The AI OSI Stack: A Governance Blueprint for Scalable and Trusted AI

Daniel P. Madden

Independent AI Researcher

https://danielpmadden.com

Originally Conceived: September 2025

Expanded and Revised: November 2025

Major Release Notice

This document is the complete Version 4 release of the *AI OSI Stack*. It replaces all earlier versions and serves as the single official reference for every related framework and implementation.

Abstract

Artificial intelligence now underpins communication, commerce, and public infrastructure, yet its governance remains fragmented and reactive. The **AI OSI Stack** provides a structural and procedural model for understanding, auditing, and governing intelligent systems at scale. Version 4 expands and consolidates earlier releases, establishing a seven–layer architecture that aligns technical components with institutional accountability. The layers are: (1) Physical and Hardware Foundations, (2) Model Architecture, (3) Training and Optimization, (4) Instruction and Control, (5) Interface and Protocol, (6) Application, and (7) Governance and Trust. Each layer defines its scope, risks, governance levers, and required audit artifacts. The framework integrates with more than sixty existing national and international standards, enabling regulators, enterprises, and researchers to trace accountability across the entire AI lifecycle. By embedding transparency and human dignity as design constraints, the Stack advances a model of "governance as architecture," in which trust becomes a measurable system property rather than a marketing claim.

Layer Overview

Layer	Governance Focus
1. Physical	Resource accountability, energy disclosure, and security baselines.
2. Architecture	Interpretability, safety constraints, and reproducibility.
3. Training	Data provenance, fairness, and optimization transparency.
4. Instruction	Human intent capture, role boundaries, and refusal logic.
5. Interface	Secure access, protocol interoperability, and open APIs.
6. Application	Deployment monitoring, sectoral compliance, and user safety.
7. Governance	Oversight structures, transparency reporting, and ethical review.

Suggested Citation:
Madden, D. (2025). *The AI OSI Stack: A Governance Blueprint for Scalable and Trusted AI.* Zenodo. https://doi.org/10.xxxxx/zenodo.xxxxx

Contents

1 Introduction and Purpose

1.1 Rationale for Layered Governance

Artificial intelligence has evolved from isolated products into a foundational layer of global infrastructure. It now influences communication, health care, education, finance, and state capacity. Despite this ubiquity, oversight mechanisms remain inconsistent, fragmented, and largely reactive. Existing governance models tend to treat AI as a singular artifact rather than an interconnected system of hardware, data, models, interfaces, and institutions. The result is persistent opacity, concentration of control, and a lack of verifiable accountability.

The **AI OSI Stack** was developed to resolve this structural opacity by decomposing artificial intelligence into discrete, auditable layers of responsibility. The approach is inspired by the Open Systems Interconnection (OSI) model in computer networking, which achieved global scalability by defining clear interfaces and boundaries between layers. In the same way, this framework allows each layer of the AI ecosystem—from physical compute to social governance—to be examined, regulated, and improved without destabilizing the whole.

By treating governance as a *design feature* rather than a post–hoc compliance task, the Stack enables targeted intervention and portable trust. It provides a shared vocabulary for engineers, policymakers, and auditors to describe where risk originates, who controls it, and how evidence of integrity can be generated and verified.

1.2 Objectives of Version 4

Version 4 expands the earlier architecture into a comprehensive governance standard. Its primary objectives are:

a) To define explicit scope, functions, and risks for each of the seven layers of the AI OSI Stack.

b) To provide formal definitions for all governance artifacts, including Interpretive Trace Packages, Decision Rationale Records, Governance Decision Summaries, Oversight Audit Memos, and Integrity Ledger Entries.

c) To align the Stack with more than sixty established national and international frameworks, including ISO/IEC 42001, NIST AI RMF, OECD AI Principles, UNESCO Ethics of AI, and the EU AI Act.

d) To refine mechanisms for temporal integrity and semantic version control, ensuring that changes in systems and definitions remain auditable through time.

e) To present implementation pathways for enterprises, regulators, researchers, and public institutions.

The framework aims to make accountability as modular and testable as the technologies it governs.

1.3 Relation to Associated Frameworks

The AI OSI Stack integrates three complementary architectures authored by the same researcher:

- The **Role–Bound Cognitive System Design Framework**, which defines bounded cognitive agents with explicit ethical and operational mandates.
- The **Reasoning Integrity Engineering Methodology**, which embeds epistemic transparency and justification into system design.
- The **Epistemic Interoperability and Audit Protocol**, which specifies how reasoning artifacts are serialized, exchanged, and verified across systems.

Together, these frameworks establish a unified ecosystem for designing, documenting, and auditing intelligent systems in accordance with measurable standards of transparency, integrity, and dignity.

1.4 Intended Audience

This document is written for:

- **Policy and Regulatory Bodies:** to target oversight precisely and align legislation with technical boundaries.
- **Enterprise Architects and Developers:** to integrate governance artifacts into existing AI development pipelines.
- **Auditors and Standards Organizations:** to create interoperable certification and reporting systems.
- **Researchers and Educators:** to extend the Stack as a living framework for accountable innovation.

1.5 Scope of Application

The Stack applies to all artificial–intelligence systems, whether centralized or distributed, proprietary or open–source, and across sectors such as finance, health, education, creative industries, and public administration. Its design supports integration with existing risk–management, cybersecurity, and quality–assurance processes. The framework is intentionally technology–neutral and adaptable to emerging paradigms, including multimodal models, autonomous agents, and hybrid human–machine systems.

2 Definitions and Acronyms (Normative)

For the purposes of this document, the following terms and definitions apply.

AI OSI Stack

The Layered Architecture for Accountable Artificial Intelligence Systems. A seven–layer structural model that decomposes AI into Physical/Hardware, Model Architecture, Training/Optimization, Instruction/Control, Interface/Protocol, Application, and Governance/Trust.

Interpretive Trace Package (ITP)

A structured record of interpretive reasoning generated by a model or agent, including input provenance, contextual factors, and explanation metadata.

Decision Rationale Record (DRR)

A concise, auditable document capturing the reasoning basis, alternatives considered, constraints, and accountability metadata for a single governance decision.

Governance Decision Summary (GDS)

A standardized summary of a decision's context, risks, and outcomes, used for reporting and certification.

Oversight Audit Memo (OAM)

A formal note from an auditing or supervisory body recording observations, compliance findings, and required follow–up actions.

Integrity Ledger Entry (ILE)

A timestamped, cryptographically verifiable record linking a governance artifact to a specific version and time window.

Epistemic Interoperability and Audit Protocol (EIA Protocol)

The communication and data–exchange standard that enables portable, verifiable reasoning records across AI systems.

Reasoning Integrity Engineering Methodology

The design discipline ensuring that systems reveal how they know, not merely what they output.

Role–Bound Cognitive System Design Framework

The specification for constructing AI agents with defined roles, boundaries, and refusal logic, preventing ambiguity in responsibility.

Semantic Version Control (SVC)

A method for tracking and documenting changes in definitions, models, or governance artifacts using consistent version numbering and provenance logs.

Additional abbreviations follow standard usage: ISO (International Organization for Standardization), NIST (National Institute of Standards and Technology), OECD (Organisation for Economic Co–operation and Development), EU (European Union), and RMF (Risk Management Framework).

3 Architectural Foundations and Methodology

3.1 Historical Context and Analogy

The design of the AI OSI Stack follows the logic of the Open Systems Interconnection (OSI) model that enabled global computer networking in the late twentieth century. That model achieved stability and scalability by establishing clear boundaries and responsibilities between layers, ensuring that innovation at one layer did not destabilize the others. Artificial intelligence requires a similar architecture: one that decomposes complex, adaptive systems into components that can be independently audited, certified, and improved.

The seven layers of the Stack define a structured path from physical computation to social governance. They clarify where technical risks concentrate, where power accumulates, and where ethical and regulatory duties reside. Each layer communicates with those directly above and below it through well–defined interfaces, creating a chain of accountability that extends from data center to decision outcome.

3.2 Governance as Design Infrastructure

Governance is typically introduced after systems are built, in the form of compliance audits or policy retrofits. This approach is insufficient for artificial intelligence, which adapts and scales faster than regulatory cycles. In the AI OSI Stack, governance is reinterpreted as part of the design infrastructure. Each layer includes built–in mechanisms for traceability, auditability, and human oversight. Rather than relying solely on external regulation, the framework embeds accountability directly into technical and organizational processes.

This architecture makes trust a measurable system property. A well–designed AI system is one in which every layer can produce evidence of integrity appropriate to its function, and in which those evidences connect coherently across layers.

3.3 Methodological Steps

The Stack can be applied using a consistent six–step method:

1. **Layer Identification:** Determine which of the seven layers a component or process occupies.

2. **Risk Mapping:** Identify the main operational, ethical, and systemic risks present at that layer.

3. **Governance Lever Selection:** Choose the appropriate mechanisms of control or oversight, including technical, procedural, and institutional measures.

4. **Artifact Generation:** Produce the governance artifacts relevant to that layer (e.g., DRR, GDS, ITP, OAM, ILE).

5. **Verification:** Validate that the artifacts are complete, consistent, and linked to upstream and downstream layers.

6. **Integration:** Record all outputs through the Epistemic Interoperability and Audit Protocol so that reasoning and accountability can be traced end–to–end.

Applying this method converts abstract principles of transparency and ethics into specific engineering and documentation practices.

3.4 Comparative Alignment with Existing Frameworks

The AI OSI Stack does not replace existing standards; it provides the structure that allows them to interoperate. Its seven layers can host and coordinate the control requirements already defined by ISO, NIST, OECD, and other institutions. For example:

- The **NIST AI Risk Management Framework** primarily operates at Layers 3, 4, and 7, focusing on mapping, measuring, and governing risks.

- **ISO/IEC 42001:2024** aligns with Layers 5–7, defining requirements for AI management systems and organizational governance.

- The **EU AI Act** establishes legal obligations that correspond mainly to Layers 6 and 7.

- **IEEE P7000–series** ethics standards provide guidance for Layer 4 (Instruction/Control) and Layer 7 (Governance/Trust).

By providing a shared structure, the Stack allows these frameworks to be applied in concert rather than in parallel, reducing duplication and regulatory friction.
Each layer builds on the foundations beneath it while exposing governance data to the layers above. The Physical/Hardware layer ensures resource accountability; the Model Architecture layer defines interpretability and control boundaries; the Training/Optimization layer records data provenance; the Instruction/Control layer encodes human intent; the Interface/Protocol layer governs access and interoperability; the Application layer mediates impact and user experience; and the Governance/Trust layer establishes oversight and transparency obligations.

4 Layer Interdependence and Power Geometry

4.1 Cross–Layer Dependencies

Although each layer can be analyzed separately, none exists in isolation. Hardware constraints determine feasible model architectures; model architectures influence training data requirements; training regimes shape control parameters; and application interfaces define what forms of governance are possible. Understanding these dependencies prevents the misallocation of responsibility—a common failure in AI governance where blame is assigned to the wrong technical layer.

4.2 Concentration of Power

Power in AI ecosystems tends to concentrate at three junctures:

i. **Layer 1–3 (Infrastructure Control):** ownership of compute, data, and proprietary architectures.

ii. **Layer 5 (Interface Gatekeeping):** control of APIs, platforms, and user access.

iii. **Layer 7 (Governance Framing):** the ability to define what counts as "ethical" or "compliant."

Effective governance requires counterbalancing these concentrations by ensuring that each layer has independent verification paths and transparent reporting.

4.3 Mitigation Strategies

The framework recommends several cross–layer mitigations:

- Diversify control of physical and model resources to avoid monopolization.
- Require open or auditable interfaces at Layer 5 to prevent platform lock–in.
- Establish multi–stakeholder governance at Layer 7 to prevent capture by a single interest group.

These structural interventions convert abstract fairness goals into actionable design features.

5 Layer-by-Layer Framework (Expanded)

Each layer of the AI OSI Stack defines a distinct domain of responsibility. This section describes the scope, objectives, principal risks, and governance mechanisms associated with each of the seven layers. An optional eighth layer, addressing public policy and societal adaptation, is also noted. For every layer, the framework identifies where accountability must be assigned and what audit artifacts are required to demonstrate conformance.

5.1 Layer 1: Physical and Hardware Foundations

Scope. The foundational layer encompasses data centers, compute hardware, energy systems, and supply chains required to produce and operate AI models. It includes semiconductor fabrication, data storage, networking infrastructure, and environmental dependencies such as power and cooling.

Primary Objectives.

- Ensure transparency in energy use and environmental impact.
- Maintain supply–chain integrity and traceability of critical components.
- Secure physical and firmware environments against tampering or unauthorized modification.

Principal Risks. Hardware monopolies, opaque sourcing, and unverified energy footprints create systemic vulnerability. Physical attacks or unaccountable resource consumption can undermine higher–layer integrity.

Governance Mechanisms.

- Hardware Bill of Materials and attestation reports.
- Environmental and energy disclosure aligned with ISO 14001 and ESG standards.
- Integration with cybersecurity baselines such as ISO 27001 and NIST SP 800–53.

Required Artifacts. Integrity Ledger Entries (ILEs) documenting hardware provenance and operational footprints.

5.2 Layer 2: Model Architecture

Scope. The architecture layer defines the structure, mathematical principles, and control parameters of models. This includes topology, objective functions, interpretability methods, and safety boundaries.

Primary Objectives.

- Document architectural choices and their governance implications.
- Embed interpretability and constraint mechanisms within model design.
- Support independent evaluation of model capabilities and limitations.

Principal Risks. Opaque model structures, undocumented parameters, and lack of reproducibility can conceal bias or unsafe behavior.

Governance Mechanisms.

- Model Cards and Architecture Specification Sheets.
- Independent review or certification against defined safety criteria.
- Conformance to standards including ISO/IEC 42001 and IEEE P7009.

Required Artifacts. Interpretive Trace Packages (ITPs) linking model outputs to architecture decisions.

5.3 Layer 3: Training and Optimization

Scope. This layer governs data selection, preprocessing, optimization procedures, and validation pipelines used to train models.

Primary Objectives.

- Maintain full lineage of data sources and transformations.
- Ensure fairness, representational balance, and legal compliance in datasets.
- Record hyperparameter configurations and optimization criteria for reproducibility.

Principal Risks. Data contamination, unlicensed data use, or optimization toward unintended objectives may lead to harmful outcomes or regulatory breach.

Governance Mechanisms.

- Dataset documentation and Data Cards compliant with the EU AI Act's data governance obligations.
- Provenance tracking systems consistent with OECD AI Principles.
- Bias detection and fairness evaluation protocols.

Required Artifacts. Decision Rationale Records (DRRs) detailing dataset inclusion choices and trade–offs.

5.4 Layer 4: Instruction and Control

Scope. The Instruction and Control layer represents the human interface with model behavior: prompts, reinforcement rules, persona definitions, and operational policies that determine what a system is authorized to do.

Primary Objectives.

- Capture human intent and value alignment in machine–readable form.
- Define explicit refusal logic and role boundaries for AI agents.
- Ensure that control systems are auditable and reversible.

Principal Risks. Ambiguous instructions or uncontrolled delegation can produce emergent behavior without clear responsibility.

Governance Mechanisms.

- Implementation of Role–Bound Cognitive System Design.
- Ethical control kernels integrating human approval checkpoints.
- Logging of all instruction updates through Semantic Version Control (SVC).

Required Artifacts. Governance Decision Summaries (GDS) and Oversight Audit Memos (OAM) for all major control updates.

5.5 Layer 5: Interface and Protocol

Scope. The Interface and Protocol layer defines how models and governance systems communicate with each other and with external actors. It governs all input–output channels, application programming interfaces (APIs), and middleware responsible for transmitting reasoning evidence and audit records.

Primary Objectives.

- Provide a secure and transparent mechanism for exchanging governance artifacts between systems.

- Prevent interface monopolies and lock–in effects that constrain innovation or oversight.
- Ensure that reasoning evidence and accountability data are interoperable across institutions and jurisdictions.

Governance Mechanisms.

- Open API documentation and standardized access policies.
- Implementation of the **Epistemic Interoperability and Audit Protocol (EIA Protocol)**, specified in Section 6.
- Alignment with cybersecurity and data–protection standards such as ISO/IEC 27018 and NIST SP 800–53.

Required Artifacts. Integrity Ledger Entries for all interface versions and access–policy revisions. Each exchange of governance artifacts shall comply with the handshake and validation procedures defined in the EIA Protocol.

5.6 Layer 6: Application

Scope. The application layer captures the deployment context of AI models in products or services, including sector–specific integration and user interaction.

Primary Objectives.

- Translate governance principles into operational behavior.
- Protect users from deceptive, unsafe, or discriminatory outcomes.
- Maintain transparent reporting of performance and limitations.

Principal Risks. Unsafe deployment, poor user disclosure, or unmonitored adaptation can transform technical risk into social harm.

Governance Mechanisms.

- Conformance to domain–specific standards (e.g., FDA AI/ML Guidance, HIPAA, ISO 13485 for medical systems).
- Post–deployment monitoring and incident reporting.
- Periodic third–party audits of impact and compliance.

Required Artifacts. Governance Decision Summaries aggregating operational metrics and incident data.

5.7 Layer 7: Governance and Trust

Scope. The highest layer defines institutional oversight, ethical review, accountability assignment, and external transparency.

Primary Objectives.

- Establish enduring mechanisms for oversight and redress.
- Codify transparency obligations in governance policy.
- Ensure traceability from social outcomes back to technical decisions.

Principal Risks. Capture of governance processes by vested interests or failure to act on audit findings.

Governance Mechanisms.

- Independent review boards and multi–stakeholder councils.
- Public reporting consistent with OECD and UNESCO principles.
- Integration with national or regional AI governance frameworks.

Required Artifacts. Oversight Audit Memos (OAMs) summarizing review outcomes and policy adjustments.

6 Epistemic Interoperability and Audit Protocol

6.1 Purpose

The EIA Protocol defines a minimal communication standard for exchanging governance artifacts between independent systems. Its purpose is to guarantee that evidence of reasoning and accountability can be verified and synchronized across institutional boundaries without relying on a single vendor or platform.

6.2 Core Concepts

Node Any participant that generates, transmits, or verifies governance artifacts.

Artifact A structured record conforming to one of the five canonical types defined in Section 6 (ITP, DRR, GDS, OAM, ILE).

Ledger A persistent record of artifact metadata and corresponding Integrity Ledger Entries.

Exchange
A transaction in which artifacts and confirmations are transmitted between nodes using the EIA message set.

6.3 Canonical JSON Schema

```
{
  "artifact_type": "DRR",
  "version": "1.0",
  "uuid": "550e8400-e29b-41d4-a716-446655440000",
  "issuer": {"organization": "Example Org"},
  "timestamp": "2025-11-01T18:00:00Z",
  "layer": 4,
  "summary": "Policy update for language model v3.2",
  "linked_artifacts": ["ILE:12345","ITP:67890"],
  "signature": "BASE64-ENCODED-SIGNATURE"
}
```

6.4 Exchange Sequence (Handshake Model)

1. **Offer:** Node A sends an `ARTIFACT_OFFER` header announcing the artifact hash, type, and version.

2. **Acknowledge:** Node B validates the header and issues `ACK_REQUEST`.

3. **Transmit:** Node A delivers the signed artifact; Node B verifies the hash and signature, records an ILE, and returns `ACK_CONFIRM`.

6.5 Error States

- **INVALID_HASH** — advertised and computed hashes differ.

- **SIGNATURE_FAIL** — digital signature cannot be validated.

- **TIMEOUT** — expected acknowledgment not received within threshold.

Each error triggers an Oversight Audit Memo documenting the cause and corrective action.

6.6 Namespace and Versioning

Every artifact shall be identified by the tuple:

```
<LayerID>-<ArtifactType>-<Major.Minor>-<UUID>
```

Example: `4-DRR-1.0-550e8400e29b41d4a716446655440000`. Version numbers follow the Semantic Version Control scheme in Section 7.

6.7 Security and Integrity Requirements

- All payloads shall be signed using an asymmetric key pair registered to the issuing organization.
- Timestamps must conform to ISO 8601 UTC and be verified by a trusted time source.
- Ledger entries may store only cryptographic hashes where confidentiality is required.
- Transport security shall employ mutual TLS or equivalent encryption.

6.8 Conformance Criteria

An implementation is conformant when it:

a) Generates valid JSON artifacts for all five artifact types.

b) Implements the three–step handshake and handles defined error states.

c) Maintains a ledger of ILEs referencing artifact hashes and timestamps.

d) Passes signature and version–matching verification.

6.9 Relation to the AI OSI Stack

The EIA Protocol operates primarily at Layers 5 and 7, connecting technical operations with institutional oversight. It forms the exchange mechanism through which global frameworks (see Section 8) can share verified governance data.

6.10 Layer 8 (Optional): Policy and Society

Although not part of the core technical stack, many implementations benefit from an eighth layer representing legislative, educational, and societal feedback processes. This layer captures how governance insights feed back into culture, law, and public understanding of AI. Its inclusion ensures that governance remains adaptive to social change rather than static to a single era.

7 Governance Artifacts and Audit Infrastructure

Governance artifacts are the tangible evidences through which accountability is demonstrated. Each artifact serves a specific purpose within the Stack, linking a defined layer to verifiable documentation. Version 4 formalizes five core artifact types and their interrelations. Together they provide a continuous record of reasoning, action, and oversight throughout the system lifecycle.

7.1 Interpretive Trace Package (ITP)

The preceding Section 6 defines the exchange protocol that enables interoperability across standards. The mappings in this section describe how those external frameworks align with the layers of the Stack once connected through the EIA Protocol.

Purpose. The ITP captures the interpretive process by which an AI system produces an output. It records contextual data, intermediate representations, and explanatory metadata.

Structure. An ITP typically includes:

- Unique identifier and timestamp.
- Model version and configuration hash.
- Description of input context and constraints.
- Feature importance or attention mapping.
- Generated rationale or explanation text.
- Validation metrics or uncertainty indicators.

Governance Function. ITPs enable post–hoc and real–time interpretability. They allow independent reviewers to understand how an output was derived, forming the evidence base for Decision Rationale Records. ITPs shall be retained according to applicable data–protection requirements and versioned under the Semantic Version Control scheme.

7.2 Decision Rationale Record (DRR)

Purpose. A DRR summarizes the reasoning process and justification for a decision or design choice at any layer of the Stack. It formalizes how constraints, trade–offs, and human judgments were considered.

Content Requirements.

- Decision description and context.
- Alternatives evaluated and rationale for selection.
- Stakeholders consulted.
- Associated risks and mitigations.
- Cross–references to relevant ITPs or data sources.
- Signature of responsible party and date.

Governance Function. DRRs provide traceable accountability for discrete decisions, allowing audits to verify that ethical and regulatory criteria were applied. They are mandatory for significant model updates, dataset changes, or control modifications.

7.3 Governance Decision Summary (GDS)

Purpose. The GDS consolidates multiple DRRs or operational decisions into a single structured report. It is used for organizational reporting, regulatory submissions, or public transparency disclosures.

Required Elements.

- Executive overview of decisions within the reporting period.
- Key performance and compliance indicators.
- Summary of identified issues and corrective actions.
- Linkage to Oversight Audit Memos.

Governance Function. GDS documents provide management and external stakeholders with a concise overview of system governance status. They also act as the gateway artifact connecting technical documentation to institutional oversight.

7.4 Oversight Audit Memo (OAM)

Purpose. The OAM records findings from internal or external audits. It certifies whether governance artifacts are complete, coherent, and compliant with declared standards.

Required Elements.

- Audit scope and objectives.
- Methodology and evidence reviewed.
- Summary of findings and risk grading.
- Recommendations and deadlines for remediation.

Governance Function. OAMs establish a verifiable feedback loop between auditors and system owners. They shall be archived in accordance with organizational policy and linked to affected DRRs and GDS entries.

7.5 Integrity Ledger Entry (ILE)

Purpose. The ILE provides temporal and cryptographic assurance that each artifact has not been altered since issuance.

Structure.

- Artifact identifier and version reference.
- Cryptographic hash of artifact content.
- Timestamp issued through trusted time service.
- Optional blockchain or distributed–ledger anchor.

Governance Function. By connecting artifacts through ILEs, organizations create an immutable chain of custody for reasoning and accountability data. This supports independent verification and long–term trust portability.

7.6 Artifact Lifecycle and Interrelations

Each ITP generated by a model instance may give rise to one or more DRRs when human or automated decisions are made. GDS reports aggregate DRRs across time, while OAMs provide external validation. All artifacts are linked and versioned through corresponding ILEs. This layered artifact infrastructure creates a continuous, auditable trace from technical operations to institutional oversight.

8 Temporal Integrity and Semantic Version Control

8.1 Purpose of Temporal Governance

Temporal integrity ensures that both data and governance artifacts remain traceable through time. Every AI system evolves; models are retrained, policies are revised, and definitions shift. Without temporal accountability, evidence of intent and justification is lost. Version 4 introduces an enhanced Semantic Version Control (SVC) scheme to manage temporal change across layers.

8.2 Semantic Version Hierarchy

Each artifact, dataset, and model component shall carry a version identifier in the format `major.minor.patch`. Major changes represent structural or policy alterations requiring new DRRs; minor changes correspond to parameter updates or clarifications; patch increments document small corrections or typographical fixes. Every version increment must generate an Integrity Ledger Entry with a timestamp and reference to the preceding version.

8.3 Temporal Drift Monitoring

Temporal drift occurs when the meaning or effect of a model or artifact diverges from its original specification without formal update. Monitoring shall include scheduled reviews comparing current outputs with baseline behavior and threshold indicators. Detected drift triggers an immediate governance review and, if necessary, creation of new DRRs and ILEs.

8.4 Memory Governance and Forgetting by Design

To comply with privacy and data–protection laws, systems must support selective forgetting while maintaining audit integrity. The Stack prescribes that deletions or redactions be logged as new ILEs referencing the removed material by hash only, preserving proof of deletion without retaining personal data.

8.5 Attestation Workflow and Provenance Verification

All governance artifacts and versions are verified through an attestation workflow:

1. Artifact generation and internal validation.

2. Signing by responsible actor using approved cryptographic keys.

3. Ledger registration and timestamping.

4. Periodic external verification by independent auditor.

This workflow ensures the authenticity, integrity, and continuity of the governance record.

8.6 Intellectual Stewardship and Temporal Custody

Each version of the Stack is timestamped and archived to preserve provenance and authorship. Any subsequent modification must carry a new custodial signature and DOI, ensuring that the original epistemic lineage remains intact. This temporal custody mechanism upholds both technical and intellectual integrity across the lifecycle of the standard.

9 Global Framework Interoperability Map

9.1 Purpose and Overview

Artificial intelligence governance is already represented in dozens of overlapping regulatory, technical, and ethical frameworks. The AI OSI Stack does not replace these efforts. Instead, it provides a structure in which their principles and controls can be coherently applied. This section outlines how the seven layers of the Stack align with the major international frameworks that currently guide AI oversight. The intent is to show where existing standards operate, where gaps remain, and how a layered architecture can serve as a common reference.

9.2 Methodology

Mapping between the Stack and other frameworks follows three principles:

a) **Functional correspondence:** identify which Stack layer performs a comparable role or control function to the referenced framework.

b) **Audit alignment:** specify which governance artifacts or metrics can serve as evidence of compliance with that framework.

c) **Non–duplication:** treat external standards as complementary, avoiding redundancy of terminology or reporting requirements.

All framework citations in this section refer to publicly released documents as of November 2025.

9.3 Framework Categories

For clarity, international standards are grouped into five broad categories: *(i)* Regulatory and Legal, *(ii)* Management and Organizational, *(iii)* Technical and Security, *(iv)* Ethical and Societal, and *(v)* Sector–Specific. Each category contributes controls relevant to one or more layers of the Stack.

9.4 Category I: Regulatory and Legal Frameworks

- **EU Artificial Intelligence Act (2024)** – defines risk tiers and obligations corresponding to Layers 6 and 7.
- **OECD AI Principles (2019)** – provide overarching policy guidance that informs all layers, particularly 7.

- **UNESCO Recommendation on the Ethics of AI (2021)** – establishes human rights and dignity criteria reflected at Layer 7.
- **U.S. Executive Order on Safe, Secure, and Trustworthy AI (2023)** – operationalizes national oversight, mainly for Layers 4–7.

9.5 Category II: Management and Organizational Standards

- **ISO/IEC 42001:2024** – Management system standard for AI; applies to Layers 5–7.
- **ISO/IEC 38507:2022** – Governance of IT for organizational leadership; relates to Layers 6–7.
- **NIST AI Risk Management Framework (2023)** – cross–cutting risk processes spanning Layers 3–7.
- **CIS Critical Security Controls v8 (2023)** – foundational cybersecurity practices relevant to Layers 1–5.

9.6 Category III: Technical and Security Standards

- **ISO/IEC 27001:2022** – Information security management; foundational to Layers 1–5.
- **ISO/IEC 23894:2023** – AI risk management; provides metrics and methods for Layers 3–7.
- **IEEE P7000 Series** – suite of ethical and technical standards informing Layers 4–7.
- **MITRE ATLAS and MITRE ATT&CK** – adversarial threat modeling used for risk identification at Layers 2–5.

9.7 Category IV: Ethical and Societal Frameworks

- **Design Justice Network Principles (2020)** – inclusive design considerations at Layers 4–7.
- **UN Sustainable Development Goals (2015)** – long–term societal alignment, guiding Layer 8.
- **World Economic Forum AI Governance Toolkit (2022)** – policy translation for Layers 6–7.

- **Partnership on AI Best Practices** – collaborative governance models across Layers 5–7.

9.8 Category V: Sector–Specific Standards

- **FDA Good Machine Learning Practice Guidance (2023)** – applicable to medical AI at Layers 6–7.
- **HIPAA Privacy Rule (U.S.)** – health data governance for Layers 3 and 6.
- **FINRA AI Guidelines (2024)** – financial compliance requirements corresponding to Layers 5–7.
- **CMMC 2.0 and NIST SP 800–53** – defense and contractor security controls at Layers 1–5.

9.9 Cross–Layer Concordance

Table 2 summarizes how the major frameworks align with each layer of the AI OSI Stack. The table is descriptive and non–exhaustive; organizations should tailor mappings to their operational environment.

Layer	Primary Framework Alignments
1	ISO/IEC 27001, CIS v8, CMMC 2.0, ESG Reporting Standards.
2	ISO/IEC 23894, IEEE P7009, MITRE ATLAS.
3	NIST AI RMF (Map/Measure), OECD AI Principles, EU AI Act Articles 9–10.
4	IEEE P7000–P7010, NIST AI RMF (Manage), Design Justice Network Principles.
5	ISO/IEC 42001, ISO/IEC 27018, EIA Protocol (as defined herein).
6	FDA GMLP, HIPAA, FINRA, EU AI Act Annex III, OECD AI Impact Assessment Guidelines.
7	OECD AI Principles, UNESCO Ethics of AI, ISO/IEC 38507, ISO/IEC 42001, WEF Toolkit.
8	UN Sustainable Development Goals, national AI strategies, educational and public awareness initiatives.

Table 2: Illustrative cross–layer concordance between the AI OSI Stack and major international frameworks.

9.10 Interpretation

This mapping demonstrates that most global frameworks address overlapping aspects of governance but rarely specify how technical and institutional controls interact. The AI OSI Stack provides that missing connective tissue. It clarifies the boundaries between layers, identifies where existing frameworks already provide sufficient coverage, and highlights where additional standards or artifacts are required. By embedding external standards into a single layered architecture, the Stack facilitates regulatory interoperability and reduces compliance complexity.

10 Foresight and Adaptive Governance

10.1 Purpose

Artificial intelligence systems evolve faster than conventional governance cycles. Foresight methods provide structured anticipation of technological, social, and environmental change. This section introduces approaches that enable the AI OSI Stack to remain adaptive across short, medium, and long time horizons. Foresight is not prediction; it is disciplined preparation for multiple plausible futures.

10.2 Adaptive Governance Principles

Adaptive governance treats policies and standards as living instruments that must learn from their own implementation. Within the Stack, adaptation occurs through three feedback mechanisms:

a) **Temporal Review:** scheduled re–evaluation of governance artifacts and model behavior at defined intervals.

b) **Cross–Layer Feedback:** communication of audit results from higher layers to technical layers where remediation occurs.

c) **Scenario Updating:** integration of foresight results into design and regulatory planning.

These feedback channels transform governance from static compliance into continuous learning.

10.3 Panarchic Adaptive Cycles

The concept of panarchy, drawn from resilience theory, describes systems that cycle through phases of growth, accumulation, release, and reorganization. AI ecosystems follow similar dynamics: rapid innovation (growth), consolidation by large actors (accumulation), disruptive events or failures (release), and emergence of new paradigms (reorganization). Effective governance must anticipate these cycles and design safeguards for each phase.

- During **growth**, governance emphasizes experimentation and documentation of assumptions.
- During **accumulation**, attention shifts to transparency and preventing monopolistic control.

- During **release**, post–incident review and learning procedures are critical.
- During **reorganization**, foresight and inclusive participation guide rebuilding.

Embedding these stages into review policies helps institutions avoid both regulatory lag and over–correction.

10.4 Futures Cone and Three Horizons Analysis

Foresight practice often uses the futures cone, representing possible, plausible, and preferable futures, and the Three Horizons model, distinguishing present systems (H1), transitional innovations (H2), and long–term transformations (H3). Applied to AI governance:

- **H1:** Current systems—incremental improvements to existing oversight methods.
- **H2:** Transitional—emergence of hybrid human–AI decision structures and new audit technologies.
- **H3:** Transformative—institutionalization of continuous, machine–assisted governance.

Periodic workshops or simulation exercises can position an organization's policies within these horizons and plan transitions deliberately.

10.5 Design Justice and Inclusive Oversight

Resilient governance requires participation from those most affected by AI systems. Design justice principles provide criteria for inclusive decision–making:

1. Include affected communities in early design stages.
2. Distribute benefits and burdens equitably.
3. Recognize and value diverse forms of expertise.
4. Maintain transparency about decision procedures and representation.

Within the Stack, these principles are implemented primarily at Layers 4–7, ensuring that control logic, applications, and oversight bodies represent the public interest.

10.6 Temporal Feedback Loops

Every governance artifact contains temporal metadata that allows retrospective evaluation. Auditors and policymakers can query the sequence of ILEs, DRRs, and OAMs to analyze how decisions evolved over time. This creates a measurable feedback loop between policy and technical implementation, supporting continuous improvement rather than episodic reform.

10.7 Adaptive Governance Timeline

Organizations adopting the Stack should establish review intervals matched to system criticality:

- **Short–term (0–12 months):** operational monitoring, artifact validation, and immediate risk mitigation.
- **Medium–term (1–3 years):** architectural review, regulatory alignment, and stakeholder consultation.
- **Long–term (3+ years):** strategic foresight studies, research investment, and revision of governance doctrine.

These cycles ensure that AI governance remains synchronized with both technological advancement and social expectation.

11 Implementation and Adoption Pathways

11.1 Enterprise Integration

Organizations can incorporate the AI OSI Stack into their development and compliance pipelines without restructuring entire operations. Implementation typically begins with a gap analysis: mapping existing controls to Stack layers, identifying missing governance artifacts, and establishing documentation procedures. Enterprises may phase adoption layer by layer, starting where risk exposure is highest.

11.2 Policy and Regulatory Adoption

Regulators can use the Stack as a reference taxonomy for aligning laws and oversight functions. Mapping legal obligations to layers clarifies jurisdictional boundaries and prevents duplication of enforcement. For instance, infrastructure agencies may focus on Layers 1–3, while consumer–protection bodies concentrate on Layers 6–7.

11.3 Research and Education

Academic institutions can employ the Stack as a teaching framework for interdisciplinary AI governance curricula. Research programs can test new methods for artifact generation, audit automation, and cross–framework interoperability. Because the Stack is open and modular, empirical findings can be contributed back through version updates.

11.4 Standards and Auditing Bodies

Auditors and standards organizations may reference this document to define certification procedures. The governance artifacts formalized herein provide concrete evidence requirements for audits, enabling consistent assessments across sectors.

11.5 Public and Civil Society Engagement

Civil society groups can use the Stack as an interpretive tool to evaluate claims of transparency or ethical compliance. By translating technical governance into accessible layers, the framework supports informed public dialogue and accountability.

11.6 Certification and Verification

Entities may voluntarily declare conformance with the Stack by producing the required artifacts and undergoing independent review. A minimal conformance statement should identify applicable layers, list associated artifacts, and include signatures of responsible officers.

12 Authorship and Custodianship

12.1 Statement of Authorship

All conceptual, architectural, and terminological frameworks—*The AI OSI Stack*, *Persona Architecture*, *Epistemology by Design*, and *AEIP*—constitute the **AI Governance Trinity**, authored and owned by **Daniel P. Madden**. This version formalizes their interdependence as an integrated epistemic governance architecture.

12.2 Custodianship Commitments

The custodian undertakes to:

a) Maintain and publish authenticated updates of this standard using timestamped DOIs.

b) Preserve integrity of terminology and framework relations across all derivative references.

c) Promote research and educational use while protecting against unauthorized commercialization or alteration.

12.3 License

12.4 Licensing and Authorship Clarification

Earlier drafts of the *AI OSI Stack* (Versions 1–3) were distributed under Creative Commons Attribution (CC BY) to ensure academic accessibility and historical transparency. As of Version 4, this framework is licensed under CC BY–NC–ND 4.0 to prevent unauthorized commercialization, modification, or derivative reuse. Users are invited to cite freely but may not alter, adapt, or sell the material. Requests for partnership or licensing should be directed through the author's contact page at https://danielpmadden.com/contact.

12.5 Citation and Referencing

The canonical citation for this document is:

> Madden, D. P. (2025). *The AI OSI Stack: Layered Architecture and Governance Framework for Accountable Artificial Intelligence (Version 4.0 Expanded).* Zenodo. https://doi.org/10.xxxxx/zenodo.xxxxx

Citations should reference the version number to maintain provenance in scholarly and regulatory use.

13 Closing Reflection

13.1 Architecture as Accountability

Accountability in artificial intelligence cannot rely solely on after–the–fact investigation. It must be built into the architecture of the systems themselves. The AI OSI Stack advances this idea by making governance coextensive with design. Every layer of the architecture carries its own evidence of integrity, and those evidences connect across interfaces to form a complete audit chain. In this structure, accountability is no longer a policy layer applied from above but an intrinsic property of the system.

13.2 Trust as Infrastructure

Trust emerges when each participant in a complex system can verify the reliability of others without requiring perfect knowledge or constant oversight. The Stack operationalizes this through portable governance artifacts and verifiable reasoning protocols. Transparency and dignity are treated as infrastructural design constraints rather than optional virtues. When implemented consistently, this approach transforms trust from a rhetorical objective into an auditable engineering outcome.

13.3 Future Development

The author intends to continue refining the Stack through open consultation, empirical testing, and alignment with evolving international standards. New applications, sectoral profiles, and interoperability demonstrations will extend the framework while preserving its core structure of layered accountability.

Appendices

A Glossary of Core Concepts

Accountability Assurance Mechanism (AAM)
A process ensuring that all decisions can be traced to responsible actors with documented rationale.

Audit Artifact
Any structured record produced to demonstrate compliance or integrity within a layer.

Dignity as Constraint
Design principle ensuring that AI systems never instrumentalize human beings or diminish human agency.

Transparency as Infrastructure
Treating transparency as a built–in capability of systems, not as an external reporting function.

Trust Portability
The ability to transfer verified trust and audit records across systems and institutions without loss of integrity.

Semantic Version Control (SVC)
System for maintaining time–linked accountability across evolving artifacts, models, and policies.

B Layer-to-Standard Concordance (Summary)

For full details, see Table 2 in the body of the document. This appendix provides a compact overview linking each layer to primary governance frameworks and example artifacts.

Layer	Representative Frameworks	Typical Artifacts
1	ISO/IEC 27001, CIS v8	Hardware bills of materials, energy audits.
2	ISO/IEC 23894, IEEE P7009	Architecture specification sheets, interpretability documentation.

3	NIST AI RMF, OECD Principles	Dataset lineage reports, fairness evaluations.
4	IEEE P7000, Design Justice Principles	Role definitions, refusal logic policies.
5	ISO/IEC 42001, EIA Protocol	API policies, interface attestation records.
6	FDA GMLP, HIPAA	Deployment audits, incident reports.
7	OECD AI Principles, UNESCO Ethics of AI	Oversight Audit Memos, transparency reports.

C Version Change Log

- **Version 1.0 (2025 Q3)** – Initial three–layer concept draft.
- **Version 2.0 (2025 Q3)** – Integration of governance artifact prototypes and foresight principles.
- **Version 3.0 (2025 Q4)** – Full seven–layer architecture with preliminary audit schema.
- **Version 4.0 (2025 Q4)** – Consolidation of previous material, updated definitions, global framework mapping, and refined temporal integrity model.

D References

Key public standards and publications referenced in this document include:

- ISO/IEC 42001:2024, *Artificial Intelligence – Management System Standard.*
- NIST AI Risk Management Framework, Version 1.0, 2023.
- OECD AI Principles, 2019.
- UNESCO Recommendation on the Ethics of Artificial Intelligence, 2021.
- EU Artificial Intelligence Act, 2024.
- IEEE P7000 Series on Ethical Design of Autonomous and Intelligent Systems.
- Design Justice Network Principles, 2020.

E About the Author

Daniel P. Madden is an independent researcher focused on artificial intelligence governance, epistemic infrastructure, and design ethics. His work integrates systems architecture with human–centered policy design to produce verifiable, transparent, and accountable AI ecosystems. He has developed a family of frameworks including the AI OSI Stack, the Role–Bound Cognitive System Design Framework, the Reasoning Integrity Engineering Methodology, and the Epistemic Interoperability and Audit Protocol. Further information and publications are available at https://danielpmadden.com.

Contact:
Daniel P. Madden
Website: https://danielpmadden.com

License