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
Revised and expanded October 2025

Abstract

Artificial intelligence is moving from discrete products to systemic infrastructure. Yet most AI is still designed and governed as if it were a single opaque system, producing concentration risk, weak accountability, and interface monopolies. This paper introduces the **AI OSI Stack** — a seven-layer architectural and governance framework that clarifies how AI is built, where risks concentrate, and how trust can be made portable. Inspired by the OSI model in computer networking, the Stack separates AI systems into: (1) Physical / Hardware, (2) Model Architecture, (3) Training / Optimization, (4) Instruction / Control, (5) Interface / Protocol, (6) Application, and (7) Governance / Trust. By treating governance as a design-time feature rather than a compliance afterthought, the framework enables targeted regulation, auditable decision artifacts, persona-based safety, and protection against interface monopolies. It is intended for AI labs, policymakers, enterprise architects, and standards bodies seeking operational clarity for trustworthy AI.

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

Contents

1 Architectural Methodology: How to Use the Stack

The AI OSI Stack functions not only as a descriptive model but as a procedural method for mapping, auditing, and governing complex AI systems. Its purpose is to make architecture actionable. Practitioners in research, policy, or enterprise environments can apply the same sequence of steps to locate risks, assign accountability, and trace decisions through the system. The method mirrors the structure of the Stack itself: clear boundaries, documented interfaces, and evidence at every layer.

1.1 Step 1: Define the System Boundary

The first requirement is to identify the boundaries of the system under review. A boundary is not only technical but institutional: it marks where control and responsibility begin and end. A boundary definition should specify:

- The type of system (e.g., large language model, agent runtime, sectoral application).
- The primary delivery surface (API, interface, or embedded module).
- The actors who control training, deployment, and governance functions.

This boundary statement becomes the reference point for all subsequent layer analysis.

1.2 Step 2: Trace Downstream Impact

The second step is to identify where the system produces value and where it may cause harm. This inquiry typically resides in Layers 6 and 7: the application and governance layers. It asks who is affected, how outputs are used, and whether there are social, ethical, or regulatory implications. Downstream tracing translates technical change into human consequence.

1.3 Step 3: Trace Upstream Dependencies

Once the outputs are mapped, the analysis moves downward through the stack. Each layer is interrogated for its dependencies:

- Which API or protocol (Layer 5) mediates access?
- Which model architecture (Layer 2) and training corpus (Layer 3) underlie the system?
- Which hardware and data-center infrastructure (Layer 1) host it, and under what jurisdiction or export controls?

This upstream trace exposes where concentrated risk or dependency may reside and establishes the causal chain linking lower-layer design to higher-layer outcomes.

1.4 Step 4: Produce a Stack-Aligned Report

Each layer is documented using a uniform schema of four fields:

(1) **Actor:** The entity controlling that layer (lab, vendor, regulator, or standards body).

(2) **Risk:** The dominant failure mode or concentration hazard.

(3) **Evidence:** The audit artifact, model card, or decision record verifying behavior.

(4) **Regulator or Standard:** The governing authority or framework relevant to that layer.

The completed report forms a multidimensional audit log that can be attached to procurement reviews, compliance filings, or incident investigations.

1.5 Step 5: Version the System

AI systems evolve. To preserve accountability over time, each system maintains semantic versioning across its layers. Major changes (architecture redesigns at Layer 2) increment the first digit. Minor changes (alignment or control updates at Layer 4) increment the second. Documentation improvements or governance updates (Layer 7) increment the third. This practice ensures that when a behavior shifts, the corresponding artifact can be traced to its exact version context.

1.6 Step 6: Generate Cross-Layer Governance Maps

Using the data collected, practitioners create a Governance Map linking technical components to their governing authorities. Each node corresponds to a layer; each edge represents a dependency or regulatory relationship. The resulting map visualizes where law, policy, or oversight currently lacks coverage and highlights opportunities for interoperability between frameworks such as NIST AI RMF, ISO/IEC 42001, and the EU AI Act.

Outcome

Applied consistently, the methodology produces an evidence chain connecting infrastructure to impact. It transforms governance from a reactive audit into an active design

process. When organizations apply these steps, they can identify whether a failure originates in architecture, training, control, or deployment context. Regulators can pinpoint where to intervene. Most importantly, trust becomes portable because accountability is no longer inferred; it is recorded, versioned, and verifiable.

2 Layer Interdependence and Power Geometry

The seven layers of the AI OSI Stack are designed to be modular, but they are not independent. Real systems exhibit feedback loops, dependency chains, and power asymmetries that link the behavior of one layer to another. Understanding these relationships is essential for designing regulation and organizational policy that target the true levers of control rather than their symptoms. This section outlines the dominant forms of layer interdependence and the geometry of power that results from them.

2.1 Cross-Layer Dependencies

Although each layer serves a distinct purpose, technical and institutional dependencies create predictable pathways of influence:

- **L1 → L3: Physical Constraint.** Compute availability and energy capacity determine what can be trained. A shortage of accelerators or network bandwidth limits experimentation, reinforcing dependency on a few hyperscale providers.
- **L2 → L4: Architectural Constraint.** Design choices in the base model restrict how alignment and control can be implemented. For example, architectural opacity can render post-training ethical constraints unverifiable.
- **L5 → L6: Interface Lock-in.** Control of the API layer allows dominant firms to define which tools, plug-ins, and telemetry data are available to application developers. This creates structural dependency that stifles competition and innovation.
- **L7 → L3: Governance Feedback.** Policy decisions, such as data provenance mandates or reproducibility requirements, directly alter how models are trained and documented.
- **Cross-Layer Coupling:** A change in any layer reverberates across others. Hardware scarcity reshapes training incentives; interface monopolies influence which applications reach users; governance fragmentation undermines international auditability.

These dependencies form a dynamic equilibrium. Regulation or design that ignores them risks misalignment between capability, accountability, and public legitimacy.

2.2 Power Concentration and Chokepoints

The Stack reveals that power in the AI ecosystem is not evenly distributed. It pools at specific layers where control over scarce or irreplaceable resources provides leverage over the entire system.

- **Layer 1: Hardware Monopolies.** A handful of chip manufacturers and hyperscale cloud providers determine global access to training capacity. Export controls and proprietary interconnects create geopolitical dependencies.
- **Layer 2–3: Model and Data Concentration.** Frontier labs and large-scale data owners control both capability and epistemic scope. Without open benchmarks and lineage transparency, knowledge becomes gated property.
- **Layer 5: Interface Gatekeeping.** Platform firms that own the API surface can set pricing, dictate usage terms, and revoke access unilaterally. This layer is the modern equivalent of the telecommunications switchboard—a chokepoint where information and authority converge.
- **Layer 7: Regulatory Capture.** Institutions may adopt governance frameworks authored or lobbied for by the same entities they are meant to oversee. When compliance becomes a barrier to entry rather than a guarantee of safety, ethics turns into theater.

Power geometry is therefore layered rather than linear. Actors in lower layers (hardware and infrastructure) exert physical and economic control; actors in middle layers (models, APIs) control capability and access; actors in upper layers (governance) influence legitimacy and compliance narratives. Effective governance must align interventions across these dimensions rather than treating them separately.

2.3 Governance Levers and Counterbalances

The Stack translates structural power into specific governance levers:

- **Diversification.** Incentivize multiple suppliers at the hardware and model levels to prevent dependency collapse.
- **Transparency.** Mandate logging and public disclosure of Layer 5 access policies and Layer 3 data provenance.
- **Portability.** Require exportable audit trails so that Layer 6 applications can migrate between providers without losing accountability evidence.

- **Oversight Independence.** Establish independent auditing authorities and shared schemas for reporting incidents and control policies at Layer 7.

When implemented collectively, these levers convert asymmetrical control into a balanced ecosystem of shared accountability. They also enable a global trust architecture in which no single actor owns the definition of safety.

Interpretive Note

Power geometry is not an accident of technology but an outcome of design. Each layer carries not only technical functions but political and ethical weight. Recognizing this geometry allows governance to move from reactive crisis management to proactive architectural design. The AI OSI Stack transforms power analysis from a sociological critique into an operational map that policymakers, engineers, and institutions can act upon.

3 Integration with Governance Frameworks

The AI OSI Stack is designed to interoperate with existing international standards and ethical frameworks rather than replace them. It provides an architectural substrate on which these instruments can be anchored, ensuring that high-level principles correspond to concrete system boundaries. By mapping regulatory intent to technical layers, the Stack transforms abstract guidance into enforceable governance practice.

3.1 Mapping of Major Frameworks

OECD AI Principles (2019). These principles emphasize inclusive growth, human-centered values, transparency, robustness, and accountability. Within the Stack, they correspond primarily to Layer 7 (Governance / Trust) as overarching benchmarks for institutional legitimacy. The principle of transparency aligns with the Stack's structural demand for auditability, while robustness maps to the cross-layer discipline of reproducibility and control traceability.

UNESCO Recommendation on the Ethics of AI (2021). UNESCO's framework stresses cultural diversity, data sovereignty, human oversight, and environmental sustainability. These align with Layers 3 through 7. Data sovereignty governs training provenance (Layer 3) and deployment jurisdiction (Layer 1). Human oversight aligns with the Instruction and Control layer (Layer 4). Environmental sustainability embeds within the hardware and infrastructure layer (Layer 1), ensuring that energy use and ecological cost are treated as governance metrics, not operational afterthoughts.

Council of Europe Framework Convention on AI, Human Rights, and Democracy (2024 draft). This binding convention introduces obligations on explain-

ability, redress, and accountability. It occupies Layers 6 and 7, operationalizing human rights principles as system requirements. The Stack proposes that compliance artifacts such as Clarity Packages and Decision Cards can serve as standardized evidence for explainability and redress.

NIST AI Risk Management Framework (AI RMF 1.0, 2023). NIST's four core functions—govern, map, measure, and manage—can be explicitly distributed across the Stack. Mapping and measurement occur in Layers 1 through 5; management in Layers 6 and 7; and governance integrates across the entire vertical chain. The Stack therefore turns NIST's procedural logic into a layered execution plan.

ISO/IEC 42001:2023 (AI Management System Standard). This standard defines a management system for AI, establishing organizational controls and certification processes. The Stack provides the architectural surface for binding these controls. For example, ISO control clauses on design validation map to Layers 2 and 3, documentation controls to Layer 7, and audit logging to Layer 5.

ISO/IEC 38507 (Governance of IT and AI). This framework defines board-level responsibilities for AI oversight. Its prescriptions align with Layer 7, providing fiduciary structure for governance bodies that must reconcile technical transparency with institutional accountability.

COSO Enterprise Risk Management Framework. COSO provides enterprise risk tiers and internal control guidance. Its operational flow of identify–assess–respond–monitor integrates naturally with Layers 6 and 7, where organizational risk assessment and response mechanisms are encoded as part of Decision Insurance and Guardian Notes.

EU Artificial Intelligence Act (2024). The Act introduces a risk-based classification of AI systems. High-risk applications (such as healthcare, education, and employment) correspond to Layer 6. Documentation and logging requirements map to Layer 5, while data transparency obligations reach down to Layer 3. The Stack allows these obligations to be traced precisely, ensuring that risk classification remains connected to underlying design assumptions.

Singapore Model AI Governance Framework. This model emphasizes practical guidance for accountability, transparency, and human-centricity. It supports Layer 6 deployment standards and Layer 7 auditing, promoting sector-specific governance maturity.

3.2 Philosophical and Cognitive Foundations

Beyond institutional frameworks, the Stack is grounded in several theoretical traditions that illuminate cognition, agency, and temporality in artificial systems.

Extended Mind Theory (Clark & Chalmers). This theory views cognition as distributed across tools and environments. It underpins Layer 4b, where Persona Architecture formalizes how human reasoning extends into aligned AI systems. Each

persona is a bounded cognitive extension designed to preserve accountability and value coherence.

Actor-Network Theory (Latour). Agency emerges not from individuals but from networks of human and non-human actors. This informs the Stack's concept of Layer Blurring: power is not confined to technical nodes but flows through connections and dependencies. Governance, therefore, becomes the management of relationships rather than the policing of artifacts.

Temporal Epistemology and Media Archaeology (Parikka, Zielinski). Media systems store and distort time; AI intensifies this effect. This framework grounds the Stack's concept of Temporal Integrity, addressing how synthetic memory and version drift alter the chronology of trust. It justifies the creation of Semantic Version Control and temporal governance artifacts at Layer 7.

3.3 Foresight and Systems Design Frameworks

Foresight disciplines extend the Stack beyond present governance into future adaptation cycles.

Design Justice (Costanza-Chock). Centering marginalized voices ensures that design processes account for those historically excluded from technical decision-making. Within the Stack, this principle infuses Layers 4 through 7, shaping inclusivity and legitimacy in both control design and governance processes.

Panarchy and Adaptive Cycles (Gunderson & Holling). Systems evolve through cycles of growth, conservation, collapse, and renewal. This ecological lens aligns with the Stack's vision of AI mitosis: the branching of model lineages in a high-mortality environment. Governance must anticipate these cycles rather than resist them.

Futures Cone and Three Horizons Framework. Governance must account for possible, plausible, and preferable futures. Layer 7 incorporates foresight analytics into Decision Insurance frameworks, ensuring that strategic planning encompasses emerging risks and long-tail consequences.

Synthesis

The AI OSI Stack acts as a Rosetta Stone between disparate governance instruments. It translates principle into architecture, architecture into evidence, and evidence into trust. Where existing frameworks provide moral or procedural direction, the Stack provides spatial coordinates. This integration ensures that the global governance ecosystem for AI remains interpretable, compatible, and grounded in verifiable structure rather than rhetoric.

4 Layer-by-Layer Framework

The AI OSI Stack consists of seven layers that decompose artificial intelligence into discrete domains of responsibility. Each layer defines a class of components, dominant risks, and governance levers. Together they form an auditable map of how intelligence systems are built, deployed, and governed.

4.1 Layer 1: Physical / Hardware

Scope. Specialized accelerators (GPUs, TPUs, ASICs), networking, storage, energy infrastructure, and data-center geography.

Why It Matters. All higher layers depend on compute capacity and network resilience. Hardware concentration determines who can participate in frontier research. Geopolitical dependencies create hidden leverage points where national policy intersects with private infrastructure.

Dominant Risks.

- Supply-chain concentration among a handful of chipmakers.
- Export control volatility and geopolitical constraints.
- Environmental strain from energy and cooling requirements.

Governance Levers.

- Hardware provenance and lifecycle transparency.
- Incentives for open instruction set architectures and regional fabrication diversity.
- Environmental impact reporting and sustainability disclosure.

Audit Artifacts. Infrastructure bill of materials, data-center siting reports, supply diversity indices.

4.2 Layer 2: Model Architecture

Scope. Model families and structural innovations—transformers, diffusion models, neurosymbolic hybrids, sparse expert mixtures, and long-context variants.

Why It Matters. Architecture encodes capability ceilings and interpretability. It defines what kinds of control are possible in later layers.

Dominant Risks.

- Research centralization around a few frontier labs.
- Open-weight leakage without safety scaffolding.

- Opaque inductive biases that hinder verification.

Governance Levers.

- Capability and safety reporting per release.
- Funding for interpretable architectures and public benchmarks.
- Model cards and architectural diagrams as certification evidence.

Audit Artifacts. Model specification sheets, explainability evaluations, interpretability scorecards.

4.3 Layer 3: Training / Optimization

Scope. Data pipelines, dataset curation, deduplication, optimizer selection, fine-tuning, reinforcement learning from human feedback, distillation, and efficiency techniques.

Why It Matters. Training defines data provenance, model bias, and reproducibility. It is the layer where copyright, privacy, and labor ethics intersect.

Dominant Risks.

- Opaque data lineage and unlawful data inclusion.
- High training costs producing economic barriers to entry.
- Post-training fine-tuning erasing prior safety alignment.

Governance Levers.

- Training run registries and reproducibility standards.
- Dataset disclosure categories and lineage graphs.
- Structured model cards including training provenance.

Audit Artifacts. Hyperparameter logs, dataset summaries, reproducibility attestations.

4.4 Layer 4: Instruction / Control

Scope. Prompts, system messages, embeddings, tool policies, role-based controllers, RLHF, Constitutional AI, and persona architectures.

Why It Matters. This is where human intent meets model capability. Instruction design determines alignment, safety, and ethical coherence.

Sub-Layers.

- **L4a Control:** Prompt engineering, context management, refusal logic, tool selection, jailbreak resistance.
- **L4b Ethical / Value Reasoning:** Persona Architecture, Heartwood Safety Core, Dignity as Constraint, Decision Insurance.

Dominant Risks.

- Misalignment between policy and control logic.
- Prompt injection and covert persuasion.
- Hidden influence through tone or persona manipulation.

Governance Levers.

- Independent audits of control schemas.
- Publication of high-level refusal and escalation policies.
- Red-team evaluation of ethical controllers.

Audit Artifacts. Persona manifests, red-team results, control policy declarations, refusal logs.

4.5 Layer 5: Interface / Protocol

Scope. APIs, SDKs, middleware, orchestration frameworks, and agent runtimes that mediate model access.

Why It Matters. The interface layer is the new chokepoint of power. Whoever owns the runtime governs access, pricing, and telemetry.

Dominant Risks.

- Interface monopolies that constrain competition.
- Opaque logging and unverified execution traces.
- Automatic tool use without human visibility.

Governance Levers.

- Open protocol standards.
- Mandatory execution transparency and audit log exportability.
- Right to portability for agent workflows.

Audit Artifacts. API call histories, protocol compliance reports, execution trace maps.

4.6 Layer 6: Application

Scope. End-user systems: copilots, chat interfaces, decision aids, sector-specific tools in health, education, finance, manufacturing, and government.

Why It Matters. This is the point of contact between AI and society. Harm and value creation occur here.

Dominant Risks.

- Misuse in high-stakes domains without contextual oversight.
- Shadow AI deployments within enterprises.
- Overreliance on model output without disclosure of limitations.

Governance Levers.

- Context-specific deployment policies and red lines.
- Human-in-the-loop requirements for critical applications.
- User-facing transparency reports and limitation disclosures.

Audit Artifacts. Decision Insurance briefs, application risk assessments, human review logs.

4.7 Layer 7: Governance / Trust

Scope. The institutional layer where audits, compliance, oversight, and social legitimacy converge.

Why It Matters. Trust is not a property of models but of systems. It must be constructed through evidence, disclosure, and institutional memory.

Dominant Risks.

- Regulatory capture and ethics theater.
- Fragmented international governance regimes.
- Lack of long-term monitoring and accountability.

Governance Levers.

- Shared audit schemas and incident registries.
- Clarity Packages and Solomon Briefs as portable decision evidence.
- Governance Maps linking standards to system layers.
- Cross-jurisdictional alignment and temporal provenance tracking.

Audit Artifacts. Guardian Notes, Governance Maps, Stack-Aligned Reports, cryptographic attestations.

Synthesis

The seven layers together define the anatomy of control. Each is independently auditable and collectively accountable. The Stack's purpose is to shift governance from post hoc compliance to design-time architecture. By treating transparency and dignity as built-in constraints, it enables trust to move across systems—portable, verifiable, and durable through time.

5 Governance Artifacts and Audit Infrastructure

The upper layers of the AI OSI Stack require concrete evidence to transform governance from theory into practice. Governance artifacts are the operational outputs that make accountability portable. They ensure that decisions, assumptions, and trade-offs are recorded, interpretable, and transferable across institutions. This section defines the principal artifacts—Clarity Packages, Solomon Briefs, Governance Maps, and Stack-Aligned Reports—and explains how they collectively enable Trust Portability.

5.1 Clarity Packages

Definition. A Clarity Package is a structured artifact produced by interpretive engines that convert raw text or data into auditable meaning. Originally prototyped within the Compass workflow, it converts unstructured input into a structured JSON schema containing: (1) diagnostic patterns, (2) narrative structures, (3) semantic drift notes, and (4) recommended actions.

Purpose. To make interpretation auditable and repeatable. It operationalizes the principle of *Transparency as Infrastructure* by embedding provenance and interpretive trace directly into the analytical process.

Governance Function. Serves as Layer 7 evidence for interpretive decisions, policy diagnostics, and regulatory reporting. In cross-layer audits, Clarity Packages function as semantic black boxes that can be replayed and verified.

Example Fields.

```
{
  "source": "Policy Draft V3",
  "patterns": ["risk aggregation", "alignment drift"],
  "narratives": ["centralization → fragility"],
  "actions": ["publish L5 audit schema", "require L4 persona manifests"],
  "timestamp": "2025-10-31"
}
```

5.2 Solomon Briefs

Definition. A Solomon Brief is a concise, one-page decision artifact documenting rationale, trade-offs, timing, constraints, and ownership for any governance or deployment decision. It is a Decision Card designed to capture the deliberative logic behind a choice, not merely its outcome.

Purpose. To protect reasoning integrity and prevent post-hoc rationalization. A Solomon Brief ensures that a decision's epistemic lineage—the why and how—is as transparent as the result.

Governance Function. Used primarily in Layers 6 and 7. In enterprise or regulatory audits, Solomon Briefs form part of *Decision Insurance*, ensuring that reasoning failures can be diagnosed through reconstructable evidence rather than conjecture.

Core Fields.

- Decision context and objectives.
- Constraints (budgetary, ethical, temporal).
- Alternatives considered and rationale for rejection.
- Decision owner and review authority.
- Timestamp and version reference.

Analogy. In aviation, the flight recorder captures mechanical history; the Solomon Brief captures cognitive history.

5.3 Governance Maps

Definition. A Governance Map is a structured linkage between decisions, risk controls, and regulatory frameworks. It connects persona-level reasoning to compliance standards, showing which controls satisfy which obligations and where residual gaps remain.

Purpose. To visualize accountability. Governance Maps act as connective tissue between technical and institutional layers, enabling both vertical traceability (from application to hardware) and horizontal interoperability (across organizations and jurisdictions).

Governance Function. They are core to Layer 7 and used to generate comparative audits across organizations. By mapping decisions to standards such as NIST AI RMF or ISO/IEC 42001, Governance Maps ensure that oversight remains intelligible to both engineers and regulators.

Structure.

```
Decision: Enable autonomous data labeling
Linked Risks: L3 provenance, L4 control bias
```

```
Linked Standards: NIST-MEASURE, ISO 42001-5.3
Audit Artifacts: Solomon Brief #014, Clarity Package #058
```

5.4 Stack-Aligned Reports

Definition. A Stack-Aligned Report (SAR) is the formal audit deliverable defined by the AI OSI Stack methodology. It documents each layer's actor, risk, evidence, and governing standard in a unified template.

Purpose. To make systemic accountability feasible. SARs enable enterprises and auditors to reason about the full AI system rather than isolated components.

Standard Format.

```
Layer: 4 - Instruction / Control
Actor: Internal Alignment Team
Risk: Misalignment between policy and prompt logic
Evidence: Persona Manifest, Red-Team Report #22
Regulator / Standard: ISO/IEC 42001-6.x, EU AI Act Annex IV
```

Governance Function. SARs serve as living documentation for AI systems. They support versioning, cross-comparison, and root-cause analysis in the event of incidents or compliance reviews.

5.5 Decision Insurance and Trust Portability

Concept. Decision Insurance is the governance mechanism that ensures decisions remain auditable, interpretable, and defensible over time. It formalizes the process of documenting uncertainty, alternatives, and context. Its function is preventive: to cushion judgment where cognition or context may fail.

Mechanics. Each high-stakes decision at Layer 6 or 7 must be accompanied by:

1. A Solomon Brief documenting rationale.
2. A Clarity Package capturing interpretive trace.
3. A Governance Map linking the decision to compliance frameworks.
4. A Stack-Aligned Report consolidating evidence for review.

Together these form a portable chain of trust evidence. They convert ephemeral decisions into persistent governance memory, enabling what the Stack calls *Trust Portability*—the ability of trust to move across organizations, timeframes, and technologies without degradation.

Interpretive Note

Governance artifacts are not bureaucracy; they are the grammar of accountability. They give structure to institutional memory and make ethics operational. Where regulation defines what must be done, governance artifacts show how it can be proven. The Stack's innovation lies in designing these artifacts as native system outputs rather than external compliance paperwork, turning governance into infrastructure.

6 Temporal Integrity and Version Governance

AI systems evolve continuously. Their components update, retrain, fine-tune, and recombine in ways that erode any single "moment of truth." Temporal governance addresses this loss of chronological stability by embedding time itself into the design of accountability. The AI OSI Stack introduces explicit mechanisms—Semantic Version Control, Temporal Drift Tracking, and Memory Governance—to manage how systems, definitions, and trust evolve.

6.1 Semantic Version Control (SVC)

Purpose. To make change transparent and traceable across time. Semantic Version Control is a federated documentation protocol that tracks updates in both technical systems and interpretive frameworks.

Rationale. Without versioning, every governance artifact risks temporal ambiguity. It becomes impossible to determine which model, data pipeline, or control schema produced an outcome. SVC restores temporal integrity by providing a consistent record of modification and intent.

Versioning Convention. The Stack adopts a semantic hierarchy:

- **Major Increment:** Structural or architectural change (Layer 2).
- **Minor Increment:** Control schema or behavioral change (Layer 4).
- **Patch Increment:** Documentation or governance refinement (Layer 7).

Governance Function. Each update triggers an automatic Stack-Aligned Report regeneration, ensuring that governance artifacts remain synchronized with system reality. SVC extends versioning beyond code to include definitions, assumptions, and ethical constraints.

6.2 Temporal Drift and Synthetic Memory

Definition. *Temporal Drift* refers to the nonlinear decay of trust caused by asynchronous updates, data drift, or unlogged fine-tuning. Synthetic memory, the persistence of learned

patterns within models, amplifies this effect because models retain traces of prior states even after retraining.

Governance Challenge. Unlike traditional media, AI does not forget. Its memory is statistical and persistent, meaning that even deleted or "unlearned" data may survive in weights. This persistence produces what the Stack terms *Chronological Displacement*: outputs that appear current but are in fact temporally mixed.

Governance Levers.

- **Temporal Provenance.** All model artifacts must declare training cutoff dates and fine-tuning intervals.
- **Semantic Drift Logs.** Track shifts in reasoning or definition between model versions and alignment updates.
- **Versioned Governance Artifacts.** Each Solomon Brief, Clarity Package, and Governance Map carries a timestamp and version reference.

Audit Artifacts. Semantic drift registers, retraining logs, and provenance declarations create an auditable temporal chain that can reconstruct how an output relates to prior states.

6.3 Memory Governance and Forgetting by Design

Core Problem. AI systems exhibit a structural bias toward persistence. Forgetting is costly and partial. Deletion claims often amount to cosmetic removal rather than genuine epistemic erasure.

Ethical Principle. If memory is never neutral, then forgetting must be deliberate. The question is not "Can AI forget?" but "Who decides what it forgets?" The Stack frames this as an issue of dignity and agency: memory curation is a governance act, not a technical patch.

Design Requirements.

1. Explicit policies for retention, redaction, and unlearning.
2. Provenance attestations showing when and how data or model parameters were altered.
3. Temporal firewalls that isolate legacy models from current decision loops.

Governance Instruments. Semantic Version Control and Decision Insurance collectively act as memory governance protocols. The former ensures that change is visible; the latter ensures that forgotten reasoning can still be reconstructed through archived briefs.

6.4 Temporal Governance in Practice

Temporal integrity mechanisms integrate across the Stack:

- Layer 1–3: Training data lineage and compute logs timestamped and reproducible.
- Layer 4–5: Versioned control manifests and protocol schemas.
- Layer 6–7: Time-aware audit trails and drift-aware risk reports.

Governance institutions can use these artifacts to anchor accountability in time as well as structure. A model's ethical and operational status must always be referenced to a specific version and date.

Interpretive Note

Temporal integrity converts time from an adversary into an audit axis. By embedding chronology directly into governance, the Stack ensures that meaning, capability, and accountability remain synchronized even as systems evolve. The passage of time no longer erodes trust; it becomes the framework through which trust is maintained.

7 Foresight, Adaptation, and System Evolution

Artificial intelligence evolves in cycles of acceleration, collapse, and renewal. The AI OSI Stack must therefore serve not only as a diagnostic map but also as an adaptive governance ecosystem capable of surviving continuous transformation. This section synthesizes foresight methodologies—Design Justice, Panarchy, and the Futures Cone—into a coherent theory of adaptive governance. It reframes AI evolution as ecological process rather than linear progress, ensuring that accountability mechanisms remain resilient under uncertainty.

7.1 Adaptive Cycles and Panarchic Structure

Concept. Borrowing from ecological systems theory (Gunderson and Holling), *Panarchy* describes complex systems as interlinked adaptive cycles of growth, conservation, collapse, and renewal. These cycles occur across scales: a local innovation may disrupt global equilibrium, while a systemic collapse can enable localized renewal.

Application to AI. The Stack interprets AI evolution as undergoing *mitosis*—rapid branching of model lineages, most of which fail. Surviving lineages become infrastructure. Each layer participates in its own adaptive cycle:

- Hardware (L1) grows through technological scaling and contracts through resource scarcity.

- Model architectures (L2) evolve through research diversification and consolidate around stable paradigms.
- Training pipelines (L3) collapse under cost pressure and renew through efficiency innovation.
- Governance frameworks (L7) adapt through feedback from public trust and crisis response.

Recognizing these cycles allows policymakers to plan for resilience rather than stability. Governance, in this model, is a process of ecological stewardship—maintaining the conditions for renewal rather than enforcing permanence.

7.2 Futures Cone and Temporal Horizon Design

Concept. The *Futures Cone* and *Three Horizons* frameworks distinguish between possible, plausible, probable, and preferable futures. The Stack integrates these temporal perspectives directly into governance workflows.

Operationalization.

- **Horizon 1 (Present Systems):** Existing Layer 5 monopolies and Layer 6 deployments. Governance priority—stabilize and audit.
- **Horizon 2 (Emergent Systems):** Transitional architectures, agentic runtimes, and hybrid ecosystems. Governance priority—experiment and document drift.
- **Horizon 3 (Transformative Futures):** Post-centralized, distributed cognition systems. Governance priority—prototype adaptive oversight and multi-jurisdictional coordination.

In this model, each governance artifact—Solomon Briefs, Governance Maps, Stack-Aligned Reports—acts as a time capsule, allowing future auditors to reconstruct both context and intent.

7.3 Design Justice and Inclusive Foresight

Principle. Design Justice (Costanza-Chock) holds that those most affected by technology must have a voice in its design. It reframes inclusivity as a structural requirement for resilience.

Stack Integration.

- Layer 4: Persona Architecture must encode cultural and ethical diversity.
- Layer 6: Applications in high-stakes contexts must include participatory feedback loops with impacted communities.

- Layer 7: Governance bodies must publish inclusion audits—transparency on who participated in shaping safety and policy.

In this view, inclusivity is not moral ornamentation but epistemic necessity: systems that hear fewer voices collapse under narrow assumptions.

7.4 Governance as a Living Practice

From Static to Dynamic Oversight. Traditional compliance frameworks treat governance as periodic certification. The Stack replaces this with *Living Governance*—a continuous feedback process connecting data, behavior, and legitimacy. Living Governance requires:

1. Continuous monitoring of cross-layer indicators.
2. Iterative updating of Stack-Aligned Reports based on drift and adaptation.
3. Integration of foresight models into strategic planning and regulatory sandboxes.

Temporal Feedback. Each audit cycle produces a *Temporal Loop*: new evidence informs updated assumptions, which in turn redefine acceptable risk thresholds. This creates governance with memory, learning from prior failure rather than merely recording it.

7.5 Adaptive Governance Timeline

Short Term (Operational Readiness).

- Publish Stack-aligned model cards.
- Establish open schemas for Layer 5 audit logs.
- Conduct red-team exercises for ethical controllers at Layer 4b.

Medium Term (Institutional Integration).

- Embed Stack language in procurement, contracting, and RFP requirements.
- Develop open-source Governance Map generators and shared audit schema repositories.
- Coordinate pilot programs with national standards bodies and regulatory sandboxes.

Long Term (Structural Resilience).

- Implement cryptographic attestation for Layers 4 and 5 artifacts.

- Converge on cross-border Layer 7 schemas to align regulatory expectations.
- Establish multi-actor oversight boards with rotating public representation.

Interpretive Note

Governance that cannot evolve becomes theater; governance that only evolves without structure becomes chaos. The AI OSI Stack aims to resolve this tension by creating *structured adaptability*. It combines ecological literacy with institutional rigor, ensuring that as systems change, accountability does not dissolve but regenerates. The Stack thus becomes both infrastructure and organism—a living architecture of trust.

8 Stakeholder Implementation Pathways

The AI OSI Stack is designed for use by multiple stakeholder classes—enterprises, policymakers, researchers, and standards bodies. Each operates at different scales but shares a common requirement: to translate principles into operational control. This section defines how each group applies the Stack in practice, what artifacts they produce, and how cross-layer governance coordination can be institutionalized.

8.1 Enterprises

Purpose. Enterprises face the dual challenge of integrating AI responsibly while maintaining competitiveness. The Stack provides a structured method for embedding governance within corporate decision cycles.

Implementation Pathway.

1. **Layer Audit:** Conduct internal audits using Stack-Aligned Reports for all deployed AI systems, identifying the actor, risk, evidence, and applicable regulator per layer.
2. **Decision Insurance:** Mandate Solomon Briefs for all high-stakes deployments, recording rationale, trade-offs, and responsible owners.
3. **Governance Integration:** Map corporate risk frameworks (e.g., COSO ERM) to Layers 6 and 7.
4. **Continuous Oversight:** Establish cross-functional review committees to evaluate alignment drift and compliance evolution quarterly.

Deliverables.

- Stack-Aligned Reports per system.

- Clarity Packages summarizing interpretive risk.
- Governance Maps linking corporate controls to standards.
- Annual "Governance Digest" documenting temporal drift and version history.

8.2 Policymakers and Regulators

Purpose. To target regulation precisely, avoiding both overreach and fragmentation. The Stack allows regulators to map legal obligations to concrete technical layers, improving enforceability.

Implementation Pathway.

1. **Layer Targeting:** Use the Stack to locate chokepoints—Layer 1 (hardware concentration) and Layer 5 (interface control).
2. **Standard Alignment:** Reference existing instruments such as NIST AI RMF and ISO/IEC 42001, assigning them to corresponding layers.
3. **Evidence Schema:** Require certified audit artifacts (Solomon Briefs, Clarity Packages) as compliance evidence.
4. **Temporal Oversight:** Incorporate Semantic Version Control into regulatory filings to maintain continuous accountability.

Deliverables.

- Policy briefs referencing Stack layers for targeted intervention.
- Open audit schema registries enabling international compatibility.
- Annual cross-jurisdictional "Trust Portability" reports tracking governance interoperability.

8.3 Researchers and Builders

Purpose. To innovate safely and transparently within defined boundaries. Layer separation ensures that research at one level can proceed without destabilizing others.

Implementation Pathway.

1. **Layer Discipline:** Limit experimentation to a clearly defined layer (e.g., alignment at L4b, optimization at L3).
2. **Audit Readiness:** Maintain detailed provenance logs for data, code, and models.

3. **Ethical Encoding:** Apply Persona Architecture to reflect Dignity as Constraint and Transparency as Infrastructure.

4. **Interoperability Testing:** Ensure compatibility through L5 protocol conformance checks.

Deliverables.

- Reproducible model cards and training logs.
- Persona manifests documenting ethical and control design.
- L5 conformance reports ensuring interface stability.

8.4 Standards Bodies and Auditing Institutions

Purpose. To harmonize fragmented regulatory regimes and technical standards through a common reference model. The Stack provides the meta-architecture to align certification, audit, and compliance systems.

Implementation Pathway.

1. **Standard Mapping:** Align certification criteria (ISO/IEC, OECD, UNESCO) to specific Stack layers.

2. **Schema Standardization:** Publish interoperable templates for Solomon Briefs and Stack-Aligned Reports.

3. **Cross-Border Collaboration:** Develop mutual recognition agreements for governance artifacts.

4. **Temporal Certification:** Incorporate version histories and drift tracking into audit cycles.

Deliverables.

- Layer-to-standard concordance tables.
- Audit schema specifications.
- Global governance repository integrating versioned artifacts.

8.5 Public and Civil Society Organizations

Purpose. To represent affected communities and advocate for accountability where market and state oversight may fall short. The Stack offers a transparent lens through which civic actors can evaluate risk concentration and ethical compliance.

Implementation Pathway.

1. **Civic Auditing:** Use Stack-Aligned Reports and public Clarity Packages to interpret system accountability.
2. **Participatory Feedback:** Collaborate in Layer 6 evaluations to capture real-world harm and bias.
3. **Transparency Monitoring:** Demand publication of L5 interface policies and L7 incident registries.

Deliverables.

- Public trust dashboards visualizing governance metrics.
- Independent oversight reports benchmarking compliance.
- Open libraries of Decision Insurance cases for public education.

Interpretive Note

The AI OSI Stack establishes a shared accountability grammar across domains. Each stakeholder class speaks a different dialect—policy, engineering, management, or advocacy—but the Stack provides a lingua franca. It turns governance from an act of translation into a structured conversation. When used collectively, these pathways transform AI oversight from isolated compliance silos into a distributed, living system of trust.

9 Conclusion: Architecture as Accountability

Artificial intelligence has reached the stage where it no longer behaves like a discrete product but like infrastructure. It underpins communication, finance, education, creativity, and state capacity. Yet governance of AI remains structurally immature. Most oversight efforts still treat "AI" as a singular object, collapsing hardware, models, data, and institutional context into one black box. The AI OSI Stack resolves that collapse by restoring architecture to governance.

9.1 From Black Boxes to Layered Systems

The Stack separates the anatomy of control. Each layer isolates a domain of responsibility, allowing policymakers, engineers, and organizations to locate failure precisely rather than generically. Layering does not fragment responsibility; it clarifies it. Hardware concentration, model opacity, and interface monopolies become visible as distinct problems rather than indistinguishable complexity.

This clarity transforms governance from reaction to design. Instead of treating regulation as a corrective mechanism, the Stack treats it as a form of architecture—an engineering of accountability.

9.2 Trust as Infrastructure

The guiding principles—*Dignity as Constraint* and *Transparency as Infrastructure*—anchor the Stack's ethics. Dignity ensures that human welfare constrains design, not the reverse. Transparency ensures that every system reveals how it knows, not only what it outputs. Together they produce *Trust as Infrastructure*: the conviction that trust should emerge from systemic legibility, not marketing or compliance theatre.

When trust becomes infrastructural, it ceases to be fragile. It is encoded through audit trails, semantic version control, and governance artifacts that render reasoning and evidence portable across time and institutions.

9.3 Governance as a Design-Time Feature

Governance cannot remain an afterthought appended to finished systems. The Stack makes it intrinsic: Layer 7 is as essential as Layer 1. This transforms the development lifecycle itself. Each layer must ship not only with functional documentation but with proof of integrity—capability reports, data lineage logs, persona manifests, audit schemas, and decision briefs. The result is an AI ecosystem that is both interoperable and accountable by construction.

9.4 Temporal Resilience

Versioning and semantic integrity convert time into a governance dimension. Temporal drift, once a source of uncertainty, becomes a measurable signal. By integrating semantic versioning across all artifacts, the Stack ensures that change itself is legible. This capacity to preserve meaning through evolution is the difference between trust that erodes and trust that endures.

9.5 Strategic Implications

The Stack reframes competition, compliance, and innovation as elements of a single system. For enterprises, it enables credible self-regulation backed by evidence. For policymakers, it provides precision targeting of chokepoints. For researchers, it preserves the freedom to experiment without destabilizing the larger ecosystem. For society, it offers the prospect of verifiable alignment between power, capability, and legitimacy.

9.6 Closing Reflection

The lesson of the internet's evolution still applies: global systems scale only when they know where their boundaries lie. The OSI model gave the world an interoperable network. The AI OSI Stack aspires to give the world an interoperable conscience—a framework where intelligence can grow without erasing accountability.

If AI is to become the substrate of human infrastructure, it must adopt the same discipline that built the physical and informational systems before it. Architecture is not metaphor here; it is governance made visible. The Stack is not merely a model but a practice—a way of ensuring that progress remains intelligible, responsibility remains assignable, and trust remains alive through time.

Appendices

Appendix A: Glossary of Core Concepts

- **AI OSI Stack:** A seven-layer framework that decomposes AI into Physical/Hardware, Model Architecture, Training/Optimization, Instruction/Control, Interface/Protocol, Application, and Governance/Trust. It clarifies where risk concentrates and how accountability can be verified.
- **Trust Portability:** The property that allows confidence in one system to be transferred to another through verifiable evidence, audit artifacts, and governance interoperability.
- **Decision Insurance:** A Layer 6–7 governance mechanism ensuring that every consequential decision carries structured documentation of rationale, trade-offs, and context, enabling post-decision accountability.
- **Persona Architecture:** The structured design of bounded AI roles that encode ethical constraints, epistemic assumptions, and refusal logic; the foundation of Layer 4b (Ethical/Value Reasoning).

- **Guardian Notes:** Layer 7 governance artifacts summarizing oversight findings, intended to serve as living records of governance decisions and safety outcomes.
- **Epistemology by Design:** The principle that systems must encode how they know—embedding epistemic transparency into architectures rather than treating it as external documentation.
- **Dignity as Constraint:** The principle that human welfare and autonomy set hard design boundaries; no function may trade off human dignity for efficiency.
- **Transparency as Infrastructure:** The principle that systems must be built to reveal their reasoning, lineage, and provenance; trust emerges from visible structure, not asserted integrity.
- **Semantic Version Control (SVC):** The formal system of time-aware governance introduced by the Stack to track shifts in definitions, artifacts, and reasoning over time.
- **Temporal Drift:** The gradual misalignment between a system's current behavior and its recorded intent due to unlogged fine-tuning, data drift, or contextual re-use.
- **Clarity Package:** Structured interpretive output (often JSON) generated by analytical pipelines to surface meaning, bias, and drift in text or model output.
- **Solomon Brief:** One-page decision record capturing the rationale, constraints, alternatives, and version metadata for any governance-relevant decision.
- **Governance Map:** The linkage matrix connecting decisions, controls, and regulatory standards, establishing traceability across the Stack.
- **Stack-Aligned Report (SAR):** The formal audit deliverable documenting actors, risks, evidence, and governing standards for each layer of the Stack.
- **Temporal Integrity:** The assurance that a system's outputs can be chronologically and semantically traced to a known version and governance context.
- **Living Governance:** The discipline of continuous oversight, iterative auditing, and dynamic adaptation in response to evolving AI systems.
- **AI Mitosis:** The ecological metaphor describing AI's tendency to branch into specialized lineages—some of which stabilize as infrastructure while others perish.
- **Layer Blurring:** The phenomenon in which decisions or constraints at one layer (e.g., architecture) affect governance viability at another (e.g., control or application).

Appendix B: Layer-to-Standard Concordance

Stack Layer	Primary Standards / Frameworks	Governance F
L1 Physical / Hardware	ISO/IEC 27001, NIST SP 800-171	Supply chain se
L2 Model Architecture	OECD AI Principles, ISO/IEC 42001	Safety by design
L3 Training / Optimization	NIST AI RMF (Map, Measure)	Data provenanc
L4 Instruction / Control	ISO/IEC 38507, RLHF and alignment methods	Ethical alignme
L5 Interface / Protocol	EU AI Act (Article 52), OpenAPI / MCP standards	Access governan
L6 Application	EU AI Act (Annex III), FDA Guidance (2025)	Contextual risk,
L7 Governance / Trust	NIST AI RMF (Govern), UNESCO Ethics of AI	Auditability, acc

Appendix C: Governance Artifact Lifecycle

1. **Design Stage:** Each project defines intended Stack layers and produces initial Governance Maps.
2. **Development Stage:** Persona manifests (L4b) and control policies are documented.
3. **Deployment Stage:** Solomon Briefs and Decision Insurance reports are completed.
4. **Operation Stage:** Continuous audit produces Stack-Aligned Reports and Clarity Packages.
5. **Post-Incident Stage:** Guardian Notes are issued and archived with Semantic Version updates.

Appendix D: Layer Interaction Summary

- **L1 → L3:** Compute and energy constraints determine feasible training regimes.
- **L2 → L4:** Model architecture limits the scope of alignment and interpretability.
- **L5 → L6:** Interface control dictates which applications can exist and how they are monitored.
- **L7 → L3:** Governance rules on data provenance shape how and where training occurs.
- **Cross-Layer Coupling:** Governance feedback cycles ensure vertical accountability throughout the system.

Appendix E: Implementation Timeline (Extended)

Phase 1 — 2025–2026: Infrastructure Readiness

- Publish Stack-aligned model cards and L5 audit schemas.
- Create open databases of Stack-Aligned Reports for regulatory sandboxes.
- Establish working groups for open semantic logging protocols.

Phase 2 — 2026–2028: Institutional Integration

- Embed Stack language in international standards, procurement, and certification criteria.
- Launch cross-sector pilot programs linking enterprise and regulatory audits.
- Develop an open-source repository for Governance Maps and Decision Insurance templates.

Phase 3 — 2028–2030: Global Convergence

- Adopt cryptographic attestation for L4 and L5 artifacts.
- Harmonize cross-border governance schemas under OECD and UNESCO principles.
- Establish multi-actor oversight boards to maintain semantic and temporal integrity.

About the Author

Daniel P. Madden is an independent AI researcher and IT specialist focused on feasibility, governance, and reasoning systems. His ongoing work explores the intersection of epistemology, design, and safety through live experiments in persona architecture, decision auditing, and cognitive governance. His AI Lab Notebook and associated essays document iterative frameworks for aligning artificial systems with human values. More at danielpmadden.com.

He can be reached for research collaborations or policy consultations through the contact form on his website or via the professional networks linked therein.

Final Note

The **AI OSI Stack** is not simply a conceptual model. It is a living governance architecture designed to evolve with the systems it describes. Each layer defines a distinct locus of power, responsibility, and transparency. Together, they form the scaffolding for portable trust — a structure capable of making accountability as modular as the technologies it governs. If AI is to become infrastructure for human life, it must be built like infrastructure: layered, transparent, and accountable.

References

References

[1] National Institute of Standards and Technology. *AI Risk Management Framework (AI RMF 1.0)*. U.S. Department of Commerce, 2023. Available at: https://www.nist.gov/itl/ai-risk-management-framework.

[2] International Organization for Standardization. *ISO/IEC 42001:2023 — Artificial Intelligence Management System*. ISO/IEC, 2023. Available at: https://www.iso.org/standard/81230.html.

[3] European Commission. *Regulation (EU) 2024/1689 — Artificial Intelligence Act*. Official Journal of the European Union, 2024. Available at: https://eur-lex.europa.eu/eli/reg/2024/1689/oj.

[4] International Organization for Standardization. *ISO/IEC 7498-1:1994 — Information Technology: Open Systems Interconnection Basic Reference Model*. ISO, 1994. Available at: https://www.iso.org/standard/20269.html.

[5] Partnership on AI. *Guidance for Safe Foundation Model Deployment*. 2024. Available at: https://partnershiponai.org/modeldeployment/.

[6] Organisation for Economic Co-operation and Development. *OECD Principles on Artificial Intelligence*. 2019. Available at: https://oecd.ai/en/ai-principles.

[7] UNESCO. *Recommendation on the Ethics of Artificial Intelligence*. 2021. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000381137.

[8] Council of Europe. *Framework Convention on Artificial Intelligence, Human Rights and Democracy (Draft)*. Council of Europe, 2024. Available at: https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence.

[9] International Organization for Standardization. *ISO/IEC 38507:2022 — Governance of IT, AI and Related Digital Technologies*. ISO, 2022. Available at: https://www.iso.org/standard/82890.html.

[10] Committee of Sponsoring Organizations of the Treadway Commission (COSO). *Enterprise Risk Management: Integrating with Strategy and Performance*. AICPA, 2017. Available at: https://www.coso.org/Pages/erm-integrating-with-strategy-and-performance.aspx.

[11] White House Office of Science and Technology Policy. *Blueprint for an AI Bill of Rights*. 2022. Available at: https://www.whitehouse.gov/ostp/ai-bill-of-rights/.

[12] U.S. Food and Drug Administration. *Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices*. 2025. Available at: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices.

[13] Clark, A., and Chalmers, D. "The Extended Mind." *Analysis*, vol. 58, no. 1, 1998, pp. 7–19. DOI: https://doi.org/10.1093/analys/58.1.7.

[14] Latour, B. *Reassembling the Social: An Introduction to Actor-Network Theory*. Oxford University Press, 2005.

[15] Parikka, J. *What is Media Archaeology?* Polity Press, 2012.

[16] Gunderson, L. H., and Holling, C. S. *Panarchy: Understanding Transformations in Human and Natural Systems*. Island Press, 2002.

[17] Costanza-Chock, S. *Design Justice: Community-Led Practices to Build the Worlds We Need*. MIT Press, 2020.

[18] Madden, D. P. *Persona Architecture Notes*. Unpublished Lab Notebook Entry, September 2025.

[19] Madden, D. P. *Decision Insurance: Preliminary Design Specification*. Unpublished Working Paper, October 2025.

[20] Madden, D. P. *Guardian Notes and Trust Portability Framework*. Internal Research Memorandum, October 2025.

[21] Madden, D. P. *Epistemology by Design*. Private Essay Draft, October 2025.

[22] Wu, S. et al. "BloombergGPT: A Large Language Model for Finance." *arXiv preprint*, 2023. Available at: https://arxiv.org/abs/2303.17564.

[23] Singhal, K. et al. "Towards Expert-Level Medical Question Answering with Large Language Models." *arXiv preprint*, 2023. Available at: https://arxiv.org/abs/2305.09617.