ZIRU LABS · STACK POSITION PRIMER

The AI Infrastructure Stack and the Trust Layer Position

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

ABSTRACT

> The AI infrastructure stack consists of layers that descend from end-user applications through software runtime, model framework, accelerator silicon, and silicon fabrication. The trust layer for AI is the substrate position in the stack at which properties of AI computation become cryptographically verifiable from hardware up. This primer maps the eleven layers of the AI infrastructure stack at the level of category granularity adopted by current standards bodies and industry references, identifies where the trust layer sits in the stack, and surveys what changes across the adjacent layers when the trust layer position is occupied.

> The primer is a stack-position complement to The Trust Layer for AI: A Reference Framework for the Category (Ziru Labs, 2026). The Reference Framework establishes the trust layer as a substrate category through diagnostic criteria and category-formation analysis. The present primer establishes the trust layer's position in the actual AI infrastructure stack through layer enumeration, adjacent-layer mapping, and cascading-consequences analysis. The two documents are designed to be read together; either independently is coherent but the pair is stronger than either alone.

> The companion Physics-Layer Threat Taxonomy for AI Infrastructure (Ziru Labs, 2026) catalogs the threat classes the substrate addresses. The companion Runtime Verification Gap in Federal AI Deployment: A Reference Primer (Ziru Labs, 2026) concretizes the substrate in the federal classified-deployment context.

> Ziru Labs authored this primer as one operator within the trust layer for AI category. The primer maps the stack at category level and is offered as a reference rather than as a description of any single product; it does not depend on Ziru Labs' specific architecture. Where Ziru Labs technology is referenced, it is referenced as one operational response to the substrate.

1. The AI Infrastructure Stack

The AI infrastructure stack at the level of category granularity adopted by current standards bodies, industry references, and academic analysis consists of approximately eleven distinguishable layers. The exact number varies by source, with some references collapsing or subdividing specific layers depending on analytical purpose. The eleven-layer framing below corresponds to the operational granularity at which the trust layer category is most legibly situated.

Each layer has a distinct function, a distinct set of dominant participants, and a distinct industry-recognized category mapping. The eleven layers, in stack order from physical silicon at the bottom through governance at the top:

Layer	Function	Dominant Participants and Categories
1. Fabrication	Physical silicon manufacturing	TSMC, Samsung Foundry, Intel Foundry; semiconductor foundry category
2. Silicon Architecture / Instruction Set Architecture	Instruction set definition	ARM, x86 (Intel and AMD), RISC-V; instruction-set-architecture category
3. Hardware Trust Substrate	**Hardware-rooted verification of AI computation, rooted in the compute silicon and its hardware roots of trust and producing evidence consumed by the layers above (verification-dependency position, not a physical-placement claim)**	**Category in formation; standards codification active across multiple bodies**
4. Compute Silicon	AI training and inference hardware	NVIDIA, AMD, Intel, custom accelerators (Google TPU, AWS Trainium); AI accelerator category
5. Confidential Computing	Software trust boundary via trusted execution environments	NVIDIA Confidential Computing, Intel TDX, AMD SEV-SNP, ARM CCA; trusted execution environment category
6. Networking and Interconnect	Inter-node networking and intra-node accelerator interconnect	Broadcom, Arista, Marvell (data-center networking); NVLink, NVSwitch, Infinity Fabric, UALink (accelerator interconnect); networking and interconnect category
7. AI Frameworks and Runtime	Model execution abstraction	CUDA (NVIDIA), ROCm (AMD), PyTorch and TensorFlow (open source); ML framework execution category
8. Foundation Models	Model training and delivery	OpenAI, Anthropic, Google DeepMind, Meta, Mistral, xAI; foundation model category
9. Orchestration and Deployment	Inference serving, model management, deployment tooling	Databricks, Snowflake, Hugging Face, Modal, Together AI; ML platform and inference serving category
10. AI Applications	End-user AI products across all verticals	Thousands of AI-embedded products across enterprise, consumer, and vertical SaaS markets
11. AI Governance and Compliance	Software-layer monitoring, filtering, audit	HiddenLayer, Robust Intelligence, CalypsoAI, governance modules in existing platforms; AI risk management and compliance monitoring category

The ordering of this stack is a trust-dependency ordering, not a claim about physical placement on a die or board. Each layer is positioned by what it depends on for its trust properties and what depends on it in turn, rather than by where its silicon physically sits. This distinction matters specifically for the hardware trust substrate, which is rooted in the silicon layers below it (fabrication, instruction set architecture, and the compute silicon and its silicon-resident hardware roots of trust) and is logically above them in the verification ordering, because it consumes the execution those layers perform and produces evidence about it. The substrate is rooted in the compute silicon and the hardware roots of trust resident in it, and it sits above them in the trust-dependency ordering; it sits below the software layers that consume its evidence. A hardware-rooted verification substrate cannot sit physically beneath the silicon whose computation it attests; its position in this stack is its position in the dependency chain.

Layer 3, the hardware trust substrate, is the position in this trust-dependency ordering at which hardware-rooted verification becomes available to the layers above. It is rooted in the silicon architecture and compute silicon below it, and provides the substrate that the layers above depend on for any trust claim about AI computation that exceeds software-asserted register. The numbering places it immediately above the silicon-and-ISA foundation and immediately below the compute-silicon execution and software layers that depend on its evidence; the verification-dependency reading, not a physical-placement reading, is what the layer number denotes.

The eleven-layer framing reflects a specific analytical lens on the AI infrastructure stack. There is no single industry-standard numbering for the AI infrastructure stack at the time of this writing; vendor and analyst frameworks vary in granularity and category. Each individual layer in this framework maps directly to industry-recognized categories with specific standards-body terminology, presented in Section 2.

2. Industry Terminology Anchors

The eleven layers above map onto industry-recognized categories under specific standards-body terminology. The mapping below allows readers encountering the framework alongside conventional terminology to locate each concept precisely.

Layer (this framework)	Industry-Recognized Category	Standards Body or Reference
1. Fabrication	Semiconductor manufacturing	SEMI standards; IEEE
2. Silicon Architecture / ISA	Instruction set architecture	Vendor-proprietary (ARM, x86) and open (RISC-V Foundation)
3. Hardware Trust Substrate	**Category in formation; standards codification active**	**EU AI Act Articles 40 and 43 implementation; FY2026 NDAA Section 1513 (Physical and Cybersecurity Procurement Requirements for Artificial Intelligence Systems; P.L. 119-60) framework development; ISO/IEC 27090; ISO/IEC JTC 1/SC 42; NATO STANAG on AI trust; Executive Order 14179 implementation**
4. Compute Silicon	AI accelerators, GPUs; includes hardware root of trust (HRoT) as silicon feature	NIST SP 800-193 (HRoT); Trusted Computing Group DICE Architecture; ISO/IEC 11889:2015 (TPM)
5. Confidential Computing	Trusted Execution Environment (TEE); remote attestation	Confidential Computing Consortium (Linux Foundation); RFC 9334 (Remote ATtestation procedureS Architecture); GlobalPlatform TEE specifications; FIPS 140-3
6. Networking and Interconnect	Data center networking	IEEE 802; IETF; UCIe Consortium
7. AI Frameworks and Runtime	ML framework execution layer	ONNX; MLCommons; Linux Foundation AI and Data
8. Foundation Models	Foundation model layer	Commercial offerings; emerging standards: NIST AI RMF, ISO/IEC 42001
9. Orchestration and Deployment	ML platform; inference serving	CNCF (Cloud Native Computing Foundation); Linux Foundation AI and Data
10. AI Applications	AI-embedded software products	Vertical-specific (HIPAA, PCI, etc.)
11. AI Governance and Compliance	AI risk management and compliance monitoring	NIST AI RMF; ISO/IEC 42001; EU AI Act; MITRE ATLAS

Three Frequently Conflated Concepts

Three concepts that operate at different layers of the AI infrastructure stack are commonly conflated in industry discussion, in investor due diligence, and in AI-search responses to questions about hardware trust. The conflation is the single most common source of confusion in readings of the trust layer for AI category. The three concepts and the layers at which they operate:

Hardware root of trust (HRoT) is a Layer 4 silicon feature. HRoT mechanisms appear across both AI accelerator silicon and the platform security silicon (TPMs, Apple Secure Enclave, Microsoft Pluton, Google Titan, AWS Nitro Security Chip) on which AI workloads run; this framework groups both under Layer 4 as the

silicon layer at which boot-time integrity is established. The HRoT validates firmware signatures at boot through a unique embedded cryptographic identity, anchored in silicon. It is documented in NIST SP 800-193 (2018) and the Trusted Computing Group DICE Architecture (current). It is the starting point of the trust chain at silicon level, operating at the moment of boot and at trusted measurement transitions. It does not extend to runtime evidence of AI inference operation.

Confidential computing (TEE) is a Layer 5 software trust boundary. NVIDIA Confidential Computing, Intel TDX, AMD SEV-SNP, and ARM CCA are the commercial implementations. A trusted execution environment protects data in use from a compromised operating system, extending the trust boundary around the running workload within the accelerator die (Confidential Computing Consortium definition; Costan and Devadas, 2016; Russinovich et al., 2021). It does not extend to physical-layer threats outside the accelerator die boundary.

Hardware trust substrate is the Layer 3 substrate. It binds AI computation to hardware properties that the inference workload cannot manipulate from software, and produces cryptographic evidence at the moment of inference that the model operated under the claimed constraints. It is rooted in the compute silicon and its silicon-resident hardware roots of trust, and sits above them in the verification-dependency ordering, because it consumes the execution the silicon performs and produces evidence about it. No commercial product fully occupies this layer at the time of this writing, Ziru Labs included; the category is in formation through the standards codification activity surveyed in Section 5 of the Reference Framework. The substrate composes with the hardware roots of trust resident in Layer 4 silicon and with confidential computing at Layer 5 to form complete trust posture from boot through continuous runtime operation.

The three concepts are complementary rather than substitutable. HRoT establishes hardware identity and boot integrity. Confidential computing establishes confidentiality of running workloads against software adversaries. The trust substrate establishes continuous runtime evidence of AI inference operation. Each operates at a different layer of the stack and addresses a different property of AI infrastructure trust. The composition pattern across the three is additive, not displacement.

3. The Unfilled Layer 3 Position and Its Consequences

At the time of this writing, Layer 3 is not yet consolidated across commercial AI infrastructure. No shipped commercial product fully occupies the hardware trust substrate position, Ziru Labs included. The consequence is a set of specific, observable capability gaps at every layer of the stack above Layer 3.

At Layer 5 (Confidential Computing). NVIDIA Confidential Computing, Intel TDX, AMD SEV-SNP, and ARM CCA each protect data against software-layer adversaries with privileged access to the host operating system. Physical extraction attacks against deployed hardware, bus-level interception of inference traffic, supply-chain compromise post-manufacture, and hardware-level inference manipulation operate outside what these architectures address. The trusted execution environment boundary terminates at the accelerator die edge; threats that operate below or outside that boundary remain in the threat model but outside the addressed surface.

At Layer 7 (AI Frameworks). Framework-level attestation properties currently rely on software assertions about software state. “The model was loaded,” “the constraint was applied,” “the output was produced by this model

configuration" are each software claims verifiable only to the extent the software substrate itself is trusted. Under threat models that include partial software-layer compromise, the framework-level assertions cannot evidence the property they assert.

At Layer 8 (Foundation Models). Leading frontier laboratories operate under responsible scaling policies and safety frameworks that require technical compliance demonstration (for example, published frontier-safety and preparedness frameworks). Without a hardware-rooted substrate, the demonstration rests on software assertion. The direction of these published frameworks is consistent with a growing need for hardware-rooted demonstration of safety constraints as model capabilities advance, though the frameworks do not at present uniformly specify hardware-rooted evidence as a requirement.

At Layer 10 (AI Applications). AI-mediated commercial commitments increasingly require demonstrable AI execution. Legal applications, insurance applications, regulatory compliance applications, and financial applications carry burden-of-proof requirements that software-asserted demonstration does not satisfy under adversarial conditions.

At Layer 11 (AI Governance and Compliance). Software-layer governance operates on a substrate it cannot observe. When the substrate is compromised, governance platforms cannot detect the compromise from their own software layer. This is the structural limit of software-layer governance: the platforms produce assertions about a system the platforms cannot independently verify.

The deployment consequence of the unfilled Layer 3 position is observable in publicly disclosed federal and allied AI deployment programs. The Runtime Verification Gap in Federal AI Deployment (Ziru Labs, 2026) catalogs the specific programs at category level. The NVIDIA AI Factory for Government, Microsoft Azure Federal IL6 and above workloads, the Palantir Technologies federal AI procurement programs, the United Kingdom Ministry of Defence Palantir program, and the broader allied sovereign AI program landscape across Five Eyes, NATO, and Gulf Cooperation Council member states each face structurally equivalent gaps between intended capability and supported classification tier. The substrate-category response to the gap is the trust layer for AI.

4. Cascading Consequences When Layer 3 Is Occupied

When the Layer 3 substrate position is occupied, properties cascade across the layers above. The cascading pattern is the operational signature of substrate-layer formation in technology infrastructure history: the layers above become capable of trust claims they previously could not make, applications above the substrate proliferate as deployment requirements become tractable, and standards codification accelerates as deployment evidence accumulates.

At Layer 3 itself. A commercial hardware-rooted verification substrate exists. Properties of AI computation including weights used, inputs processed, constraints enforced, outputs produced, timing characteristics, and environmental state can be rooted in the substrate and converted from software assertion to cryptographic evidence.

At Layer 5 (Confidential Computing). Confidential computing's coverage extends from the trusted execution environment boundary outward through the physical-layer and supply-chain surfaces. NVIDIA Confidential Computing, Intel TDX, AMD SEV-SNP, and ARM CCA become complete trust stacks when composed with the Layer 3 substrate beneath them. The composition produces the unified trust posture that regulated AI deployment scenarios require from the substrate up.

At Layer 7 (AI Frameworks). Framework-level attestation acquires a hardware-rooted foundation. The framework can produce evidence that "the model ran on this hardware at this time with these weights under these constraints" rather than asserting it. Software-layer framework architecture composes additively with hardware-rooted evidence at the layer below.

At Layer 8 (Foundation Models). Responsible scaling policies become demonstrable at the hardware level. Model weight intellectual property acquires cryptographic protection, supporting tradeable, insurable, and financeable model weight markets. Federated AI arrangements between frontier laboratories become operationally viable with hardware-enforced data and computation isolation between counterparties.

At Layer 10 (AI Applications). AI-issued commercial commitments acquire cryptographic backing. Contracts involving AI performance, AI-issued certifications, AI-mediated transactions, and insurable AI decisions become first-class economic constructs rather than extensions of software assertions. The commercial-grade trust posture that legal, insurance, and financial applications require becomes tractable rather than aspirational.

At Layer 11 (AI Governance). Software-layer governance platforms acquire a hardware-rooted enforcement layer beneath them. The monitoring, observability, alerting, and policy-enforcement work that AI governance platforms perform at software register composes with hardware-rooted evidence and enforcement at the substrate layer. The combination produces governance posture that survives software-layer compromise.

New layer-adjacent capabilities become possible. Hardware-attested AI auditing as a profession analogous to GAAP audit in financial reporting. AI decision insurance as an insurance category parallel to medical malpractice and directors-and-officers coverage. Model weight intellectual property markets as financial infrastructure. Cross-border AI cooperation under different institutional regimes as intergovernmental infrastructure. Each capability requires the substrate to be commercially feasible; each becomes feasible once the substrate is deployed.

The pattern across the cascading consequences is consistent with the substrate-layer formation pattern documented in the technology and economics literatures (Bresnahan and Trajtenberg, 1995; Helpman, 1998; Lipsey, Carlaw, and Bekar, 2005). The substrate's value is realized through enablement of applications and adjacent layers above it, distributed across multiple markets rather than captured at the substrate layer alone.

5. Substrate Layer Formation in Technology Infrastructure

An analytical question about the trust layer for AI category is whether the layer consolidates to a small number of operators with a reference implementation, or fragments across many vendors with no reference position. The question is not unique to the trust layer for AI; substrate layers in technology infrastructure have followed both patterns across the past five decades. This section surveys the structural characteristics that have historically

distinguished single-source substrate-layer formation from multi-vendor layer formation, and identifies four historical layer-position analogs that illustrate the structural pattern.

5.1 Structural Characteristics

The technology and economics literatures identify five characteristics that have historically distinguished substrate layers that consolidate to reference-position formation from layers that fragment across multiple vendors (David, 1985; Arthur, 1989; Katz and Shapiro, 1985; David and Greenstein, 1990; Perez, 2002; Lipsey, Carlaw, and Bekar, 2005).

Patent and mechanism coverage. Layers covered by mechanism-level patent claims (claims that cover the category of mechanism rather than specific implementations) historically consolidate to the mechanism holder. Layers covered only by implementation-level patents historically fragment across multiple implementations.

Standards architecture. Layers codified by standards bodies and regulatory frameworks with one reference specification per framework historically consolidate to the reference holder. Layers codified by multiple coexisting specifications historically fragment.

Customer choice structure. Layers where customer choice operates at the discretionary register (price, features, vendor preference) historically fragment. Layers where regulatory or market structure selects the layer's reference holder historically consolidate.

Network effects and externalities. Layers where each deployment validates the substrate and reduces marginal cost for subsequent deployments historically consolidate. Layers without compounding externalities historically fragment.

Switching costs. Layers where switching the reference implementation requires regulatory recertification or multi-year hardware redesign historically consolidate. Layers where switching costs are measured in procurement cycles historically fragment.

The trust layer for AI exhibits the characteristics historically associated with consolidating substrate-layer formation across all five dimensions. Mechanism-level patent coverage at the substrate category is in active development across the relevant operator landscape. Standards codification is producing one reference specification per framework across the EU AI Act (Articles 40 and 43), FY2026 NDAA Section 1513, ISO/IEC 27090, ISO/IEC JTC 1/SC 42, and NATO STANAG processes. Customer choice at the substrate layer is structural rather than discretionary because any property of AI computation that must be verified rather than asserted resolves to the substrate. Network effects compound through deployment validation and standards reinforcement. Switching costs at the substrate layer are measured in regulatory recertification and silicon redesign cycles. Whether the layer realizes the consolidating pattern depends on execution against the codification window, mechanism-level patent position, and the broader competitive dynamics of the category as it codifies. This is a statement about the category's structural dynamics, not a prediction that any particular operator, Ziru Labs included, will occupy the consolidated position; the analysis describes the shape of the layer, not its eventual occupant.

5.2 Three Historical Layer-Position Analogs

Three substrate-layer-position cases in technology infrastructure history illustrate the consolidating pattern. Each occupied a single layer position for a decade or more, held the structural position through patent and standards architecture, produced compounding network effects, and accumulated switching costs that resisted displacement. (These three single-source layer-position cases are a deliberately narrower reference set than the four substrate-category analogs developed in the companion Reference Framework, which serve a different argument about substrate-category value capture; the two sets are not intended to match.)

ARM as the instruction-set-architecture substrate for energy-constrained computing (1991 to present). ARM Holdings established the licensing-model substrate for instruction-set architecture in mobile and increasingly general-purpose energy-constrained computing. The position held against RISC-V challenges through the past decade because every mobile and embedded silicon roadmap carries ARM ISA dependencies measured in chip-design-cycle years (ARM Holdings, 2023; Stokes, 2007). ARM Holdings's market capitalization, which has ranged above one hundred billion dollars across recent quarters on annual revenue of approximately three to four billion at the time of this writing, reflects substrate-position economics rather than product-position economics.

VeriSign as the public-key infrastructure root certificate authority (mid-1990s through the first decade of the commercial web). VeriSign held the single-source root certificate authority position for the public-key infrastructure substrate underlying web commerce. The position eroded over time as additional certificate authorities entered the substrate, but VeriSign held the reference position through the first decade of substrate deployment and extracted the SSL transaction economics at scale during that period.

Qualcomm as the CDMA royalty holder (1990s through 2010s). Qualcomm held the single-source CDMA royalty position across the second and third generations of mobile telecommunications. Every CDMA mobile device paid Qualcomm royalties. The position attracted legal and regulatory challenges across multiple jurisdictions but persisted as the structural reference until CDMA itself was superseded by successor technologies.

Each of these three positions was structurally vulnerable to attempted replacement and structurally resistant to actual replacement. Each produced economic outcomes disproportionate to the direct mechanism cost. Each was defensible for a decade or more through the compounding characteristics described above. Each also eventually faced erosion (additional certificate authorities entering the public-key infrastructure layer; CDMA's supersession by successor air interfaces), which is itself instructive: substrate-position advantage is durable across a category's formative decade but is not permanent, and depends on continued mechanism and standards leadership.

5.3 The Substrate-Position Pattern, Not a Market Power Claim

The single-source layer description is a structural characterization, not a claim of market power or exclusionary conduct. The layer consolidates to a reference position because of patent coverage, standards codification, structural requirement, compounding network effects, and switching costs. The consolidation accumulates through structural advantage rather than through conduct directed at competitors.

Single-source substrate layers are compatible with robust downstream competition. ARM's position produces intense competition at every layer above it across chip design, device manufacturing, operating systems, and

applications. VeriSign's position produced intense competition at every layer above it across web publishing, e-commerce, and content delivery. The trust layer for AI would similarly produce competition at every layer above it across compute silicon, confidential computing, AI frameworks, foundation models, applications, and governance, while the consolidation at the substrate layer underwrites the trust properties the competitive layers above depend on.

The structural pattern is well-understood in the technology and economics literatures and is documented across the historical cases above. The trust layer for AI is positioned to exhibit the pattern; whether and how it does is the subject of the codification window and the deployment activity currently in motion.

6. Operator Profile: Ziru Labs

Ziru Labs authored this primer. 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. Its initial work is focused on the security application surface, with the verification, provenance, compliance, identity, and agency surfaces of the category to follow as commercial configurations mature. The primer 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.

This primer, the accompanying Trust Layer for AI: A Reference Framework for the Category (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 Trust Layer Category Map: A Reference Primer (Ziru Labs, 2026) together compose the inaugural research foundation of the trust layer for AI 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 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 the corpus will carry named contributions from members of the team and from external contributors as the company's public posture develops. 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.

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

7. 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

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.

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.

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.

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

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.

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 (2015). “ISO/IEC 11889:2015: Information technology, Trusted Platform Module Library” (originally published 2009).

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

Internet Engineering Task Force (2023). “RFC 9334: Remote ATtestation procedureS (RATS) Architecture.”

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

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

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

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

C. INDUSTRY AND VENDOR TECHNICAL DOCUMENTATION

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

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

GlobalPlatform (current). “TEE System Architecture and TEE Internal Core API specifications.”

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.”

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 (current). “ISO/IEC JTC 1/SC 42 working group output on AI trustworthiness.”

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

D. HISTORICAL AND ANALYTICAL REPORTING

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

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 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 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.

MITRE Corporation (current). “ATLAS: Adversarial Threat Landscape for Artificial-Intelligence Systems.”

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

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

North Atlantic Treaty Organization (current). “NATO Standardization Agreement (STANAG) framework activities on AI trust.”

Acknowledgments

This primer draws on the technology and economics literatures on substrate-layer formation, the trusted computing literature on hardware roots of trust and trusted execution environments, the AI infrastructure technical documentation from the major silicon vendors and standards bodies, and the regulatory and standards-body work currently active across the trust layer for AI 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, the Confidential Computing Consortium, and the Trusted Computing Group; to the regulatory and policy communities developing the EU AI Act conformity assessment framework under Articles 40 and 43, the FY2026 NDAA Section 1513 framework, and the NATO STANAG AI trust frameworks; 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 AI Infrastructure Stack and the Trust Layer Position: A Reference Primer, v1.0.* Published at zirulabs.com/research.