Bresnahan and Trajtenberg, 1995; Helpman, 1998; Lipsey, Carlaw, and Bekar, 2005). Each substrate occupied a single foundational layer. Each layer’s applications proliferated above it across decades. Each had modest direct-market revenue at inception. Each produced enabled economic value at ratios substantially exceeding direct-market revenue. Each was codified, over its first decade in market, by standards bodies and by regulatory frameworks that have proven structurally resistant to subsequent displacement.

The trust layer for AI occupies the analogous structural position in the AI infrastructure stack. The remainder of this paper develops the analytical case that the structural parallel produces the category-determining properties documented in the four historical cases.

3. Diagnostic Criteria for Foundational Substrate Categories

The technology and economics literatures have developed diagnostic criteria for distinguishing foundational substrate categories from commoditized product categories. The five characteristics introduced in Section 1.1 resolve, for diagnostic purposes, into seven discrete tests. The criteria below synthesize the literature into these seven tests, each of which the four historical analogs in Section 2 satisfied at their respective inceptions. A category that satisfies all seven is analytically eligible for the substrate-category value-capture pattern. Eligibility is necessary but not sufficient; timing, execution, and competitive dynamics determine whether a category that is eligible realizes the pattern.

This section applies the seven criteria to the trust layer for AI as a category. The application demonstrates that the category is analytically eligible for the substrate-category pattern. Whether any specific operator within the category realizes the pattern is a separate question, treated in Section 7.

3.1 Criterion 1: New Substrate Capability or Trust Model

The category establishes a new capability rather than improving an existing one. The historical analogs each introduced a capability that did not exist in the preceding stack: asymmetric trust without shared secret (public-

key cryptography); packet-switched inter-network communication (TCP/IP); civilian-accessible verified position (GPS); energy-constrained instruction-set execution at general-purpose performance (ARM).

The trust layer for AI introduces a new capability: cryptographic evidence at the moment of inference that AI computation operated under specific configuration, on specific hardware, against specific inputs, producing specific outputs, all rooted in hardware properties that the inference workload cannot manipulate from software. This capability does not exist in the preceding AI infrastructure stack. Confidential computing addresses workload confidentiality against software adversaries but does not extend to runtime physical-layer verification (Trusted Computing Group, 2003; Costan and Devadas, 2016). Hardware roots of trust as conventionally defined attest at boot and at configuration transitions but do not produce continuous runtime evidence of AI workload operation (NIST, 2018). AI governance platforms operate at the software layer above the substrate (OWASP Foundation, 2025) and depend on the substrate's integrity for their own evidence integrity.

The trust layer for AI satisfies Criterion 1.

3.2 Criterion 2: Irreducibility to Higher Layers

The category occupies its layer by structural necessity rather than by competitive selection. The historical analogs each occupied positions in their respective stacks that the applications above could not substitute for at higher layers. Software-layer cryptography cannot establish the trust public-key cryptography establishes without the substrate. Application-layer protocols cannot route packets across heterogeneous networks without the substrate TCP/IP provides. Software-layer position estimation cannot verify position without the GPS signal substrate. Software-layer emulation of an instruction set is not the instruction set.

The trust layer for AI is irreducible to higher layers because the integrity of any evidence produced at a software layer above the substrate depends on the integrity of the substrate itself. A software-layer attestation mechanism is part of the substrate being attested; under substrate compromise, the attestation is compromised with it. The structural property is documented across the trusted computing literature and is not a limitation of any specific implementation. The trust layer for AI must be hardware-rooted to perform the function it performs at category level.

The trust layer for AI satisfies Criterion 2.

3.3 Criterion 3: Application Dependence

The category enables applications that require it structurally. The historical analogs each became indispensable to application surfaces above them across decades.

The application surfaces that depend on the trust layer span: hardware-attested compliance frameworks under the EU AI Act (Articles 40 and 43) and successor regulations; cryptographic provenance for AI-generated content under Coalition for Content Provenance and Authenticity specifications (C2PA, 2024); verifiable AI execution in regulated commerce contexts requiring auditable evidence; AI identity infrastructure for federated and machine-to-machine commerce arrangements; sovereign AI deployment frameworks across Five Eyes, NATO, and Gulf Cooperation Council national programs requiring hardware-rooted independence verification; and AI safety mechanisms requiring constraint enforcement that survives software-layer compromise. Section 4 develops these application surfaces in greater detail. Each surface requires the trust-layer substrate to be realized fully; software-

layer substitutes do not produce equivalent evidence under the threat models the applications must address. These are the surfaces the category is positioned to serve as it matures; the regulatory and standards activity driving them is surveyed in Section 5.

The trust layer for AI satisfies Criterion 3.

3.4 Criterion 4: Positive Externalities

The category's value compounds with adoption. The historical analogs each exhibited network and standardization externalities such that each deployment validated the substrate, reduced marginal deployment cost for subsequent adopters, and accelerated standards codification (Arthur, 1989; Katz and Shapiro, 1985).

The trust layer for AI is structured to exhibit the same externality pattern. Each deployment of the substrate produces evidence about the substrate's operational characteristics, reduces integration cost for subsequent silicon-vendor and platform-vendor partnerships, and accelerates standards-body codification by demonstrating operational viability. The category is at an early stage: the externality is a structural property the category will exhibit as deployments accumulate, not a track record this paper claims has already been established. The mechanism is the same one the four historical analogs exhibited at the corresponding stage of their development.

The trust layer for AI satisfies Criterion 4.

3.5 Criterion 5: New Commercial and Institutional Arrangements

The category enables new professions, markets, and commercial arrangements that did not exist in the preceding stack. The historical analogs each generated commercial categories that the substrate enabled: the digital certificate authority industry (public-key cryptography); the internet service provider industry (TCP/IP); the location-based services industry (GPS); the mobile silicon licensing industry (ARM).

The trust layer for AI is positioned to produce identifiable new commercial arrangements: hardware-attested AI auditing professions parallel to the auditing professions that GAAP enabled in financial reporting; AI decision insurance categories that depend on hardware-rooted evidence of inference configuration; model weight intellectual property markets that depend on hardware-rooted protection of weight assets; cross-border AI cooperation arrangements that depend on hardware-rooted trust between counterparties operating under different institutional regimes; sovereign AI licensing categories that depend on hardware-rooted independence verification. Each arrangement requires the trust-layer substrate to be commercially feasible, and each is at an early and largely prospective stage rather than an established market this paper claims already exists.

The trust layer for AI satisfies Criterion 5.

3.6 Criterion 6: Multiplying Integration Surfaces

The category's integration footprint grows with deployment rather than saturates. The historical analogs each exhibited expanding integration surfaces across decades; the integration count for TCP/IP, public-key cryptography, GPS, and ARM today is larger than at any prior measurement point.

The trust layer for AI integration surface expands with each application surface it serves. The same substrate supports security applications, verification applications, provenance applications, compliance applications,

identity applications, and emerging economic-agency applications. Each application surface generates its own integration footprint with silicon vendors, with platform vendors, with regulatory frameworks, and with standards bodies. The integration count grows with each surface added rather than saturating at a single surface.

The trust layer for AI satisfies Criterion 6.

3.7 Criterion 7: Standards Codification

The category is codified by standards bodies and by regulatory frameworks as the reference for the function it performs. The historical analogs each were codified across the first decade of their respective adoption: IETF RFCs for the internet protocol suite; ISO/IEC standards for public-key cryptography; ICAO and IEEE standards for civilian GPS; ARM Architecture Reference Manual for the ARM instruction set architecture.

The trust layer for AI is currently in active codification across multiple standards-and-regulatory bodies. Under the EU AI Act, Articles 40 (harmonised standards) and 43 (conformity assessment procedures) are developing the technical requirements and conformity assessment procedures for high-risk AI systems, within which hardware-rooted attestation is one available mechanism for demonstrating conformity (European Commission, 2024). NATO STANAG work on AI trust is developing equivalent specifications. ISO/IEC 27090 (in final draft as of early 2026), the AI security guidance standard under ISO/IEC JTC 1/SC 27, addresses security threats to AI systems at the software and data layers, including adversarial machine learning, data poisoning, and model extraction; it is an adjacent standard the trust layer composes with rather than a hardware-substrate standard. ISO/IEC JTC 1/SC 42 working groups on AI trustworthiness are developing related specifications including ISO/IEC 23894:2023 (AI risk management) and ISO/IEC 5338:2023 (AI system lifecycle). On the U.S. federal side, FY2026 NDAA Section 1513 (Physical and Cybersecurity Procurement Requirements for Artificial Intelligence Systems, P.L. 119-60) directs procurement-level requirements directly relevant to a hardware-rooted trust substrate, within a broader federal AI policy environment shaped by Executive Order 14179 and Executive Order 14365. The standards codification window for the category is open as of the publication date of this paper, with primary codification activity expected to complete across the eighteen to thirty-six months from publication.

The trust layer for AI satisfies Criterion 7.

3.8 Summary

The trust layer for AI satisfies all seven diagnostic criteria for foundational substrate categories. The category is analytically eligible for the substrate-category value-capture pattern documented in the four historical analogs. Whether any specific operator within the category realizes the pattern depends on the operator’s strategic execution against the standards codification window, the operator’s mechanism-level intellectual property position, the operator’s partnership architecture, and the broader competitive dynamics of the category as it codifies.

4. Applications of the Trust Layer Category

The trust layer category supports application surfaces across six identifiable domains. The domains are described below at category level. Specific commercial configurations within each domain take shape across multi-year

horizons as standards codification, regulatory frameworks, and silicon-vendor integration mature.

4.1 Security

The security application surface is the one closest to deployment and the focus of the company's initial work. Hardware-rooted attestation that AI workloads are operating under authorized configuration, on authorized hardware, against authorized inputs, producing authorized outputs, addresses the threat classes documented in the accompanying Physics-Layer Threat Taxonomy for AI Infrastructure (Ziru Labs, 2026). The security surface includes deployment of AI in classified federal environments at impact levels above the current commercial stack ceiling, deployment of AI in regulated commercial environments under fiduciary and compliance requirements, and deployment of AI in sovereign and edge environments where physical custody of the silicon cannot be continuously assured.

4.2 Verification of Execution

The verification surface establishes cryptographic evidence that an AI workload ran as specified: processed the claimed inputs, enforced the claimed constraints, used the claimed model and weights, and produced the claimed outputs. The verification surface is required by regulatory conformity assessment regimes including those under the EU AI Act (Articles 40 and 43; European Commission, 2024) and successor frameworks, by commercial contract enforcement involving AI-mediated decisions, by safety-critical auditing requirements, and by legal evidence integrity standards. Software-layer verification mechanisms produce evidence that depends on the integrity of the substrate the software runs on; hardware-rooted verification produces evidence at a layer the software cannot fabricate.

4.3 Provenance

The provenance surface establishes cryptographic attribution that an AI output originated from a specific hardware instance at a specific time using a specific model configuration. The provenance surface is required by content authenticity frameworks including the Coalition for Content Provenance and Authenticity specifications (C2PA, 2024), by synthetic media attribution requirements in regulated jurisdictions, by supply chain provenance requirements for AI-mediated decisions, and by emerging civic information infrastructure for distinguishing AI-generated from human-authored content. The provenance surface scales in importance as AI-generated content saturates the information environment to thresholds at which attribution infrastructure becomes economically and politically necessary.

4.4 Compliance

The compliance surface establishes regulatory compliance through hardware-rooted attestation rather than software assertion. The compliance surface is required by EU AI Act high-risk application provisions, by U.S. federal AI procurement frameworks at FIPS 199 moderate and high categorization levels (NIST, 2004), by NIST AI Risk Management Framework GOVERN and MAP function implementations (NIST, 2023), and by equivalent frameworks under development across allied jurisdictions. Hardware-attested compliance produces assurance properties software-asserted compliance frameworks cannot match under threat models that include partial software-layer compromise.

4.5 Identity

The identity surface establishes cryptographically verifiable AI identity for federated AI arrangements, machine-to-machine commerce, and cross-jurisdictional AI cooperation under different institutional regimes. Hardware-rooted identity provides the attestation that an AI instance is the instance it claims to be, operating under the configuration it claims to operate under, in a form that counterparties can verify independently of the operator's own software assertions.

4.6 Economic Agency

The economic-agency surface supports emerging machine-economy infrastructure in which AI agents act as economic counterparties whose commitments require verification beyond their own software assertions. The agency surface composes with the identity surface: verified identity establishes which agent is acting, and agency attestation establishes the authority and constraints under which the agent transacts. Specific commercial configurations of the agency surface depend on the trajectory of AI agent deployment, which is itself subject to ongoing regulatory and operational development across the technology, financial services, and legal sectors.

4.7 Cross-Application Composition

The six application surfaces compose across deployments. A regulated-vertical operator deploying AI for a specific business function may simultaneously require the security surface (against physical-layer threats), the verification surface (for regulatory audit), the provenance surface (for output attribution), the compliance surface (for framework attestation), the identity surface (for federated deployment with counterparties), and the economic-agency surface (for AI-mediated transactions). The trust-layer substrate supports all six surfaces from a single architectural foundation. The application footprint of the substrate at any deployment grows with the number of surfaces the deployment requires.

5. The Trust Layer in the Standards Landscape

The standards and regulatory frameworks that will enclose the trust layer for AI are in active development across multiple bodies as of the publication date of this paper. The category itself is not yet named as a distinct object in any of these workstreams; what is developing is the set of adjacent and enclosing standards within which hardware-rooted attestation is one available mechanism, and within which the category's reference implementation will be established across the next two to three years. The landscape is surveyed below at category level.

5.1 European Frameworks

The European Union AI Act establishes the conformity assessment regime for high-risk AI applications. Article 40 (harmonised standards and standardisation deliverables) provides the presumption-of-conformity route, and Article 43 sets out the conformity assessment procedures themselves (European Commission, 2024). The Article 40 harmonised standards are in development through CEN-CENELEC JTC 21 in response to the European Commission's standardisation request, with technical implementation work proceeding through the European

Commission AI Office. Hardware-rooted attestation is among the technical mechanisms under consideration for compliance demonstration under this framework.

5.2 United States Federal Frameworks

The U.S. federal AI standards landscape is shaped by the broader federal AI policy environment, including Executive Order 14179 (Removing Barriers to American Leadership in Artificial Intelligence) and Executive Order 14365 (Ensuring a National Policy Framework for Artificial Intelligence), which together set the administration's posture toward AI innovation and a uniform national regulatory framework. The procurement- and security-specific requirements most directly relevant to a hardware-rooted trust substrate sit within the National Institute of Standards and Technology AI Risk Management Framework and accompanying Playbook (NIST, 2023; NIST, 2024), the Department of Defense framework development activities under FY2026 National Defense Authorization Act Section 1513, and the General Services Administration FedRAMP AI procurement requirements. Hardware-rooted attestation is among the technical mechanisms relevant to the security and procurement requirements developing in these activities.

5.3 Allied and International Frameworks

The North Atlantic Treaty Organization is developing Standardization Agreement (STANAG) frameworks for AI trust across member states. The Five Eyes signals intelligence community is developing equivalent frameworks for classified AI deployment. The Gulf Cooperation Council member states are developing sovereign AI frameworks that include hardware-rooted trust requirements. The Organisation for Economic Co-operation and Development AI Principles (OECD, 2019, updated 2024) and the Bletchley and Seoul successor frameworks (United Kingdom Government, 2023; Republic of Korea Government, 2024) establish high-level principles that the technical standards bodies are operationalizing.

5.4 Industry Standards Bodies

ISO/IEC 27090 (in final draft as of early 2026) is the AI-system security-threat guidance standard under ISO/IEC JTC 1/SC 27. The ISO/IEC JTC 1/SC 42 working groups are developing AI trustworthiness specifications including ISO/IEC 23894:2023 (AI risk management) and ISO/IEC 5338:2023 (AI system lifecycle processes). The Trusted Computing Group is developing trusted-platform specifications applicable to AI infrastructure. The Coalition for Content Provenance and Authenticity is developing content authenticity standards that intersect with the provenance application surface (C2PA, 2024).

5.5 The Codification Window

Standards codification at the substrate-category layer typically occurs in the first decade of a category and produces reference architectures that are structurally resistant to subsequent displacement. The standards codification window for the trust layer for AI category is currently open. Primary codification activity is expected to complete across the eighteen to thirty-six months from the publication date of this paper. The codification window determines the reference architecture under which the category's deployment proceeds.

6. Adjacent Categories: Composition with Existing Layers

The trust layer for AI composes additively with adjacent categories rather than substituting for them. Three adjacent categories are identifiable, each operating at a layer the trust layer for AI composes with rather than replaces.

6.1 Confidential Computing

Confidential computing addresses workload confidentiality against software-layer adversaries with privileged access to the host operating system (Costan and Devadas, 2016; Russinovich et al., 2021). The category includes Intel TDX, AMD SEV-SNP, NVIDIA Confidential Computing, ARM CCA, and the broader trusted-execution-environment architecture as developed by the silicon vendors and the Confidential Computing Consortium.

The trust layer for AI composes with confidential computing by extending the attested surface below the trusted execution environment boundary. Confidential computing protects the running workload within the accelerator die against software-layer compromise; the trust layer for AI attests properties of the workload that operate outside the trusted execution environment boundary, including physical-layer threats, bus-level threats, supply-chain threats, and runtime properties that the trusted execution environment does not address. The two categories together cover the AI workload from inside the trusted execution environment outward through the physical-layer and supply-chain surfaces.

6.2 Hardware Roots of Trust

Hardware roots of trust as conventionally defined establish cryptographic identity for hardware platforms and attest measured-boot integrity (Trusted Computing Group, 2003; NIST, 2018). The category includes the Trusted Platform Module (TPM), the Apple Secure Enclave, the Google Titan family, the Microsoft Pluton processor, and the equivalent silicon-vendor implementations across the platform-security industry.

The trust layer for AI composes with hardware roots of trust by extending the attested surface from boot-time and configuration-time measurement to continuous runtime evidence of AI workload operation. Hardware roots of trust establish the identity and configuration of the platform at trusted transitions; the trust layer for AI produces evidence of workload behavior between transitions, addressing the runtime-verification gap that the measured-boot architecture does not cover. The two categories together cover the platform from boot integrity through continuous runtime operation.

6.3 AI Governance Platforms

AI governance platforms operate at the software layer, monitoring, filtering, and analyzing AI behavior across deployments (OWASP Foundation, 2025). The category includes the emerging commercial AI governance industry providing monitoring, observability, alerting, content moderation, and policy enforcement at software register.

The trust layer for AI composes with AI governance platforms by providing hardware-rooted attestation that the platforms' policy enforcement actually held during inference under the threat conditions the enforcement is intended to address. AI governance platforms provide the broad monitoring, analytics, and alerting layer across

the AI estate; the trust layer for AI provides hardware-rooted enforcement and evidence at the inference layer. The two categories together cover the AI estate from operational dashboard through silicon, with the trust layer addressing the threats the governance platforms cannot enforce against under software-layer compromise.

6.4 The Composition Pattern

The trust layer category, confidential computing, hardware roots of trust, and AI governance platforms each address a property of AI infrastructure trust that the others do not. The composition pattern across the four categories is additive rather than substitutive: deployments requiring rigorous AI trust posture stack all four categories in defense-in-depth architectures. The trust layer category occupies the specific position in the stack at which hardware-rooted attestation operates continuously across the runtime envelope and below the trusted execution environment boundary.

7. Operator Profile: Ziru Labs

Ziru Labs authored this framework. The company is an early-stage operator within the trust layer for AI category, building a hardware-rooted substrate with a minimum working prototype targeted for the second half of 2026 and an intellectual property position established through filed provisional patent applications. The company has not shipped a commercial product. The framework is offered in a convening posture: Ziru Labs is proposing a reference for the category and inviting silicon vendors, platform vendors, standards bodies, regulators, and other operators to engage with it, refine it, and build against it.

The company’s authority to propose the framework rests on the work of building the substrate and on the background of the team doing that work. The founding team’s prior experience spans U.S. federal classified AI, signals intelligence, and cryptographic warfare, including senior service in the U.S. intelligence community and the U.S. Navy, alongside experience in financial services and AI infrastructure. Future versions of this framework and its companion documents will carry named technical contributions from members of the team and from external contributors as the category’s reference material develops.

This framework, the accompanying Physics-Layer Threat Taxonomy for AI Infrastructure (Ziru Labs, 2026), the Runtime Verification Gap in Federal AI Deployment: A Reference Primer (Ziru Labs, 2026), the AI Infrastructure Stack and the Trust Layer Position: A Reference Primer (Ziru Labs, 2026), and the Trust Layer Category Map: A Reference Primer (Ziru Labs, 2026) together establish the analytical foundation for the category. The Ziru Labs Capability Posture: A Reference Primer (Ziru Labs, 2026) documents the company’s specific posture within the category at category level, describing what the substrate addresses without disclosing implementation detail held under intellectual property protection and export-control discipline. The planned quarterly State of Physics-Layer AI Trust report tracks the category’s evolution between major-version updates of these foundational documents. Ziru Labs intends to contribute to the standards work codifying the category and is preparing submissions to relevant bodies. Continued research is published at zirulabs.com/research.

Ziru Labs is one operator within the category, not its sole or definitive occupant. Other operators are expected to occupy adjacent positions across silicon-vendor partnerships, platform-vendor partnerships, standards-body participation, and sovereign-program engagements. The category is structurally large enough to support a

substantive ecosystem of operators; this framework is designed to serve as a shared analytical reference for that ecosystem rather than as the strategic positioning of any single operator within it, Ziru Labs included.

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

8. Bibliography

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

A. ACADEMIC LITERATURE

Abbate, J. (1999). *Inventing the Internet.* MIT Press.

Arthur, W.B. (1989). "Competing Technologies, Increasing Returns, and Lock-In by Historical Events." *The Economic Journal*, 99(394), 116-131.

Bresnahan, T.F. and Trajtenberg, M. (1995). "General Purpose Technologies: Engines of Growth?" *Journal of Econometrics*, 65(1), 83-108.

Cerf, V.G. and Kahn, R.E. (1974). "A Protocol for Packet Network Intercommunication." *IEEE Transactions on Communications*, 22(5), 637-648.

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

David, P.A. (1985). "Clio and the Economics of QWERTY." *American Economic Review*, 75(2), 332-337.

David, P.A. and Greenstein, S. (1990). "The Economics of Compatibility Standards: An Introduction to Recent Research." *Economics of Innovation and New Technology*, 1(1-2), 3-41.

Diffie, W. and Hellman, M.E. (1976). "New Directions in Cryptography." *IEEE Transactions on Information Theory*, 22(6), 644-654.

Garnsey, E. and Heffernan, P. (2005). "High-Technology Clustering through Spin-out and Attraction: The Cambridge Case." *Regional Studies*, 39(8), 1127-1144.

Helpman, E. (Ed.) (1998). *General Purpose Technologies and Economic Growth.* MIT Press.

Katz, M.L. and Shapiro, C. (1985). "Network Externalities, Competition, and Compatibility." *American Economic Review*, 75(3), 424-440.

Lipsey, R.G., Carlaw, K.I., and Bekar, C.T. (2005). *Economic Transformations: General Purpose Technologies and Long-Term Economic Growth.* Oxford University Press.

Perez, C. (2002). *Technological Revolutions and Financial Capital: The Dynamics of Bubbles and Golden Ages.* Edward Elgar Publishing.

Postel, J. (1981). "Internet Protocol." RFC 791, Internet Engineering Task Force; and "Transmission Control Protocol." RFC 793, Internet Engineering Task Force.

Rivest, R.L., Shamir, A., and Adleman, L. (1978). "A Method for Obtaining Digital Signatures and Public-Key Cryptosystems." *Communications of the ACM*, 21(2), 120-126.

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

Sailer, R., Zhang, X., Jaeger, T., and van Doorn, L. (2004). "Design and Implementation of a TCG-based Integrity Measurement Architecture." Proceedings of the USENIX Security Symposium.

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.

Executive Order 14179 (2025). “Removing Barriers to American Leadership in Artificial Intelligence.” The White House, signed 23 January 2025; 90 FR 8741.

Executive Order 14365 (2025). “Ensuring a National Policy Framework for Artificial Intelligence.” The White House, signed 11 December 2025; 90 FR 58499.

Federal Aviation Administration (2000). “Selective Availability Removal.” Federal Register notice and accompanying Department of Defense and Department of Transportation announcements.

FY2026 National Defense Authorization Act, Section 1513 (2025). “Physical and Cybersecurity Procurement Requirements for Artificial Intelligence Systems.” Public Law 119-60.

International Organization for Standardization / International Electrotechnical Commission (1984). “ISO/IEC 7498-1, Information processing systems, Open Systems Interconnection, Basic Reference Model.”

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.”

ISO/IEC (2023). “ISO/IEC 42001:2023, Information technology, Artificial intelligence, Management system.”

National Institute of Standards and Technology (2004). “FIPS 199: Standards for Security Categorization of Federal Information and Information 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.”

Organisation for Economic Co-operation and Development (2019, updated 2024). “OECD AI Principles.”

Republic of Korea Government (2024). “Seoul Declaration for Safe, Innovative and Inclusive AI,” AI Seoul Summit.

Trusted Computing Group (2003 to present). “TCG TPM Specifications” and successor specifications.

United Kingdom Government (2023). “The Bletchley Declaration by Countries Attending the AI Safety Summit.”

C. INDUSTRY AND CONSORTIUM TECHNICAL DOCUMENTATION

ARM Holdings (2023). “ARM Holdings plc Initial Public Offering Prospectus and Annual Report.” Filed with the U.S. Securities and Exchange Commission.

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

Confidential Computing Consortium (current). “Technical Documentation and Member Working Group Output.”

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

Trusted Computing Group (current). “Platform Configuration Register (PCR) usage guidance.”

D. HISTORICAL AND ANALYTICAL REPORTING

National Research Council (1995). *The Global Positioning System: A Shared National Asset.* National Academy Press.

O’Connor, A.C., Gallaher, M.P., Clark-Sutton, K., Lapidus, D., Oliver, Z., Scott, T.J., Wood, D.W., and Brown, E.G. (2019). *Economic Benefits of the Global Positioning System (GPS).* RTI International, sponsored by the National Institute of Standards and Technology (NIST), U.S. Department of Commerce.

Stokes, J. (2007). *Inside the Machine: An Illustrated Introduction to Microprocessors and Computer Architecture.* No Starch Press.

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

Ziru Labs (2026). *The Trust Layer Category Map: 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

MITRE Corporation (2024). "ATT&CK Framework, Enterprise Matrix," Version 14.

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

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

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

ISO/IEC (2023). "ISO/IEC 42001:2023, Information technology, Artificial intelligence, Management system" (full citation in Section B).

ISO/IEC (2022). "ISO/IEC 27001:2022, Information security, cybersecurity and privacy protection, Information security management systems, Requirements."

ISO/IEC (2022). "ISO/IEC 15408:2022, Information security, cybersecurity and privacy protection, Evaluation criteria for IT security."

IEC (2018-ongoing). "IEC 62443, Security for industrial automation and control systems."

National Institute of Standards and Technology (2019). "FIPS 140-3: Security Requirements for Cryptographic Modules."

Acknowledgments

This framework draws on the academic literature on general-purpose technologies, the technology and economics literatures on substrate categories and standardization, the trusted computing literature, the AI risk and governance literatures, and the regulatory and standards-body work currently active across the category. Specific acknowledgments to the academic communities at the IEEE Symposium on Security and Privacy, the USENIX Security Symposium, the ACM Conference on Computer and Communications Security, and the IEEE International Symposium on Hardware Oriented Security and Trust; to the standards-body working groups at NIST, ISO/IEC JTC 1/SC 42, ISO/IEC JTC 1/SC 27, and the Confidential Computing Consortium; to the regulatory and policy communities developing the EU AI Act conformity assessment framework under Articles 40 and 43 and the NATO STANAG AI trust frameworks; and to the prior working communities of the founding team 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 for AI: A Reference Framework for the Category, v1.0.* Published at zirulabs.com/research.

