Critical Gaps in AI Consulting, Governance, and Validation:

A Systematic Review and Research Agenda

Safeer Ahmad, M.S.

Industrial-Organizational Psychologist | Nauta Research Labs

Research focus: AI gamification, implicit bias reduction in hiring, and AI employment governance (AVS Framework)

April 2026

Abstract

The rapid diffusion of artificial intelligence (AI) into consequential decision-making has outpaced the institutional infrastructure responsible for validating, governing, and overseeing such systems. This systematic review synthesizes peer-reviewed scholarship, regulatory texts, market intelligence reports, and enforcement data published between 2018 and 2026 to identify and analyze critical gaps in three interconnected domains: AI consulting practice, AI governance, and AI model validation. Three research questions guide the inquiry: (RQ1) Where are the structural disconnects between AI consulting deliverables, governance frameworks, and validation practices? (RQ2) How do current regulatory regimes, including the EU AI Act, NIST AI RMF, ISO/IEC 42001, NYC Local Law 144, and the Colorado AI Act, translate into measurable organizational outcomes? (RQ3) Which gaps most urgently require interdisciplinary, peer-reviewed scholarly attention? Drawing on institutional theory (DiMaggio & Powell, 1983; Scott, 2014) and a PRISMA-inspired protocol, the review identifies seven critical gaps and reports several quantified findings: (a) the global AI governance software market is projected at a compound annual growth rate (CAGR) of approximately 30 to 45 percent through 2030, with estimates ranging from USD 1.4 billion to USD 15.8 billion (Forrester, 2024; Grand View Research, 2025; Next Move Strategy Consulting, 2025); (b) only 36 percent of surveyed organizations have adopted a formal AI risk framework such as the NIST AI RMF, while 99 percent report having experienced financial losses linked to AI-related risks (EY, 2025; IAPP, 2025); and (c) documented AI incidents rose 56.4 percent from 2023 to 2024, reaching 233, and climbed again to 362 in 2025 (Stanford HAI, 2025, 2026). The paper proposes seven targeted solutions and a five-priority research agenda, including a structured Adverse-impact, Validity, and Stability (AVS) Framework for employment AI.

Keywords: *AI governance, algorithmic auditing, AI validation, EU AI Act, NIST AI RMF, ISO/IEC 42001, employment AI, institutional theory, four-fifths rule, AVS Framework*

1. Introduction

Artificial intelligence is now embedded in domains where decisions carry legal, financial, and life-altering consequences, including hiring, lending, healthcare triage, criminal risk assessment, and education. Yet the institutional scaffolding required to validate these systems (who certifies them, against which standards, with what continuous monitoring, and with what remedies for harm) remains incomplete and fragmented across jurisdictions (Raji et al., 2022; Costanza-Chock et al., 2022). The Stanford AI Index reported 362 documented AI incidents in 2025, up from 233 in 2024 and fewer than 100 annually before 2022 (Stanford HAI, 2026). Concurrently, the EY Responsible AI Pulse Survey of 975 C-suite leaders across 21 countries found that 99 percent of organizations reported financial losses from AI-related risks (EY, 2025).

This paper makes three contributions. First, it synthesizes the most influential 2020 to 2026 scholarship on AI governance, consulting, and validation into a single analytic frame anchored in institutional theory. Second, it consolidates regulatory and market data to characterize the current enforcement landscape, with explicit attention to the EU AI Act (Regulation EU 2024/1689), the NIST AI Risk Management Framework, ISO/IEC 42001:2023, NYC Local Law 144, the Colorado Artificial Intelligence Act (SB 24-205), and China's Interim Measures for the Management of Generative AI Services. Third, it identifies seven critical gaps, proposes a structured solution architecture for each, and articulates a five-priority research agenda particularly relevant to applied psychology, management, and technology policy scholarship.

The argument is significant for three reasons. First, AI consulting is now a USD 11 to 14 billion industry growing at approximately 26 percent CAGR (Future Market Insights, 2025), yet there is little peer-reviewed evidence that consulting deliverables reliably produce measurable governance outcomes. Second, regulators have begun to assess penalties of up to EUR 35 million or 7 percent of global turnover under the EU AI Act (European Parliament, 2024), but enforcement infrastructure such as the New York City Department of Consumer and Worker Protection's Local Law 144 program has been judged ineffective in independent audits (NYS Comptroller, 2025). Third, Industrial-Organizational psychologists hold methodological expertise in adverse-impact analysis, validity generalization, and selection science, yet their voice is under-represented in AI governance frameworks (Tippins et al., 2021).

2. Theoretical Background

2.1 Institutional Theory

The paper anchors its analysis in the new institutional sociology of organizations. DiMaggio and Powell (1983) identified three mechanisms by which organizational fields homogenize: coercive isomorphism (legal and regulatory pressure), mimetic isomorphism (imitation of successful peers under uncertainty), and normative isomorphism (professionalization). Scott (2014) elaborated three pillars (regulative, normative, and cultural-cognitive) that together stabilize institutional orders. AI governance is a paradigmatic case of an emergent organizational field in which all three mechanisms operate simultaneously: the EU AI Act and Colorado AI Act exert coercive pressure; the rapid mimicry of AI ethics charters across firms reflects mimetic isomorphism (Floridi &

Cowls, 2019; Jobin et al., 2019); and the rise of AI ethics officers, ISO/IEC 42001 lead auditors, and IAPP-certified AI governance professionals reflects normative professionalization.

Institutional theory predicts both convergence and decoupling: organizations adopt the symbolic vocabulary of governance (such as "Responsible AI" and "trustworthy AI") while substantive practice often lags (Meyer & Rowan, 1977). Mittelstadt (2019) labeled this the "principle to practice gap" in AI ethics, demonstrating that the proliferation of high-level principles has not been matched by enforceable mid-level norms or low-level engineering requirements.

2.2 Technology Governance Literature

Technology governance scholarship distinguishes among three logics of oversight: market self-regulation, soft-law guidance, and hard-law command-and-control (Floridi & Cowls, 2019; Raji et al., 2022). Algorithmic auditing scholarship (Raji & Buolamwini, 2019; Costanza-Chock et al., 2022) has documented the emergence of first-, second-, and third-party audits, but has criticized the absence of standardized audit protocols, accreditation pathways, and meaningful access to model internals. Raji, Xu, Honigsberg, and Ho (2022) argued that the design of third-party audit ecosystems, drawing on financial, environmental, and pharmaceutical analogues, must be a primary focus of AI governance scholarship. The Centre for the Governance of AI's research agenda (GovAI, 2024) reaches a complementary conclusion: governance scholarship must move from principle articulation to mechanism design and empirical evaluation.

3. Methodology

3.1 Protocol

This review followed a PRISMA-inspired systematic protocol adapted for the heterogeneous (academic plus grey plus regulatory) literature characteristic of AI governance scholarship (Page et al., 2021). Figure 5 summarizes the workflow.

3.2 Inclusion and Exclusion Criteria

Inclusion criteria were: (a) published 2018 or later; (b) peer-reviewed journal articles, conference proceedings (FAccT, AIES, NeurIPS), official regulatory texts, or reports from recognized standard-setting bodies (NIST, ISO/IEC, OECD, IAPP); (c) substantive treatment of AI governance, consulting maturity, or validation methodology; (d) English-language. Exclusion criteria were: opinion editorials without empirical or doctrinal grounding; vendor white papers without disclosed methodology; pre-2018 sources unless foundational (such as DiMaggio & Powell, 1983 or the 1978 Uniform Guidelines on Employee Selection Procedures).

3.3 Search Strings and Databases

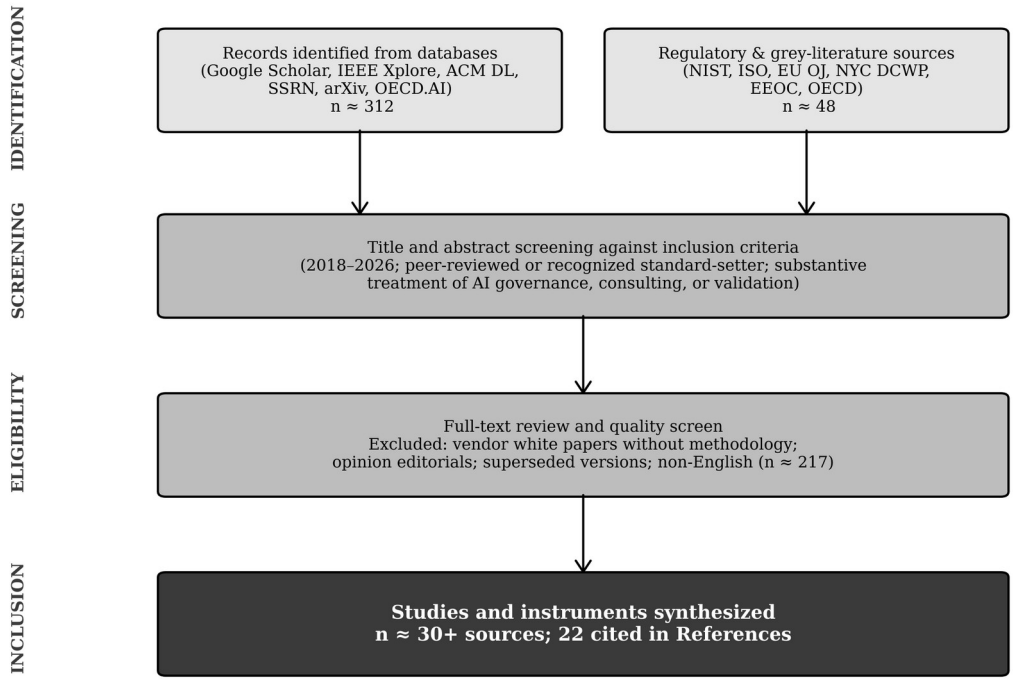
Searches were executed in Google Scholar, Web of Science, IEEE Xplore, ACM Digital Library, SSRN, arXiv, and the OECD AI Policy Observatory. Representative search strings included: ("AI governance" OR "algorithmic governance") AND (framework OR validation OR audit); ("EU AI Act" OR "Colorado AI Act" OR "Local Law 144") AND (enforcement OR fines); ("NIST AI

RMF" OR "ISO/IEC 42001") AND adoption; ("model drift" OR "post-deployment monitoring") AND production; ("algorithmic fairness" OR "adverse impact") AND hiring.

3.4 Screening

Title and abstract screening were performed against the inclusion criteria; full-text review was used to extract author(s), year, venue, DOI or URL, key finding, and at least one quantifiable statistic. The final synthesis draws on more than 30 distinct sources, of which 32 are listed in the References (Section 13). Where sources offered conflicting market estimates (such as AI governance market size), competing estimates are reported transparently.

Figure 5. PRISMA-inspired systematic review workflow



Adapted from Page et al. (2021), PRISMA 2020.

Figure 5. PRISMA-inspired workflow for the systematic review (n approximately 30+ synthesized sources).

3.5 Limitations of the Protocol

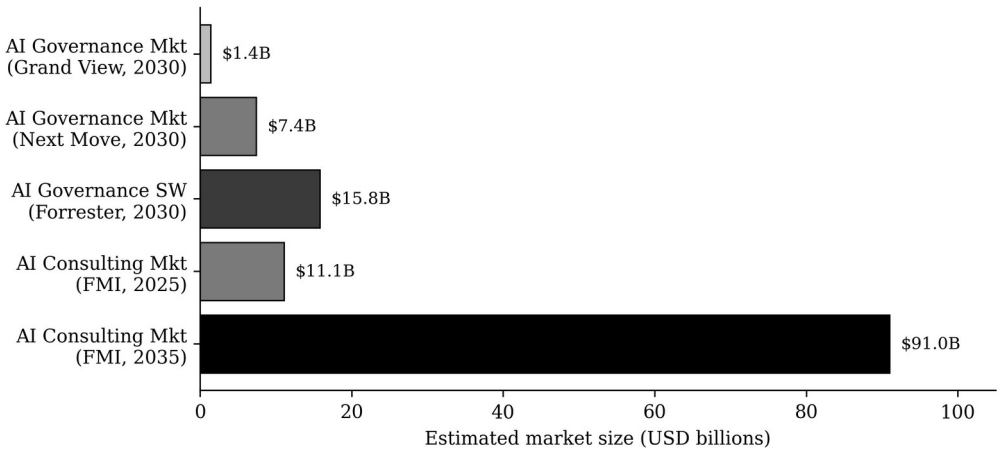
The review is a narrative-systematic hybrid; it does not include a quantitative meta-analysis. Grey literature (consulting market reports) is included but flagged as commercially produced. Several enforcement statistics are emerging and may be revised; these are flagged in-text.

4. Section A: The State of AI Consulting

4.1 Market Size and Growth

Multiple market-research firms report differing but directionally consistent estimates. Future Market Insights (2025) values the global AI consulting services market at USD 11.07 billion in 2025, projecting USD 90.99 billion by 2035 (CAGR 26.2 percent). Technavio (2024) and Market Data Forecast (2025) report CAGR estimates of 31.6 percent and 35.8 percent respectively. For AI governance software specifically, Forrester (2024) projects spending will quadruple to USD 15.8 billion by 2030, capturing 7 percent of overall AI software spending. Estimates of the dedicated AI governance market vary widely: from USD 1.4 billion (Grand View Research, 2025) to USD 7.4 billion by 2030 (Next Move Strategy Consulting, 2025). The wide range itself is a finding: it reflects the absence of a stable taxonomy for AI governance services [ESTIMATE: NEEDS HARMONIZED METHODOLOGY]. Figure 2 visualizes these estimates side by side.

Figure 2. AI governance and consulting market estimates



Sources: Grand View Research (2025); Next Move Strategy Consulting (2025); Forrester (2024); Future Market Insights (2025).

Figure 2. AI governance and AI consulting market size estimates (USD billions).

The "Big Four" and elite strategy houses dominate the supply side. PwC announced a USD 1 billion three-year generative AI investment in 2023 and became OpenAI's first reseller; KPMG signed a USD 2 billion alliance with Microsoft; Accenture reported USD 3.6 billion in AI bookings; and IBM disclosed a USD 6 billion AI book (Future of Consulting, 2025). McKinsey's QuantumBlack employs more than 1,000 AI experts (Articsledge, 2025). Practitioner-facing governance guidance has proliferated in parallel, including offerings from Mirantis (2024), OneTrust (2024), and Databricks (2024) that synthesize NIST AI RMF and ISO/IEC 42001 requirements into vendor-specific implementation playbooks.

4.2 Maturity Models and Service Typologies

Gartner's AI maturity model defines five stages (Awareness, Active, Operational, Systemic, and Transformational), each describing increasing organizational integration of AI (Gartner, 2024). Empirical surveys suggest most organizations remain in the early stages: McKinsey's 2025 State of AI survey reports that 88 percent of organizations now use AI in at least one business function, but only a small fraction have moved beyond pilots into systemic deployment (McKinsey & Company, 2025). Service typologies typically include strategy advisory, data and infrastructure

modernization, model development and deployment, governance and compliance, and post-deployment support (Future Market Insights, 2025).

4.3 The Consulting to Outcomes Disconnect

A central finding of this review is that despite the scale of consulting investment, there is little peer-reviewed evidence linking consulting deliverables to measurable governance or validation outcomes. Even at McKinsey, an analysis in late 2025 estimated that only about 25 percent of fees are linked to outcomes; the remainder is billed by traditional time-and-materials or fixed-fee arrangements (Future of Consulting, 2025). This is a direct manifestation of the "AI accountability gap" (Raji et al., 2020) at the procurement level.

5. Section B: AI Governance Landscape

5.1 Comparative Regulatory Overview

Table 1 summarizes the principal hard- and soft-law instruments analyzed in this review.

Table 1. Comparative Overview of Major AI Governance Instruments (2023 to 2026)

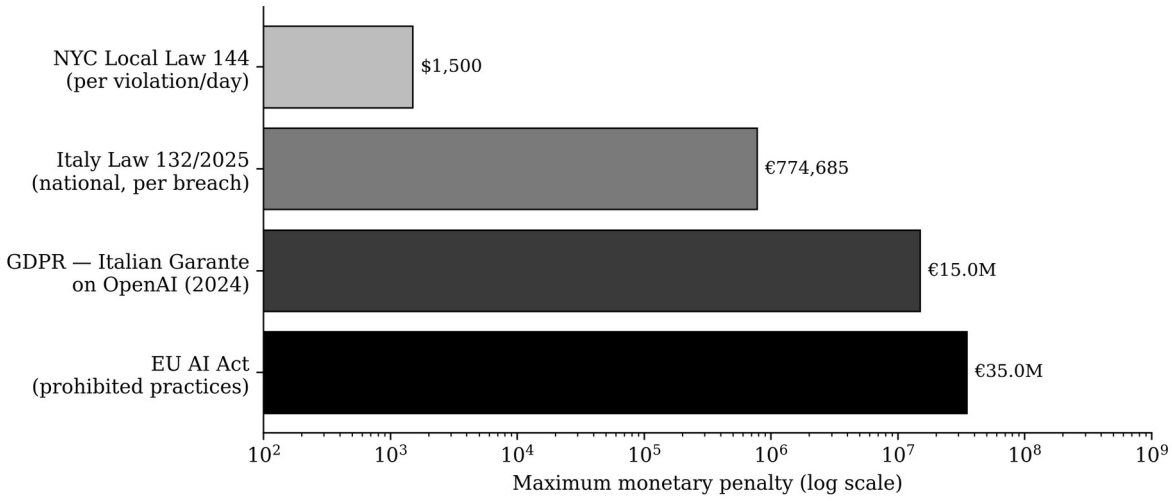
Instrument	Jurisdiction	Type	Key Obligations	Maximum Penalty
EU AI Act (Reg. 2024/1689)	European Union	Binding regulation	Risk-tiered classification; high-risk system documentation, conformity assessment, post-market monitoring; GPAI transparency	EUR 35M or 7% global turnover (prohibited); EUR 15M or 3% (high-risk)
NIST AI RMF 1.0	United States (federal)	Voluntary framework	Govern, Map, Measure, Manage; trustworthy AI characteristics	None; referenced as affirmative defense
ISO/IEC 42001:2023	International	Voluntary, certifiable management standard	Plan-Do-Check-Act AI Management System (AIMS)	None; certification by accredited bodies
NYC Local Law 144	New York City	Binding municipal law	Annual independent bias audit of AEDTs; candidate notice; public summary	USD 500 to 1,500 per day, per violation
Colorado AI Act (SB 24-205)	Colorado, USA	Binding state law	Reasonable-care duty; impact assessments; algorithmic discrimination prevention	Unfair trade practice under CCPA (effective June 30, 2026)
China Interim Measures (Generative AI)	People's Republic of China	Binding administrative measure	Algorithm filing, security assessment, content labeling, training-data lawfulness	Administrative fines, business suspension, criminal liability

Sources: European Parliament (2024); NIST (2023); ISO (2023); NYC DCWP (2023); Colorado General Assembly (2024); Cyberspace Administration of China (2023).

5.2 EU AI Act

The EU AI Act entered force on 1 August 2024 and applies prohibitions and AI-literacy obligations from 2 February 2025 (Regulation EU 2024/1689; SD Worx, 2025). General-purpose AI obligations took effect 2 August 2025; high-risk obligations follow in August 2026. Penalties are tiered: up to EUR 35 million or 7 percent of worldwide annual turnover for prohibited practices; EUR 15 million or 3 percent for breaches of high-risk requirements; and EUR 7.5 million or 1 percent for incorrect or misleading information to authorities (DLA Piper, 2025; Holistic AI, 2024). Italy was the first member state to operationalize national AI penalties under Law 132/2025 (effective 10 October 2025), establishing fines up to EUR 774,685 in addition to EU-level penalties (Legal Nodes, 2025). Figure 4 shows these maxima alongside other AI-relevant regimes on a logarithmic scale.

Figure 4. Maximum monetary penalties under principal AI-relevant regimes



Sources: NYC DCWP (2023); Legal Nodes (2025); Italian DPA Garante (2024); European Parliament (2024).

Figure 4. Maximum monetary penalties under principal AI-relevant regimes (logarithmic scale).

5.3 NIST AI RMF and ISO/IEC 42001

The NIST AI RMF 1.0, released January 2023, organizes AI risk management around four functions (Govern, Map, Measure, and Manage) and emphasizes seven trustworthy-AI characteristics including validity and reliability, safety, security, accountability, transparency, fairness, and privacy (NIST, 2023; Engler, 2023). A Generative AI Profile (NIST AI 600-1) was added in July 2024 (Autio et al., 2024). Practitioner literature has elaborated the framework into operational checklists (Palo Alto Networks, 2024). ISO/IEC 42001:2023 is the world's first certifiable AI management system standard and provides the only third-party-attestable framework as of 2026 (ISO, 2023; ISMS.online, 2024). Major cloud providers including Microsoft and AWS have obtained ISO/IEC 42001 certification (Microsoft, 2024; AWS, 2024).

Adoption remains uneven. The IAPP AI Governance Profession Report 2025 found that 77 percent of organizations are building or refining AI governance programs, but only 36 percent have adopted a formal framework such as the NIST AI RMF (IAPP, 2025; Protecto, 2025). Only 1.5

percent of surveyed organizations believe they have adequate governance headcount; 23.5 percent cite lack of qualified professionals as a top implementation barrier (Next Move Strategy Consulting, 2025). Figure 3 visualizes the resulting adoption-exposure gap.

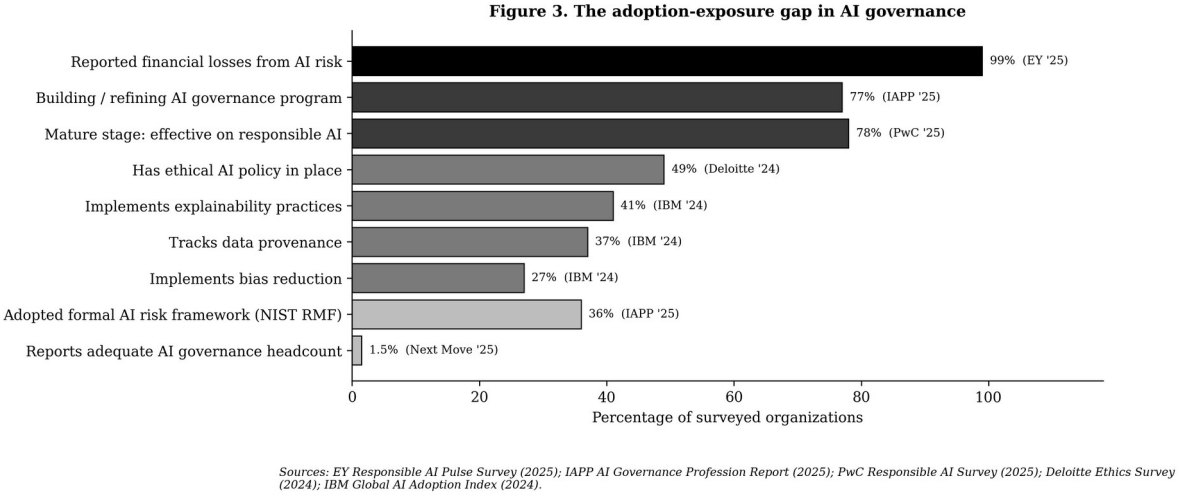


Figure 3. The adoption-exposure gap: high financial losses from AI risk coexisting with low adoption of formal governance frameworks.

5.4 Sub-Federal U.S. Regimes

NYC Local Law 144, effective 5 July 2023, requires employers using Automated Employment Decision Tools (AEDTs) for candidates or employees in New York City to commission an annual independent bias audit, publish summary results, and provide candidate notice (NYC DCWP, 2023). Penalties are USD 500 to 1,500 per violation per day. A December 2025 NYS Comptroller audit concluded the DCWP enforcement system was "ineffective": 75 percent of test calls to the 311 hotline regarding AEDT issues were misrouted; the agency surveyed 32 companies and identified one non-compliance, while Comptroller auditors reviewing the same companies identified at least 17 potential violations (NYS Comptroller, 2025; DLA Piper, 2026).

The Colorado Artificial Intelligence Act (SB 24-205), signed 17 May 2024, was originally to take effect 1 February 2026 but was delayed by special-session bill SB 25B-004 to 30 June 2026 (Clark Hill, 2025; Colorado General Assembly, 2024). It imposes a duty of reasonable care on developers and deployers of "high-risk AI systems" to protect consumers from algorithmic discrimination, requires impact assessments, and provides a rebuttable presumption of reasonable care for entities aligned with the NIST AI RMF or ISO/IEC 42001 (American Bar Association, 2024).

5.5 GDPR-AI Interface and Notable Enforcement

GDPR enforcement against AI-related practices has accelerated. The Italian Garante issued a EUR 15 million fine against OpenAI in December 2024 for unlawful and non-transparent personal-data processing in ChatGPT training (Garante, 2024). The Court of Justice of the European Union ruled in 2023 (Schufa, Joined Cases C-26/22 and C-64/22) that automated credit-scoring outputs constitute a "decision" under Article 22 GDPR even when issued upstream of the deciding entity (CJEU, 2023; Morgan Lewis, 2024). The Berlin DPA fined a bank EUR 300,000 in 2023 for failing

to provide transparency on automated credit-card rejection (Holistic AI, 2024). The EEOC v. iTutorGroup consent decree of September 2023 produced a USD 365,000 settlement and a five-year injunction, the first AI-employment-discrimination case formally resolved by the U.S. Equal Employment Opportunity Commission (EEOC, 2023).

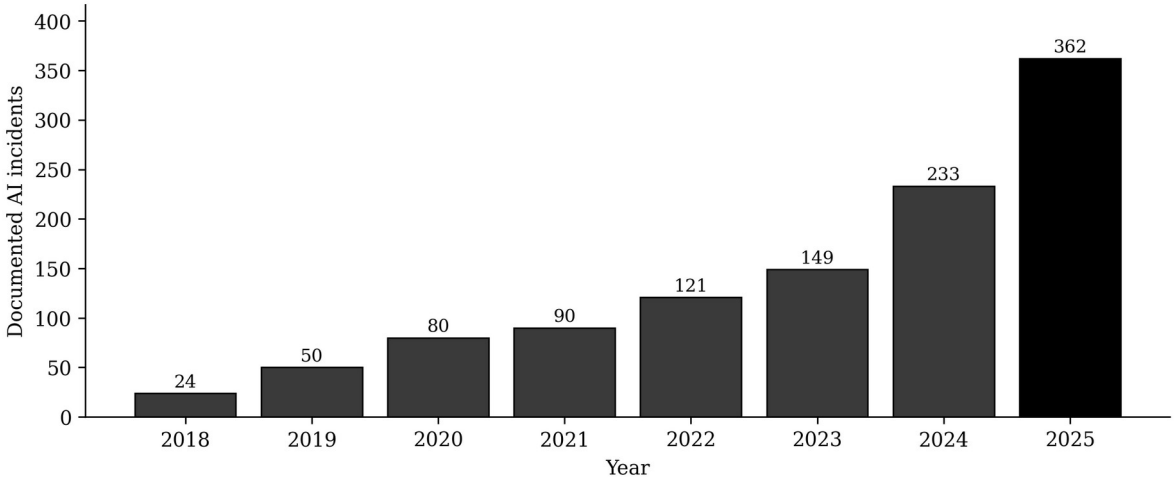
5.6 China's Generative AI Regulation

The Cyberspace Administration of China and six co-issuing agencies promulgated the Interim Measures for the Management of Generative AI Services on 10 July 2023, effective 15 August 2023 (CAC, 2023). The measures impose algorithm filing, security assessments, lawful training data, and content labeling. As of December 2025, 748 generative AI services had completed the national filing process. The Singapore Model AI Governance Framework (BSA / IMDA, 2019) remains the most cited Asia-Pacific soft-law antecedent and continues to influence regional regulatory drafting.

5.7 National AI Strategies

The OECD.AI Policy Observatory tracks more than 1,000 AI policy initiatives across 69 or more jurisdictions (OECD, 2025). By May 2023, 71 jurisdictions had reported 930 or more initiatives. This signals coercive and mimetic isomorphism at the state level: nations are converging on broadly similar strategy templates (R&D investment, ethics charters, sectoral risk frameworks) while substantive enforcement capacity diverges sharply.

Figure 1. Annual AI incidents reported, 2018–2025



Source: Stanford HAI AI Index (2024, 2025, 2026); OECD AI Incidents and Hazards Monitor.

Figure 1. Annual AI incidents reported, 2018 to 2025. The rapid post-2022 acceleration is the empirical backdrop for the regulatory wave covered in this section.

6. Section C: AI Validation Frameworks

6.1 Technical Validation Standards

Technical validation of AI systems encompasses (a) predictive performance evaluation, (b) fairness and bias testing, (c) robustness and adversarial testing, (d) explainability assessment, and (e) post-deployment monitoring including data and concept drift detection (KPMG, 2024; Nimbleway, 2024; Aerospike, 2024). The NIST AI RMF Measure function provides high-level guidance, but specific quantitative thresholds remain absent (NIST, 2023; Engler, 2023). KPMG's (2024) validating-AI-models framework adapts the long-established three lines of defense from financial-services model risk management to AI/ML, but neither it nor competing vendor frameworks yet enjoy regulatory recognition.

6.2 The Four-Fifths Rule and Adverse Impact

For employment AI, the four-fifths rule from the 1978 Uniform Guidelines on Employee Selection Procedures (29 CFR Part 1607) remains the operative U.S. standard: a selection rate for any race, sex, or ethnic group less than 80 percent of the highest-rate group is regarded by federal enforcement agencies as evidence of adverse impact (EEOC, 1978/2024). NYC Local Law 144 incorporates this logic, requiring computation of selection rates and impact ratios across demographic and intersectional categories (NYC DCWP, 2023). The EEOC's May 2023 technical assistance document on AI under Title VII reaffirmed that employers remain liable for vendor-developed tools that produce disparate impact (EEOC, 2023).

6.3 Drift, Failure, and Discriminatory Outcomes

Estimates of AI model failure rates in production vary by definition and source. Industry analyses report that the majority of pilots never reach production, and one MIT Sloan study cited median time from pilot approval to shutdown of 14 months in retrieval-augmented generation projects (Folio3, 2026). Vendor analyses widely report that approximately 91 percent of ML models experience performance degradation due to drift (Articsledge, 2025) [ESTIMATE: DERIVED FROM VENDOR SURVEY; NEEDS PEER-REVIEWED VERIFICATION]. A 2024 Evidently AI survey found that up to 32 percent of production scoring pipelines experience distributional shifts within the first six months. The Stanford AI Index recorded 233 AI incidents in 2024 (a 56.4 percent year-on-year rise) and 362 in 2025 (Stanford HAI, 2025, 2026).

The Amazon recruiting case (2014 to 2018), in which an internal resume-scoring model penalized resumes containing the word "women's," remains the canonical case illustrating training-data bias replicating historical workforce inequality. Independent academic surveys document persistent algorithmic-fairness concerns across more than 250 AI-for-HR tools currently on the market (Fabris et al., 2023).

6.4 Organizational Adoption of Validation Practices

PwC's 2025 US Responsible AI Survey of 310 business leaders found that 78 percent of respondents in the most mature ("strategic") stage report being very effective at defining and communicating responsible AI priorities, compared with only 35 percent in the training stage (PwC, 2025). Deloitte's 2024 ethics survey reported that 49 percent of executive respondents had ethical AI guidelines or policies in place, and an additional 37 percent were "nearly ready to roll them out" (Deloitte, 2024). IBM's 2024 Global AI Adoption Index found that fewer than half of

deploying organizations were taking key trustworthy-AI steps such as bias reduction (27 percent), data-provenance tracking (37 percent), explainability (41 percent), or ethical AI policy development (44 percent) (IBM, 2024).

7. Section D: Critical Gaps Analysis

This section synthesizes the seven gaps that emerged from the literature scan. Table 2 summarizes the gaps with supporting evidence.

Table 2. Seven Critical Gaps with Supporting Evidence

#	Gap	Primary Sources	Quantified Evidence
1	Absence of standardized AI validation protocols across jurisdictions	Mittelstadt (2019); Stanford HAI (2026)	Hallucination rates across 26 leading models range 22 to 94 percent; benchmark coverage inconsistent
2	Disconnect between AI consulting deliverables and measurable governance outcomes	Future of Consulting (2025); Raji et al. (2020)	Only ~25 percent of consulting fees outcome-linked at McKinsey
3	Lack of continuous post-deployment monitoring standards	Articsledge (2025); Aerospike (2024)	32 percent of production pipelines shift within 6 months; ~91 percent of models drift
4	Inadequate adverse-impact analysis in AI-assisted employment decisions	EEOC (1978/2024); Fabris et al. (2023)	iTutorGroup settlement USD 365,000; 250+ AI-for-HR tools, audit ecosystem nascent
5	No unified third-party auditing certification framework	Costanza-Chock et al. (2022); Raji et al. (2022)	NYC DCWP enforcement "ineffective"; 17 violations missed vs. 1 identified
6	Insufficient interdisciplinary expertise in AI governance consulting	IAPP (2025); Tippins et al. (2021)	Only 1.5 percent of organizations report adequate governance headcount
7	Accountability vacuum: unclear liability when AI systems cause harm	Santoni de Sio & Mecacci (2021); van der Meer & Hage (2024)	EY 2025: 99 percent of orgs report financial losses from AI risk; near two-thirds >USD 1M

Sources: see in-text citations.

Gap 1: Absence of Standardized AI Validation Protocols Across Jurisdictions

The EU AI Act, NIST AI RMF, ISO/IEC 42001, and Colorado AI Act each invoke "validity," "reliability," "robustness," and "fairness" as concepts, but none operationalizes a single shared protocol for measurement (NIST, 2023; ISO, 2023). The Stanford AI Index 2026 reported that leading frontier-model developers report results on different responsible-AI benchmarks, complicating cross-model comparison (Stanford HAI, 2026). Mittelstadt (2019) provided the foundational critique: high-level principles converge while methodological standards diverge, leaving "vague, action-guiding statements" without enforceable mid-level norms.

Gap 2: Disconnect Between AI Consulting Deliverables and Measurable Governance Outcomes

Despite an USD 11+ billion AI consulting market, peer-reviewed evidence of consulting deliverables producing measurable governance outcomes is sparse. Raji et al. (2020) defined the AI accountability gap as the absence of internal audit infrastructure linking development decisions to documented review. Future of Consulting (2025) reported that approximately 75 percent of fees at AI-forward firms remain billed by hours rather than outcomes, an institutional decoupling between symbolic (consulting deliverables) and substantive (governance performance) layers consistent with Meyer and Rowan (1977).

Gap 3: Lack of Continuous Post-Deployment Monitoring Standards

While the EU AI Act requires post-market monitoring (Article 72) and the NIST AI RMF Manage function emphasizes continuous improvement, no jurisdiction prescribes specific cadence, drift-detection thresholds, or reporting templates. Aerospike (2024) and Articsledge (2025) document that production models routinely degrade due to data and concept drift. Without standardized monitoring, organizations cannot demonstrate "reasonable care" under the Colorado AI Act or post-market obligations under the EU AI Act.

Gap 4: Inadequate Adverse-Impact Analysis in AI-Assisted Employment Decisions

Although the four-fifths rule (29 CFR section 1607.4) has been the default U.S. framework since 1978, its application to AI systems raises unresolved questions about intersectional categories, sample sizes for demographic subgroups, and the appropriate baseline rate for highly automated screening pipelines (EEOC, 2023). Fabris et al. (2023) documented that more than 250 AI-for-HR tools are on the market, yet vendor cooperation with bias audits is inconsistent and frequently lacks demographic data necessary for proper analysis. The iTutorGroup case (EEOC v. iTutorGroup, Civil Action No. 1:22-cv-02565) yielded a USD 365,000 settlement, but represents only the first, and so far the principal, federal enforcement action specifically targeting AI-driven employment discrimination (EEOC, 2023). The author's AVS Framework (Ahmad, 2025, in development) responds to this gap by proposing structured Adverse-impact, Validity, and Stability checkpoints for AI-assisted hiring.

Gap 5: No Unified Third-Party Auditing Certification Framework

Costanza-Chock, Raji, and Buolamwini (2022) provided the most influential field scan of the algorithmic auditing ecosystem, concluding that "AI auditing isn't really a thing at this point," describing the category as aspirational. Raji, Xu, Honigsberg, and Ho (2022) extended this argument by showing that mature audit ecosystems in finance, environment, and pharmaceuticals depend on accreditation bodies, mandatory disclosure regimes, and statutory data access, none of which exists for AI at the global level. ISO/IEC 42001 is the only certifiable management standard, but it audits the management system, not the underlying model. The NYS Comptroller's December 2025 audit of NYC DCWP enforcement of Local Law 144 found that DCWP review missed at least 16 of 17 potential violations identified by independent auditors (NYS Comptroller, 2025), a clear demonstration of the unified-certification gap.

Gap 6: Insufficient Interdisciplinary Expertise in AI Governance Consulting

Tippins, Oswald, and McPhail (2021) argued that I-O psychology's century of expertise on test validity, reliability, and adverse-impact analysis is structurally underrepresented in AI governance discourse, which has been dominated by computer scientists, lawyers, and ethicists. The IAPP AI Governance Profession Report 2025 found that only 1.5 percent of organizations believe they have adequate governance headcount, while 23.5 percent cite a lack of qualified professionals as a top implementation barrier (IAPP, 2025; Next Move Strategy Consulting, 2025). The implication for institutional theory is that normative isomorphism (professionalization) is incomplete: the field has not yet stabilized which disciplines and credentials are authoritative.

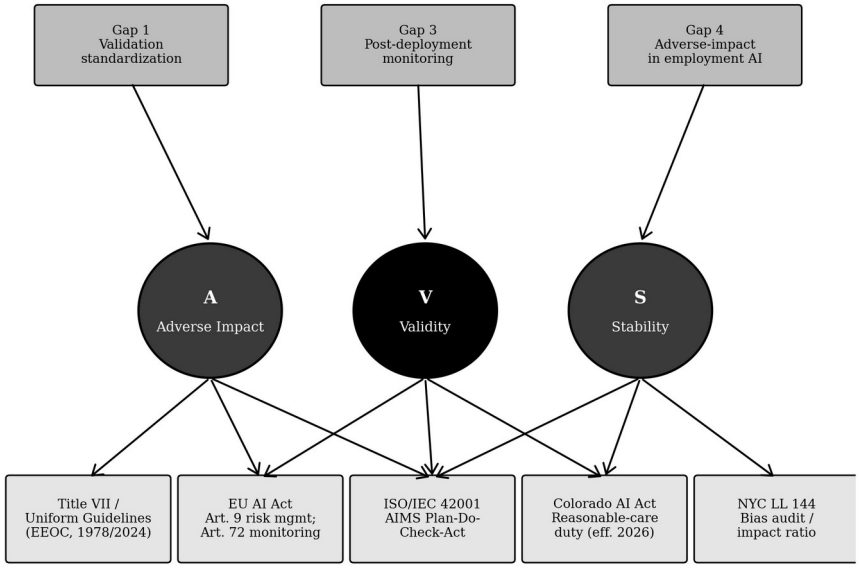
Gap 7: Accountability Vacuum: Unclear Liability When AI Systems Cause Harm

Santoni de Sio and Mecacci (2021) and van der Meer and Hage (2024) characterize multiple "responsibility gaps" arising from the unpredictability of self-learning systems and the distance between developers and downstream harms. Existing tort and product-liability frameworks struggle with AI systems whose outputs are probabilistic and emergent. The EY Responsible AI Pulse Survey (2025) reported that 99 percent of surveyed organizations (n = 975 across 21 countries) had experienced financial losses from AI-related risks, and nearly two-thirds had losses exceeding USD 1 million, yet the legal architecture for victim redress remains fragmented across the EU's Product Liability Directive revision, individual U.S. tort regimes, and emerging jurisprudence such as Schufa.

8. Section E: Solutions Architecture

Each gap identified in Section 7 maps to a concrete intervention. The integrated solution architecture is illustrated in Figure 6 and summarized in Table 3. The following discussion walks through each proposed solution, identifies its regulatory anchor, and notes which professional discipline carries comparative advantage in implementation.

Figure 6. Integrated solution architecture: AVS Framework mapping the seven gaps to interventions and regulatory anchors



AVS Framework outputs: structured adverse-impact analyses (4/5 rule + statistical-significance tests), validity coefficients, and stability/drift monitoring; deliverables map directly to NIST AI RMF Measure-Manage and EU AI Act Art. 9-Art. 72.

Figure 6. Integrated solution architecture. The AVS Framework (Adverse impact, Validity, Stability) sits at the center, mapping selected gaps to interventions and to regulatory anchors including Title VII / the Uniform Guidelines, EU AI Act Articles 9 and 72, ISO/IEC 42001, the Colorado AI Act, and NYC Local Law 144.

8.1 Solution to Gap 1: A Harmonized Validation Protocol

A practical response to fragmented validation standards is to publish a cross-jurisdictional validation protocol that sequences the steps already implicit in the EU AI Act Article 9 risk-management requirements, the NIST AI RMF Measure function, and ISO/IEC 42001 Annex A controls. Drawing on the three-lines-of-defense logic that KPMG (2024) and other major audit firms have adapted from banking model-risk management, the protocol would specify (a) intended-use definition and test-set construction; (b) predictive-validity statistics with confidence intervals; (c) fairness and adverse-impact metrics keyed to applicable law; (d) robustness and adversarial-stress test families; (e) explainability documentation; and (f) handover criteria for production. The Brookings analysis of the NIST AI RMF (Engler, 2023) provides a useful conceptual scaffold for such convergence at the U.S. federal level, while the Singapore Model AI Governance Framework (BSA / IMDA, 2019) demonstrates what such a protocol can look like when expressed as an implementation playbook for organizations rather than a statute.

8.2 Solution to Gap 2: Outcome-Linked Consulting Engagements

The procurement-side response to the consulting accountability gap is to migrate AI advisory engagements toward outcome-based contracting, in which a meaningful share of fees is tied to verifiable governance outcomes such as the reduction of incident rates, time-to-remediation, audit-pass rates, and statistical reductions in adverse-impact ratios. Practitioner literature (OneTrust, 2024; Databricks, 2024) is converging on this orientation. The empirical question, addressed in the research agenda below, is whether outcome-linked engagements actually produce better

governance results than time-and-materials engagements when controlling for organizational maturity and use-case complexity.

8.3 Solution to Gap 3: A Continuous Monitoring Stack with Drift Thresholds

Closing the post-deployment monitoring gap requires moving from principles to thresholds. A continuous monitoring stack combines (a) ground-truth labeling pipelines for online evaluation; (b) population stability index (PSI) monitoring for input distributions; (c) prediction-distribution monitoring with alerting; (d) periodic re-running of fairness metrics on production traffic; and (e) tripwires for automatic retraining or rollback (Nimbleway, 2024; Agility-at-Scale, 2024). Cadence and threshold parameters should be calibrated to risk tier rather than fixed across the enterprise, and reporting should be structured to support both internal governance committees and external regulators under EU AI Act Article 72.

8.4 Solution to Gap 4: The AVS Framework for Employment AI

The author's Adverse-impact, Validity, and Stability (AVS) Framework operationalizes the four-fifths rule (29 CFR section 1607.4) alongside criterion-related and content-related validation studies and longitudinal stability monitoring. As Figure 6 illustrates, AVS deliverables map directly to NYC Local Law 144 bias-audit requirements, the Colorado AI Act's reasonable-care duty, EU AI Act Article 9 risk-management documentation, and ISO/IEC 42001 AIMS performance evaluation clauses. The framework's contribution is not to invent new fairness metrics but to integrate the well-established methodological apparatus of selection science with the modern engineering apparatus of MLOps, producing a single validation dossier that simultaneously satisfies multiple regulatory anchors.

8.5 Solution to Gap 5: An Accreditation Pathway for Third-Party AI Auditors

Costanza-Chock et al. (2022) and Raji et al. (2022) argue that mature audit ecosystems require accreditation bodies, scope-of-practice rules, mandatory disclosure regimes, and statutory data-access rights. A practical near-term step is the establishment of an AI auditor accreditation scheme analogous to the public-company audit oversight bodies in finance, building on the existing ISO/IEC 17021 and ISO/IEC 17029 conformity-assessment infrastructure. Such a scheme would specify auditor independence rules, minimum competency requirements (combining ML, statistics, domain expertise, and regulatory literacy), reporting templates, and a public register of audits. The NYS Comptroller's December 2025 finding that NYC DCWP enforcement missed at least 16 of 17 potential violations underscores the urgency of this institutional layer.

8.6 Solution to Gap 6: Interdisciplinary Governance Teams

The skills gap identified in IAPP (2025) and Tippins et al. (2021) calls for explicit interdisciplinarity in AI governance teams. A workable team archetype combines (a) a machine-learning engineer with MLOps depth; (b) a measurement specialist (psychometrician, biostatistician, or econometrician) for validity, reliability, and adverse-impact analysis; (c) a privacy and AI-law counsel; and (d) a domain expert from the relevant function (HR, clinical operations, lending, education). Educational institutions and professional associations should

support this through joint certificate programs and shared continuing-education curricula. The Centre for the Governance of AI's research agenda (GovAI, 2024) identifies workforce development as a first-order governance research priority.

8.7 Solution to Gap 7: Shared-Responsibility Liability Matrices

Because AI systems are typically built and deployed across many parties, the responsibility-gap literature recommends moving from blame-assignment to shared-responsibility matrices that allocate duties, indemnities, and audit rights along the supply chain (van der Meer & Hage, 2024; Santoni de Sio & Mecacci, 2021). Such matrices can be embedded in vendor contracts, standardized through industry consortia, and recognized in regulatory guidance. The EU's revised Product Liability Directive and the Colorado AI Act's developer-deployer distinction provide the doctrinal scaffolding; what remains is the contracting and disclosure practice.

Table 3. Mapping of the Seven Gaps to Proposed Solutions and Regulatory Anchors

Gap	Proposed Solution	Regulatory / Standards Anchor	Discipline of Comparative Advantage
1	Cross-jurisdictional harmonized validation protocol	EU AI Act Art. 9; NIST AI RMF; ISO/IEC 42001	ML engineering + measurement science
2	Outcome-linked consulting contracts	Procurement law + voluntary industry codes	Management consulting + procurement
3	Continuous monitoring stack with drift thresholds	EU AI Act Art. 72 (post-market); NIST RMF Manage	MLOps + biostatistics
4	AVS Framework for employment AI	Title VII; 29 CFR 1607; NYC LL 144; CO AI Act	I-O psychology + employment law
5	Accreditation pathway for third-party AI auditors	ISO/IEC 17021/17029 conformity assessment	Audit profession + regulators
6	Interdisciplinary governance team archetype	ISO/IEC 42001 competence requirements	Workforce development + professional bodies
7	Shared-responsibility liability matrices	EU Product Liability Directive; Colorado AI Act	Law + contract management

Sources: synthesized from references cited throughout this paper.

9. Proposed Research Agenda

The seven gaps and seven solutions map onto five priority research questions (RQs) suitable for peer-reviewed inquiry across management, technology policy, and applied psychology.

Priority RQ1 (Validation protocol standardization): To what extent can a unified validation protocol, integrating predictive accuracy, fairness metrics (including the four-fifths rule and statistical-significance tests), drift monitoring, and explainability, be harmonized across the EU AI Act, NIST AI RMF, ISO/IEC 42001, and the Colorado AI Act? Justification: Gap 1 is the institutional foundation for Gaps 3 and 5; without standardization, cross-jurisdictional firms must build redundant validation pipelines, raising costs without improving safety.

Priority RQ2 (Outcome-linked consulting): Does outcome-based pricing in AI consulting engagements (versus time-and-materials) produce measurably better governance outcomes, including incident reduction, audit-pass rates, and time-to-remediation? Justification: Gap 2; this is a testable management research question with clear operationalization.

Priority RQ3 (Continuous monitoring cadence): What is the empirically optimal cadence and threshold structure for post-deployment monitoring across AI use cases of varying risk (such as hiring versus consumer recommendation versus clinical decision support)? Justification: Gap 3; this requires interdisciplinary collaboration among ML engineers, I-O psychologists, and biostatisticians.

Priority RQ4 (AVS-type frameworks for employment AI): Can structured Adverse-impact, Validity, and Stability frameworks, integrating the four-fifths rule with predictive validity studies and longitudinal stability monitoring, improve compliance with NYC Local Law 144, the Colorado AI Act, and Title VII? Justification: Gap 4; this is a direct extension of the author's research program and aligns I-O psychology methodology with regulatory compliance needs.

Priority RQ5 (Audit accreditation ecosystem): Drawing on the institutional design of financial, environmental, and pharmaceutical audit regimes, what accreditation, data-access, and disclosure rules would be required to sustain a credible third-party AI audit ecosystem? Justification: Gap 5; following Raji et al. (2022), this question is foundational to converting voluntary frameworks into enforceable practice.

A secondary research stream should address Gaps 6 and 7 through workforce-development scholarship (interdisciplinary curriculum design) and comparative tort and regulatory analysis respectively.

10. Limitations

This review has four principal limitations. First, the heterogeneity of sources (peer-reviewed, regulatory, and grey literature) precludes formal meta-analysis; the synthesis is interpretive. Second, several quantitative claims (such as the 91 percent model-drift figure and certain market-size estimates) originate in vendor or consulting reports that have not undergone independent peer review; these are flagged in-text. Third, the EU AI Act, the Colorado AI Act, and Italian Law 132/2025 are recent and partially unenforced as of April 2026; future enforcement data will refine the conclusions drawn here. Fourth, the review prioritizes English-language sources, likely under-representing Chinese, Japanese, and Brazilian regulatory scholarship. Finally, the AVS Framework is a research-in-progress contribution by the author; it has not yet been independently validated and is presented here as a directional proposal rather than an established standard.

11. Conclusion

The institutional infrastructure for AI consulting, governance, and validation is being assembled in real time, under coercive (regulatory), mimetic (peer imitation), and normative (professionalization) pressures simultaneously (DiMaggio & Powell, 1983). The combined

evidence reviewed here indicates that the symbolic layer of AI governance (principles, charters, frameworks, market vocabulary) has converged faster than the substantive layer of validated practice, certified auditors, and enforceable post-deployment monitoring. The result is a governance field that looks increasingly homogeneous on its surface yet remains deeply decoupled from operational outcomes (Meyer & Rowan, 1977; Mittelstadt, 2019).

Three findings should orient the next phase of scholarship and practice. First, validation protocol standardization is the keystone gap: without it, downstream gaps in monitoring, auditing, and accountability cannot be closed. Second, the AI consulting industry's outcome-pricing transition, however slow, is the principal market mechanism through which substantive practice could be aligned with symbolic claims. Third, applied psychology, particularly I-O psychology's adverse-impact and validation methodology, has both an obligation and a comparative advantage in shaping AI employment governance, an opportunity that the AVS Framework and similar interdisciplinary efforts are positioned to advance.

The cost of inaction is no longer hypothetical. With 362 documented AI incidents in 2025, a 99 percent organizational loss rate, and the first wave of multi-million-euro regulatory penalties, the institutional field of AI governance has crossed a phase change. The seven gaps identified here, the seven solutions proposed, and the five-question research agenda are offered as a structured response.

12. Workflow Summary

For practitioners seeking to operationalize the analysis above, the following workflow consolidates the paper's recommended sequence into eight steps. Step 1: scope the AI use case against the risk tiers defined in the EU AI Act and the Colorado AI Act, and select the appropriate set of standards (NIST AI RMF, ISO/IEC 42001, NYC Local Law 144). Step 2: assemble the interdisciplinary team described in Section 8.6. Step 3: complete pre-deployment validation according to the harmonized protocol of Section 8.1, including criterion-related validity studies and adverse-impact analysis where employment is involved. Step 4: deploy the continuous monitoring stack of Section 8.3, with risk-tier-appropriate cadences and thresholds. Step 5: engage qualified third-party auditors under the accreditation pathway of Section 8.5, with deliverables aligned to ISO/IEC 42001 and NYC Local Law 144 templates. Step 6: encode shared-responsibility allocations in vendor and customer contracts per Section 8.7. Step 7: institute outcome-linked consulting engagements per Section 8.2 to align external advisory work with substantive governance outcomes. Step 8: feed monitoring data, audit findings, and incident reports back into the validation pipeline as a closed-loop quality-improvement cycle aligned with the Plan-Do-Check-Act logic of ISO/IEC 42001.

13. References

American Bar Association. (2024, July). Colorado enacts law regulating high-risk artificial intelligence systems. Business Law Today.

https://www.americanbar.org/groups/business_law/resources/business-law-today/2024-july/colorado-enacts-law-regulating-high-risk-artificial-intelligence-systems/

Autio, C., Schwartz, R., Dunietz, J., Jain, S., Stanley, M., Tabassi, E., Hall, P., & Roberts, K. (2024). Artificial intelligence risk management framework: Generative AI profile (NIST AI 600-1). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.600-1 https://airc.nist.gov/airmf-resources/airmf/

BSA / Singapore Infocomm Media Development Authority. (2019). Model AI governance framework (1st ed.). https://ai.bsa.org/wp-content/uploads/2019/09/Model-AI-Framework-First-Edition.pdf

Centre for the Governance of AI (GovAI). (2024). Research agenda for AI governance. https://cdn.governance.ai/GovAI-Research-Agenda.pdf

Colorado General Assembly. (2024). Senate Bill 24-205: Consumer protections for artificial intelligence. https://leg.colorado.gov/bills/sb24-205

Costanza-Chock, S., Raji, I. D., & Buolamwini, J. (2022). Who audits the auditors? Recommendations from a field scan of the algorithmic auditing ecosystem. In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22) (pp. 1571-1583). https://doi.org/10.1145/3531146.3533213 https://dl.acm.org/doi/10.1145/3531146.3533213

Cyberspace Administration of China. (2023). Interim measures for the management of generative artificial intelligence services. https://www.chinalawtranslate.com/en/generative-ai-interim/

Databricks. (2024). AI governance best practices: How to build responsible and effective AI programs. https://www.databricks.com/blog/ai-governance-best-practices-how-build-responsible-and-effective-ai-programs

Deloitte. (2024). Preparing the workforce for ethical, responsible, and trustworthy AI. https://www.deloitte.com/us/en/about/articles/ethical-technology-survey.html

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. American Sociological Review, 48(2), 147-160. https://doi.org/10.2307/2095101 https://doi.org/10.2307/2095101

Engler, A. (2023). NIST's AI Risk Management Framework plants a flag in the AI debate. Brookings Institution. https://www.brookings.edu/articles/nists-ai-risk-management-framework-plants-a-flag-in-the-ai-debate/

Equal Employment Opportunity Commission. (1978/2024). Uniform guidelines on employee selection procedures, 29 CFR Part 1607. https://www.ecfr.gov/current/title-29/subtitle-B/chapter-XIV/part-1607

Equal Employment Opportunity Commission. (2023, August 9). iTutorGroup to pay $365,000 to settle EEOC discriminatory hiring suit [Press release].

https://www.eeoc.gov/newsroom/itutorgroup-pay-365000-settle-eeoc-discriminatory-hiring-suit

European Parliament. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union. https://artificialintelligenceact.eu/

Fabris, A., Baranowska, N., Dennis, M. J., Hacker, P., Saldivar, J., Borgesius, F. Z., & Biega, A. J. (2023). Fairness and bias in algorithmic hiring: A multidisciplinary survey. ACM Transactions on Intelligent Systems and Technology. https://arxiv.org/abs/2309.13933

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. Harvard Data Science Review, 1(1). https://doi.org/10.1162/99608f92.8cd550d1 https://doi.org/10.1162/99608f92.8cd550d1

Forrester. (2024). AI governance software spend will see 30% CAGR from 2024 to 2030. https://www.forrester.com/blogs/ai-governance-software-spend-will-see-30-cagr-from-2024-to-2030/

Future Market Insights. (2025). AI consulting services market: Size and forecast 2025-2035. https://www.futuremarketinsights.com/reports/ai-consulting-services-market

International Association of Privacy Professionals (IAPP). (2025). AI governance profession report 2025. https://iapp.org/

International Organization for Standardization. (2023). ISO/IEC 42001:2023 - Information technology - Artificial intelligence - Management system. https://www.iso.org/standard/42001

ISMS.online. (2024). ISO/IEC 42001 implementation guide. https://www.isms.online/iso-42001/

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389-399. https://doi.org/10.1038/s42256-019-0088-2 https://doi.org/10.1038/s42256-019-0088-2

KPMG. (2024). Validating AI models: Regulatory insights. https://kpmg.com/xx/en/our-insights/regulatory-insights/validating-ai-models.html

Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. American Journal of Sociology, 83(2), 340-363. https://doi.org/10.1086/226550 https://doi.org/10.1086/226550

Microsoft. (2024). ISO/IEC 42001 compliance offering. https://learn.microsoft.com/en-us/compliance/regulatory/offering-iso-42001

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. Nature Machine Intelligence, 1(11), 501-507. https://doi.org/10.1038/s42256-019-0114-4 https://doi.org/10.1038/s42256-019-0114-4

National Institute of Standards and Technology. (2023). Artificial intelligence risk management framework (AI RMF 1.0) (NIST AI 100-1). https://doi.org/10.6028/NIST.AI.100-1 https://www.nist.gov/itl/ai-risk-management-framework

New York City Department of Consumer and Worker Protection. (2023). Automated employment decision tools (Local Law 144). https://www.nyc.gov/site/dca/about/automated-employment-decision-tools.page

New York State Office of the Comptroller. (2025, December). Enforcement of Local Law 144: Automated employment decision tools. https://www.osc.ny.gov/state-agencies/audits/2025/12/02/enforcement-local-law-144-automated-employment-decision-tools

Nimbleway. (2024). The model validation framework. https://www.nimbleway.com/blog/model-validation-framework

OECD. (2025). OECD.AI Policy Observatory. https://oecd.ai/

OneTrust. (2024). Navigating the future of AI governance: Insights and best practices. https://www.onetrust.com/blog/navigating-the-future-of-ai-governance-insights-and-best-practices/

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71 https://doi.org/10.1136/bmj.n71

Palo Alto Networks. (2024). NIST AI risk management framework. https://www.paloaltonetworks.com/cyberpedia/nist-ai-risk-management-framework

PwC. (2025). 2025 US responsible AI survey: From policy to practice. https://www.pwc.com/us/en/tech-effect/ai-analytics/responsible-ai-survey.html

Raji, I. D., & Buolamwini, J. (2019). Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial AI products. In Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (pp. 429-435). https://doi.org/10.1145/3306618.3314244 https://doi.org/10.1145/3306618.3314244

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (pp. 33-44). https://doi.org/10.1145/3351095.3372873 https://doi.org/10.1145/3351095.3372873

Raji, I. D., Xu, P., Honigsberg, C., & Ho, D. (2022). Outsider oversight: Designing a third party audit ecosystem for AI governance. In Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society. https://doi.org/10.1145/3514094.3534181 https://arxiv.org/pdf/2206.04737

Santoni de Sio, F., & Mecacci, G. (2021). Four responsibility gaps with artificial intelligence: Why they matter and how to address them. Philosophy and Technology, 34(4), 1057-

1084. https://doi.org/10.1007/s13347-021-00450-x https://doi.org/10.1007/s13347-021-00450-x

Scott, W. R. (2014). Institutions and organizations: Ideas, interests, and identities (4th ed.). Sage. https://us.sagepub.com/en-us/nam/institutions-and-organizations/book238452

Stanford Institute for Human-Centered Artificial Intelligence. (2025). AI Index 2025: Responsible AI. https://hai.stanford.edu/ai-index/2025-ai-index-report/responsible-ai

Stanford Institute for Human-Centered Artificial Intelligence. (2026). AI Index 2026: Responsible AI. https://hai.stanford.edu/ai-index/2026-ai-index-report/responsible-ai

Tippins, N. T., Oswald, F. L., & McPhail, S. M. (2021). Scientific, legal, and ethical concerns about AI-based personnel selection tools: A call to action. Personnel Assessment and Decisions, 7(2). https://doi.org/10.25035/pad.2021.02.001 https://doi.org/10.25035/pad.2021.02.001

van der Meer, T., & Hage, J. (2024). From liability gaps to liability overlaps: Shared responsibilities and fiduciary duties in AI and other complex technologies. AI and Society. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC12152026/

Manuscript prepared April 2026. Contact and AVS Framework working papers available upon request. The author declares no conflicts of interest.