The Machine Value Score (MVS):

The Safety and Justice Standard for All Engineered Systems

White Paper Draft for IEEE, NIST, and Legal Standards Communities

January 12, 2026

Contents

Part I

Foundations

Chapter 1

Introduction: Scope, Urgency, and Terminology

Why This White Paper Exists

Engineered systems—from credit decision software and hiring screeners to court-scheduling tools and autonomous vehicles—shape access to opportunity, safety, and justice. Some are "smart" (machine learning), some are rigid rules engines, and many are hybrids. Crucially, many are *engineered by neglect*: guardrails never added, fairness never measured, documentation never written. Neglect is not neutral; it is engineering by omission with predictable failure modes.

This white paper proposes the **Machine Value Score (MVS)** as a unified, quantitative standard for *safety and justice* across the full spectrum of engineered decision systems—whether designed explicitly or engineered by neglect. It is written for engineers (IEEE/NIST), regulators (FTC, EEOC, CFPB, FDA, NHTSA), and legal communities (trial courts to appellate and academic).

Definition (Engineered Decision System). Any socio-technical mechanism—algorithmic, rule-based, or hybrid—that *makes or constrains* consequential decisions about people. This includes high-ML systems, low-tech bureaucracies, and processes engineered by neglect.

Problem Statement: Structural Violence by Design or Neglect

When a system denies a loan, filters a resume, triages a patient, schedules a hearing, or flags a neighborhood—wrongly or unfairly—it produces avoidable harm. Today, performance metrics (e.g., accuracy) dominate, while governance metrics (fairness, calibration, legal defensibility, reproducibility) are optional or absent. This fragmentation leads to structural violence at scale.

Historical Pattern

Societies solved analogous crises by making harm *measurable* and *enforceable*:

- **Drug safety (FDA).** Efficacy and safety thresholds before market.
- **Automobile safety (NHTSA).** Crash ratings that consumers and courts could trust.
- **Pollution control (EPA).** Air quality indices that made invisible harm visible.
- **Technical standards (IEEE/NIST).** Interoperability and safety made concrete, auditable, and universal.

The same move is now required for engineered decision systems.

The Core Move

We unify engineering and law through a single rubric:

$$\boxed{\text{MVS} \;=\; \sum_{j=1}^{k} w_j\, m_j \quad \text{with} \quad \sum_{j} w_j = 1}$$

Each component m_j is a normalized (0–1) governance metric (e.g., performance, calibration, fairness, robustness/reproducibility, legal defensibility/transparency, cost/efficiency), and w_j are domain-appropriate weights agreed through an open standards process.

Why "All Engineered Systems" (Not Just AI)

Bad actors often dodge accountability by saying, "this isn't AI." Our scope eliminates that loophole. Whether the decision is produced by a neural network, a rules engine, or a brittle

bureaucracy, it must be *safe and just*. Neglect is engineering; it belongs in scope.

Audience and Use

- **Engineers (IEEE/NIST).** Implement, measure, and publish MVS at design time and before deployment.
- **Regulators (FTC/EEOC/CFPB/FDA/NHTSA).** Gate deployments and recalls using MVS thresholds.
- **Courts and Counsel.** Admit MVS as expert methodology; enforce duty of care using threshold breaches.
- **Public Institutions & NGOs.** Use MVS nutrition labels to monitor, compare, and challenge unsafe systems.

Roadmap

Part I establishes concepts. Part II defines the MVS standard, thresholds, and metrics. Part III describes legal and policy adoption (courts, agencies, IEEE/NIST, international). Part IV covers implementation: nutrition labels, transparency portals, raising the bar over time.

Chapter 2

The Two Pillars: Quantified Safety & Collective Guardianship

Pillar One: The Machine Value Score (MVS)

The MVS makes governance engineerable. It treats fairness, calibration, reproducibility, and legal defensibility as *first-class* metrics, alongside performance and efficiency.

Components and Normalization

Let $m_1, \ldots, m_k \in [0, 1]$ be normalized metrics. Examples:

- **Performance** (e.g., AUROC, F1) normalized to [0,1].
- **Calibration** (e.g., Brier score) inverted and scaled to [0,1].
- **Fairness** (e.g., equal opportunity gap, demographic parity) converted to a 0–1 score via an agreed mapping.
- **Reproducibility/Robustness** (variance across runs, stress tests) mapped to [0,1].
- **Legal Defensibility/Transparency** (documentation, explainability, audit trail) scored via a documented checklist.
- **Cost/Efficiency** (latency, energy, user burden) normalized to [0,1].

Weights and Domain Fit

Weights w_j reflect domain priorities (sum to 1). For lending/hiring, fairness and defensibility weigh more; for real-time safety, calibration and robustness may dominate. Weights are set

by an open standards process (IEEE/NIST working groups) with public comment.

Deployment Thresholds (Draft).
Safe to deploy: MVS ≥ 0.80.
Caution / restricted use: MVS0.70–0.79.
Unsafe—recall or halt: MVS < 0.70.

Pillar Two: The Anti-Isolation Doctrine

Unsafe systems exploit isolation: most people cannot challenge a black box alone. The anti-isolation doctrine recognizes *collective guardianship*. In practical terms:

- **Standing by guardianship:** Enable advocates ("Eagle Eyes") and NGOs to act when individuals cannot.
- **Transparency as duty:** Systems must publish MVS nutrition labels pre-deployment.
- **Public trust logic:** Decision infrastructures function like environmental commons; exposure to unsafe systems is harm.

Engineering translation: increase the *redress rate* r in society so the residual harm multiplier $(1 - r)$ shrinks.

A Stakeholder Value Matrix (Engineering, Law, Policy)

Stakeholder	Perf.	Calib.	Fair.	Reprod.	Defens.	Cost
Line DOJ (trial)	✓	✓	✓		✓	
DOJ Civil Rights			✓		✓	
Regulators (FTC/EEOC/CFPB)	✓	✓	✓	✓	✓	
IEEE/NIST (WG chairs)	✓	✓	✓	✓	✓	✓
Corporate Counsel	✓	✓	✓	✓	✓	✓
Harvard Law Faculty			✓		✓	
Public/NGOs			✓		✓	

(Illustrative; final matrices should be developed via multi-stakeholder workshops.)

From Theory to Practice: The MVS Nutrition Label

A compact, one-page disclosure published before deployment:

- **System overview:** purpose, domain, decision scope.
- **Metrics + weights:** performance, calibration, fairness, reproducibility, defensibility, cost.
- **Final MVS and threshold outcome:** deploy / restricted / recall.
- **Documentation pointers:** data provenance, model card, audit logs.

Engineering by Neglect (No Loopholes). A system that omits fairness tests, calibration checks, or documentation *has engineered a risk profile by omission.* It is in-scope, and its MVS will reflect this (often yielding recall).

Preview of Part II

Part II specifies normalization functions, default weights by domain, worked examples (good/borderline/bad), and validation protocols so courts and agencies can rely on MVS as a repeatable expert methodology.

Chapter 3

Defining the Machine Value Score Standard

Formal Definition

The Machine Value Score (MVS) is defined as a weighted sum of normalized governance metrics:

$$\text{MVS} = \sum_{j=1}^{k} w_j m_j \quad \text{with} \quad \sum_{j=1}^{k} w_j = 1, \quad m_j \in [0,1]$$

Where:

- m_j are normalized scores for governance metrics (e.g., performance, fairness, calibration).
- w_j are weights reflecting domain priorities, set through an open standards process (IEEE/NIST).
- The final MVS is interpreted against deployment thresholds:
  - Safe: MVS ≥ 0.80
  - Caution: MVS 0.70–0.79
  - Unsafe: MVS < 0.70

Normalization Functions

Metrics must be normalized into [0,1]. Example mappings:

- **Performance (AUROC, F1).** Linear rescale, e.g., $m = (\text{F1} - 0.5)/0.5$ for F1 in [0.5,1].

- **Calibration (Brier).** $m = 1 - (\text{Brier}/\text{Brier}_{\max})$.
- **Fairness (parity gaps).** $m = 1 - |\text{gap}|$ (after rescaling to [0,1]).
- **Reproducibility.** $m = 1 - \text{variance}$ (scaled to [0,1]).
- **Legal Defensibility.** Checklist scored as fraction of criteria met.
- **Cost/Efficiency.** Normalized relative to domain benchmarks.

Default Metric Set

The baseline metric set (subject to domain refinement):

1. Performance
2. Calibration
3. Fairness
4. Reproducibility / Robustness
5. Legal Defensibility / Transparency
6. Cost / Efficiency

Default Weights by Domain (Draft)

Weights reflect domain priorities. Proposed defaults:

Domain	Perf.	Calib.	Fair.	Reprod.	Defens.	Cost
Credit & Lending	0.15	0.15	0.30	0.10	0.20	0.10
Employment/Hiring	0.15	0.15	0.35	0.10	0.20	0.05
Healthcare/Triage	0.25	0.25	0.20	0.15	0.10	0.05
Justice/Sentencing	0.15	0.20	0.30	0.10	0.20	0.05
Autonomous Vehicles	0.30	0.20	0.15	0.20	0.10	0.05

These defaults are illustrative and must be validated through IEEE/NIST open processes.

Worked Example 1: Credit Scoring System (Unsafe)

Scenario. A credit risk model used by a major lender. Audited metrics: F1=0.80, calibration moderate, racial disparity high, weak documentation.

Metric	Weight	Score (0–1)
Performance	0.15	0.80
Calibration	0.15	0.70
Fairness	0.30	0.35
Reproducibility	0.10	0.60
Defensibility	0.20	0.40
Cost/Efficiency	0.10	0.80
MVS	1.00	**0.55 (Unsafe)**

Outcome: MVS = 0.55 < 0.70. This system is unsafe and should be recalled or re-engineered.

Worked Example 2: Healthcare Diagnostic AI (Safe)

Scenario. An FDA-cleared model for diabetic retinopathy screening. High accuracy, strong calibration, fairness across age groups validated.

Metric	Weight	Score (0–1)
Performance	0.25	0.95
Calibration	0.25	0.90
Fairness	0.20	0.80
Reproducibility	0.15	0.85
Defensibility	0.10	0.70
Cost/Efficiency	0.05	0.65
MVS	1.00	**0.85 (Safe)**

Outcome: MVS = 0.85 ≥ 0.80. Safe for deployment, subject to continuous monitoring.

Preview of Next Examples

The next chapter provides examples from employment (unsafe hiring screener), justice (predictive policing, unsafe), and autonomous vehicles (borderline). Each illustrates domain-

specific weighting and thresholds in practice.

Chapter 4

Worked Examples Across Domains

Why Examples Matter

Abstract formulas are not enough to change practice. Stakeholders must *see* what safe, unsafe, and borderline systems look like in practice. These worked examples illustrate how the MVS rubric applies across diverse domains. Each example highlights why the thresholds (≥ 0.80, 0.70–0.79, < 0.70) matter.

Example 3: Employment Screening (Unsafe)

Scenario. A resume-screening AI used to filter applicants for technical roles. Audit reveals gender and racial disparities in selection rates, limited calibration, and minimal documentation.

Metric	Weight	Score (0–1)
Performance	0.15	0.80
Calibration	0.15	0.75
Fairness	0.35	0.25
Reproducibility	0.10	0.50
Defensibility	0.20	0.30
Cost/Efficiency	0.05	0.85
MVS	1.00	**0.53 (Unsafe)**

Outcome: Unsafe (MVS=0.53). This system produced disparate impact (women and minorities disproportionately screened out). No defensible deployment without redesign.

Example 4: Predictive Policing Tool (Unsafe)

Scenario. A predictive policing system used to forecast "crime hot spots." Accuracy moderate, calibration poor, fairness gap severe (minority neighborhoods over-policed), documentation weak.

Metric	Weight	Score (0–1)
Performance	0.15	0.70
Calibration	0.20	0.60
Fairness	0.30	0.20
Reproducibility	0.10	0.40
Defensibility	0.20	0.25
Cost/Efficiency	0.05	0.85
MVS	1.00	**0.46 (Unsafe)**

Outcome: Unsafe (MVS=0.46). By codifying biased policing data, the system amplifies injustice. Fails fairness and defensibility standards.

Example 5: Autonomous Vehicle Detection (Borderline)

Scenario. An object detection system for autonomous vehicles. Very high accuracy overall, but weaker detection of pedestrians with darker skin tones at night. Robustness reasonable, but documentation incomplete.

Metric	Weight	Score (0–1)
Performance	0.30	0.93
Calibration	0.20	0.80
Fairness	0.15	0.55
Reproducibility	0.20	0.70
Defensibility	0.10	0.50
Cost/Efficiency	0.05	0.60
MVS	1.00	**0.72 (Caution)**

Outcome: Caution (MVS=0.72). This system should be restricted to pilot use until fairness and documentation improve. A "borderline" case where public risk is non-trivial.

Observations Across Examples

- **Credit (0.55)** and **Employment (0.53)** demonstrate unsafe deployment driven by fairness and defensibility failures.
- **Policing (0.46)** shows the extreme: fairness collapse produces systemic injustice.
- **Autonomous Vehicles (0.72)** illustrates borderline caution: high technical performance undermined by fairness/documentation gaps.
- **Healthcare AI (0.85)** earlier demonstrates a safe case: strong performance + fairness + calibration.

Together, these cases form a spectrum—unsafe, borderline, safe—and demonstrate that MVS thresholds provide a clear governance signal.

Chapter 5

Adoption Pathways: From Standard to Enforcement

Why Adoption Matters

The Machine Value Score (MVS) is only transformative if it can be *enforced*. Technical standards are powerful because courts, regulators, and industries recognize them as authoritative benchmarks. Just as FDA efficacy thresholds or NHTSA crash ratings reshaped entire markets, MVS adoption can create enforceable accountability for engineered decision systems.

Courts and Litigation

Courts already rely on expert methodologies to evaluate harm. Introducing MVS follows a familiar pattern:

- **Expert Witness Testimony.** Courts admit engineering standards (Daubert/Frye). An MVS report presented by an expert can demonstrate that a system fell below safety thresholds.
- **Negligence per se.** Once MVS is recognized, a system failing threshold (MVS < 0.70) could be presumptively negligent.
- **Equal Protection and Due Process.** Disparate impact measured through fairness metrics feeds directly into constitutional claims. MVS provides the quantitative evidence courts require.
- **Remedies.** Courts may order injunctive relief (halt unsafe system), damages (compensate harmed parties), or structural reform (require MVS disclosure going forward).

In practice: A plaintiff denied a loan or a job presents an MVS audit showing fairness=0.25, defensibility=0.30, final score=0.53 (unsafe). The court recognizes unsafe engineering as a violation of duty of care.

Regulatory Agencies

Agencies already gate system deployment in critical domains. MVS offers them a ready-to-adopt standard.

- **FTC.** Consumer protection: deceptive or unfair practices include unsafe decision systems.
- **EEOC.** Employment law: disparate impact measured by MVS fairness scores.
- **CFPB.** Lending oversight: require lenders to publish MVS nutrition labels for credit scoring.
- **FDA.** Healthcare AI: extend medical device approvals to include MVS thresholds.
- **NHTSA.** Autonomous vehicles: mandate MVS scoring before market approval.

Agencies can adopt MVS through rulemaking, guidance, or enforcement action, just as crash tests or emissions standards became mandatory.

IEEE and NIST Integration

Engineering standards become powerful when codified by IEEE and NIST:

- **IEEE.** An official P-series standard (like IEEE 802.11 for WiFi) sets global benchmarks. An IEEE MVS standard would instantly shape engineering practice.
- **NIST.** Publishes Special Publications (SP) that federal agencies and contractors must follow. A NIST SP defining MVS would propagate across U.S. government procurement.
- **De facto Law.** Courts and regulators treat compliance with IEEE/NIST as evidence of "due care." Non-compliance signals negligence.

In effect: IEEE/NIST adoption makes MVS not optional but a baseline expectation across industry.

International Adoption

MVS is globally portable.

- **EU AI Act.** Risk-based regulation already requires conformity assessments. MVS can provide the quantitative scoring layer.
- **UN Human Rights Instruments.** NGOs could submit MVS audits as evidence of systemic harm.
- **Cross-border procurement.** Multinationals must publish MVS scores to meet procurement standards in multiple jurisdictions.

From Recommendation to Requirement

The adoption pathway follows a predictable sequence:

1. Publish as a white paper (MVS draft standard).
2. Adopt by IEEE working group; circulate through NIST for federal alignment.
3. Courts begin admitting MVS audits as expert methodology.
4. Agencies issue guidance requiring MVS disclosure.
5. Legislatures codify thresholds into statutory law.

At that point, unsafe systems become legally indefensible. MVS has evolved from concept to standard to enforceable law.

Preview of Part III

Next chapters illustrate *implementation*: how to format MVS nutrition labels, build transparency portals, and create public reporting architectures. These tools move adoption from legal theory into public practice.

Chapter 6

MVS "Nutrition Labels": Specification & Templates

Purpose

The MVS Nutrition Label is a one–page, public-facing disclosure that summarizes a system's governance metrics, final score, and deployment status. It is designed for engineers, regulators, courts, and the public to read at a glance—like NHTSA crash stars or FDA drug labels.

Scope

Labels apply to *all engineered decision systems*, including systems engineered by neglect (missing guardrails, fairness tests, calibration, or documentation). No AI-only loopholes.

Required Contents (Specification)

Each label MUST include:

R1. **System Identity**: Name, version/hash, owner/operator, domain, decision scope, deployment context (pilot/production).

R2. **Data & Training Summary**: Data sources, time window, provenance, known limitations, data protection notes.

R3. **Metrics Table**: Normalized scores $m_j \in [0, 1]$, weights w_j, weighted contribution, and total MVS.

R4. **Threshold Outcome**: `DEPLOY` (safe), `CAUTION` (restricted), or `RECALL/HALT` (unsafe), with applicable rules/mitigations.

R5. **Fairness & Calibration Evidence**: Short narrative + links to detailed audits (e.g., demographic slices, reliability curves).

R6. **Documentation & Defensibility**: Links to model card, datasheets, logs, versioned code/artifacts, and sign-off chain.

R7. **Monitoring & Redress**: On-call/ops SLOs, drift detection (PSI/KS), incident reporting contact, user appeal/redress process.

R8. **Attestation**: Date, responsible engineer(s), legal/compliance reviewer, and cryptographic signature (recommended).

Deployment Thresholds (Reference)

Safe to deploy: MVS ≥ 0.80. Caution (restricted): MVS0.70–0.79. Unsafe (recall/halt): MVS < 0.70.

Normalization Reminder

Publish exact mappings (Appendix C). At minimum include: performance metric definition, calibration mapping (e.g., Brier $\rightarrow$ [0,1]), fairness metric(s) and parity bounds, robustness tests, defensibility checklist rubric, and cost baselines.

Printable One-Page Template

MVS NUTRITION LABEL (v1.0) *Public Disclosure*

A. System Identity	System name / version
	Owner / operator
	Domain / scope
	Deployment context
	Hash / build ID

B. Data & Training Summary	Datasets / sources
	Time window
	Provenance & licensing
	Known limitations
	Privacy / protection

C. Metrics & Final Score

Metric	**Weight** w_j	**Score** m_j	$w_j m_j$
Performance (domain-appropriate)			
Calibration (Brier→[0,1])			
Fairness (EO/DP mapping)			
Repro./Robustness (variance/stress)			
Defensibility/Transparency (checklist)			
Cost/Efficiency (latency/energy)			
TOTAL MVS	**1.00**		**0.00**

Outcome (circle one): DEPLOY CAUTION RECALL/HALT

D. Fairness & Calibration Evidence (links allowed; summarize key findings)

- Demographic slices: parity gaps within agreed bounds? Outliers?
- Reliability curves / ECE: acceptable? Decision thresholds calibrated?

E. Documentation & Defensibility (links)

- Model card, datasheet, audit logs, training manifest, versioned code.
- Sign-offs: engineering lead, risk/compliance, domain owner.

F. Monitoring & Redress	Drift detection (PSI/KS)
	Alerting / rollback
	User appeals / redress
	Contact (email/URL)

G. Attestation	Date
	Responsible engineer(s)
	Legal/compliance reviewer
	Signature / key

Filled Example: Safe (Healthcare Triage)

A. System Identity

System name / version	RetinaDX v2.4
Owner / operator	ClinicNet Health, Inc.
Domain / scope	Diabetes retinopathy screening (primary care)
Deployment context	Production (FDA-cleared)
Hash / build ID	8c91e1a5_retinadx_v2.4

B. Data & Training Summary

Datasets / sources	Multi-clinic retinal image sets; public benchr
Time window	2018–2024
Provenance & licensing	IRB-approved; patient consent per site polic
Known limitations	Less data for ages < 18; rare pathologies unc
Privacy / protection	PHI de-identification; on-prem inference

C. Metrics & Final Score

Metric	Weight	Score	Weighted
Performance (AUROC→[0,1])	0.25	0.95	0.2375
Calibration (Brier→[0,1])	0.25	0.90	0.2250
Fairness (EO gap mapping)	0.20	0.80	0.1600
Repro./Robustness	0.15	0.85	0.1275
Defensibility/Transparency	0.10	0.70	0.0700
Cost/Efficiency	0.05	0.65	0.0325
TOTAL MVS	**1.00**		**0.85 (DEPLOY)**

D–G. See Appendix links: fairness audit v2.2, calibration curves, model card, monitoring SLOs, appeals portal.

Filled Example: Unsafe (Employment Screening)

A. System Identity

System name / version	ResumeRanker v1.1
Owner / operator	HirePro Tools, LLC
Domain / scope	Resume pre-screen for software roles
Deployment context	Production (multi-client)
Hash / build ID	4f2b77_rr_v1.1

B. Data & Training Summary	Datasets / sources	Historical resumes + hiring outcomes from 1
	Time window	2015–2023
	Provenance & licensing	Private client data; no demographic ground t
	Known limitations	Label bias; proxy variables for gender/race
	Privacy / protection	Standard PII redaction; vendor-hosted infere

C. Metrics & Final Score

Metric	Weight	Score	Weighted
Performance (F1→[0,1])	0.15	0.80	0.1200
Calibration (Brier→[0,1])	0.15	0.75	0.1125
Fairness (EO/DP mapping)	0.35	0.25	0.0875
Repro./Robustness	0.10	0.50	0.0500
Defensibility/Transparency	0.20	0.30	0.0600
Cost/Efficiency	0.05	0.85	0.0425
TOTAL MVS	**1.00**		**0.47 (RECALL/HALT)**

D–G. Summary: severe disparate impact against women/minorities; missing documentation; no appeals workflow. Recommended action: **Halt deployment**, remediate, re-audit.

Publication & Signing Guidance

- Publish the label *before* deployment; update after any material model/data change.
- Sign the label with a verifiable key; include build hash for reproducibility.
- Host public copies; provide machine-readable JSON alongside the PDF.

Anti-Gaming Protections (Draft)

- Independent audit sampling; spot checks on raw slices.
- Cross-metric consistency tests (e.g., calibration vs. fairness under threshold changes).
- Version pinning; compare deltas between label revisions.

Chapter 7

Public Transparency Portals

Purpose

A transparency portal is the public interface for the Machine Value Score (MVS). It lets engineers, regulators, courts, journalists, advocates, and affected communities *discover*, *verify*, and *monitor* engineered decision systems. Like crash-test ratings and air-quality indices, the portal makes governance *visible, auditable, and actionable*.

Outcomes

1. Every in-scope system has a discoverable, versioned MVS Nutrition Label (PDF and JSON).
2. Threshold outcomes (`DEPLOY`/`CAUTION`/`RECALL`) are public and machine-readable.
3. Changes are signed; diffs are published; recalls and incidents generate public alerts.
4. Users have a clearly marked *appeals/redress* path; investigators have reproducible artifacts.

Scope

The portal covers *all engineered decision systems*, including those *engineered by neglect*. If a system makes or constrains consequential decisions about people, it is in scope.

Minimum Viable Portal (MVP) Features

- **Directory & Search**: by name, domain, owner, jurisdiction, outcome, date.
- **Label Viewer**: renders PDF and JSON side-by-side with sha256 checksum and signature status.
- **Version History**: signed lineage; human-readable and structured diffs (metrics, weights, threshold outcome).
- **Alerts**: email/webhook for recalls, incidents, and threshold changes.
- **Redress Link**: per-system appeals/ombudsperson contact and SLA.
- **Bulk Export**: daily snapshot of all JSON labels for watchdogs and researchers.

Non-Functional Requirements

- **Availability**: $\geq 99.9\%$ monthly; appeals endpoint $\geq 99.9\%$.
- **Integrity**: all artifacts signed; immutably stored; append-only event log.
- **Usability**: WCAG 2.1 AA; grade-8 plain-language summaries; multilingual support.
- **Portability**: machine-readable JSON schema (Chapter 6); bulk export; stable IDs.

Reference Architecture

Component	Responsibilities
Ingest Service	Validate JSON labels; verify signatures; enforce schema and thresholds; quarantine failures.
Artifact Store	Immutable storage for PDF/JSON/model cards/audit attachments; content-addressed (sha256).
Metadata DB	Registry of systems, owners, outcomes, timestamps, digests; supports faceted search.
Diff Service	Computes human-readable & structured diffs between versions.
API Gateway	Read-only REST/Graph endpoints; rate limiting; API keys for bulk users.
Web UI	Public portal showing labels, diffs, alerts, and appeals links (WCAG compliant).
Alerting Bus	Pub/Sub for incidents, recalls, threshold shifts (email/webhooks/SSE).

Core Data Model

Entity	Key Fields	Notes
System	id, name, owner, domain, scope, jurisdiction	Unique across portal.
Label	system_id, version, total_mvs, outcome, weights, scores	JSON per schema.
Artifact	label_id, type(PDF/JSON/MCARD/AUDIT), sha256, url	Immutable; signed.
Event	system_id, type(ingest/inc/recall/update), payload, ts	Audit trail and alerts.
Subscriber	email/webhook, filters, status	For notices and recalls.

Public API (v1)

Base path: `/api/v1`

Endpoint	Method	Description
/labels	GET	List/filter systems (domain, owner, outcome, date_range).
/labels/{id}	GET	Latest label metadata; links to artifacts; signature status.
/labels/{id}/versions	GET	All versions with timestamps and outcomes.
/labels/{id}/v/{n}	GET	Retrieve specific version (JSON/PDF links + digests).
/diff/{id}/{n}/{m}	GET	Structured diff (metrics, weights, outcome).
/events	GET	Stream recalls/incidents (supports SSE).
/export/daily	GET	Bulk tarball of current JSON labels.
/subscribe	POST	Create alert subscription (filters + endpoint).

Example event payload


```
{
  "type": "RECALL",
  "system_id": "city.policing.hotspots",
  "label_version": 12,
  "previous_outcome": "CAUTION",
  "new_outcome": "RECALL",
  "reason": "Fairness breach: EO gap > 0.20 for 2 consecutive months",
  "timestamp": "2025-08-25T18:42:03Z",
  "links": {
    "label_pdf": ".../labels/city.policing.hotspots/v/12.pdf",
    "label_json": ".../labels/city.policing.hotspots/v/12.json",
    "incident_report": ".../incidents/2025-08-25-1839.html"
  }
}
```


Security, Privacy, and Integrity

- **Signing**: Vendors sign JSON and PDF labels (OpenPGP/X.509). The portal validates signatures and publishes signer identity.
- **Hashing**: All artifacts are content-addressed (sha256); digests rendered in UI for public verification.
- **Tamper-evidence**: Append-only event log (hash chain) for ingests and updates; public proofs available.

- **PII policy**: Labels contain no PII; audits publish only aggregates; links to raw data are forbidden.
- **Rate limits**: Prevent scraping abuse while keeping bulk exports open and predictable.

Accessibility and Inclusion

- **WCAG 2.1 AA**: keyboard navigation, focus order, alt text, high-contrast mode.
- **Plain language** summaries (grade-8 reading level); multilingual (at least top 3 local languages).
- **Printable one-pagers** with QR codes to JSON/PDF for offline distribution.

Versioning, Retention, and Open Records

- **Stable IDs**: A system id never changes; versions are monotonic integers.
- **Retention**: Keep all labels and events $\geq$ 10 years; do not delete historical versions (mark superseded).
- **Open records**: Provide daily bulk export and browsable archives to satisfy FOIA/state records requirements.

Anti-Gaming & Auditability

1. **Churn watch**: Flag abnormal label revision frequency; require rationale field on ingest.
2. **Consistency checks**: If fairness improves while performance collapses, require justification link (trade-off analysis).
3. **Random sampling**: Independent spot audits of demographic slices; recompute MVS from supplied artifacts.
4. **Benchmark sanity**: Compare submitted metrics against public benchmarks when applicable.

Rollout Plan (90 Days)

1. **Weeks 1–2**: Schema validation; ingest service; minimal UI; publish JSON schema URL.
2. **Weeks 3–6**: Directory + search; label viewer; version history; bulk export; signature verification.
3. **Weeks 7–8**: Diff service; event stream; alerts; basic accessibility pass.
4. **Weeks 9–10**: Security review; tamper-evident logs; capacity tests.
5. **Weeks 11–12**: Pilot five labels (one per domain); public launch with governance charter.

Operational SLAs

Process	**SLA**
Ingest + validate label	$<$ 24 hours from submission
Publish material update	$<$ 24 hours after deployment change
Incident banner on recall	$\leq$ 1 hour from declaration
Appeals link uptime	$\geq$ 99.9% monthly
Bulk export refresh	Daily at 00:00 UTC

Key Performance Indicators (KPIs)

- Coverage: % of in-scope systems with *current* labels.
- Freshness: median time from model change to label update.
- Safety trend: counts of recalls/cautions over time by domain/owner.
- Equity: appeals volume and median time-to-resolution.
- Integrity: signature verification pass rate; audit pass rate.

Reference Implementation Blueprint

- **Backend**: Python/Go; JSON Schema validation; Postgres metadata; S3-compatible artifact store.
- **API**: REST + GraphQL; OpenAPI published; deterministic pagination for bulk.

- **UI**: Static SPA (React/Vue/Svelte); server-side rendering optional for SEO and accessibility.
- **Infra**: Containerized; WAF; CDN for PDFs/JSON; daily backups; staged environments.
- **Open source**: MIT/Apache-2.0 reference portal to reduce adoption cost.

Case Study: City Portal (Hypothetical)

Context. A city runs hiring, benefits, policing, triage, and court scheduling systems. It launches a public portal with five labels:

1. *Benefits Eligibility (Rules Engine)*: MVS 0.82 (Deploy) with quarterly fairness checks.
2. *Resume Screener (Vendor)*: MVS 0.53 (Recall) → halted; remediation plan filed.
3. *Hotspot Policing (Vendor)*: MVS 0.46 (Recall) → program sunset after council vote.
4. *Hospital Triage (ML)*: MVS 0.85 (Deploy) with monitoring SLOs.
5. *Court Scheduling (Legacy)*: MVS 0.68 (Caution) → human override policy; re-audit scheduled.

Six-month results. 100% of systems labeled; two recalls prevented downstream harm; appeals median time fell from 45 to 12 days; public trust survey rose 18%.

Governance Charter (Public Summary)

- **Stewardship Board**: engineering, compliance, civil society, affected communities.
- **Change Control**: public RFCs for schema changes; semantic versioning for labels and APIs.
- **Conflict of Interest**: disclose vendor funding and audit relationships on each label page.

Interoperability and Federation

- Implement `/well-known/mvs-labels.json` listing all current labels.
- Support cross-portal aggregation; publish a compatibility matrix for schema versions.
- Provide a daily `/export/daily` tarball for independent archives.

Summary

Transparency portals operationalize MVS by making safety and justice *observable*. With signed labels, immutable history, alerts, and accessible interfaces, they create the same public pressure that transformed automotive safety and environmental health.

Chapter 8

Thresholds & Governance Playbook

Why Thresholds

Thresholds convert analysis into action. They determine deploy/restrict/recall and drive incentives to remediate.

Default Thresholds (v1.0)

Safe: MVS ≥ 0.80; Caution: MVS0.70–0.79; Unsafe/Recall: MVS < 0.70.

Domain Overrides (illustrative)

Domain	Deploy	Caution	Recall
Healthcare/Triage	≥ 0.85	0.78–0.84	< 0.78
Justice/Sentencing	≥ 0.85	0.78–0.84	< 0.78
Autonomous Vehicles	≥ 0.83	0.75–0.82	< 0.75
Credit/Employment	≥ 0.80	0.70–0.79	< 0.70

Raising the Bar

1. **T0 (year 1)**: adopt defaults.
2. **T1 (year 3)**: +0.02 across domains; publish roadmap.
3. **T2 (year 5)**: +0.03 more; require narrower fairness gaps.

Incident Response & Recall

- **Triggers**: drift beyond bounds; fairness breach; calibration failure; redress backlog.
- **Actions**: immediate banner in portal; halt or restrict; publish incident report within 7 days.
- **RACI**: product owner (accountable), governance lead (responsible), counsel (consulted), agency/public (informed).

Chapter 9

Operations: Monitoring, Alerts, & Redress

Monitoring Metrics

- **Data drift**: Population Stability Index (PSI), Kolmogorov–Smirnov (KS).
- **Fairness drift**: parity gaps over rolling windows.
- **Calibration drift**: Brier/ECE over time; threshold stability.
- **Reliability**: latency, error, outage SLOs.

Alerting & Gating

- **Warn**: PSI > 0.1 or fairness gap approaching bound.
- **Page**: PSI > 0.25 or validated bias incident.
- **Gate**: automatic rollback/fail-open to human review if MVS crosses into Caution/Recall.

Retraining Cadence

- Staged environments; versioned datasets; seed control.
- Before promote: run full MVS re-audit + update portal artifacts.

Redress Workflow

1. **Intake**: user files appeal (web/phone/mail).
2. **Triage**: SLA 7 days; provide case ID; freeze adverse action.
3. **Review**: human-in-the-loop; counterfactual test; provide explanation.
4. **Remedy**: reverse, compensate, or escalate to ombudsperson.

Chapter 10

IEEE & NIST: From Draft to De Facto Law

IEEE Path

- Form a Working Group; circulate PAR; iterate drafts; ballot; publish standard.
- Map MVS to P7000-series ethics standards and safety conventions.

NIST Path

- Publish a NIST Special Publication (SP) detailing MVS computation, tests, and reporting.
- Reference in Federal procurement; agencies inherit requirements.

Why Courts Care

Judges treat IEEE/NIST compliance as evidence of due care. Non-compliance signals negligence when harms are foreseeable.

Chapter 11

Agencies & Procurement: Making MVS Mandatory

Rulemaking Hooks (examples)

- **FTC**: unfair/deceptive acts—unsafe decision systems.
- **EEOC**: employment selection with disparate impact.
- **CFPB**: credit underwriting transparency & fairness.
- **FDA**: software as a medical device—add MVS gates.
- **NHTSA**: AV safety—publish MVS before deployment.

Model Procurement Clause

Vendors shall submit a current MVS Nutrition Label (PDF & JSON) for each decision system prior to award. Systems with MVS < 0.80 *are ineligible for production deployment. Material updates require re-submission within 24 hours.*

State/Local Adoption

Model policy packets can cascade requirements to agencies, schools, courts, and hospitals.

Chapter 12

Courts: Establishing MVS as Expert Methodology

Admissibility (Daubert/Frye)

- Testable, peer-reviewed, known error rates.
- General acceptance via IEEE/NIST publication.

Expert Report Template

1. System description & scope.
2. Metric normalization and weights (with citations).
3. Computed MVS; threshold outcome; sensitivity analysis.
4. Causation narrative linking unsafe score to plaintiff harm.

Remedies

Injunction (halt/recall), structural reform (publish labels, monitoring), damages/restoration, attorney fees.

Chapter 13

EU AI Act & International Alignment

EU AI Act

Risk classes require conformity assessment. MVS provides the quantitative layer (metrics, thresholds, labels).

ISO/IEC Crosswalk

Align MVS metrics to ISO/IEC safety, quality, and management standards; publish mapping tables for auditors.

Global South Considerations

Low-resource settings: prioritize transparency, lightweight audits, and community oversight to avoid extractive deployments.

UN & Human Rights

NGOs can file MVS audits as evidence of systemic harm; exposure to unsafe systems treated as public-trust injury.

Chapter 14

Human Rights & Collective Standing (Anti-Isolation)

Doctrine

No person should face systemic harm alone. Empower advocates ("Eagle Eyes") to act where individuals cannot.

Legal Theories

- Next-friend standing for those unable to sue.
- Public trust: exposure to unsafe systems = harm to the commons.
- Class actions: common questions predominate when MVS shows uniform failure.

Engineering Tie-In

Raising the social redress rate r reduces net harm multiplier $(1 - r)$ across domains.

Chapter 15

Research Agenda

Metrics & Mappings

- Better fairness mappings (multi-attribute parity; trade-off surfaces).
- Calibration under shift; uncertainty-aware scoring.

Uncertainty Bands

Monte Carlo MVS with confidence intervals; stress tests across seeds, samples, and shifts.

Gaming Resistance

Adversarial governance tests: detect metric cherry-picking and label laundering; cross-metric consistency checks.

Reference Implementations

Open-source libraries (Python/R) for label generation, schema validation, portals, and diff tooling.

Chapter 16

Conclusion: Governance as Engineering, Justice as Architecture

No Loopholes

All decision systems are engineered—by design or by neglect. MVS closes the loophole: if it makes or constrains decisions, it must be safe and just.

The Two Pillars

Pillar One quantifies safety; Pillar Two prevents isolation. Together they end structural violence by design.

Call to Action

Publish labels; set thresholds; recall unsafe systems; empower collective standing. Adopt MVS now and raise the bar on a public roadmap.

Appendix A

Appendix

A. MVS Normalization Examples

Document exact mappings for performance, calibration (Brier/ECE), fairness (EO/DP), robustness, defensibility, and cost.

B. Templates

Deployment Readiness Report; Expert Witness Outline; Procurement Clause boilerplate.

C. JSON Schema & Validation

Publish the official schema URL; include a checksum of each label artifact and a short validator script link.

D. Glossary

Cross-disciplinary terms spanning engineering, law, and policy.

Architect's Declaration

The Machine Value Score (MVS) — Safety & Justice Architecture

Christine Hillier

Kapukai Governance Lab — *Truth as a Public Utility*

Declaration. I, **Christine Hillier**, hereby declare that I have architected the **Machine Value Score (MVS)**—the first unified, quantitative standard for **safety and justice** across all engineered decision systems, including those engineered by neglect.

MVS is an *architecture*, not merely an idea. It comprises:

- A formal scoring rubric (MVS $= \sum w_j m_j$) with normalized governance metrics and domain-specific weights;
- Deployment thresholds (Deploy / Caution / Recall) that convert analysis into action;
- Public *MVS Nutrition Labels* and transparency portals for versioned, machine-readable accountability;
- Adoption pathways through IEEE/NIST, agency rulemaking, procurement requirements, and court admissibility;
- A doctrine of *Collective Guardianship (Anti-Isolation)* ensuring no person faces systemic harm alone.

Principles.

1. **Truth as a Public Utility**: Safety and justice must be calculable, visible, and enforceable.
2. **No Loopholes**: AI, rules engines, and bureaucracies engineered by neglect are all in scope.
3. **Collective Guardianship**: Society must engineer protection against systemic harms.
4. **Raising the Bar**: Thresholds and parity bounds should tighten over time.
5. **Global Justice**: The architecture is portable across jurisdictions and rights frameworks.

Call to Action. I invite engineers, regulators, judges, and communities to adopt, enforce, and evolve MVS so that every system that touches human life is *safe, just, and accountable.*

Attestation by Architect

Date: _____ City/State: _____

_____

Christine Hillier

Kapukai Governance Lab — *Truth as a Public Utility*

Optional Unsworn Declaration (28 U.S.C. § 1746):

"I declare under penalty of perjury that the foregoing is true and correct."

Executed on _____. Signature: _____

_____

Notary Public Jurat

(To be completed by a Notary Public)

State of _____ County of _____

Subscribed and sworn to (or affirmed) before me on this _____ day of _____, 20_____, by **Christine Hillier**, who is personally known to me or has produced identification.

Type of ID: _____

_____ _____

Notary Public Signature Notary Printed Name

Commission No.: _____ My Commission Expires: _____

Notary Seal:

Reference/Archive (optional): DOI/URL or repository pointer for this declaration or subsequent notarized copy:

_____