> ## Normative Language Notice
> This specification uses normative language consistent with ISO/IEC 42010 and NIST conventions.
> SHALL denotes mandatory requirements, SHOULD denotes strong recommendations, and MAY denotes optional practices.
> Interpretations of this document must preserve authorial intent fidelity to layered accountability, epistemic integrity, and human dignity as design constraints.

The AI OSI Stack v4 Test Integration Draft
A Governance Blueprint for Scalable and Trusted AI
Author: Daniel P. Madden, Independent AI Researcher
Date: November 2025
Version: Version 4.0-Test (Supersedes all previous AI OSI Stack versions in this repository.)

> **NOTE:** This v4 integrates: maturity modeling, risk catalogue, temporal legitimacy, affective constraints, cognitive diversity, distributed oversight, sectoral extensions, and AEIP alignment.

Table of Contents

1. Introduction and Purpose
The AI OSI Stack v4 consolidates Daniel P. Maddens governance architecture into a single canonical reference that aligns operational controls, epistemic integrity, and human dignity across the entire lifecycle of intelligent systems. This Test Integration Draft resolves fragmentation from earlier versions by embedding Persona Architecture, Epistemology by Design, and the AI Epistemic Infrastructure Protocol (AEIP) directly within the layered stack. The intent is to provide institutions with a deployable governance blueprint compatible with NIST AI RMF, ISO/IEC 42001, the EU AI Act, and the TRUST Framework while remaining adaptable to sector-specific constraints.

Jurisdictional Neutrality and Interpretive Authority
This framework is jurisdiction-neutral and intended for international application.
In the event of conflicting legal or linguistic interpretations, interpretive authority SHALL defer
to the authors defined principles of transparency, accountability, and dignity by design

not to localized or translated semantics.

2. Normative Definitions and Acronyms
This document adopts normative language consistent with international governance standards. SHALL denotes mandatory requirements; SHOULD indicates strong recommendations; MAY introduces optional practices. Key acronyms include: AEIP (AI Epistemic Infrastructure Protocol), AGM (Adaptive Governance Metrics), CDI (Cognitive Diversity Index), GDS (Governance Disclosure Statement), ILE (Integrity Ledger Entry), IMM (Implementation Maturity Model), OAM (Oversight Action Memorandum), OIR (Open Implementation Registry), and Persona Architecture (PA). The lexicon harmonizes with prior publications such as *Persona Architecture: Designing Role-Specific AI Systems for Accountability and Trust* and *Epistemology by Design: Embedding Reasoning Integrity in AI Systems*.

3. Architectural Foundations and Methodology
The stack is engineered as infrastructure rather than policy prose. Its methodology interleaves architectural decomposition with governance artifacts that provide audit-ready evidence. Layered controls are validated via AEIP-compliant exchanges, ensuring that reasoning states, decision artifacts, and socio-technical context remain inseparable. The methodology favors design patterns that instantiate accountability at system inception, echoing the stance articulated in *The AI OSI Stack: A Governance Blueprint for Scalable and Trusted AI*. Each layer is treated as a contract surface that binds risk controls, cognitive safeguards, and affective constraints.

4. Layer Interdependence and Power Geometry
Layer boundaries are permeable to guarantee system cohesion. Power geometry recognizes that decision authority, data provenance, and execution logic concentrate differently across layers. The stack addresses vertical power flow (from infrastructure to user interaction) and horizontal influence (across partner ecosystems) to mitigate capture and concentration risks. Persona Architecture anchors Layer 4 by establishing accountable instruction sets, while AEIP operationalizes Layer 5 as the protocol spine that transmits epistemic commitments. Governance artifacts at Layer 7 distribute power by exposing stewardship evidence to external auditors and public stakeholders.

5. Layer-by-Layer Framework
1. **Layer 0 Civic Mandate (Optional):** Establishes jurisdictional legitimacy and social license. It captures legislative authorization, community compacts, and public trust obligations.
2. **Layer 1 Physical and Compute Substrate:** Covers data centers, chip governance, and reliability engineering aligned with ISO/IEC 27001 controls. Monitoring includes resilience audits and energy accountability.
3. **Layer 2 Data Stewardship:** Ensures provenance, consent, and representational justice. Epistemology by Design applies here through provenance ledgers and reasoning lineage documentation.
4. **Layer 3 Model Development:** Governs training protocols, evaluation harnesses, and red-teaming. Includes model cards, risk statements, and fairness reviews.
5. **Layer 4 Instruction and Control:** Persona Architecture delivers structured roles, refusal logic, and affect management to ensure the AI acts within declared mandates.
6. **Layer 5 Reasoning Exchange and Interface:** AEIP enforces structured dialogue between AI services, auditors, and human supervisors using signed reasoning packets and integrity fingerprints.
7. **Layer 6 Deployment and Integration:** Covers runtime environments, incident response, and change control across organizational ecosystems.
8. **Layer 7 Governance Publication:** Releases GDS packages, ILE updates, and public accountability artifacts ensuring transparency, traceability, and trust.

6. Governance Artifacts and Audit Infrastructure
The stack relies on reusable artifacts: Governance Disclosure Statements (GDS) synthesize cross-layer obligations; Integrity Ledger Entries (ILE) notarize decisions and maturity attestations; Oversight Action Memoranda (OAM) document interventions and remediations;

Temporal Review Records (TRR) capture version checks; Interpretive Trace Packages (ITP) reconstruct evidentiary reasoning; and Decision Rationale Records (DRR) preserve architectural justification. AEIP provides the transport for these artifacts, guaranteeing interoperable audit traces. Artifacts SHALL be version-controlled, cryptographically signed, and exposed through Layer 7 portals. Canonical JSON-LD definitions are maintained in the `schemas/` directory (AEIP-Schema-1.0-2025-11), with validation exemplars residing in `tests/aeip_schema_validation.py`.

7. Temporal Integrity and Semantic Version Control
Governance is inherently temporal. Every change to Persona Architecture configurations, epistemic safeguards, or AEIP schemas requires semantic versioning anchored to custodial review. Temporal integrity is sustained through synchronized clocks, validity windows for each artifact, and automated drift alerts triggered by reasoning fingerprint deviation. Semantic version control couples model updates with governance obligations to prevent silent regression.

8. Global Framework Interoperability Map
The stack crosswalks to prevailing frameworks: Layer 2 aligns with NIST AI RMF Data and Data Governance; Layer 3 maps to ISO/IEC 42001 operational controls; Layer 4 satisfies EU AI Act transparency and instruction mandates; Layer 5 implements TRUST Framework interoperability principles. The interoperability map establishes translation guides so that organizations can certify once and comply many times, minimizing audit friction while preserving rigor.

9. Adaptive Governance and Foresight
Adaptive governance assumes AI systems encounter dynamic contexts. Scenario planning, foresight scanning, and socio-technical sensing are embedded at Layer 6 and Layer 7. Persona Architecture supplies the adjustable knobs, AEIP carries foresight signals, and Governance Councils interpret the signals to recalibrate controls. The stack mandates quarterly foresight reviews incorporating emergent risks, policy shifts, and cultural feedback loops.

10. Implementation and Adoption Pathways
Organizations SHALL follow a staged adoption: inventory existing systems, map them to the stack, implement Persona Architecture for instruction surfaces, deploy AEIP nodes, and publish the first GDS. Sectoral extensionssuch as GERDY v1 for education or ARCHY v1 for healthcareserve as reference templates. Adoption pathways emphasize co-design with affected communities, aligning with the dignity-first ethos introduced in *Persona Architecture*.

11. Custodianship and Change Management
Custodianship rests with a designated Governance Council empowered to approve changes, review audits, and steward public accountability. Change management integrates AEIP change requests, DRR updates, and OAM tracking. Custodians SHALL maintain a change ledger, ensure training for human overseers, and coordinate with regulators to synchronize compliance submissions. Custodianship is designed to be portable across organizations while

preserving Daniel P. Maddens architectural intent.

12. Implementation Maturity Model (IMM)
The IMM defines six discrete levels of governance capability:
- **Level 0 No Governance:** Ad hoc AI usage without documented controls. Required artifacts: none. Expected audit frequency: incident-driven only.
- **Level 1 Basic / Reactive Governance:** Minimal policies and reactive incident handling. Required artifacts: baseline GDS, initial DRR. Expected audit frequency: annual self-assessment.
- **Level 2 Integrated Layer Controls:** Stack-aligned controls implemented across Layers 14 with Persona Architecture instantiated. Required artifacts: Persona briefs, AEIP readiness report, TRR baseline. Expected audit frequency: semiannual internal audits.
- **Level 3 Adaptive / Temporal Governance:** Temporal integrity tooling, drift monitoring, and AEIP ledger integration across Layers 16. Required artifacts: temporal audit log, OAM register, IMM attestation. Expected audit frequency: quarterly governance reviews.
- **Level 4 Trust Infrastructure:** Full AEIP deployment, public Layer 7 disclosures, and cross-jurisdictional interoperability. Required artifacts: comprehensive GDS, signed ILE packages, resilience playbooks. Expected audit frequency: continuous monitoring with biannual external audits.
- **Level 5 Cognitive Stewardship:** Cognitive diversity orchestration, affective constraint modules, and inter-organizational oversight. Required artifacts: CDI reports, dignity compliance attestations, governance simulation results. Expected audit frequency: continuous AEIP validation with annual third-party certification.

Organizations SHALL publish their IMM level annually as an Integrity Ledger Entry (ILE) to maintain transparency and accountability.

13. Risk Taxonomy and Control Catalogue
Each layer carries characteristic risks and prescribed controls:
- **Layer 1:** Riskshardware tampering, supply disruptions. Controlscertified facilities audits, redundant energy governance. ArtifactsGDS, DRR.
- **Layer 2:** Risksdata poisoning, consent erosion. Controlsprovenance validation, consent traceability. ArtifactsITP (Interpretive Trace Package), DRR.
- **Layer 3:** Risksmodel bias, evaluation blind spots. Controlscomposite testing, adversarial benchmarking. ArtifactsOAM, GDS.
- **Layer 4:** Risksprompt injection, persona drift, unethical affect. ControlsHeartwood Core enforcement, refusal logic tests, persona regression suite. ArtifactsITP, DRR, GDS.
- **Layer 5:** Risksledger desynchronization, reasoning packet corruption. ControlsAEIP handshake validation, cross-node replay defenses. ArtifactsILE, OAM.
- **Layer 6:** Risksdeployment drift, integration conflicts. Controlschange gating, blue/green governance checkpoints. ArtifactsTRR, OAM.
- **Layer 7:** Risksopaque reporting, delayed disclosures. Controlspublic GDS cadence, audit-ready publishing pipeline. ArtifactsGDS, ILE.

Controls SHALL be catalogued with traceable references to the artifact classes:

Interpretive Trace Package (ITP), Decision Rationale Record (DRR), Governance Decision Summary (GDS), Oversight Audit Memo (OAM), Integrity Ledger Entry (ILE).

14. Temporal Legitimacy Framework
Temporal legitimacy binds authority to time. Each governance artifact receives a defined validity window during which its claims remain authoritative. Temporal amnesty clauses allow constrained grace periods to remediate violations without punitive escalation when self-disclosed promptly. Drift thresholds are codified: a 5% change in reasoning fingerprint triggers mandatory review and potential rollback. Quarterly temporal audits evaluate alignment between declared versions and deployed systems. All temporal legitimacy events SHALL be recorded as AEIP-compatible ledger entries to guarantee verifiable lineage.

15. Adaptive Governance Metrics (AGM)
Adaptive metrics provide continuous insight:
- **Transparency Ratio (TR):** Published governance artifacts required artifacts. Recorded in GDS.
- **Governance Coverage Score (GCS):** Implemented controls identified controls for the operational scope. Logged alongside IMM attestations.
- **Drift Index (DI):** Aggregated delta of performance and reasoning fingerprints normalized over time. Stored in TRR and linked via AEIP.
- **Dignity Compliance Rate (DCR):** Verified adherence to affective constraints total evaluated interactions. Captured within GDS and ILE updates.

Organizations SHALL report AGM values quarterly and expose them through Layer 7 dashboards for stakeholder review.

16. Human Dignity and Affective Constraint Module
Dignity operates as a non-negotiable constraint. Persona Architecture requires agents to declare identity and intent at session initiation, ensuring clarity of role and accountability. The module prohibits intimacy simulation, manipulative sentiment scripts, and coercive affective cues. Violations trigger automatic Oversight Action Memoranda and public disclosure through Layer 7 portals. This preserves the dignity as constraint principle articulated across Daniel P. Maddens corpus while providing enforceable levers for auditors.

17. Cognitive Diversity Index (CDI)
The CDI measures epistemic plurality. Defined as CDI = (distinct epistemic families engaged) (total governance decisions), it encourages inclusive stewardship. Inspired by the cognitive scaffolding demonstrated in CASSI v1, organizations target a CDI of 0.4 or higher for high-stakes domains. CDI reporting SHALL accompany IMM and AGM disclosures, enabling auditors to detect monocultures that jeopardize resilience.

18. Governance Simulation and Stress Testing
Annual governance fire drills validate readiness for compound failures. Simulations incorporate synthetic data breaches, sudden policy shifts, and model drift scenarios.

Outputs are consolidated into an OAM-Resilience Report that documents response quality, time-to-remediation, and lessons learned. AEIP facilitates replay and external verification, ensuring simulations translate into tangible improvements.

19. Distributed Oversight Architecture
Distributed oversight balances local autonomy with global assurance. The architecture comprises three tiers: Local Ledger Nodes maintain operational records; Regional Custodians aggregate sector or jurisdiction data; and a Global Trust Anchor synchronizes standards without centralizing control. AEIP secures inter-tier exchanges, enabling cross-jurisdictional audits and mutual recognition agreements.

20. Ethical AI Supply Chain Framework
Supply chain governance requires all vendors to publish Stack-aligned Governance Cards detailing Persona Architecture conformance, epistemic safeguards, and AEIP integration status. Procurement processes SHALL mandate Stack-Aligned Reports assessing each suppliers IMM level. All supply chain attestations are logged via ILE to preserve accountability across the lifecycle.

21. Human-in-the-Governance Loop
Human stewards remain integral. Mandatory checkpoints occur during training data approval, instruction design, deployment readiness, and audit sign-off. Escalation paths route unresolved issues to the Governance Council, which can issue OAM directives or suspend operations. AEIP records human approvals to ensure traceability and to prevent automation from eroding oversight.

22. Societal Foresight and Temporal Pluralism
The stack embeds temporal pluralism to account for divergent futures. Scenario planning explores three anchor futures for 20302040: **Fragmented Governance**, where jurisdictions diverge; **Unified AEIP**, where interoperability prevails; and **AIHuman Co-Governance**, where shared stewardship emerges. Inter-temporal arbitration mechanisms reconcile decisions that impact multiple timelines, ensuring present-day actions remain legitimate across futures.

23. Open Implementation Registry (OIR)
The OIR is a public or semi-public ledger enumerating adopters of the AI OSI Stack. Each entry records organization name, IMM level, adoption date, and referenced artifacts. Entries SHALL link to published GDS packages and ILE hashes, enabling external verification while protecting sensitive details through selective disclosure.

24. Decommissioning and End-of-Life Governance
Retiring AI systems requires deliberate closure. Organizations SHALL produce a Final GDS summarizing operational history, issue a Closure ILE indicating shutdown status, perform memory redaction consistent with privacy commitments, and publish a public AI epitaph explaining lessons learned. Ethical end-of-life practices honor affected communities and prevent orphaned systems from resurfacing without governance.

Appendix A. Glossary
- **AEIP:** Protocol for exchanging reasoning artifacts with audit guarantees. Detailed in *AEIP-00: The AI Epistemic Infrastructure Protocol*.
- **Governance Disclosure Statement (GDS):** Public artifact describing compliance posture.
- **Integrity Ledger Entry (ILE):** Signed record anchoring governance claims in time.
- **Oversight Action Memorandum (OAM):** Formal notice of intervention or remediation.
- **Persona Architecture (PA):** Role-specific design methodology ensuring accountable AI personas.

Appendix B. Layer-to-Standard Concordance
- **Layer 1:** ISO/IEC 27001, NIST CSF.
- **Layer 2:** NIST AI RMF Data function, GDPR Articles 1322.
- **Layer 3:** ISO/IEC 42001 Clause 8, NIST AI RMF Measure.
- **Layer 4:** EU AI Act transparency requirements, Persona Architecture guidelines.
- **Layer 5:** TRUST Framework interoperability principles, AEIP handshake specifications.
- **Layer 6:** ITIL Change Management, NIST SP 800-53 IR controls.
- **Layer 7:** Open Government data publication norms, public accountability statutes.

Appendix C. Version Change Log
- **v4.0-Test (November 2025):** Integrated maturity model, risk catalogue, temporal legitimacy, affective constraints, cognitive diversity, distributed oversight, supply chain governance, and new registries. Supersedes v1v3 and prior drafts.

Appendix D. References
1. Daniel P. Madden, *Persona Architecture: Designing Role-Specific AI Systems for Accountability and Trust*.
2. Daniel P. Madden, *Epistemology by Design: Embedding Reasoning Integrity in AI Systems*.
3. Daniel P. Madden, *The AI OSI Stack: A Governance Blueprint for Scalable and Trusted AI*.
4. Daniel P. Madden, *AEIP-00: The AI Epistemic Infrastructure Protocol*.
5. NIST, *AI Risk Management Framework* (2023).
6. ISO/IEC, *42001:2023 Artificial Intelligence Management System*.
7. European Union, *AI Act* (2024 provisional agreement).
8. U.S. Department of Commerce, *TRUST Framework* (2024).

Custodianship and Authorship
Daniel P. Madden retains moral and intellectual authorship of this framework.
This work SHALL NOT be modified, translated, or reissued under altered terminology without written consent.
Derived works or alternative semantic renderings that could misrepresent intent SHALL be considered
non-conformant and unauthorized under the CC BY-NC-ND 4.0 License.