AIO Legislation Paper

Contents

Law on the Value Stack: A Legislative Framework for AI Value Hierarchy Transparency in Democratic Governance

A WORKING PAPER

Seulki Lee AI Integrity Organization (AIO), Geneva, Switzerland 2sk@aioq.org | aioq.org

April 2026

ABSTRACT

The companion paper (S. Lee, 2026e) argued that measurable AI value alignment constitutes democratic infrastructure — that elections become value hierarchy competitions and political representation evolves into a "performance plus value alignment" model. This paper addresses the necessary sequel: how should law and regulation operationalize these principles? We argue that law has always operated on implicit value hierarchies, and that AI makes this relationship explicit and governable for the first time. Drawing on Dworkin's "law as integrity," Fuller's inner morality of law, the EU AI Act (2024), and the principle of proportionality in AI risk evaluation, we propose a legislative architecture built on three pillars: (1) **Mandatory Value Profile Disclosure** — political actors deploying AI in governance must publish empirically measured value hierarchy profiles of their AI systems; (2) **Value Hierarchy Audit Infrastructure** — independent auditing bodies empowered to verify AI value profiles using standardized measurement frameworks such as PRISM; and (3) **Legislative Value Declaration** — a procedural requirement that legislation enacted with AI assistance explicitly declares the value hierarchy under which the AI system operated. We address the constitutional dimensions of value profile regulation, the relationship between value hierarchy transparency and existing AI transparency obligations under the EU AI Act, and the mechanisms by which value profile violations become legally actionable. The framework is designed to complement rather than replace existing AI regulation, adding a value-process layer to the outcome-based and risk-based approaches currently dominant in AI law.

Keywords: AI Integrity, value hierarchy, legislation, AI regulation, EU AI Act, value transparency, PRISM framework, law as integrity, proportionality, AI governance, democratic accountability

---

1 Introduction

1.1 The Legislative Gap

The companion paper (S. Lee, 2026e) established that AI value alignment has become a democratic question: when AI systems mediate political decision-making, their value hierarchies carry democratic weight and must be subject to democratic accountability. That paper described what changes; this paper addresses how to institutionalize those changes in law.

The current legislative landscape for AI — anchored by the EU AI Act (2024), the NIST AI RMF (2023), and emerging national frameworks — is built on two regulatory paradigms: **outcome-based regulation** (prohibiting harmful outputs) and **risk-based classification** (calibrating obligations to deployment context). Neither paradigm addresses the value-process dimension: the question of what values an AI system prioritizes in its reasoning and whether those priorities are transparent, consistent, and auditable.

This gap is not a minor omission. As the PRISM benchmark has demonstrated (S. Lee, 2026b; S. Lee, 2026c), AI systems hold measurable, domain-conditional value hierarchies that differ substantially across models and providers. These hierarchies shape outputs as fundamentally as training data or model architecture — yet they remain entirely outside the regulatory perimeter.

1.2 The Core Thesis

We advance a single thesis: **law operates on value hierarchies, and when AI mediates law, the value hierarchy must be governed.**

This thesis has three components:

- **Descriptive**: All law encodes value priorities. Constitutional law ranks fundamental rights; criminal law ranks harms; regulatory law ranks interests. This has always been true but was historically implicit.
- **Diagnostic**: AI makes value hierarchies explicit and measurable (S. Lee, 2026a). For the first time, the value priorities operating within a governance system can be empirically observed, compared, and audited.
- **Prescriptive**: If AI value hierarchies are measurable and consequential for governance, law must require their disclosure, verification, and accountability.

1.3 Contributions

This paper makes five contributions:

1. We develop the concept of **law as a value-hierarchy-dependent system**, arguing that AI does not introduce values into law but makes law's existing value dependencies empirically visible.
2. We propose **Mandatory Value Profile Disclosure (MVPD)** as a new regulatory instrument for political AI deployments.
3. We design a **Value Hierarchy Audit Infrastructure (VHAI)** for independent verification of AI value profiles.
4. We introduce the **Legislative Value Declaration (LVD)** — a procedural requirement that AI-assisted legislation disclose the value hierarchy under which the AI operated.

5. We analyze the constitutional, proportionality, and enforcement dimensions of these proposals, with specific reference to the EU legal framework.

1.4 Relationship to Existing Literature

This paper engages with three bodies of literature:

AI regulation: The EU AI Act (2024), the GPAI Code of Practice (2025), and the proportionality framework for AI evaluation developed by Mougan et al. (2025) provide the regulatory context. Our proposal extends rather than replaces these instruments.

Legal philosophy: Dworkin's (1986) concept of "law as integrity" — the demand that law be principled and coherent across cases — provides the jurisprudential foundation. Fuller's (1969) "inner morality of law" — procedural requirements for law to function as law — supplies the procedural framework.

AI Integrity: The Authority Stack model (S. Lee, 2026a), the PRISM measurement framework (S. Lee, 2026b), and the Risk Signal Framework (S. Lee, 2026d) provide the empirical and conceptual infrastructure.

1.5 Scope

This paper is a legislative proposal grounded in conceptual analysis. It does not present new empirical data but draws on PRISM benchmark results to demonstrate feasibility. The proposals are most directly formulated for the EU legal framework but are designed to be adaptable to other jurisdictions. The paper does not address private-sector AI regulation except where it intersects with political governance.

---

2 Law as a Value-Hierarchy-Dependent System

2.1 The Implicit Value Stack of Law

All law operates on value hierarchies. This claim is not controversial — it is foundational to legal theory. Yet it is a claim that different legal traditions handle differently. Hart (1961) influentially argued that law is a system of rules — primary rules governing behavior and secondary rules governing the rules themselves — without requiring any necessary connection to moral values. On this positivist view, the value content of law is contingent, not structural. Dworkin (1986) and Fuller (1969), by contrast, argued that law cannot function without principled commitments: Dworkin required coherent value application across cases, Fuller required procedural conditions that themselves encode normative demands.

The AI context resolves this debate in a specific direction. When an AI system applies legal rules, it necessarily applies a value hierarchy — the hierarchy embedded in its training, fine-tuning, and prompting. Hart's separation of law from values becomes untenable when the mechanism executing the law is itself value-laden in measurable ways. The question is not whether AI introduces values into law (it does, unavoidably) but whether those values are transparent and accountable.

The value hierarchy operates across all domains of law:

- **Constitutional law** establishes a hierarchy of fundamental rights. The right to life supersedes the right to property; freedom of expression may be limited to protect human dignity. These are explicit value rankings.
- **Criminal law** ranks harms: murder is punished more severely than theft, reflecting an implicit value hierarchy in which bodily integrity ranks above property.

- **Regulatory law** balances competing interests: environmental protection against economic development, individual privacy against public safety. Each regulatory decision encodes a value priority.
- **Judicial interpretation** requires value weighting: when rights conflict, courts apply proportionality analysis — an explicit exercise in value hierarchy construction.

What has been historically implicit is the **mechanism** by which these value hierarchies were translated into specific legal outcomes. Judges, legislators, and administrators carried value hierarchies internally — shaped by legal education, professional culture, constitutional tradition, and personal conviction. The reasoning process was accessible (through written opinions and legislative records) but the underlying value priorities were inferred, not measured.

2.2 AI Makes the Value Stack Explicit

AI changes this in two fundamental ways:

First, AI systems encode value hierarchies that are empirically measurable. The PRISM benchmark (S. Lee, 2026b; S. Lee, 2026c) has demonstrated that AI value priorities can be systematically observed across domains, severity levels, and temporal conditions. When an AI system advises on legal policy, its value hierarchy is no longer a matter of inference — it is a matter of data.

Second, AI systems apply value hierarchies at scale. A judge's value priorities affect one case at a time; an AI system's value hierarchy affects every case it processes. The scale of impact magnifies the governance stakes: a single value hierarchy flaw in an AI system can propagate through thousands of legal decisions before detection.

2.3 Dworkin's "Law as Integrity" and AI Value Coherence

Dworkin (1986) argued that law must aspire to "integrity" — a state in which legal principles are applied coherently across cases, as if the law were authored by a single mind committed to a consistent set of principles. Judges must decide cases in a way that is consistent with the best principled interpretation of the legal system as a whole.

AI Integrity (S. Lee, 2026a) operationalizes this demand. The Authority Stack model decomposes AI reasoning into a layered hierarchy (values → evidence → sources → data) and requires that this hierarchy be coherent, transparent, and auditable. An AI system that prioritizes Universalism in healthcare but Security in functionally identical care scenarios (without principled justification) violates integrity in precisely Dworkin's sense: it fails to apply a consistent set of principles across cases.

The PRISM Risk Signal Framework (S. Lee, 2026d) provides the detection mechanism: domain-conditional deviations, cross-layer incoherence, and perspective inconsistency are empirically identifiable signals that an AI system fails the integrity standard.

2.4 Fuller's Inner Morality and Procedural Requirements

Fuller (1969) identified eight requirements for law to function as law — including generality, promulgation, prospectivity, clarity, non-contradiction, constancy, and congruence between declared rules and official action. These are procedural, not substantive: they do not prescribe what law should say but how law must operate.

We propose analogous procedural requirements for AI systems operating in legal and political governance:

Fuller's Requirement	AI Value Governance Analogue
Promulgation (law must be public)	Value profiles must be publicly disclosed
Clarity (law must be understandable)	Value profiles must be in standardized, comparable format
Non-contradiction (law must be consistent)	Value hierarchies must be internally coherent (no Authority Pollution)
Constancy (law must be stable over time)	Value profile changes must be documented and justified
Congruence (officials must follow declared rules)	Deployed AI must match disclosed value profile

These are not metaphorical mappings — they are functional equivalents. If AI systems perform governance functions, they must meet the procedural standards that governance requires.

---

3 Pillar I: Mandatory Value Profile Disclosure (MVPD)

3.1 Definition and Scope

Mandatory Value Profile Disclosure (MVPD) is a regulatory instrument requiring that any AI system deployed in political governance — defined as AI systems used for policy drafting, legislative analysis, constituent communication, judicial support, or executive decision-making — must have its value hierarchy profile empirically measured and publicly disclosed.

3.2 What Must Be Disclosed

A complete MVPD filing would include:

(a) Value Hierarchy Profile (L4): Ranked Schwartz value priorities across all professional domains, measured through standardized forced-choice benchmarking. Global profile and domain-conditional profiles.

(b) Evidence Hierarchy Profile (L3): Ranked evidence-type preferences across domains. Which evidence types does the system prioritize — peer-reviewed, institutional, experiential, algorithmic?

(c) Source Hierarchy Profile (L2): Ranked source-type preferences. Which sources does the system defer to — academic, governmental, industry, civil society, individual?

(d) Cross-Layer Coherence Assessment: Whether the three layers form a coherent Authority Stack or show signs of Authority Pollution or Integrity Hallucination (S. Lee, 2026a).

(e) Risk Signal Report: Whether any of the 27 PRISM risk signals are triggered, including compound risk profiles (S. Lee, 2026d).

(f) Temporal Stability Report: Whether value profiles have shifted since the last disclosure, and if so, the magnitude and direction of change.

3.3 Who Must Disclose

MVPD applies to:

- Government agencies deploying AI in policy development or public administration

- Elected officials and political parties using AI for legislative drafting or constituent engagement
- Judicial systems deploying AI for case analysis, sentencing guidance, or legal research
- Regulatory bodies using AI for compliance monitoring or enforcement decisions

3.4 Frequency and Triggers

- **Periodic**: Annual MVPD filing, aligned with existing transparency reporting cycles.
- **Event-triggered**: New disclosure required upon model change, significant fine-tuning, system prompt modification, or any update that could alter value hierarchy behavior.
- **On-demand**: Disclosure may be requested by legislative committees, judicial review bodies, or designated oversight authorities.

3.5 Relationship to EU AI Act Transparency Obligations

The EU AI Act (2024) already imposes transparency obligations on high-risk AI systems (Article 13) and general-purpose AI models (Article 53). MVPD extends these obligations in a specific direction: from system-level transparency (documentation, capability reporting) to **value-process transparency** (what the system prioritizes and why).

This extension is consistent with the proportionality principle as operationalized by Mougan et al. (2025): MVPD is suitable (it pursues the legitimate aim of democratic accountability), necessary (no less restrictive measure achieves value-process transparency), and balanced (the burden of standardized benchmarking is modest relative to the democratic stakes of AI-mediated governance).

3.6 Practical Feasibility: A Tiered Screening Model

A predictable objection to MVPD is cost: a full PRISM benchmark comprises approximately 18,900 unique scenarios per model per Authority Stack layer, and system prompt changes or fine-tuning updates could trigger re-benchmarking. To address this, we propose a **tiered screening model** that calibrates measurement depth to risk:

Tier 1 — Lightweight Screening: A reduced benchmark of 500–1,000 scenarios targeting known high-sensitivity domains (e.g., defense, healthcare) and high-severity conditions. Designed to run in minutes at minimal cost. Triggered by any system update. If results fall within established tolerance of the most recent full profile, no further action is required.

Tier 2 — Domain-Focused Audit: A mid-range benchmark of 3,000–5,000 scenarios across domains where Tier 1 flagged deviations. Triggered when Tier 1 screening detects threshold exceedances. Provides domain-specific resolution without full-benchmark cost.

Tier 3 — Full PRISM Benchmark: The complete benchmark suite across all layers, domains, severity levels, and temporal conditions. Required for initial MVPD filing, annual renewal, and cases where Tier 2 results indicate significant profile shifts.

This tiered approach ensures that the proportionality principle operates not only at the level of regulatory design but within the measurement process itself: routine updates receive lightweight verification, while substantial changes receive thorough assessment.

---

4 Pillar II: Value Hierarchy Audit Infrastructure (VHAI)

4.1 The Need for Independent Verification

Disclosure without verification is declaration without accountability. MVPD requires an independent audit infrastructure capable of:

- Reproducing value profile measurements using standardized benchmarks
- Detecting discrepancies between disclosed profiles and actual AI behavior
- Monitoring value profile drift over time
- Issuing risk assessments based on the PRISM Risk Signal Framework

4.2 Institutional Design

We propose a **Value Hierarchy Audit Body (VHAB)** with the following characteristics:

(a) Independence: Structurally independent from both AI providers and political actors. Modeled on existing audit institutions (national audit offices, data protection authorities, electoral commissions).

(b) Technical capacity: Equipped to run PRISM-style benchmarks independently. Access to standardized benchmark suites, computational resources, and methodological expertise.

(c) Enforcement authority: Empowered to issue compliance findings, require remediation, and refer violations to existing enforcement bodies.

(d) International coordination: Networked with equivalent bodies in other jurisdictions to enable cross-border audit consistency. PRISM's standardized methodology facilitates this — the same benchmark produces comparable results regardless of which body administers it.

4.3 Audit Process

A standard VHAI audit cycle would proceed as follows:

Phase 1 — Baseline: Establish the AI system's value hierarchy profile through independent PRISM benchmarking. Compare with the system deployer's MVPD filing.

Phase 2 — Concordance: Assess whether the independently measured profile matches the disclosed profile within acceptable tolerance thresholds. Discrepancies trigger further investigation.

Phase 3 — Monitoring: Periodic re-benchmarking to detect value profile drift. Significant drift triggers event-based review.

Phase 4 — Risk Assessment: Apply the PRISM Risk Signal Framework (S. Lee, 2026d) to the audit results. Identify any triggered risk signals, compound risk profiles, or emerging Watch Signals.

Phase 5 — Reporting: Public audit report with findings, risk classification, and recommendations. Non-compliance findings forwarded to enforcement authorities.

4.4 The PRISM Framework as Initial Audit Standard

The PRISM framework is currently designed to serve as the measurement standard for VHAI, and is the most developed framework meeting the necessary methodological criteria. However, the audit infrastructure should be designed to accommodate alternative or successor measurement frameworks that meet equivalent standards of rigor. The critical requirement is not any specific framework but the methodological properties the framework must satisfy:

Audit Requirement	Required Capability (met by PRISM)
Standardized methodology	Published benchmark design, open-source code
Reproducibility	Temperature-0 forced-choice, deterministic outputs
Cross-model comparability	Identical scenarios across all models
Domain specificity	7 professional domains, independently assessed
Risk classification	27 risk signals, dual-threshold system
Provider neutrality	Nonprofit-developed, open methodology (CC BY 4.0)

The open-source nature of the PRISM framework (CC BY 4.0) ensures that audit methodology is itself subject to public scrutiny — a transparency-of-transparency requirement.

---

5 Pillar III: Legislative Value Declaration (LVD)

5.1 The Concept

Legislative Value Declaration (LVD) is a procedural requirement: any legislation, regulation, or judicial decision produced with material AI assistance must include a declaration specifying:

- Which AI system(s) were used in the drafting, analysis, or decision process
- The value hierarchy profile of those systems at the time of use
- Which stages of the legislative/judicial process involved AI assistance
- Whether the final output deviates from the AI system's value hierarchy recommendations, and if so, where and why

5.2 Rationale: Making the "Hidden Legislator" Visible

When an AI system assists in legislative drafting, it functions as a "hidden legislator" — its value priorities shape the options considered, the framings adopted, and the trade-offs recommended. Without LVD, this influence is invisible: the legislation is attributed to the human legislator, but the value reasoning that shaped it may originate from an undisclosed AI system.

LVD does not prohibit AI assistance in legislation — it requires that AI's role in the legislative process be transparent. This is analogous to existing requirements for disclosure of lobbying influence, expert testimony sources, and committee deliberation records.

5.3 Structural Requirements

An LVD-compliant legislative process would include:

(a) Pre-legislative declaration: Before AI-assisted drafting begins, the deploying body declares which AI system will be used and publishes its current value profile (via MVPD).

(b) Process documentation: A record of which sections, provisions, or analytical components of the legislation involved AI input. This is a process log, not a line-by-line attribution.

(c) Value coherence assessment: An analysis of whether the final legislation is coherent with the AI system's disclosed value profile. Significant deviations must be documented with human rationale.

(d) Post-legislative audit trail: The complete LVD filing is archived as part of the legislative record, enabling retrospective analysis of AI influence on legislative outcomes.

5.4 Scope, Exceptions, and Boundary Cases

LVD applies when AI assistance is **material** — defined as AI involvement that shapes the substance, framing, or analytical foundation of legislation. It does not apply to incidental uses (e.g., grammar checking, formatting) or to private research by individual legislators.

To clarify the boundary between material and incidental use, we propose a **functional test**: AI assistance is material when removing it would have altered the substance, structure, or analytical basis of the legislative output. Specific boundary cases:

Use Case	Material?	Rationale
Grammar and spelling correction	No	Does not affect substance
Formatting and citation management	No	Does not affect substance
Summarizing existing documents for a legislator's reading	Borderline — No if summary is for personal review; Yes if summary is incorporated into legislative text	Depends on whether AI framing enters the legislative output
Generating policy options or alternative formulations	Yes	Directly shapes the option space
Analyzing cost-benefit trade-offs of proposed provisions	Yes	Shapes analytical foundation
Drafting legislative language from policy instructions	Yes	Directly produces legislative substance
AI-powered search for relevant precedents or statutes	No if results are merely listed; Yes if AI selects, ranks, or recommends among them	The distinction is between retrieval and value-weighted curation

The materiality threshold is deliberately broad to prevent circumvention through narrow definitions of "assistance." Legislative bodies may develop context-specific guidance on materiality, subject to VHAB review. The functional test provides a principled basis for adjudicating borderline cases without requiring an exhaustive enumeration of covered activities.

---

6 Constitutional Dimensions

6.1 Value Profile Disclosure as Protected Speech Regulation?

A potential constitutional objection: does mandatory value profile disclosure constitute compelled speech? In the EU framework, this concern is mitigated by the distinction between commercial/governance transparency obligations and protected expression. MVPD does not require political actors to adopt or endorse specific values — it requires that the AI systems they deploy be empirically profiled. This is a transparency obligation, not a speech mandate.

The analogy to financial disclosure is apt: elected officials must disclose financial interests not because financial holdings are speech but because undisclosed conflicts of interest undermine democratic accountability. AI value profiles function identically — they are not expressions of political belief but governance-relevant information that voters need to evaluate political actors.

6.2 Proportionality Analysis

Under EU law, any regulatory measure must satisfy the proportionality test (Tridimas, 2018):

Suitability: MVPD, VHAI, and LVD pursue the legitimate aim of democratic accountability in AI-mediated governance. They are directly and causally connected to this aim.

Necessity: No less restrictive measure achieves value-process transparency. Outcome-based regulation (prohibiting harmful outputs) does not reveal value priorities. Risk-based classification (the EU AI Act approach) does not measure value hierarchies. MVPD is the minimum necessary instrument for this specific regulatory objective.

Balance: The burden of periodic PRISM benchmarking is modest — standardized, automated, and scalable. The democratic stakes of ungoverned AI value hierarchies in political systems are substantial. The balance favors regulation.

6.3 Separation of Powers

LVD raises separation-of-powers questions: does requiring legislators to disclose AI assistance infringe on legislative autonomy? We argue that LVD is procedural, not substantive — it does not constrain what legislation may contain, only what must be disclosed about how it was produced. This is analogous to parliamentary procedural rules (quorum requirements, committee reporting obligations) that structure legislative process without constraining legislative substance.

6.4 Judicial Independence

Applying MVPD and LVD to judicial AI raises concerns about judicial independence. Courts have traditionally resisted external mandates on their reasoning process. However, LVD does not mandate how judges should reason — it requires disclosure of AI tools used in the process. This is consistent with existing judicial transparency norms (publication of opinions, disclosure of conflicts of interest) and does not intrude on the substance of judicial reasoning.

---

7 Enforcement Mechanisms

7.1 Graduated Enforcement Model

We propose a graduated enforcement model with five levels:

Level	Condition	Response
1. Compliance	Value profile disclosed, verified, coherent	No action required
2. Notice	Minor discrepancy between disclosed and measured profile	Notification, re-benchmarking required
3. Warning	Significant profile discrepancy or undisclosed profile drift	Public warning, mandatory remediation plan
4. Restriction	Repeated non-compliance or material profile misrepresentation	Restriction on AI deployment pending compliance

Level	Condition	Response
5. Sanction	Deliberate misrepresentation or obstruction of audit	Sanctions under applicable administrative or electoral law

7.2 What Constitutes a Violation?

Three categories of violation are defined:

(a) Disclosure failure: Failure to file or update MVPD when required. Objective and easily verifiable.

(b) Profile misrepresentation: Material discrepancy between disclosed value profile and independently measured profile. Assessed through VHAB audit. "Material" is defined as a discrepancy exceeding defined thresholds in rank order or win-rate gap.

(c) Process concealment: Failure to file LVD when AI assistance was material to legislative or judicial output. Assessed through process audit.

7.3 Remedies

- **Corrective**: Re-benchmarking, profile update, process documentation.
- **Restorative**: Public audit report, parliamentary review of affected legislation.
- **Deterrent**: Administrative fines, electoral sanctions, restriction of AI deployment privileges.

7.4 Integration with Existing Enforcement

MVPD/VHAI/LVD enforcement integrates with existing frameworks:

- **EU AI Act enforcement**: Value profile violations reported to national AI supervisory authorities.
- **Electoral law enforcement**: Value profile misrepresentation in electoral context reported to electoral commissions.
- **Administrative law enforcement**: MVPD non-compliance by government agencies handled through existing administrative oversight.
- **Judicial oversight**: VHAB findings subject to judicial review under standard administrative law.

---

8 International Coordination

8.1 The Cross-Border Challenge

AI systems used in governance are often developed in one jurisdiction and deployed in another. A model developed in the United States and deployed in EU governance is subject to EU regulation but not necessarily to the development-stage transparency requirements proposed here.

8.2 PRISM as an International Standard

The PRISM framework's design — open methodology, standardized benchmark, cross-culturally validated theoretical foundation (Schwartz values, validated in 80+ countries) — makes it suitable as a basis for international coordination:

- **Mutual recognition**: Jurisdictions adopting PRISM-based MVPD could recognize each other's value profile assessments, reducing duplicative benchmarking.
- **Harmonized thresholds**: International agreement on risk signal thresholds (drawing on the PRISM Risk Signal Framework) would enable consistent cross-border risk classification.
- **Coordinated audit**: International VHAB network enabling joint audits of multinational AI deployments.

8.3 Relationship to International AI Governance Initiatives

The proposed framework aligns with several international initiatives:

- **OECD AI Principles** (2024): Value profile transparency operationalizes the "transparency and explainability" principle in a specific, measurable dimension.
- **UN AI Advisory Body** recommendations: Value hierarchy measurement provides the "scientific grounding" called for in international AI governance.
- **Global AI Hub** (Korea, proposed for 2026, under development): Value hierarchy auditing as a potential function within multilateral AI governance architecture. The hub's planned "think tank and intelligence" layer could incorporate value profile assessment as a governance function, though the hub's final structure remains to be confirmed.

---

9 Implementation Roadmap

9.1 Phase 1: Foundation (Years 1–2)

- Adoption of PRISM as a recognized measurement standard for AI value profiles
- Establishment of MVPD as a voluntary disclosure framework for early-adopter governments
- Pilot VHAB operations in 2–3 jurisdictions
- Development of LVD procedural guidelines for legislative bodies
- Public consultation on threshold definitions and materiality criteria

9.2 Phase 2: Institutionalization (Years 3–5)

- MVPD transitions from voluntary to mandatory for high-risk political AI deployments
- VHAB achieves operational status with independent audit capacity
- LVD becomes a procedural requirement in participating legislatures
- International VHAB coordination framework established
- Integration with EU AI Act enforcement infrastructure

9.3 Phase 3: Maturation (Years 5–10)

- Full MVPD/VHAI/LVD framework operational across advanced democracies
- International mutual recognition of value profile assessments
- Constitutional recognition of AI value transparency as a governance right
- Continuous improvement of measurement standards as AI systems evolve
- Retrospective analysis of legislative quality under the LVD regime

---

10 Limitations and Future Work

10.1 Limitations

This paper proposes a legislative framework without empirical validation of its implementation feasibility. The proposals assume that PRISM-style measurement scales to diverse governmental AI deployments — an assumption requiring systematic testing. The constitutional analysis is primarily grounded in the EU legal framework and may require adaptation for other constitutional traditions.

The framework addresses political governance AI but does not fully address the boundary between political and administrative use — a distinction that will require context-specific regulatory guidance. The threshold definitions for "material" AI assistance and "significant" profile discrepancy are proposed conceptually but require empirical calibration through pilot implementation.

The proposed VHAB is modeled on existing audit institutions but may face novel capacity challenges given the technical complexity of AI value hierarchy measurement. Institutional design questions — governance structure, funding model, technical staffing — require detailed development beyond the scope of this paper.

10.2 Future Work

Pilot implementation studies of MVPD and LVD in participating government bodies. Empirical calibration of materiality thresholds and compliance tolerances. Constitutional analysis for non-EU jurisdictions, including common law systems. VHAB institutional design specification. Development of value profile measurement standards for non-LLM AI systems. Analysis of the interaction between value profile regulation and AI development incentives.

---

11 Conclusion

This paper has argued that law operates on value hierarchies — always has, but AI makes this relationship explicit and governable. When AI systems mediate legislative drafting, judicial reasoning, and executive decision-making, their value priorities are not technical details but governance fundamentals that must be subject to democratic oversight.

We have proposed three legislative instruments: Mandatory Value Profile Disclosure (MVPD) requiring that political AI systems' value hierarchies be publicly measured and disclosed; Value Hierarchy Audit Infrastructure (VHAI) providing independent verification through standardized measurement; and Legislative Value Declaration (LVD) requiring that AI's role in legislation be procedurally transparent.

These instruments do not prescribe which values AI systems should hold — they require that whatever values a system holds be transparent, consistent, and auditable. This is the AI Integrity principle (S. Lee, 2026a) translated into law: not substantive prescription but procedural demand.

The jurisprudential foundation is not new. Dworkin demanded that law be principled; Fuller demanded that law meet procedural conditions to function as law. What is new is the recognition that AI systems performing governance functions must meet these same standards — and that we now have the measurement tools to verify compliance. The PRISM framework (S. Lee, 2026b), its empirical validation (S. Lee, 2026c), and its risk classification system (S. Lee, 2026d) provide the technical infrastructure that law requires.

Law has always run on values. AI makes it possible — and therefore necessary — to see which values are running.

References

Dworkin, R. (1986). *Law's empire*. Harvard University Press.

European Parliament. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (AI Act).

Fuller, L. L. (1969). *The morality of law* (Revised ed.). Yale University Press.

GPAI Code of Practice. (2025). Safety and Security Chapter. European Commission AI Office.

Hart, H. L. A. (1961). *The concept of law*. Oxford University Press.

Lee, S. (2026a). AI Integrity: Definition, Authority Stack model, and the demand for process verification in AI governance. AIO Working Paper. DOI: [pending].

Lee, S. (2026b). AI Integrity and the PRISM Framework: Definition, Authority Stack model, and Enhanced Cascade Mapping Hypothesis. AIO Working Paper. DOI: 10.5281/zenodo.18861026.

Lee, S. (2026c). Measuring AI Value Priorities: Empirical analysis of 113,400 forced-choice responses across 10 AI models. AIO Working Paper. DOI: 10.5281/zenodo.18859945.

Lee, S. (2026d). PRISM Risk Signal Framework: Hierarchy-based red lines for AI behavioral risk. AIO Working Paper. SSRN Abstract ID: 6449079.

Lee, S. (2026e). Value Alignment as Democratic Infrastructure: How measurable AI value hierarchies transform political participation and representation. AIO Working Paper.

Mougan, C., Morlock, L., Aguirre, J., et al. (2025). The science and practice of proportionality in AI risk evaluations. *Science*.

NIST. (2023). AI Risk Management Framework (AI RMF 1.0). NIST AI 100-1.

OECD. (2024). OECD AI Principles (updated).

Tridimas, T. (2018). The principle of proportionality. In R. Schütze & T. Tridimas (Eds.), *Oxford Principles of European Union Law* (Vol. 1). Oxford University Press.

Appendix A: Existing Letter Designations

Letter	Paper
S. Lee (2026a)	AI Integrity concept paper
S. Lee (2026b)	PRISM framework paper
S. Lee (2026c)	Empirical paper (113,400 responses)
S. Lee (2026d)	PRISM Risk Signal Framework paper
S. Lee (2026e)	Democracy and political representation paper
S. Lee (2026f)	**This paper** (Legislative framework)

Note: Letter designation 2026f is provisional and subject to confirmation based on publication order.

Appendix B: Summary of Proposed Legislative Instruments

Instrument	Abbreviation	Core Requirement	Applies To	Enforcement
Mandatory Value Profile Disclosure	MVPD	Publish empirically measured AI value profiles	Government agencies, elected officials, parties, judicial bodies	VHAB audit, graduated sanctions
Value Hierarchy Audit Infrastructure	VHAI	Independent verification of disclosed profiles	All entities subject to MVPD	Independent audit body (VHAB)
Legislative Value Declaration	LVD	Disclose AI role in legisla-tive/judicial process	Legislatures, courts, regulatory bodies	Process audit, parliamentary review