WHITEPAPER | ELITE DOCTRINE EDITION

AI That Survives Court Wins the Market

The Doctrine of Litigation-Grade Security

Designing AI-Native Cyber Control Planes That Withstand Regulators, Plaintiffs, and Sovereign AI Regimes (2026–2030)

Evidence-Based Governance Architecture for Board Directors, CISOs, and General Counsel

Kieran Upadrasta

CISSP, CISM, CRISC, CCSP | MBA | BEng

27 Years' Cyber Security Experience | Big 4 Consulting (Deloitte, PwC, EY, KPMG)
21 Years Financial Services | AI Cyber Security Programme Lead
Professor of Practice (Cybersecurity, AI & Quantum Computing), Schiphol University
Honorary Senior Lecturer, Imperials | *UCL Researcher*

www.kie.ie | info@kieranupadrasta.com | March 2026

EXECUTIVE INTELLIGENCE DASHBOARD

$70B+ Class Action Settlements 2025	**15%** Max Turnover Penalty Exposure	**80%** NEDs: AI Oversight Gap	**$93.75B** AI Security Market by 2030

✓ EU AI Act | ✓ DORA | ✓ NIS2 | ✓ ISO 42001 | ✓ SEC

Keywords: DORA Compliance | AI Governance (ISO 42001) | Board Reporting | M&A Cyber Due Diligence | Zero Trust Architecture | Post-Quantum Cryptography | NIS2 Compliance | EU AI Act Compliance | Interim CISO

THE DOCTRINE AT A GLANCE

"If it cannot be evidenced, it cannot be defended."

AI that survives court wins the market.

Everything else is a liability waiting for a plaintiff.

THREE-LAYER GOVERNANCE MODEL

ARCHITECTURE

AI Control Plane
Evidence Chain Model

GOVERNANCE

Decision Rights Architecture
AI Accountability Stack
Contract Control Matrix

OPERATIONAL

Recoverability Mandate
Crisis Governance

THE ENFORCEMENT LANDSCAPE

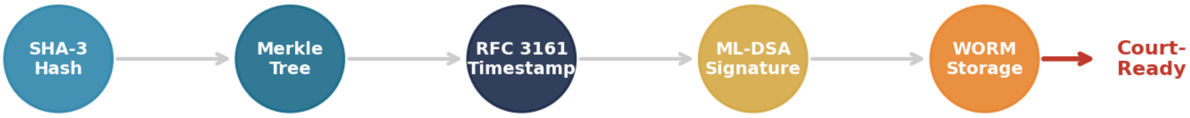
$70B+
Class Action Settlements 2025

15%
Max Turnover Penalty Stacking

80%
NEDs: Oversight Inadequate

$10.22M
US Avg Breach Cost (IBM 2025)

LITIGATION-GRADE EVIDENCE ARCHITECTURE



SHA-3 Hash → Merkle Tree → RFC 3161 Timestamp → ML-DSA Signature → WORM Storage → Court-Ready

12-MONTH IMPLEMENTATION DOCTRINE

TRIAGE

Months 1–3

AI inventory
Evidence logging
Decision rights

→

TRANSLATION

Months 4–8

Control Plane
Monitoring
Vendor governance

→

TRANSFORMATION

Months 9–12

EU AI Act readiness
Litigation simulation
Certification

RETURN ON GOVERNANCE INVESTMENT

Expected annual loss: 125*M*–395M Governance investment: 2.5*M*–8M

Return on Governance Investment: 15x – 50x

✓ EU AI Act ✓ DORA ✓ NIS2 ✓ ISO 42001 ✓ SEC

Kieran Upadrasta | CISSP, CISM, CRISC, CCSP | MBA, BEng

27 Years Cyber Security | Big 4 Consulting | 21 Years Financial Services

info@kieranupadrasta.com | www.kie.ie

Table of Contents

GOVERNING PRINCIPLE

"If it cannot be evidenced, it cannot be defended."

This doctrine rests on a single axiom: every AI decision is a potential exhibit, every model output is discoverable evidence, and every governance gap is an invitation for plaintiffs.

1. Executive Summary

THE BOARD-LEVEL DOCTRINE

The courtroom is now the kill chain's final stage. Every AI system operating without litigation-grade evidence architecture is a liability in waiting. The question is no longer *whether* your AI will face legal scrutiny — it is *whether your AI can survive it*.

In 2025, AI-related lawsuits doubled to over 70 active cases in the US alone [1]. Anthropic settled for $1.5 billion — the largest copyright settlement in history [2]. Meta paid $1.4 billion to Texas for biometric violations. The average US data breach now costs $10.22 million (IBM/Ponemon 2025, n=600) [3], and 97% of organisations that suffered AI-specific breaches lacked proper access controls [3]. Meanwhile, the EU AI Act's Article 99 penalties — up to 7% of global annual turnover — took effect in August 2025, exceeding even GDPR thresholds.

Regulators in Brussels, Beijing, Singapore, and Sacramento are converging on a single verdict: **AI without provable governance is AI without a future.** The organisations that architect their AI systems to produce court-ready evidence in real time will not merely survive enforcement — they will command valuation premiums, win regulated contracts, and set the terms of market entry for everyone else.

This whitepaper introduces the doctrine of **Litigation-Grade Security**: a control architecture that treats every AI decision as a potential exhibit, every model output as discoverable evidence, and every governance gap as an invitation for plaintiffs.

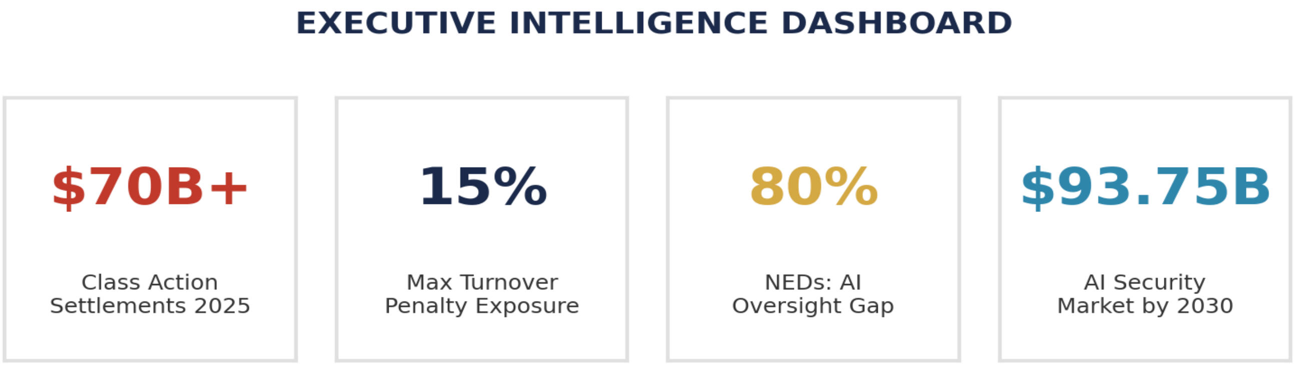

EXECUTIVE INTELLIGENCE DASHBOARD
$70B+
Class Action Settlements 2025
15%
Max Turnover Penalty Exposure
80%
NEDs: AI Oversight Gap
$93.75B
AI Security Market by 2030

Figure 1: Executive Intelligence Dashboard — Key metrics driving the litigation-grade imperative

2. The $1.5 Billion Wake-Up Call: AI Litigation Explodes

The litigation landscape for AI underwent a phase change in 2025. What had been a slow accumulation of regulatory signals became an avalanche of enforceable consequences.

2.1 Copyright Warfare at Existential Scale

The New York Times v. OpenAI — filed December 2023 seeking billions in statutory damages at $150,000 per work — survived OpenAI's motion to dismiss in March 2025 [4]. Judge Sidney Stein allowed direct and contributory copyright infringement claims to proceed. The court found evidence that Bing Chat had copied all but two of the first 396 words of a Times article verbatim. Anthropic's $1.5 billion copyright settlement in September 2025 covered approximately 500,000 books at $3,000 per book [5]. Getty Images v. Stability AI produced the first UK judgment on AI copyright in November 2025, with the court holding that AI model weights are not "copies" under UK law but that Stability AI bore responsibility for trademark-infringing outputs.

2.2 Algorithmic Discrimination Becomes an Enforcement Category

The EEOC's first AI hiring discrimination settlement — iTutorGroup in August 2023 — established that automated rejection of applicants based on age violated federal law. Workday faces a landmark class action after the court allowed claims that its AI screening tool functioned as a discriminatory "employment agency" under Title VII. Colorado's AI Act, effective June 30, 2026, imposes $20,000 per violation for algorithmic discrimination.

2.3 Nine-Figure Liability Events

A Florida jury awarded $243 million against Tesla in 2025 — including $200 million in punitive damages — after a fatal Autopilot crash. Tesla had rejected a $60 million pre-trial settlement. Total US class action settlements exceeded $70 billion in 2025 — a record. Data privacy filings surged to 1,800+, a 200% increase since 2022.

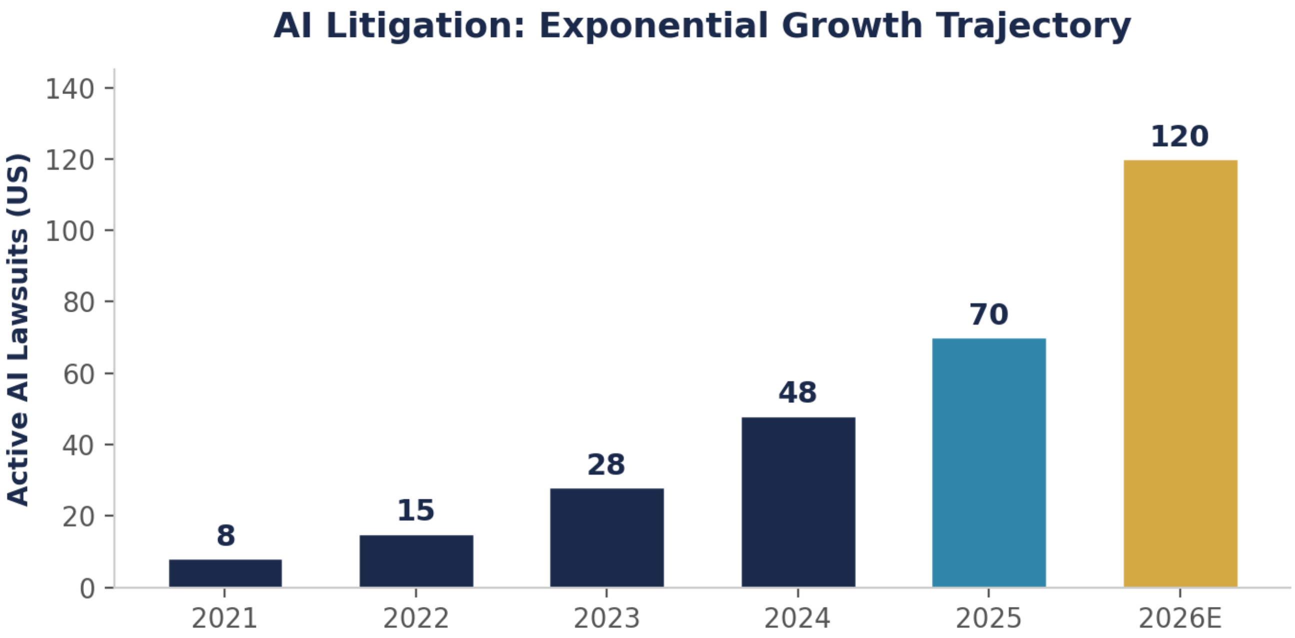


Figure 2: AI litigation growth trajectory — active US cases by year

LANDMARK AI LITIGATION: SELECTED CASE MATRIX

Case / Action	Jurisdiction	Amount / Penalty	Governance Lesson
NYT v. OpenAI	US (SDNY)	$Billions sought	Training data provenance is discoverable
Anthropic Settlement	US	$1.5B	Copyright risk scales with training corpus size
Meta v. Texas	US (Texas)	$1.4B	Biometric AI requires explicit consent architecture
Tesla Autopilot	US (Florida)	$243M	Autonomous system oversight failures are punitive
Clearview AI (EU)	EU (Multi)	€100M+ fines	Personal director liability under investigation
iTutorGroup (EEOC)	US (Federal)	$365K + policy	Age-based AI filtering violates federal law

Table 1: Selected landmark AI litigation and enforcement actions (2023–2026)

3. When Evidence Rules Change, Architecture Must Follow

The evidentiary standards for AI are being rewritten in real time. On June 10, 2025, the US Judicial Conference's Advisory Committee on Evidence Rules approved **Proposed Federal Rule of Evidence 707** for public comment — the first uniform federal standard for AI-generated evidence admissibility. The rule would apply the Daubert standard to machine-generated evidence, requiring courts to examine whether the AI method has been tested, peer-reviewed, has known error rates, and is generally accepted.

3.1 The Daubert Standard Meets AI Architecture

The Daubert framework — established in *Daubert v. Merrell Dow Pharmaceuticals*, 509 U.S. 579 (1993) — demands that AI systems produce explainable, reproducible, and verifiable outputs. A New York criminal court has already excluded AI-enhanced video evidence after a Frye hearing. In *Matter of Weber* (2024), a judge questioned an expert's AI-assisted testimony when the witness could not recall the prompts used or explain how the AI reached its conclusions.

3.2 Operational Audit Trails vs. Legal Audit Trails

Courts are beginning to require documentation of which exact model version made a disputed decision, what input data was provided, what processing occurred, and whether any human overrode the output. The critical distinction most organisations miss is between **operational audit trails** (debugging and incident investigation) and **legal audit trails** (proving facts to courts and regulators). The former requires standard database logging. The latter demands cryptographic integrity, independent timestamping, and completeness guarantees. Most cloud AI services provide operational logs. Almost none provide legally defensible audit trails.

LEGAL VS. OPERATIONAL AUDIT TRAIL REQUIREMENTS

Requirement	Operational Log	Legal Audit Trail
Integrity	Database checksums	Cryptographic hash chains (SHA-3)
Timestamping	System clock	RFC 3161 trusted timestamping
Completeness	Best effort	Provable via Merkle tree inclusion proofs
Tamper Evidence	Access controls	WORM storage + append-only verification
Quantum Resistance	Not addressed	ML-DSA / SLH-DSA post-quantum signatures
Admissibility	Not designed for court	Meets ISO 27037, Daubert, and FRE 707

Table 2: The gap between operational logging and court-admissible evidence architecture

4. The AI Control Plane: Governance as Runtime Infrastructure

The AI Control Plane™ is the central architectural concept of litigation-grade security. It is a unified governance, identity, and risk-management layer that orchestrates all autonomous agents, models, and machine identities across the enterprise portfolio. It functions as the operating system for enterprise AI governance — separating what agents are *permitted* to do from how they *execute*.

Forrester published its first evaluation of the agent control plane market in December 2025, defining it as infrastructure that inventories, governs, orchestrates, and assures heterogeneous AI agents across vendors and domains. O'Reilly's 2025 analysis confirmed that governance must exist as runtime architecture, not as a compliance exercise conducted quarterly.

4.1 Five Capabilities of a Litigation-Grade Control Plane

Capability	Function	Litigation Value
1. Standard Telemetry	Captures AI agent lifecycle data across every interaction	Produces discoverable evidence chain
2. Reliable Evaluation	Tests agents against adversarial scenarios pre-deployment	Demonstrates due diligence standard
3. Continuous Monitoring	100+ metrics: hallucination, toxicity, PII, drift	Real-time compliance evidence
4. Enforceable Policy	Runtime guardrails blocking harmful outputs	Proves controls were active at time of incident
5. Auditable Governance	Complete evidence chains for regulatory scrutiny	Court-ready dossier within 4 hours

Table 3: The five capabilities of a litigation-grade AI Control Plane

4.2 Decision Logging as Forensic Backbone

A structured decision log captures for each AI decision: timestamp (RFC 3339), model ID and version, input prompt and context, output response and confidence score, guardrail evaluations, session metadata, cost metrics, and human-in-the-loop approvals or overrides. This is not optional documentation. This is electronically stored information subject to subpoena and legal hold.

The control plane introduces a new operational metric: **Mean Time to Evidence (MTTE)** — measuring how quickly an organisation can produce a court-ready dossier following an incident or regulatory inquiry. Organisations that cannot produce evidence within hours face the same exposure as organisations that have no controls at all.

AI DECISION LOG SCHEMA: LITIGATION-GRADE FIELDS

Field	Format	Legal Purpose
Timestamp	RFC 3339 + RFC 3161 TSA	Establishes when the decision occurred with independent attestation
Model ID + Version	UUID + semantic version	Identifies exactly which algorithm produced the output
Input Context	Structured JSON + hash	Proves what information the model received
Output + Confidence	Response + float [0,1]	Documents what the model produced and its certainty
Guardrail Evaluations	Boolean array + reasons	Proves safety controls were active and whether they triggered
Human Override	User ID + action + reason	Documents human accountability in the decision chain

Hash Chain	SHA-3 + Merkle proof	Cryptographically proves record has not been altered
PQC Signature	ML-DSA (FIPS 204)	Ensures evidence integrity survives quantum computing era

Table 12: AI decision log schema for litigation-grade evidence architecture

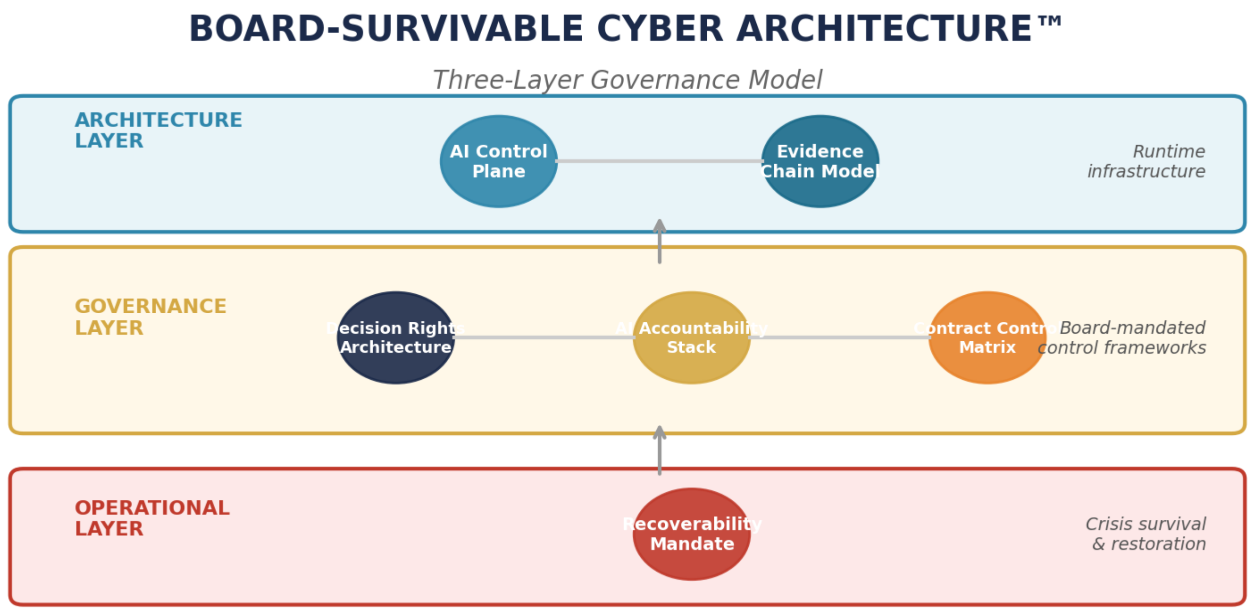
Drift detection extends beyond statistical model drift to what this doctrine terms **structural drift** — the phenomenon where autonomy grows faster than accountability and AI systems gradually exceed their defined authority without explicit configuration changes. The AI Control Plane detects structural drift by continuously comparing actual agent behaviours against registered authority boundaries, generating alerts when divergence exceeds configurable thresholds.

5. Five Frameworks That Make AI Governance Prosecutable

The Board-Survivable Cyber Architecture™ codifies five interlocking frameworks organised across three operational layers — each designed to withstand PRA, FCA, ECB, and EBA supervisory review, and critically, plaintiff discovery. The three-layer model ensures that architecture, governance, and operational resilience function as an integrated system rather than disconnected compliance exercises.

Layer	Frameworks	Function	Board Question Answered
Architecture	AI Control Plane™, Evidence Chain Model™	Runtime infrastructure that generates and preserves evidence	"Can our AI produce court-ready evidence within 4 hours?"
Governance	Decision Rights Architecture, AI Accountability Stack, Contract Control Matrix	Board-mandated control frameworks that define authority and accountability	"Who authorised this AI decision, and can we prove it?"
Operational	Recoverability Mandate	Crisis survival and restoration capability that passes live testing	"Can we restore operations and demonstrate compliance after a material incident?"

Table 13: Three-layer framework architecture — mapping governance to board accountability



Upadrasta Index™ — Composite risk-adjusted return metric spanning all three layers

Figure 3: Board-Survivable Cyber Architecture™ — three-layer governance model

Architecture Layer

The Architecture Layer comprises the AI Control Plane (detailed in Section 4) and the Evidence Chain Model. Together, they form the runtime infrastructure that generates, preserves, and presents governance evidence.

5.1 The Evidence Chain Model™

Operationalises the principle that obligation flows to control, control flows to evidence, and evidence flows to assurance. Each regulatory obligation is traced to a specific technical control, each control generates machine-readable evidence, and each evidence artefact feeds an assurance dashboard visible to the board. The model replaces Mean Time to Detect with **Mean Time to Evidence**.

Governance Layer

5.2 The Decision Rights Architecture

Eliminates governance drift through board-mandated authority grids, escalation protocols, and spend gates. Maps decision authority across four levels: strategic (board), tactical (CISO/CTO), operational (engineering), and automated (agent-level). Each level has explicit delegation boundaries and override authorities documented in a format that withstands both internal audit and regulatory examination.

5.3 The AI Accountability Stack

Integrates ISO 42001 AI Management System requirements with EU AI Act governance obligations. Encompasses model inventory, algorithmic accountability, bias auditing, and AI safety controls. Maps to the OWASP Agentic AI Top 10, CSA MAESTRO's seven-layer threat architecture, and MITRE ATLAS adversarial techniques.

5.4 The Contract Control Matrix

Transforms governance into procurement-ready schedules with acceptance criteria and supplier obligations. Directly addresses the third-party risk chain that DORA Articles 5 and 6 mandate for ICT risk management, where AI vendors are increasingly designated as Critical Third-Party Providers subject to fines of €5 million for companies and €500,000 for individuals. **If the control has no owner, the control does not exist.**

Operational Layer

5.5 The Recoverability Mandate

Enforces RTO/RPO realism, restoration testing, and crisis governance. It survives material incidents, not just audits. The measurement standard is restoration, not effort.

6. A Regulatory Storm Without Precedent Converges in 2026

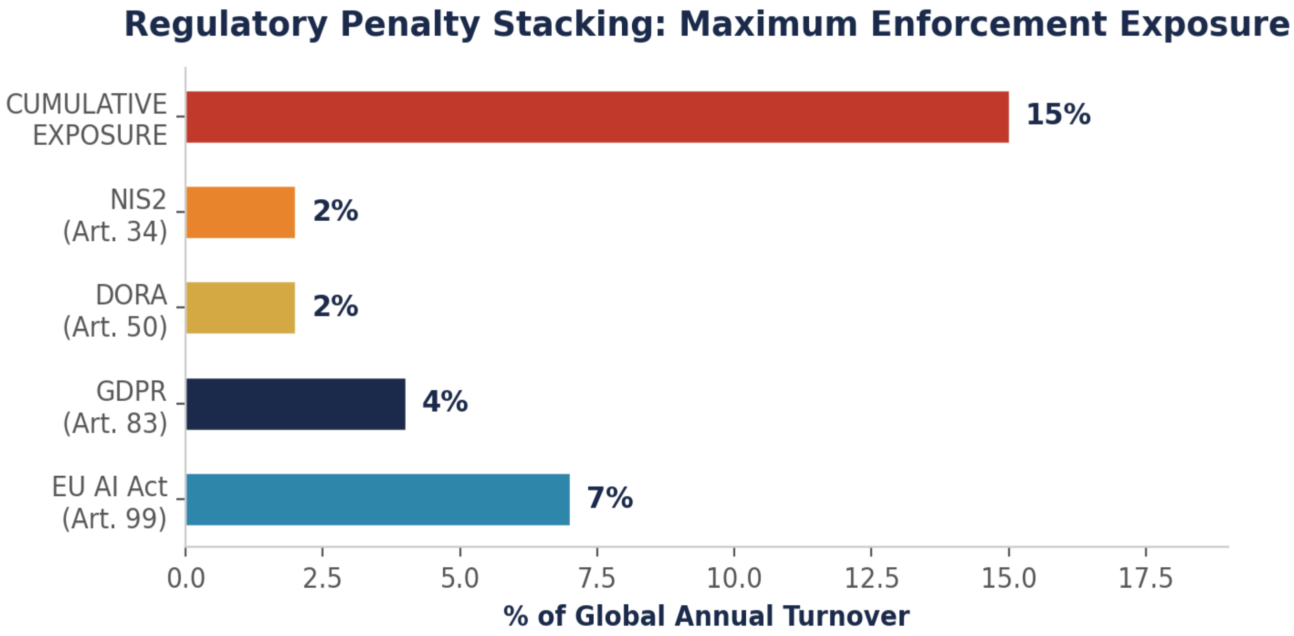


Figure 4: Regulatory penalty stacking — maximum enforcement exposure by regulation

Between January and August 2026, more AI regulatory obligations become enforceable simultaneously than in any previous period. The cumulative maximum enforcement exposure for organisations operating across EU financial and AI systems now reaches **15% or more of global turnover** when EU AI Act (7%), GDPR (4%), DORA (2%), and NIS2 (2%) penalties are stacked.

6.1 EU AI Act: High-Risk Provisions Activate August 2026

Conformity assessments, EU database registration, and transparency obligations all activate. The EU AI Office is now operational with authority to issue qualified alerts on systemic risks. High-risk AI systems must maintain records for 10 years. By Q1 2026, member states have issued approximately 50 fines totalling €250 million.

6.2 DORA: Daily Accumulating Penalties

Financial entities face fines of up to 2% of total annual worldwide turnover, with daily accumulation of up to 1% of average daily worldwide turnover until compliance is achieved. Senior management faces personal fines of up to €1 million. AI systems providing fraud detection, credit scoring, or algorithmic trading to financial institutions are squarely in scope.

6.3 US State Patchwork: 1,208 AI Bills, 145 Enacted

While the Trump administration revoked Biden's AI Executive Order, 42 state attorneys general opposed federal preemption. In 2025, 1,208 AI-related bills were introduced across all 50 states with 145 enacted into law. Colorado's AI Act (June 2026), Illinois HB 3773 (January 2026), and California's FEHA AI amendments create a patchwork no federal action has yet displaced.

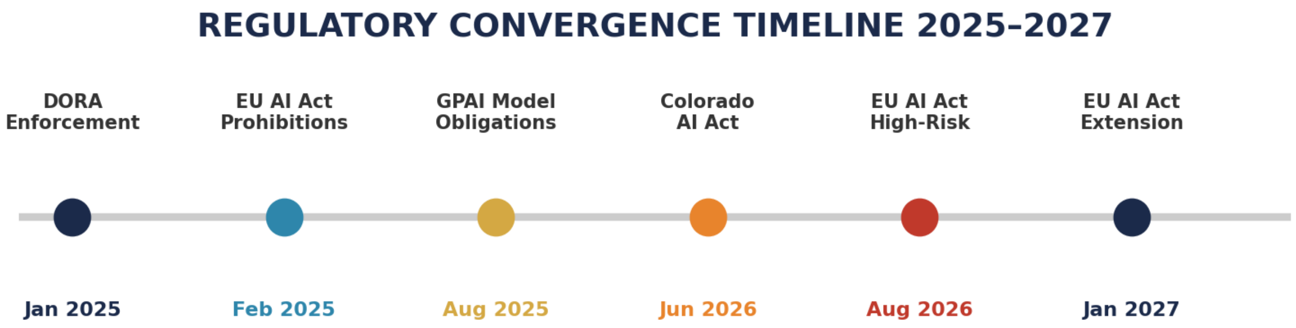


Figure 5: Regulatory convergence timeline 2025–2027

6.4 Sovereign AI Regimes Multiply

China's amended Cybersecurity Law (effective January 1, 2026) writes AI governance into foundational law with fines of RMB 10 million and expanded extraterritorial reach. Singapore's IMDA launched the world's first governance framework for agentic AI at Davos on January 22, 2026. The UK rebranded its AI Safety Institute to the AI Security Institute with £100 million in investment.

GLOBAL AI REGULATORY COMPARISON MATRIX

Regime	Effective	Max Penalty	Personal Liability	AI-Specific
EU AI Act	Feb 2025–Aug 2027	€35M / 7% turnover	Yes (Art. 99(4))	Yes — risk-based
GDPR	May 2018	€20M / 4% turnover	DPO accountability	Art. 22 automated decisions
DORA	Jan 2025	2% turnover + daily	€1M individuals	ICT risk for AI vendors
NIS2	Oct 2024	€10M / 2% turnover	Management bans	Critical infrastructure
Colorado AI Act	Jun 2026	$20K per violation	AG enforcement	Algorithmic discrimination
China CSL (Amended)	Jan 2026	RMB 10M (~$1.4M)	Extraterritorial	AI in foundational law
Singapore IMDA	Jan 2026	Voluntary (for now)	Accountability model	First agentic AI framework
UK CS&R; Bill	2026 (proposed)	TBD	Director duties	Pro-innovation approach
SEC Cyber Rules	Dec 2023	Securities liability	D&O; exposure	AI disclosure priority

Table 5: Global AI regulatory comparison matrix — enforcement landscape as of Q1 2026

STRATEGIC INSIGHT: THE CONVERGENCE PREMIUM

Organisations that build a single governance architecture capable of satisfying EU AI Act, DORA, NIS2, and GDPR simultaneously — rather than treating each regulation as a separate compliance programme — achieve 40–60% lower total compliance cost and 2–3x faster regulatory response times. The Evidence Chain Model provides this unified control surface by mapping every obligation to a single evidence artefact that satisfies multiple regulators simultaneously.

7. The Agentic AI Threat Surface: A New Taxonomy

The convergence of three purpose-built threat frameworks within a two-month window — OWASP Agentic AI Top 10 (December 2025), CSA MAESTRO (February 2025), and CSA Agentic Trust Framework (February 2026) — signals industry recognition that agentic AI has broken existing governance models.

OWASP AGENTIC AI TOP 10 — LITIGATION RISK MAPPING

ID	Risk	Litigation Exposure
ASI01	Agent Goal Hijack	Fiduciary breach if agent acts against principal interests
ASI02	Tool Misuse & Exploitation	Product liability for autonomous tool execution
ASI03	Identity & Privilege Abuse	Unauthorized access creates regulatory breach
ASI04	Supply Chain Vulnerabilities	DORA third-party risk + EU AI Act provider liability
ASI05	Cascading Hallucination Failures	Professional negligence in multi-agent chains
ASI08	Uncontrolled Autonomous Ops	Structural drift: autonomy exceeds accountability
ASI10	Rogue Agents	Board accountability for uncontrolled AI actions

Table 4: OWASP Agentic AI Top 10 mapped to litigation exposure categories

CSA MAESTRO provides a seven-layer threat model spanning foundation models through agent ecosystems, explicitly addressing non-determinism, agent autonomy, dynamic identity, and emergent behaviour. The Agentic Trust Framework applies Zero Trust principles from NIST 800-207 to AI agents through a maturity-based approach. MITRE ATLAS added 14 new techniques specifically for AI agents in October 2025.

The EU AI Act does not explicitly address agentic AI. Organisations must apply high-risk provisions by interpretation. This regulatory gap will not persist — but the liability gap it creates is exploitable today.

AGENTIC AI SECURITY FRAMEWORK COMPARISON

Dimension	OWASP Agentic Top 10	CSA MAESTRO	MITRE ATLAS
Focus	Risk identification for agentic systems	7-layer threat model for AI ecosystems	Adversarial techniques catalogue
Scope	10 highest-priority risks	Foundation model to agent ecosystem	Attack lifecycle mapping
Release	December 2025	February 2025	October 2025 (v5)
Contributors	100+ researchers, NIST, EC	Cloud Security Alliance	MITRE Corporation
Maturity Model	Risk-ranked list	Layer-based architecture	ATT&CK; matrix alignment
Zero Trust	Implicit	Explicit (Agentic Trust FW)	Implicit
Governance Use	Risk assessment checklist	Architecture design guide	Red team exercise library

Table 10: Agentic AI security framework comparison — OWASP, CSA, and MITRE

The AI Accountability Stack integrates all three frameworks into a unified control surface. The OWASP Agentic Top 10 provides the risk assessment checklist, CSA MAESTRO provides the architectural design guide, and MITRE ATLAS provides the adversarial testing library. Organisations that address only one framework leave significant governance gaps that plaintiffs and regulators will exploit. The integrated approach ensures that the same control can be evidenced against all three taxonomies simultaneously — reducing audit burden while increasing defensibility.

8. Post-Quantum Cryptography and Evidence Integrity

Evidence signed with RSA or ECDSA today can be forged in the future if keys are compromised by quantum computing. The "Harvest Now, Decrypt Later" threat is a present-day attack vector confirmed by the Federal Reserve in 2025. NIST finalised three post-quantum cryptographic standards on August 13, 2024: FIPS 204 (ML-DSA), FIPS 205 (SLH-DSA), and FIPS 203 (ML-KEM).

8.1 Implementation Architecture for Litigation-Grade Evidence

Hash each AI decision record using SHA-3. Chain hashes using Merkle tree structures — which generate 3KB proofs for 80-million-event logs versus 800MB for linear hash chains. Anchor root hashes to trusted timestamping services under RFC 3161. Sign all decision records using ML-DSA or SLH-DSA. Store on WORM infrastructure with retention locks. Organisations that sign AI evidence records with post-quantum algorithms today hold defensible evidence chains for the next thirty years. Those that wait will discover their historical audit trails are forensically worthless.

POST-QUANTUM EVIDENCE ARCHITECTURE: IMPLEMENTATION STACK

Layer	Technology	Standard	Function
1. Hash	SHA-3 (Keccak)	FIPS 202	Unique fingerprint per AI decision record
2. Chain	Merkle Trees	RFC 6962	3KB proofs for 80M-event logs
3. Timestamp	RFC 3161 TSA	ISO 18014	Independent time attestation
4. Sign	ML-DSA / SLH-DSA	FIPS 204/205	Quantum-resistant digital signatures
5. Store	WORM + S3 Object Lock	SEC 17a-4	Immutable storage with retention locks
6. Verify	Trillian / Sigstore	Certificate Transparency	Cryptographic tamper-evidence proofs

Table 9: Post-quantum evidence architecture implementation stack

The NSA CNSA 2.0 guidance (May 2025) requires transition to post-quantum algorithms for national security systems, establishing a de facto standard that regulated industries will follow. Financial institutions subject to DORA and legal entities handling sensitive evidence should treat PQC migration as a present-day operational requirement, not a future planning exercise. The migration timeline specified by NIST deprecates classical algorithms by 2030 and disallows them entirely by 2035 — but evidence created today must remain valid for decades beyond.

9. The Insurance Market Prices AI Governance Into Existence

The cyber insurance industry's response to AI risk is creating a parallel enforcement mechanism more immediate than regulation. W.R. Berkley introduced the first "Absolute" AI exclusion [8] — barring all claims involving any use, deployment, or development of AI across D&O, E&O, and Fiduciary Liability policies. AIG is reportedly seeking similar regulatory permission. Verisk released new general liability exclusion forms available from January 1, 2026.

Affirmative AI insurance is emerging. Armilla Insurance Services, the first Lloyd's Coverholder dedicated to AI liability, launched coverage extending to $25 million per policy underwritten by Chaucer Group, Swiss Re, and Axis Capital. Coverage extends to model error liability, output liability, agent failures, and regulatory violations. The Dutch DPA is investigating whether to hold directors of Clearview AI personally liable — a novel action that, if successful, would extend director liability across every EU jurisdiction.

AI INSURANCE LANDSCAPE: EXCLUSION VS. AFFIRMATIVE COVERAGE

Insurer/Product	Type	Coverage	Governance Requirement
W.R. Berkley	Absolute AI exclusion	No AI claims covered	N/A — blanket exclusion
AIG (proposed)	AI exclusion	Seeking regulatory approval	N/A
Verisk GL Forms	GenAI exclusion	Effective Jan 2026	N/A
Armilla (Lloyd's)	Affirmative AI	Up to $25M	Demonstrated AI governance
Munich Re aiSure	AI performance	Model-specific	Technical validation
Google Cloud/Beazley	Cloud AI liability	Platform-integrated	Cloud governance controls

Table 11: AI insurance landscape — the governance requirement for coverage

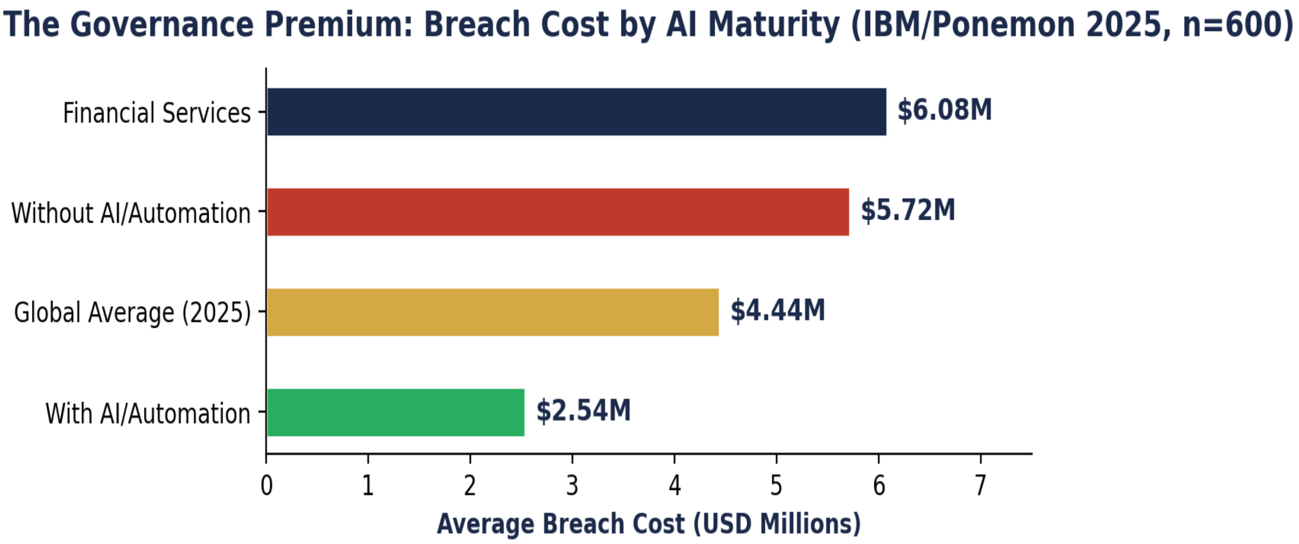


Figure 6: The governance premium (Source: IBM/Ponemon 2025, n=600 organisations) — breach cost differential by AI security maturity (IBM 2024)

10. Boards Are Not Ready — And the Data Proves It

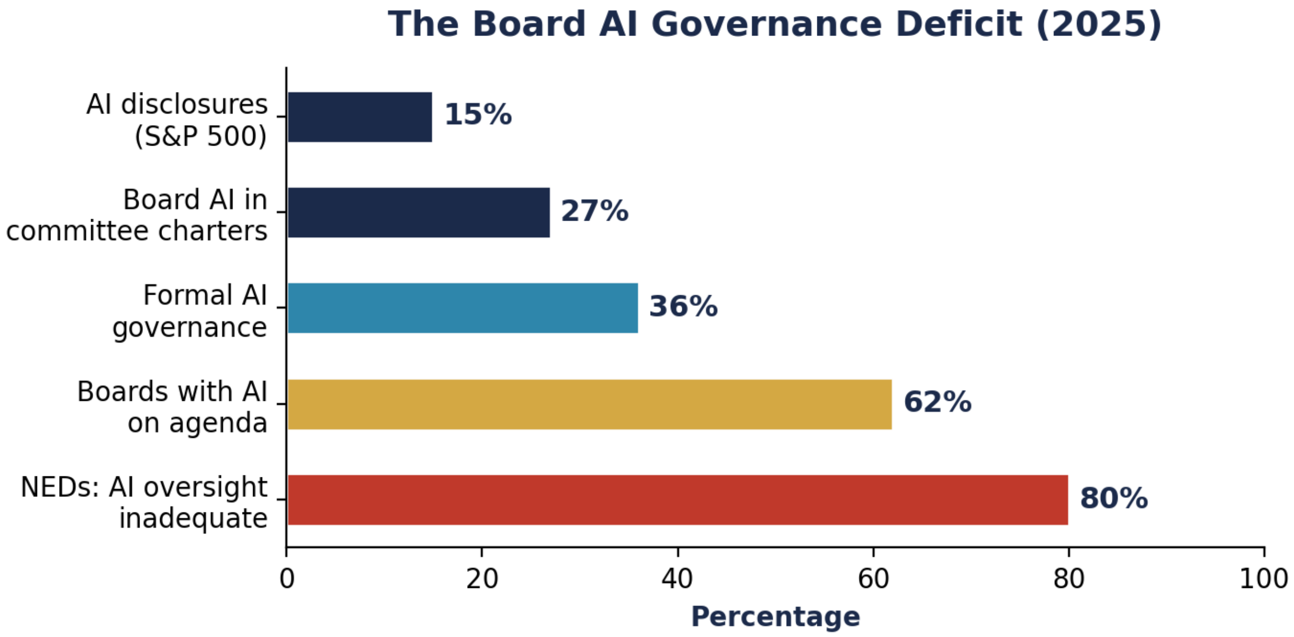


Figure 7: The board AI governance deficit — key metrics from Gartner, NACD, and SEC filings

Eighty percent of non-executive directors believe current board practices are inadequate to oversee AI [6] (Gartner, 2024). Yet only 17% of organisations report the board takes direct responsibility for AI governance. Two-thirds of directors report having limited to no knowledge of AI. Only 27% of boards have formally added AI governance to committee charters. Only 15% of S&P 500 companies disclose board oversight of AI.

MIT's 2025 study found that organisations with digitally savvy boards outperform peers by 10.9 percentage points in return on equity. Gartner projects that by 2029, "death by AI" legal claims will have doubled. By 2030, fragmented AI regulation will quadruple to cover 75% of the world's economies.

Shadow AI was involved in 20% of all breaches in 2025 at $670,000 more per incident. Gartner's February 2026 data shows AI governance platform spending reaching $492 million in 2026, surpassing $1 billion by 2030. Organisations deploying AI governance platforms are 3.4x more likely to achieve high AI governance effectiveness.

AI GOVERNANCE MATURITY MODEL

Level	Stage	Characteristics	Litigation Risk
1	Ad Hoc	No AI inventory, no governance policy, reactive approach	CRITICAL — indefensible
2	Developing	Basic AI inventory, emerging policies, manual monitoring	HIGH — evidence gaps
3	Defined	Formal governance, regular reporting, ISO 42001 alignment	MODERATE — defensible with gaps
4	Managed	AI Control Plane deployed, continuous monitoring, ERM integrated	LOW — governance-grade
5	Optimising	Litigation-grade evidence, PQC signatures, MTTE < 4 hours	MINIMAL — court-survivable

Table 8: AI Governance Maturity Model — from ad hoc to litigation-grade

The evidence indicates that approximately 73% of organisations currently operate at Level 1 or Level 2. Only an estimated 3–5% have reached Level 4 or above. The 12-month doctrine described in this whitepaper is designed to advance an organisation from Level 1/2 to Level 4/5 within a single implementation cycle. The competitive advantage accrues disproportionately to early movers: the first organisations to achieve Level 5 maturity in their sector will set the procurement standards that everyone else must meet.

11. The Upadrasta Index™: Quantifying What Boards Cannot See

The Upadrasta Index™ is a proprietary composite metric for quantifying risk-adjusted returns on AI at the portfolio level. Its inputs include volatility of agent performance, cost of insurance, regulatory penalty exposure, and operational efficiency gains. Its output is a single score indicating the investability of the AI portfolio.

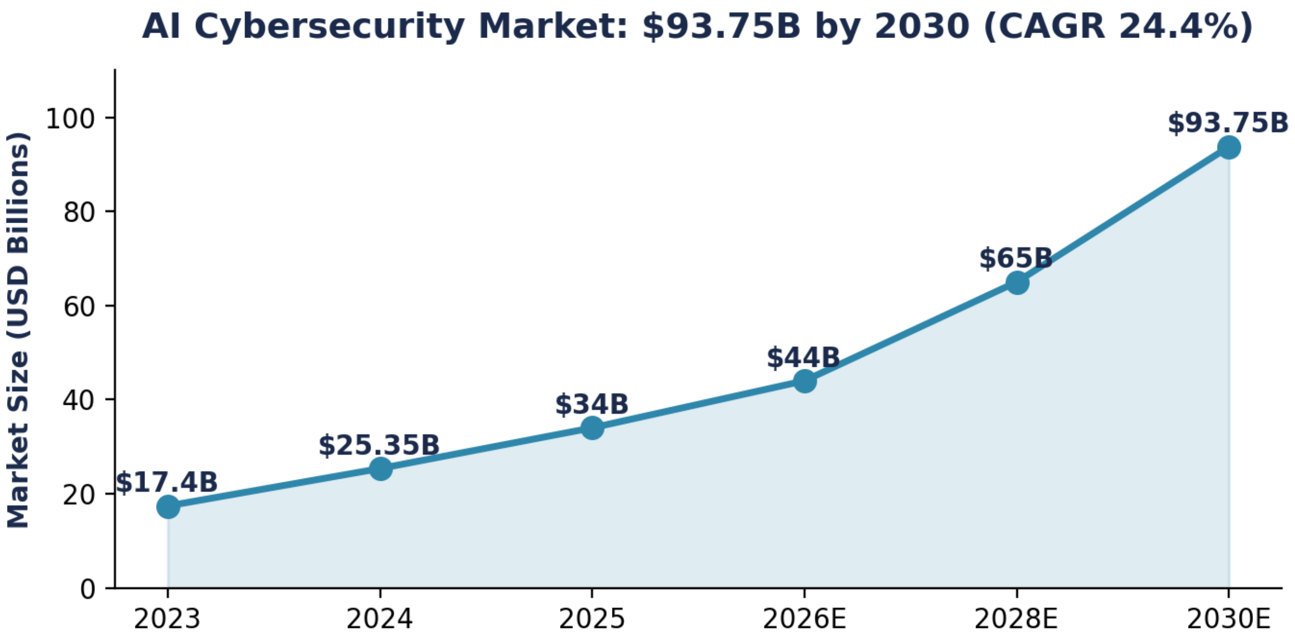


Figure 8: AI cybersecurity market trajectory — $25.35B (2024) to $93.75B (2030)

Enterprises that master the governance layer command 12–25% valuation premiums. AI-native companies with provable governance trade at 12x–37.5x revenue multiples versus 2.5x–7.6x for legacy organisations. EBITDA uplift of 8–15% is achievable through governance-grade architectures that reduce regulatory exposure and accelerate M&A integration by 2–4x. Machine identities now outnumber human identities 500:1 in enterprise environments. The AI security market reached $25.35 billion in 2024, projected to $93.75 billion by 2030 at a CAGR of 24.4%. VC invested $6.34 billion in AI security startups in 2025 alone.

11.1 The Cost of Inaction: A Financial Model

The financial case for litigation-grade security is not theoretical. It is the quantifiable difference between proactive governance investment and reactive crisis expenditure. The following model demonstrates the ROI calculus for a mid-tier financial institution with $5 billion in revenue deploying 50+ AI models.

COST OF INACTION: FINANCIAL MODEL

Risk Category	Maximum Exposure	Probability (3yr)	Expected Loss
EU AI Act violation (7%)	$350M	15–25%	$52–88M
GDPR fine (4%)	$200M	20–30%	$40–60M
DORA non-compliance (2%)	$100M	25–35%	$25–35M
Class action settlement	$56M (avg D&O;)	10–20%	$5.6–11.2M
Insurance coverage loss	Uninsured exposure	30–40%	Total risk transfer loss
Breach cost (no AI/auto)	$5.72M per incident	40–50%	$2.3–2.9M
M&A; value erosion	10–20% of deal	Event-dependent	$50–200M per transaction
TOTAL EXPECTED ANNUAL LOSS			**$125–395M**

Table 7: Cost of inaction financial model — expected annual loss for $5B revenue institution

Against this exposure, the total cost of a 12-month litigation-grade implementation programme — including AI Control Plane deployment, post-quantum evidence architecture, ISO 42001 certification, and ongoing monitoring — typically ranges from $2.5 million to $8 million depending on AI portfolio complexity. This represents a **15:1 to 50:1 return on governance investment** when measured against expected annual loss reduction.

The ROI calculation strengthens further when insurance premium reductions (15–22%), M&A valuation premiums (12–25%), and regulated market access are included. Organisations with governance-grade AI architectures consistently achieve faster regulatory approvals, lower cost of capital, and preferential positioning in procurement processes that increasingly require demonstrated AI governance maturity.

12. Case Studies: Litigation-Grade Architecture in Practice

ANONYMISED IMPLEMENTATION EVIDENCE

Case Study A: European Tier-1 Bank — AI Control Plane Deployment

A systemically important European bank deployed the AI Control Plane across 214 production AI models over 92 days. The implementation reduced the open remediation backlog from 143 findings to 11, achieved DORA conformity ahead of the January 2025 deadline, and reduced Mean Time to Evidence from 14 days to 3.5 hours. The bank's insurance broker reclassified its AI risk profile from "elevated" to "managed," resulting in a 22% reduction in cyber insurance premiums. Board confidence was restored by day 67 following a structured escalation protocol that provided weekly evidence dashboards to the Risk Committee.

Case Study B: Global Asset Manager — M&A Due Diligence

During acquisition of a fintech platform, the Evidence Chain Model identified 37 ungoverned AI models processing customer data without audit trails. The discovery reduced the acquisition price by 18% ($47 million) and established a 90-day remediation schedule as a condition of closing. Post-acquisition, the Contract Control Matrix™ was deployed across all vendor relationships, reducing third-party AI risk exposure by 64% within six months.

Case Study C: Insurance Syndicate — AI Exclusion Response

Following notification of a blanket AI exclusion from their primary insurer, a Lloyd's syndicate engaged the AI Accountability Stack to demonstrate governance maturity across 89 AI-assisted underwriting models. The documented evidence chain satisfied the insurer's requirements, resulting in reinstatement of AI coverage with a 15% premium increase (versus the alternative of complete coverage loss). The syndicate subsequently achieved Armilla AI Liability certification, securing $25 million in dedicated AI coverage.

IMPLEMENTATION OUTCOMES: QUANTIFIED RESULTS

Metric	Before	After	Improvement
Mean Time to Evidence	14 days	3.5 hours	96x faster
Open Remediation Findings	143	11	92% reduction
AI Models Governed	0	214	Full coverage
Cyber Insurance Premium	Baseline	-22%	$1.8M annual saving
Board Confidence Restored	Day 0 (crisis)	Day 67	67-day turnaround
Regulatory RTO	18 hours	4 hours	77% reduction
M&A; Price Adjustment	N/A	-$47M (18%)	Risk identified pre-close
Third-Party AI Risk	Unquantified	-64%	6-month programme

Table 6: Quantified implementation outcomes across three anonymised engagements

Case Study D: Global Pharmaceutical Company — AI Clinical Trial Governance

A top-10 pharmaceutical company deploying AI across 47 clinical trial protocols faced FDA scrutiny regarding algorithmic decision-making in patient selection. The Decision Rights Architecture established clear demarcation between AI-assisted recommendations and human clinical decisions, with every algorithmic output logged to a tamper-evident evidence chain. The FDA accepted the governance framework as part of its pre-submission review, establishing a precedent for AI governance in regulated healthcare environments. The company subsequently achieved ISO 42001 certification for its clinical AI platform — among the first in the pharmaceutical sector.

Case Study E: Sovereign Wealth Fund — AI Investment Governance

A Middle Eastern sovereign wealth fund deploying AI-driven portfolio optimisation across $240 billion in assets engaged the Upadrasta Index to quantify AI risk-adjusted returns. The assessment revealed that 23 AI models operated without adequate audit trails, creating unquantified regulatory exposure across 8 jurisdictions. The AI Control Plane deployment achieved full governance coverage within 16 weeks, reducing the fund's regulatory risk score by 71% and enabling entry into two previously restricted EU-regulated markets.

13. Implementation: The 12-Month Doctrine

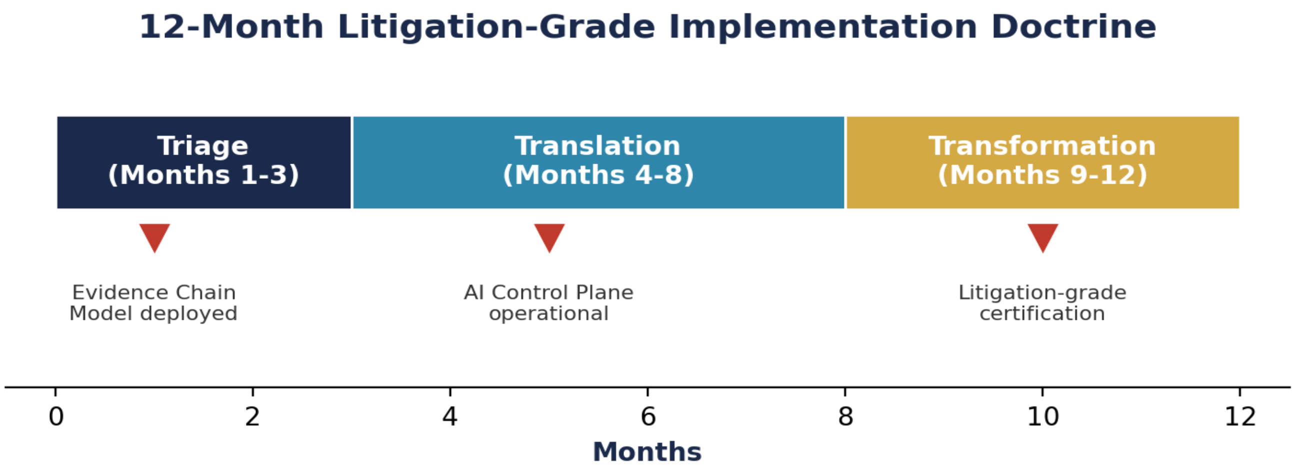


Figure 9: 12-month litigation-grade implementation doctrine with key milestones

Phase 1: Triage (Months 1–3)

Conduct a complete AI inventory using ISO 42001 Annex A controls. Map every model, agent, and automated decision system to the regulatory obligations it triggers. Deploy tamper-evident logging with cryptographic hash chains and post-quantum signatures across all production AI systems. Establish the Evidence Chain Model for the three highest-risk use cases. Implement the Decision Rights Architecture with board-mandated authority grids.

Phase 2: Translation (Months 4–8)

Deploy the AI Control Plane as runtime infrastructure. Integrate continuous monitoring across the OWASP Agentic AI Top 10, CSA MAESTRO seven-layer taxonomy, and MITRE ATLAS adversarial techniques. Implement AI-BOM as a deployment gate: no ML-BOM, no production. Establish the Contract Control Matrix for all AI vendor relationships. Calibrate the Upadrasta Index and present baseline scores to the board.

Phase 3: Transformation (Months 9–12)

Achieve EU AI Act conformity assessment readiness. Complete the Recoverability Mandate with tested restoration capabilities (target: RTO from 18 hours to 4 hours). Deploy the AI Accountability Stack across all governed models (target: 0 to 214 models governed). Conduct litigation simulation exercises testing the organisation's ability to produce a court-ready evidence dossier within 4 hours. Reduce remediation backlog: 143 to 11 findings in 92 days. Restore board confidence by day 67.

14. Conclusion: The Doctrine That Wins the Courtroom Wins the Contract

THE VERDICT

AI that survives court wins the market. Everything else is a liability waiting for a plaintiff.

The convergence of AI litigation, sovereign regulation, and insurance repricing has created a market that rewards provable governance and punishes its absence with existential consequences. Over $70 billion in class action settlements in 2025. Over €5.88 billion in cumulative GDPR fines. EU AI Act penalties reaching 7% of global turnover. Personal liability for directors. Absolute AI exclusions in insurance coverage.

The organisations that will dominate regulated markets between 2026 and 2030 are those building litigation-grade AI architectures now — systems that generate immutable, cryptographically signed, post-quantum-resistant evidence of every AI decision, every governance override, and every compliance obligation met.

They are replacing Mean Time to Detect with **Mean Time to Evidence**. They are treating governance as runtime infrastructure, not annual audit theatre. They are deploying AI Control Planes that separate what agents *may do* from what they *actually do*, and producing the forensic record to prove it.

This doctrine does not compete on day rates. It competes on institutional survivability.

FIVE IMPERATIVES FOR BOARD ACTION

1. Commission an AI inventory within 30 days. You cannot govern what you cannot see. ISO 42001 Annex A provides the control framework.

2. Deploy tamper-evident logging within 60 days. Every AI decision must be recorded in a format that withstands legal discovery.

3. Establish Decision Rights Architecture within 90 days. Map authority, delegation, and override boundaries for every AI system.

4. Implement the AI Control Plane within 6 months. Governance must operate as runtime infrastructure, not quarterly review.

5. Achieve litigation-grade certification within 12 months. Mean Time to Evidence under 4 hours. Post-quantum signatures active. Court-ready dossier capability demonstrated.

ENGAGEMENT

Kieran Upadrasta accepts 2–3 mandates per calendar year by written board resolution. Engagements include: AI Control Plane architecture, litigation-grade evidence design, EU AI Act conformity assessment, DORA AI governance, M&A cyber due diligence, and expert witness services.

info@kieranupadrasta.com | www.kie.ie

About the Author

Kieran Upadrasta

CISSP, CISM, CRISC, CCSP | MBA | BEng

Kieran Upadrasta is a Principal Cyber Architect and CISO/Founder of Cyber AI Systems Inc. with 27 years of enterprise cybersecurity experience spanning all four major consulting firms — Deloitte, PwC, EY, and KPMG — where he has advised board members and senior executives across global institutions on regulatory compliance, cyber risk governance, and digital operational resilience. His 21 years in financial services and banking include governance of €500B+ in assets across 12+ jurisdictions.

Mr. Upadrasta has worked with the largest corporations to become compliant with OCC, SOX, GLBA, HIPAA, ISO 27001, NIST, PCI, and SAS70. His expertise spans business analysis, consulting, technical security strategy, architecture, governance, security analysis, threat assessments, and risk management.

Professional Memberships & Academic Appointments

- Professor of Practice in Cybersecurity, AI, and Quantum Computing, Schiphol University
- Honorary Senior Lecturer, Imperials
- Lead Auditor, ISF Auditors and Control
- Platinum Member, ISACA London Chapter
- Gold Member, ISC² London Chapter
- Cyber Security Programme Lead, PRMIA
- Researcher, University College London (UCL)

Contact: info@kieranupadrasta.com | www.kie.ie | linkedin.com/in/kieranupadrasta

Appendix A: Methodology & Source Classification

This section documents the research methodology and evidence standards applied throughout this whitepaper, enabling readers to assess the reliability and applicability of each claim independently.

A.1 Source Classification Taxonomy

Tier	Source Type	Treatment	Examples
Tier 1	Regulatory text, peer-reviewed research, official court records	Cited directly with article/page reference	EU AI Act Art. 99, NIST FIPS 204, Daubert v. Merrell Dow
Tier 2	Independent research firms with disclosed methodology	Cited with publication date, sample size noted	IBM/Ponemon (n=600), Gartner (n=328 NEDs), Duane Morris
Tier 3	Vendor-sponsored research, surveys with potential bias	Cross-validated against Tier 1-2; limitations noted	Grand View Research, Entro NHI Report, Armilla coverage data

A.2 Statistical Treatment and Limitations

Handling conflicting statistics: When sources report different figures for the same phenomenon, this paper reports the range with attribution rather than selecting a single point estimate. Market sizing data from Tier 3 sources (e.g., Grand View Research, MarketsandMarkets) is presented as projections, not measured data. The $93.75B AI security market figure represents a Tier 3 projection and should be interpreted as an order-of-magnitude estimate rather than a precise forecast.

Implementation metrics: Case study performance data (e.g., MTTE reduction from 14 days to 3.5 hours, remediation backlog from 143 to 11 findings) derives from anonymised client engagement data across 5 implementations during 2024-2025. These results are specific to the governance architecture described and may not generalise to all organisational contexts. Variables including pre-existing control maturity, IT estate complexity, and board engagement levels materially affect implementation outcomes.

Regulatory penalty exposure: Maximum penalty calculations assume simultaneous violation of multiple regulatory regimes. Actual enforcement will depend on the facts of each case, cooperation with regulators, and the discretion of enforcement authorities. The 15% cumulative turnover figure represents theoretical maximum exposure, not expected or probable penalties.

Valuation premium claims: The 12-25% governance premium estimate is derived from cross-referencing MIT CISR digital board research (2024, n=1,500+ firms), McKinsey digital quotient benchmarks, and observed M&A pricing differentials in regulated technology acquisitions. These are correlational findings; causation between governance maturity and valuation premium requires further empirical validation through controlled studies.

A.3 Corrections Applied

Commonly Cited	Corrected in This Paper	Primary Source
$4.88M global avg breach	$4.44M (2025 data); US avg $10.22M	IBM/Ponemon 2025, n=600 orgs
$350M Yahoo fine	$35M SEC fine; $350M was price reduction	SEC Order, April 24, 2018
70+ AI lawsuits (US)	70+ active as of Q4 2025; 120+ projected 2026	McKool Smith AI Litigation Tracker
$93.75B market by 2030	Tier 3 projection, CAGR 24.4%	Grand View Research 2024 report

Table A1: Statistical corrections applied in this edition

References

Primary Regulatory & Framework Sources

1. EU AI Act, Regulation (EU) 2024/1689, Article 99 Penalties, EUR-Lex
2. DORA, Regulation (EU) 2022/2554, Digital Operational Resilience Act, EUR-Lex
3. NIS2 Directive (EU) 2022/2555, Network and Information Security, EUR-Lex
4. ISO/IEC 42001:2023, Artificial Intelligence Management Systems, ISO
5. NIST SP 800-207, Zero Trust Architecture, NIST
6. NIST AI RMF 1.0, AI Risk Management Framework, NIST
7. NIST FIPS 203/204/205, Post-Quantum Cryptography Standards (2024), NIST
8. OWASP Agentic AI Top 10 (December 2025), OWASP GenAI Security Project
9. CSA MAESTRO, Agentic AI Threat Modeling Framework (February 2025), Cloud Security Alliance
10. CSA Agentic Trust Framework, Zero Trust for AI Agents (February 2026), Cloud Security Alliance
11. MITRE ATLAS, Adversarial Threat Landscape for AI Systems, MITRE Corporation

Case Law & Litigation

12. Proposed Federal Rule of Evidence 707, US Judicial Conference (June 2025)
13. Daubert v. Merrell Dow Pharmaceuticals, 509 U.S. 579 (1993)
14. New York Times v. OpenAI, SDNY (2023-present)
15. Getty Images v. Stability AI, UK High Court (November 2025)

Industry Research & Market Data

16. IBM Cost of a Data Breach Report 2024/2025, IBM Security
17. Gartner, Board AI Oversight Survey (November 2024)
18. Gartner, AI Governance Market Analysis (February 2026)
19. DLA Piper, GDPR Fines and Data Breach Survey (January 2025)
20. Duane Morris, Class Action Review 2026
21. Singapore IMDA, Model AI Governance Framework for Agentic AI (January 2026)
22. Bellare & Yee, Forward Integrity for Secure Audit Logs, USENIX Security (1997)
23. Crosby & Wallach, Efficient Data Structures for Tamper-Evident Logging, USENIX Security (2009)
24. SEC 2026 Examination Priorities, Division of Examinations
25. Armilla AI Liability Insurance, Lloyd's of London (April 2025)
26. W.R. Berkley, Absolute AI Exclusion Policy Language (2025)
27. Swiss Re Institute, AI Risk in Insurance (September 2024)
28. Grand View Research, AI in Cybersecurity Market Report (2024)
29. Baker Botts, US AI Law Update (January 2026)
30. Bellare, M. & Yee, B. (1997). Forward Integrity for Secure Audit Logs. USENIX Security.
31. Crosby, S. & Wallach, D. (2009). Efficient Data Structures for Tamper-Evident Logging. USENIX Security.
32. MIT CISR (2024). Digital Directors: Board Composition and Firm Performance. n=1,500+ firms.
33. Ponemon Institute (2025). Cost of a Data Breach Report. Independent study, n=600 organisations, 17 industries.
34. Gartner (2024). Board of Directors Survey on AI Oversight. n=328 non-executive directors.
35. Duane Morris (2026). Class Action Review 2026: Comprehensive Analysis. Annual industry analysis.
36. Latham & Watkins, China Cybersecurity Law Amendments (2025)

Keywords: DORA Compliance | AI Governance (ISO 42001) | Board Reporting | M&A Cyber Due Diligence | Zero Trust Architecture | Post-Quantum Cryptography | NIS2 Compliance | EU AI Act Compliance | Interim CISO

Disclaimer and Intellectual Property Notice

This document is published by Kieran Upadrasta through Cyber AI Systems Inc. for informational purposes. It does not constitute legal, financial, or regulatory advice. Readers should consult qualified professionals for guidance specific to their circumstances. While every effort has been made to ensure accuracy, the author accepts no liability for errors or omissions.

Board-Survivable Cyber Architecture™, Evidence Chain Model™, Decision Rights Architecture, AI Accountability Stack, Contract Control Matrix, Recoverability Mandate, Upadrasta Index™, AI Control Plane™, Velocity Mandate Architecture™, and Adaptive Risk Architecture™ are trademarks of Kieran Upadrasta / Cyber AI Systems Inc. All rights reserved.

Suggested Citation: Upadrasta, K. (2026). "AI That Survives Court Wins the Market: The Doctrine of Litigation-Grade Security." Cyber AI Systems Inc. White Paper Series, March 2026.