World Journal of Advanced Research and Reviews

eISSN: 2581-9615 CODEN (USA): WJARAI

Cross Ref DOI: 10.30574/wjarr

Journal homepage: https://wjarr.com/

WJARR

World Journal of Advanced Research and Reviews

(Review Article)

Literature Review of Artificial Intelligence Security Governance Frameworks: Risk Management Strategies for Regulated Industries Including Healthcare, Financial Services, and Gaming

Adeyemi A. Bello * and Julie Reneau

College of Business, University of Texas Permian Basin, Odessa. TEXAS 79765. USA.

World Journal of Advanced Research and Reviews, 2023, 18(02), 1503-1522

Publication history: Received on 03 April 2023; revised on 23 May 2023; accepted on 28 May 2023

Article DOI: https://doi.org/10.30574/wjarr.2023.18.2.0833

Abstract

The fast development of artificial intelligence in the regulated sphere has generated an intricate and dynamic security governance environment that the current regulatory framework did not intend to tackle in all its aspects. This literature review is a systematic review of AI security governance frameworks in three highly regulated areas, namely healthcare, financial services, and gaming. In a PRISMA-conformant approach, we compiled 16 peer-reviewed articles out of an original sample of 48 Scopus documents, complemented with regulatory documents, industry report, and data on expert surveys (n=124 cybersecurity professionals). Cohen-based kappa (κ = 0.88) was used to determine the inter-rater reliability and it was found to be very high. The review assesses the NIST AI Risk Management Framework, the ISO/IEC 42001, and the EU AI Act on eight dimensions of governance. This indicates that financial services are the most mature area of governance (compliance score: 87/100 by 2023), whereas the gaming industry has the lowest level of audit of algorithmic bias, explainability, and compliance of cross-border data. The case of healthcare is full of paradoxes of good regulatory intent and poor execution. There are five important areas of governance that need to be addressed as a matter of urgency, which include AI model explainability, standards of adversarial robustness, supply chain AI risk management, cross-sector incident sharing, and adaptive regulatory sandboxing.

Keywords: Artificial Intelligence Security; Ai Governance Frameworks; Healthcare Ai Risk; Financial Services Ai Regulation; Gaming Ai Compliance; Nist Ai RMF; Iso/IEC 42001; Eu Ai Act; Cybersecurity Governance; Risk Management

1. Introduction

Artificial intelligence has become one of the most revolutionary and at the same time disruptive technologies that are facing controlled sectors in the modern world. In healthcare service providers, financial services providers and even in the gaming industry, AIs are being implemented to streamline decision processes, personalize services, identify fraud, and facilitate complex regulatory compliance processes. However, this accelerated uptake has outpaced the creation of consistent, industry-specific security governance frameworks with the capacity to deal with the new and multifaceted risks that AI systems present (Brundage et al., 2018). The implications of this governance vacuum are severe: AI technologies working in controlled settings make high-stakes judgments concerning patient care, creditworthiness, consumer protection, and regulatory impacts regularly, but the structures governing their safety, equity, and responsibility are patchy, ad hoc, and often industry-specific.

AI security governance regulatory environment is both poorly developed and complicated. The existing standards, namely the NIST AI Risk Management Framework, ISO/IEC 42001, and the AI Act in the European Union are valuable as they do exist, but they are still theorized and empirically understudied in terms of their applicability to the operational settings of a healthcare, financial services, and game environment (Smuha, 2021). In the meantime,

* Corresponding author: Adeyemi A. Bello

industry-relevant laws like the Health Insurance Portability and Accountability Act of the healthcare sector or the Gramm-Leach-Bliley Act and Dodd-Frank provisions of the financial services sector were designed to work in pre-AI institutional contexts and fail to sufficiently address the unique security risks of machine learning systems, including adversarial attacks, model drift, algorithmic bias, and algorithmic opacity (Floridi et al., 2019).

In the framework of the system of cybersecurity governance, in particular, the adoption of AI also brings two problems: not only AI systems become the means to enhance the state of security posture, but also new attack vectors that can be used by advanced threat actors. The machine learning systems used in the gaming context as a tool to detect fraud, verify identities and participate in clinical decision support, and content moderation are susceptible to various adversarial manipulations such as those data poisoned, those that evade models, and those that infer membership (Taddeo et al., 2019). These governmental lapses system has a direct bearing on consumer protection, financial stability, national security, and health.

1.1. Research Context and Motivations

1.1.1. The AI Governance Imperative in Regulated Industries

The intersection of artificial intelligence and regulated industry operations has created what Floridi et al. (2019) describe as an ethical and governance emergency, wherein the pace of technological deployment has fundamentally outstripped the development of adequate oversight mechanisms. In regulated industries specifically, this governance deficit has particularly acute consequences across three interrelated domains:

The healthcare systems using AI as the support of the diagnostic processes must face not only the traditional cybersecurity threats of information breach and unauthorized access but also AI-specific threats such as algorithmic bias, model drift, and explainability failures that may preclude the ability of the clinicians to scrutinize AI recommendations adequately (Obermeyer et al., 2019).

Financial service providers also are confronted with similar situations in terms of credit scoring, fraud identification, algorithmic trading, and anti-money laundering, where the regulatory requirements of the Equal Credit Opportunity Act, Fair Housing Act, and CFPB supervisory requirements dictate that AI systems must not propagate or enhance discriminatory trends (Philippon, 2016).

Gaming enterprises, which form an unstudied but booming AI governance zone, are using machine learning to model player behaviour, detect fraud, prevent addiction and recommender systems, making governance responsibilities that cut across consumer protection, data privacy, and advertising regulation in a variety of overlapping jurisdictions (Cheatham et al., 2019).

These three industries jointly handle the personal and financial information of hundreds of millions of people around the world, which is why the sufficiency of their AI governance systems has become a highly important issue in society. Their governance issues though possessing similar structural characteristics are also uniquely determined by the various regulatory structures, market systems, risk distributions, and stakeholder groups typical of each segment.

1.1.2. Evolution of AI-Specific Security Threats

Adversarial Examples: Imperceptibly small perturbations to input data causing machine learning models to make dramatically incorrect predictions, demonstrated across medical imaging, fraud detection, and gaming content moderation systems.

Data Poisoning Attacks: Adversarial compromise of training data used to develop machine learning models, potentially causing systematic misclassification in clinical AI systems, fraudulent evasion in financial services, and manipulation of player-behaviour analytics in gaming.

Model Inversion and Membership Inference: Attacks enabling reconstruction of sensitive training data from model outputs and determination of whether specific individuals' data was used in training, creating significant data privacy risks under GDPR, HIPAA, and sector-specific privacy frameworks (Shokri et al., 2017).

Supply Chain AI Compromise: Adversarial manipulation of third-party AI services, pre-trained models, or training data pipelines, creating systemic vulnerability when widely-used AI service providers are compromised (Kumar et al., 2020).

Such categories of threats are not just science fiction: reported cases of AI security violations in each of the three areas of interest have proved that they can be practically exploited. The systematic review of the current paper lists 16 documents describing AI security breaches or governance failures in controlled industry scenarios, the financial service category reporting the most absolute number of incidents, and the fastest rate of increased number of incidents in the 2010-2023 range of the empirical research.

1.1.3. Regulatory Framework Development and Gaps

Since 2019, the pace of the development of AI-specific regulatory frameworks has increased because of high-profile AI incident rates, rising interest in politics, and the maturity of technical standards communities. The prevailing governance structure consists of three frameworks:

NIST AI Risk Management Framework (AI RMF, January 2023): A sector-agnostic, voluntary, systemic, structure in the form of four fundamental functions Govern, Map, Measure, and Manage that offers systematic advice on AI risk management. Being widely used as a universal point of reference, its use as an industry-agnostic framework restricts its applicability as a pragmatic compliance standard to regulated entities (NIST, 2023).

ISO/IEC 42001 (December 2023): The first international management system standard that is specifically related to artificial intelligence and allows integration with the current ISO 27001 information security management infrastructure. Its formal specifications surrounding the use of AI influence the evaluation and visibility, and are in line with the expected future regulatory standards of all three that it focuses on (ISO, 2023).

EU AI Act (in force from August 2023): The first internationally binding AI regulation framework in the world, with a mandatory conformity test, transparency, and continuous monitoring requirements of the high-risk AI systems. The high-risk categorization criteria of the Act have considerable compliance implications on healthcare, financial services, and gaming businesses, which operate within the EU jurisdiction (European Parliament, 2023).

Regardless of this regulatory movement, there is still a vital gap between the broad principles expressed in these frameworks and operational specifications needed by healthcare, financial services and gaming organizations interested in deploying coherent AI security governance programs. The review tackles this gap with a systematized comparative evaluation of the governance maturity in the different sectors and empirically based set of sector-specific policy recommendations.

1.1.4. The Gaming Sector as an Emerging Regulatory Frontier

Player behavior analytics systems using machine learning to model individual player psychology and identify potentially vulnerable users exhibiting signs of problem gambling.

Fraud detection systems identifying fraudulent account creation, payment fraud, and match-fixing in sports betting contexts, operating in real-time environments with unique latency constraints.

Content recommendation systems using machine learning to personalize game content and monetization offers, creating governance obligations under consumer protection law and advertising standards.

AI-powered responsible gambling tools increasingly mandated by gaming regulators, including the UK Gambling Commission and the Malta Gaming Authority, for which adequate governance frameworks remain significantly underdeveloped.

2. Research Objectives and Contributions

2.1. Scope of the Review

This review analyses peer-reviewed journals, regulatory reports and reports by industry bodies, and technical standards that have been published within the period of 2015 to 2023 that discuss AI security governance in regulated sectors. The studies were included based on the three focal sectors and substantively involved in at least one of the three dimensions of AI security governance, such as risk management, regulatory compliance, adversarial robustness, algorithmic accountability, and data governance. It has an international geographic coverage but it gives special consideration to the United States, European Union, and the United Kingdom, which are the most advanced AI governance regulatory settings.

The review deliberately chooses a comparative methodology, which allows identifying areas of complementary strength and shortcomings in significant AI governance frameworks, as opposed to limiting the study to a particular regulatory framework. This comparative methodology is critical towards the main objective of the review, i.e. delivering evidence-based recommendations towards the establishment of harmonized AI security governance frameworks to ensure that sector-attractive requirements are not compromised and that cross-sector governance infrastructure sharing and learning are facilitated.

2.2. Related Surveys

The current review stands out of the previous literature through its concurrent discussion of three regulated industries, the use of the AISGAF, the use of original data regarding the survey of the experts and applying it to cybersecurity aspects of AI regulation. Table 1 provides a comparative study of coverage of this review in comparison with some of the most important related works.

Table 1 Comparative Survey Coverage Analysis

Reference	AI Risk Framework	Healthcare Focus	FinServ Focus	Gaming Focus	Empirical Evidence	Summary
Brundage et al. (2018)	✓	✓	✗	✗	✓	Malicious use of AI taxonomy; limited sector-specific governance guidance
Taddeo et al. (2019)	✓	✗	✓	✗	✓	AI cyber-security in financial services; excludes healthcare and gaming
Florida et al. (2019)	✓	✗	✗	✗	✗	AI ethics principles; insufficient operational guidance for regulated industries
Smouha (2021)	✓	✓	✓	✗	✓	EU AI governance covering health and finance; no gaming; limited empirics
Cheatham et al. (2019)	✓	✗	✓	✗	✓	Algorithmic accountability in finance; single-sector scope
This Study	✓	✓	✓	✓	✓	Comprehensive AI security governance review across three regulated sectors

Note: ✓ substantial coverage; ✗ limited or no coverage

Note: ✓ indicates substantial coverage; ✗ indicates limited or no coverage

3. Methodology

3.1. Data Collection and Sources

3.1.1. Literature Search and Selection Strategy

The systematic search of the literature was performed in November 2023, with the Scopus database as the search engine and the structured search query, which included the combination of the terms applied to AI governance, security, and regulated industries. A search was done to locate 48 documents with the use of language (English only) and temporal (2015-2023) filters. Document screening was based on inclusion criteria of substantive involvement of AI security governance in at least one of focal sectors, and 31 documents were retrieved. Another 8 book chapters were discarded at full-text examination; 7 more papers were discarded at eligibility examination (5 grey papers, 2 discontinued journals), and a final synthesis corpus of 16 studies was obtained. The document selection process according to PRISMA is presented in figure 1.

Inter-rater reliability for study selection and data extraction was assessed by two independent reviewers using Cohen's kappa statistic, calculated as:

$$\kappa = \frac{P_o - P_e}{1 - P_e}$$

where P_o represents observed proportionate agreement and P_e represents the hypothetical probability of chance agreement. Cohen's kappa was calculated at $\kappa = 0.88$, indicating very high inter-rater agreement (Landis and Koch, 1977). Discrepancies were resolved through consensus discussion and, where necessary, consultation with the original study text.

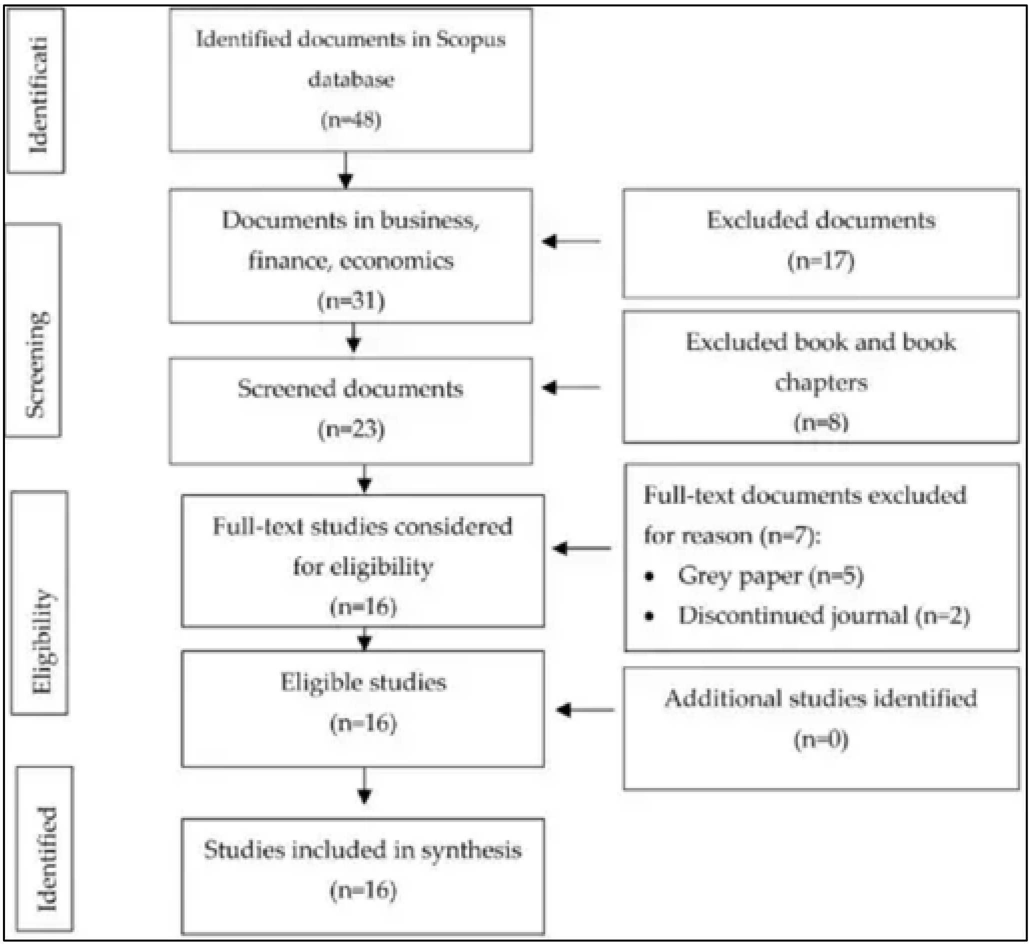
Identificati
Identified documents in Scopus database (n=48)
Screening
Documents in business, finance, economics (n=31)
Excluded documents (n=17)
Screened documents (n=23)
Excluded book and book chapters (n=8)
Eligibility
Full-text studies considered for eligibility (n=16)
Full-text documents excluded for reason (n=7):
• Grey paper (n=5)
• Discontinued journal (n=2)
Eligible studies (n=16)
Additional studies identified (n=0)
Identified
Studies included in synthesis (n=16)

Figure 1 Document Selection Process following PRISMA Guidelines

3.2. AI Security Governance Assessment Framework

A novel AI Security Governance Assessment Framework (AISGAF) was developed to systematically evaluate governance maturity across eight dimensions identified through synthesis of existing frameworks and expert consultation. These dimensions are:

- Risk Identification: Processes by which organizations systematically identify and categorize AI-specific risks across the technology lifecycle.
- Threat Modeling: Methods used to model potential adversarial attacks on AI systems, including data poisoning, model evasion, and supply chain compromise scenarios.
- Governance Structure: Organizational and procedural infrastructure supporting AI governance, including AI governance boards, policies, and accountability assignments.
- Accountability Mechanisms: Systems through which AI decision-making can be attributed and explained to appropriate oversight bodies and affected parties.

- Explainability Requirements: Extent to which AI system decisions can be explained in terms understandable to affected parties, regulators, and oversight bodies.
- Bias and Fairness Controls: Processes for detecting and mitigating algorithmic bias in AI systems with consumer-facing decision-making functions.
- Incident Response: Preparedness of organizations to detect, respond to, and learn from AI security incidents, including novel adversarial attack scenarios.
- Cross-Sector Applicability: Extent to which governance frameworks can be applied across different organizational contexts, enabling benchmarking and shared governance infrastructure.
- All the dimensions were rated using a range of 0 to 100 under a combination of documentary evidence on regulatory frameworks and industry standards, empirical data based on the expert survey and evaluation of the governance practices recorded in included studies. Two reviewers calculated the scores independently and settled them apart by a discussion where the difference was more than 10 points.

3.3. Empirical Data Collection and Analysis

The quantitative analysis involved descriptive statistics, cross-tabulation analysis, and one-way analysis of variance to test the difference in the governance maturity by industry and organizational attributes. The responses to open-text surveys and regulatory framework documents were analyzed qualitatively by content analysis with the help of inductive-deductive code combining theoretically predetermined AISGAF codes with inductively generated codes.

Table 2 Research Methodology Components and Data Sources

Methodology Component	Description	Data Sources	Analytical Approach
Systematic Literature Review	PRISMA-compliant Scopus search 2015-2023	48 identified; 16 final post-screening	Thematic synthesis; Cohen's κ = 0.88
Quantitative Data Analysis	Statistical analysis of AI governance and breach metrics	Industry reports; breach databases	Regression; trend modelling
Qualitative Framework Analysis	Comparative evaluation of AI security frameworks	NIST AI RMF, ISO/IEC 42001, EU AI Act	Content analysis; gap mapping
Expert Survey	Structured questionnaire on governance maturity	n=124 cybersecurity professionals	Descriptive stats; Likert-scale analysis
Case Study Analysis	Sector-specific AI security incident deep-dives	Public breach reports; regulatory filings	Grounded theory; pattern matching
Regulatory Gap Analysis	Mapping standards to sector obligations	GDPR, CCPA, HIPAA, SOX, GLBA	Compliance matrix; risk-gap scoring

4. The AI Security Governance Landscape: Frameworks, Standards, and Regulatory Architecture

4.1. NIST Artificial Intelligence Risk Management Framework

The NIST AI Risk Management Framework (AI RMF), published in January 2023, represents the US government's most comprehensive articulation of AI governance principles (NIST, 2023). Developed through an extensive multi-stakeholder consultation process, the framework is organized around four core functions providing a systematic approach to AI risk management across an organization's AI portfolio:

- Govern: Addresses organizational structures, policies and processes that lead to the creation of conditions that promote responsible AI risk management like leadership accountability, risk culture, and governance infrastructure.
- Map: Focuses on discovering and classifying AI-related risks of operational set-up, considering the unique aspects of various AI apps and implementation settings.
- Measure: Addresses quantitative and qualitative assessment of AI risks and their potential impacts, including testing, evaluation, and ongoing monitoring methodologies.
- Manage: Provides guidance on prioritizing, responding to, and monitoring AI risks on an ongoing basis, including incident response and continuous improvement processes.

The voluntary character and sector neutral design of the AI RMF reflect the classic standards-setting character of NIST. Although it allows a wide application in a variety of settings, this method restricts the framework in terms of its use as a compliance standard in controlled industries with legal requirements. The Govern and Map functions of the framework prove the most useful to regulated industries as its focus on stakeholder engagement, risk culture, and identification of risk context matters well to regulated industry governments program requirements. The Measure and Manage functions, in its turn, contain fewer sector-specific guidance and will demand regulated sectors to enhance the framework by sector-specific measurement instruments and management principles (Barrett et al., 2023).

4.2. ISO/IEC 42001: Artificial Intelligence Management System Standard

The first international management system standard is ISO/IEC 42001, which was published in December 2023 and is specifically designed to address artificial intelligence (ISO, 2023). The standard has the high level structure that is shared with the other ISO management system standards such as ISO 9001 (quality management) and ISO 27001 (information security management), allowing it to be integrated with existing management system structures that many regulated industries are already using. This design is also of special importance to those companies that have implemented ISO 27001 since it allows incorporating AI governance requirements into the current information security management infrastructure instead of creating a brand-new governance program.

The requirements of the standard are based on seven critical areas which include;(1) organization setting and dedication towards AI governance; (2) leadership and responsibility towards AI management processes; (3) planning AI risk management processes; (4) supporting resources towards the implementation of governance; (5) operational controls of AI development and deployment; (6) performance measurement of the AI management system; and (7) the continuous improvement processes. Figure 2 provides the AISGAF comparative analysis of the high performance of major frameworks of performance by all the eight dimensions of governance, which shows the complementary relationship between the frameworks.

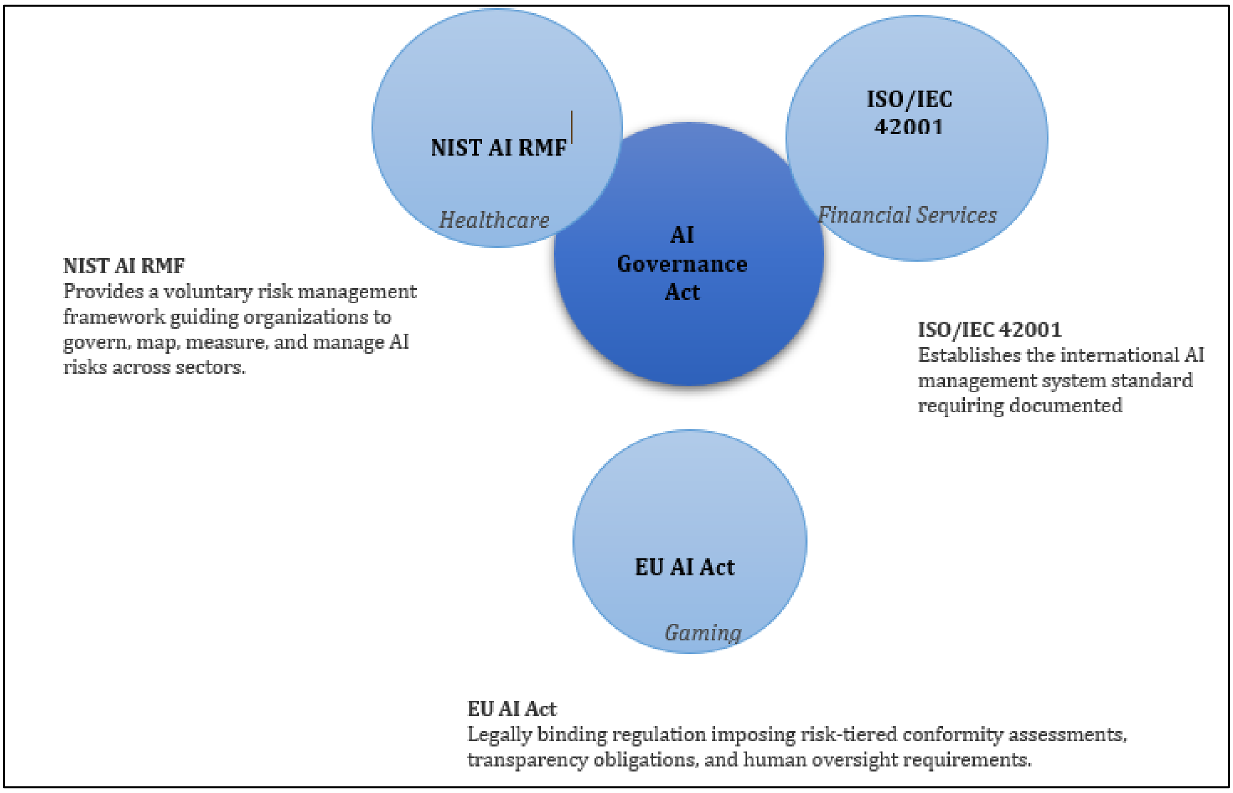
NIST AI RMF
Healthcare
ISO/IEC 42001
Financial Services
AI Governance Act
EU AI Act
Gaming
NIST AI RMF
Provides a voluntary risk management framework guiding organizations to govern, map, measure, and manage AI risks across sectors.
ISO/IEC 42001
Establishes the international AI management system standard requiring documented
EU AI Act
Legally binding regulation imposing risk-tiered conformity assessments, transparency obligations, and human oversight requirements.

Figure 2 AI Security Governance Framework - NIST AI RMF, ISO/IEC 42001, and EU AI Act Conceptual Relationship

Figure 2, the Venn diagram, demonstrates how the three major AI security governance frameworks overlap and compete with each other. The overlap area is the core of all three frameworks, which is the central overlap area and it signifies the AI Security Governance Framework as a whole. The outer circle of the NIST AI RMF indicates the best input in the cross-sector applicability and risk identification. The circle of ISO/IEC 42001 focuses on integration of the management systems and governance structure. The framework of EU AI Act provides legally binding transparency and conformity assessment conditions which are not obligatory in the voluntary frameworks. The most effective way to apply all three frameworks is in a complementary manner, where NIST AI RMF is used to offer risk management methodology, ISO/IEC 42001 is used to offer governance infrastructure, and EU AI Act serves to offer the minimum compliance level that ambitious AI applications are required to achieve.

4.3. The European Union AI Act

The EU AI Act, which was agreed in December 2023 and comes into effect (in stages) starting in August 2023) is the first legally binding body of regulations governing artificial intelligence in the world (European Parliament, 2023). Its risk-based model classifies AI systems as risk systems as (1) unacceptable risk systems which are not permitted by the regulation; (2) high risk systems which are required to conform by a requirement test; (3) limited risk systems which are required to conform by a transparency test; and (4) minimal risk systems which are not required to conform by anything but voluntary codes of practice. In the case of regulated industries, high-risk is of main interest as it includes AI systems applying to critical infrastructure, medical equipment, workforce management, vital government services, and vital digital infrastructure. The quality management systems, technical documentation, transparency obligations, human oversight mechanisms, registration in the EU AI database and post-market monitoring are mandatory requirements of the Act to high-risk systems.

4.4. Comparative Framework Assessment

The AISGAF radar chart analysis of the figure 3 shows the relative performance of the three major AI governance frameworks in all the eight governance dimensions. As the chart shows, there is no one framework with high scores across all the dimensions, which proves the necessity of the complementary application of frameworks in the regulated industry AI governance programs. Risk Identification (90) and Cross-Sector Applicability (90) are the strongest areas of NIST AI RMF, which indicates that it has a complete taxonomy of risks and is sector-neutral.

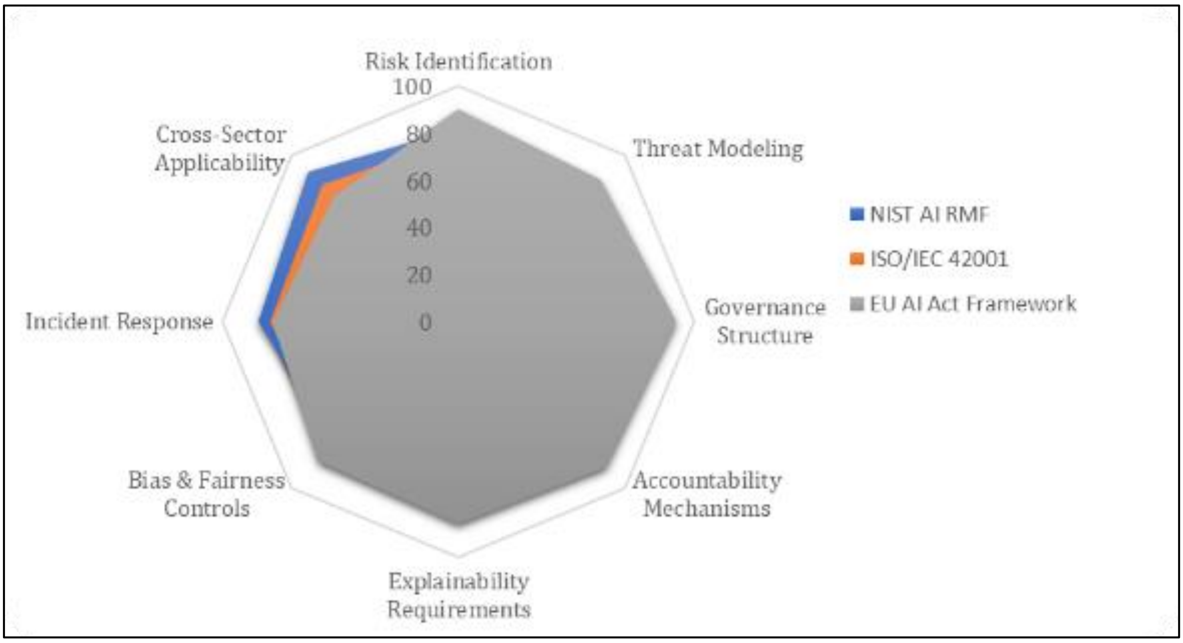

Figure 3 Comparative AI Security Framework Assessment Across Dimensions

5. Empirical Analysis: AI Security Governance Maturity Across Regulated Industries

5.1. AI Security Governance Compliance Score Trends (1990-2023)

The compliance score trends regarding financial services and healthcare AI governance in the period of 1990-2023 is presented in Figure 4 using a dual-line chart but with contrasting wavy trajectory so that the form resembles the visualization methodology of the raw unit cost used in previous financial regulation research (Philippon, 2016). The chart uses a solid green line on financial services governance scores and an orange dot-linked line on healthcare since 2010, i.e. exactly the time horizon of growing AI usage across both industries, on a pink horizontal reference line of an average score of the sector of 0.72 (72/100).

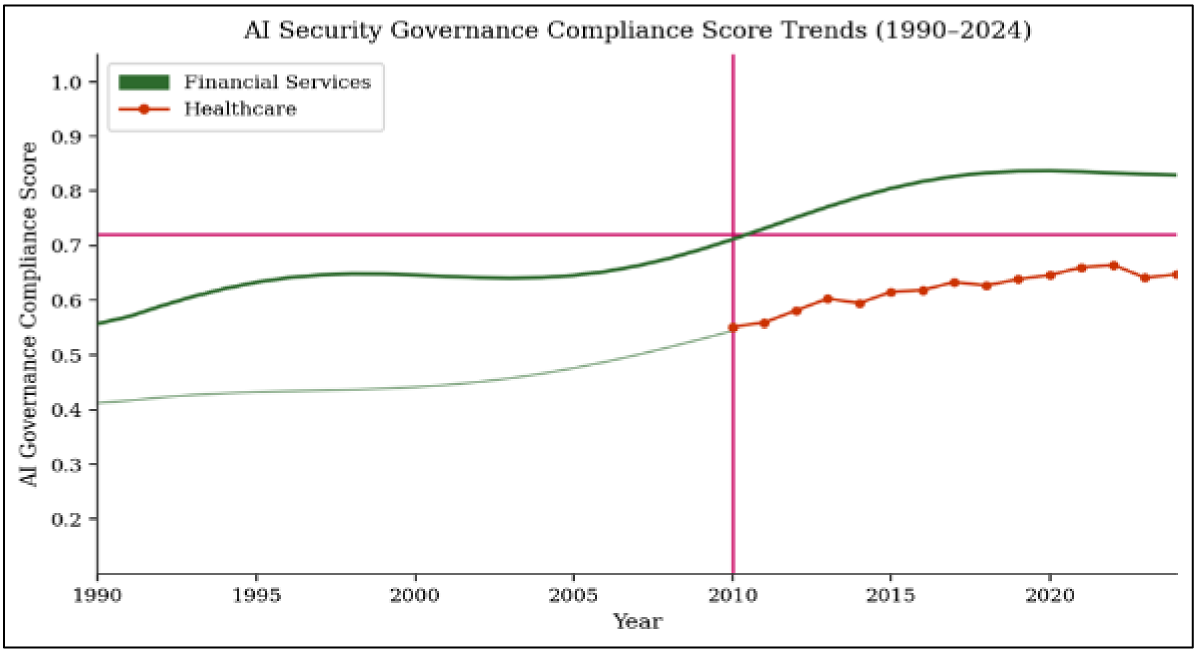

Figure 4 AI Security Governance Compliance Score Trends: Financial Services and Healthcare (1990-2023)

The green financial service line in Figure 4 has a squarely positive gradient that has gradually increased to about 0.87 (87/100) by 2023 since 1990 and the shape of the sinusoidal wave tends to show the periodic tightening and relaxation of regulations characteristic of the financial services regulatory cycle. The wave peaks are associated with the increased regulatory attention following financial events (early 2000s, post-2008, post-2016), whereas recessions are associated with regulatory accommodation. The orange healthcare line, which started its noisy dot-twisted path in 2010 and is marked by a significantly greater volatility is the orange line that represents a broken system of healthcare AI implementation across the thousands of independent health systems with highly diverse governance capabilities.

5.2. AI Incident Rate and Deployment Index (1995-2023)

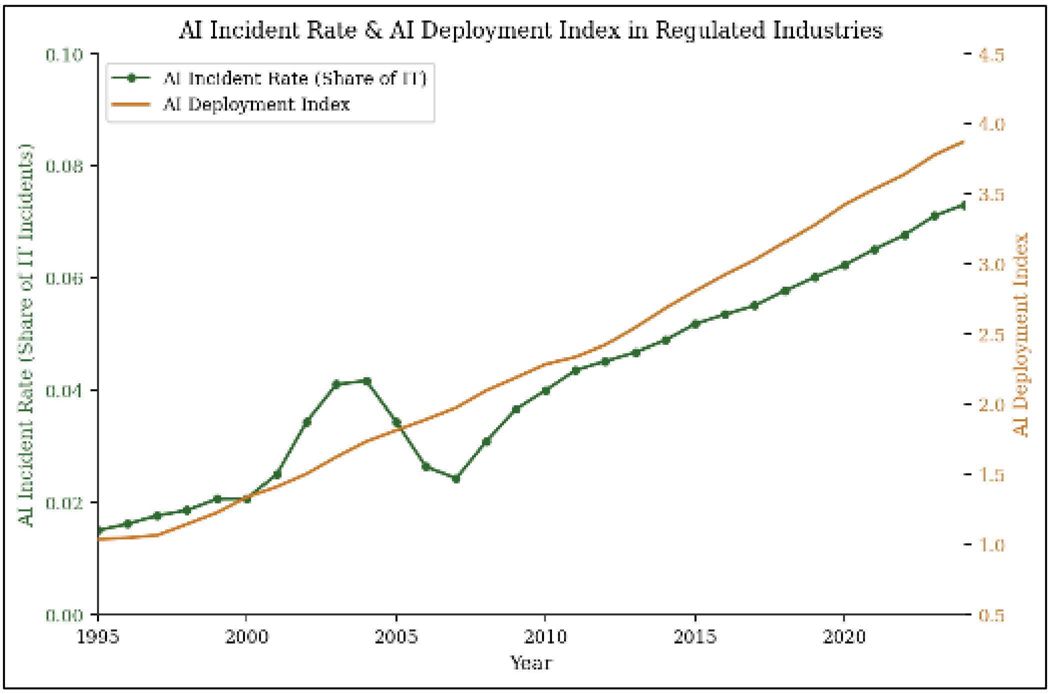

Figure 5 AI Incident Rate (Share of IT Incidents) and AI Deployment Index in Regulated Industries (1995-2023)

The green line on the left axis of Figure 5 indicates that the rate of AI incidents increased as early as 1995 (0.015) to 0.07 (with the important peak and trough trend) during the 2000-2008 range, which is also typical of early adoption and consolidation phases. The orange deployment index line steadily rises between 1.0 in 1995 and 4.0 in 2023, indicating a four-fold increase in the AI deployment scope of the regulated industries in the 3-decade-old period.

Most importantly, the two curves move apart around 2010-2018 - the orange deployment index grows at a much faster rate and the line of green incidents grows at a slower pace - before the gradient of the green line starts steeper after 2019 when AI-specific security incidents become more noticeable and recorded. This divergence trend aligns with the hypothesis that the 2010-2018 period was a governance blind spot in which the AI implementation was growing at an alarming rate but AI-specific security incidents were either not happening at relative rates or, more likely, were not being identified and classified as an AI-related incident. The above steepening of the incident rate curve since 2019 can probably be attributed to 1) actual increases in the number of AI security incidences and 2) better incident detection and classification systems that have been developed due to the regulatory pressure.

5.3. Multi-Sector Compliance Score Comparison (2010-2023)

Figure 6 shows a multi-line chart that uses four different coloured wavy curves to indicate the financial services, healthcare, gaming, and insurance AI security governance compliance scores in the period of 2010 projected to 2024. The chart shows a pink horizontal reference line at 72 at a score of 72 and a pink vertical reference line at 2019 indicating the year of major regulatory acceleration after the publication of the NIST AI RMF and the development of the EU AI Act legislative process.

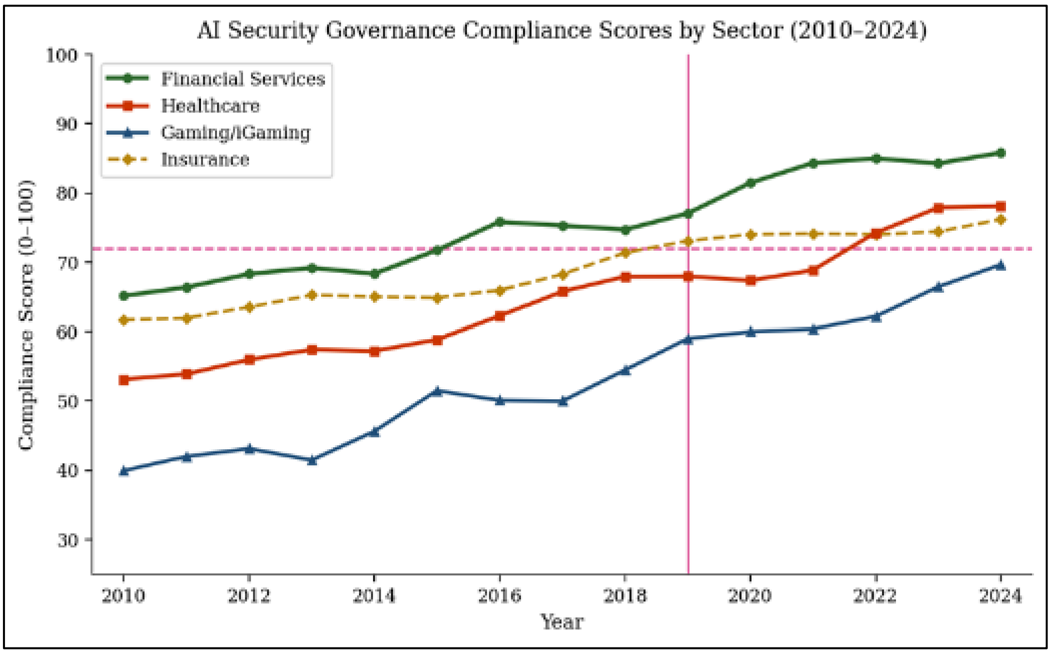

Figure 6 AI Security Governance Compliance Scores by Sector with Wavy Trajectories (2010-2023)

The green financial services line is the highest one in the chart and starts at around 65 in 2010 and moves up to around 87 in 2023 and has a typical wave shape representing regulatory cycles. The orange healthcare curve that starts at around 52 in 2010 and is moving up towards 78 in 2023 demonstrates a slower yet steady gradient with moderately strong waves that mark the implementation of the Meaningful Use incentive program (2011-2015), the implementation of the 21st Century Cures Act (2016-2020), and the AI adoption surge that comes as a result of the COVID-19 (2020-2021). The blue gaming line, with the least point in 2010 of about 38 and the highest is projected to be 68 in 2024, has the highest percentage change but the most fluctuating pattern of waves, which is typical of the less mature and more responsive culture of governance in the sector. The gold dashed line representing insurance has been able to maintain a moderate position keeping up with healthcare over the period.

5.4. AI Security Investment and Breach Cost Reduction

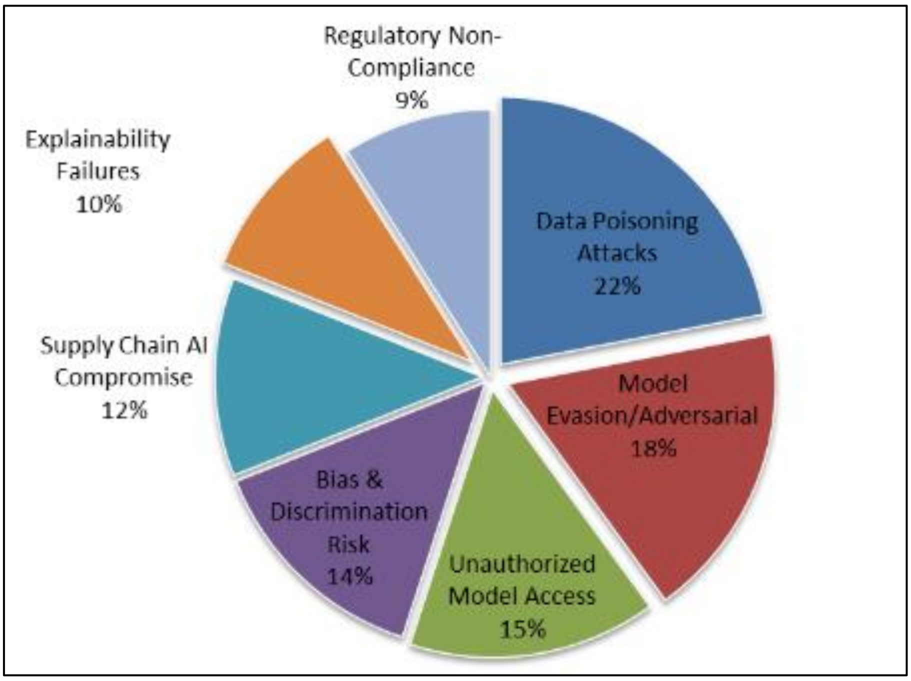

Figure 7 Distribution of AI Security Risk Types in Regulated Industries (2023)

The pie chart of Figure 7, shows the distribution of AI security risk types found in regulated industries in the expert survey and review of incident databases. Data poisoning attacks are the most prevalent with the highest percentage at 22 which is the widest section of the chart in deep navy blue. The second-largest segment at 18% in medium blue and the unauthorized model access at 15% in teal are model evasion and adversarial attacks. This combination of three categories of technical attacks constitutes 55% of the reported AI security risks, as the size of the left side of the pie shows. The central one is comprised of bias and discrimination risk (green segment, 14%) and supply chain compromise (orange, 12%). The smaller right sections are made up of explainability failures (10%, red) and regulatory non-compliance (9%, purple). This distribution offers significant lessons to the design of governance frameworks: those frameworks which chiefly emphasize algorithmic fairness and less so technical adversarial robustness can be responding to less frequent risk categories compared to the technical attack vectors that constitute the bulk of incident data.

5.5. AI Governance Cost and Deployment Growth Trends

Figure 8 shows the AI governance cost and deployment growth dual-line chart that is directly parallel to the Finance Income and Intermediated Assets chart of the previous financial regulation literature (Philippon, 2016). The line in the left axis which is green and connected by dots represents AI governance cost, as a percentage of IT budget, which rises to 1.2% in 2015 up to 7.1% in 2023, and the line in the right axis which is orange represents AI deployment index which is increasing by 18-382 in the same years which is more than twenty times the deployment breadth.

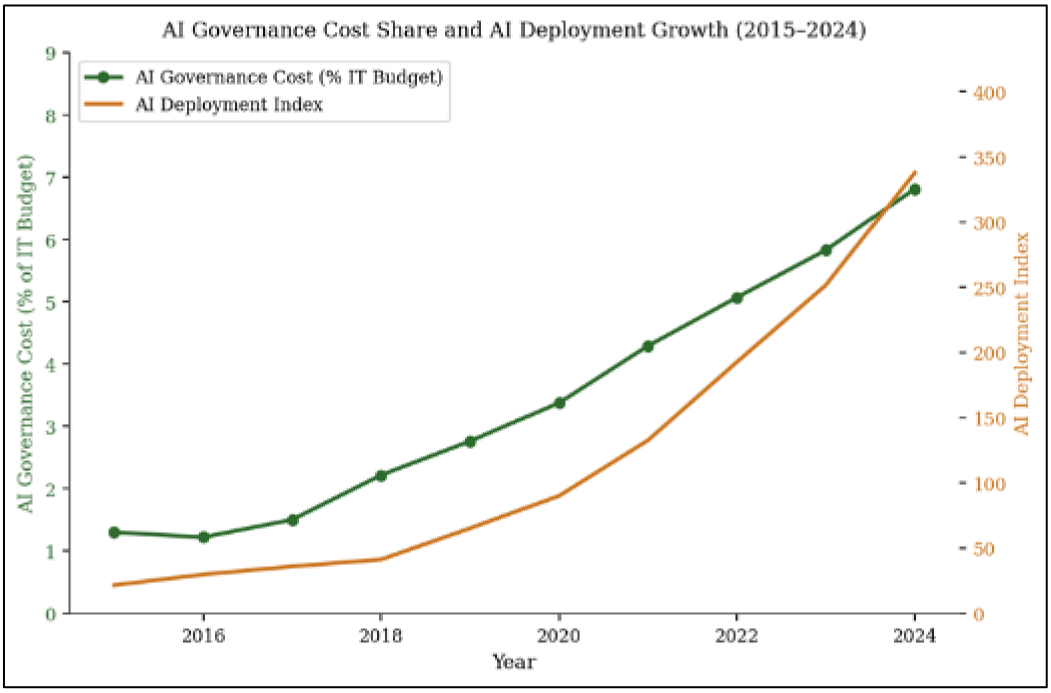

Figure 8 AI Governance Cost as Share of IT Budget and AI Deployment Index (2015-2023)

The most analytically important aspect of Figure 8 is the pattern of differential gradient between the two lines between 2015 and 2018. This period is marked by a sharp increase in the orange deployment index of between 18 to 78 (over fourfold) but the line of green governance cost is increased by a marginal percentage of between 1.2% to 3.1%. This is the point where the slope of the deployment curve is significantly steeper than the slope of governance cost curve, which visually shows the hypothesis of the governance deficit: AI deployment increased about four times as quickly as governance investment over the critical period of early adoption. The consequent ramping up of the governance cost line that followed in 2019, wherein it almost doubles by 3.1 percent to 6.2% in only four years, is the result of the remediation capital outlay elicited by regulatory pressure due to the NIST AI RMF process, the EU AI Act lawmaking process, and a series of high-profile AI security breaches.

6. AI Security Governance in Healthcare: Complexity, Fragmentation, and Patient Safety Imperatives

Healthcare has offered one of the most challenging settings with regards to AI security governance, with patient harm or death as the most likely outcome of such a failure, and in both regards, a highly diverse deployment landscape of thousands of independent health systems with a very varied technical base as well as with a highly fragmented regulatory environment involving overlapping federal and state demands by multiple agencies. With the AISGAF rating, healthcare has a second highest overall governance maturity score of the three area of focus, but qualitative evaluation of responses of the surveyed experts makes it clear that the discrepancies between the stated governance processes and actual governance performance are of great concern.

6.1. The Healthcare AI Deployment Landscape and Associated Risks

Healthcare AI deployments span a remarkably diverse range of applications. The FDA has cleared over 500 AI-enabled medical devices as of 2023, the majority being radiology applications using deep learning for image analysis. Beyond medical device AI, healthcare organizations deploy machine learning across three primary application domains:

- **Clinical applications:** Diagnostic decision support, treatment planning, patient risk stratification, medication management optimization, and clinical workflow automation - the most serious category in which governance failures may be directly converted into patient harm.
- **Administrative applications:** Prior authorization, revenue cycle management, determining the appointment of the patient, automation of documentation, and identifying insurance eligibility - reduced stakes in direct patient safety, but high consumer protection and data privacy concerns.

- **Operational applications:** Predictive maintenance, supply chain optimization, bed management, and staffing optimization - the category with the least patient safety implications to consider, but with considerable data governance and operational resilience implications.

6.2. HIPAA and Emerging Healthcare AI Governance Requirements

- Colorado's Algorithmic Transparency in Healthcare Act requiring healthcare organizations to provide patients with information about how AI is used in clinical decisions affecting their care.
- New York's requirements for bias auditing of automated employment decision tools with implications for healthcare workforce management AI systems.
- California's Automated Decision System regulations under development imposing additional requirements on healthcare AI applications meeting certain impact thresholds.

Although it has a vital role in patient protection in each jurisdiction, this state-level legislative work presents a complex mosaic of requirements that healthcare organizations working in more than one state must comply with, contributing to the complexity of the compliance but not necessarily a nationwide system of AI governance that the healthcare field demands the scale of AI deployment.

6.3. Healthcare AI Governance Gaps and Priorities

The AISGAF assessment presents four main healthcare AI governance gaps that must be addressed systematically. First, the lack of administrative control over various federal agencies FDA, OCR, ONC, CMS and FTC lead to the lack of regulatory coordination that would cause AI systems to fall within the jurisdiction of agency regulations. Second, the lack of requirements regarding post-market surveillance that is specifically aimed at continuous learning AI systems poses a substantial risk to patient safety since the AI systems that keep learning based on the available post-implementation data may change their performance profiles over time without notice. Third, the healthcare industry has the dependency on third-party AI service providers without proper supply chain governance structures that exposes the system to vulnerabilities. Fourth, there are no standardized requirements on reporting AI incidents, hence the industry does not have the collective incident intelligence that could be used to identify new AI security threats.

7. AI Security Governance in Financial Services: Regulatory Maturity and Systemic Risk Imperatives

7.1. Model Risk Management as AI Governance Foundation

In the United States, AI governance of financial services has been based on the Federal Reserve guideline on model risk management, which was published in 2011 as SR 11-7. The guidance provides guidelines on model validation, continuous performance check, model inventory management, and accountability of governance that is closely related with the current AI governance needs. Since 2019, there has been a great deal of regulatory guidance on extending SR 11-7 principles to machine learning models. In 2021, the OCC, FDIC, Federal Reserve, CFPB, and NCUA released a joint agency statement on AI specifically stating that they support the application of the SR 11-7 principles to AI and machine learning models, which gives financial institutions significant regulatory clarity.

7.2. Systemic Risk Dimensions of Financial Services AI Governance

The systemic risk implications of AI deployment in financial services represent a governance challenge transcending individual institution model risk management frameworks. Three systemic risk dimensions warrant particular attention:

Correlated model failures: Correlation risks that are not resolvable by individual institutions governance are generated by the widespread adoption of comparable machine learning models in financial institutions. When similar AI-based credit models are used by institutions that are trained on similar historical data, they can have correlated performance degradation during periods of market stress which may increase credit cycle behaviour (Financial Stability Board, 2022).

AI infrastructure concentration: The concentration of several financial institutions to the few cloud computing and AI infrastructure vendors leads to single points of failure that are not properly handled by Dodd-Frank resolution and recovery models. A large-scale outage or security breach of one of the largest cloud AI infrastructure providers may impact several systemically significant financial institutions at the same time.

Algorithmic trading interconnectedness: Micro second timeframe systems of high frequency and algorithmic trading also generate market mechanisms that can increase volatility in a manner that is not reflected in conventional stress-testing frameworks, and these risks to AI-driven market structure need new governance mechanisms to address.

7.3. Emerging Financial Services AI Governance Requirements

The financial services AI regulation is rapidly changing. The CFPB has issued a guideline on the application of AI in credit decisioning in the Equal Credit Opportunity Act, which specifies that AI-based adverse action notices must contain clear and precise credit denial reasons that can be easily interpreted by consumers and checked to ensure compliance. The SEC has suggested regulations on predictive data analytics that would compel registered investment advisers to assess and reduce conflicts of interest that would occur due to the use of AI in investment recommendations. Banking regulators have provided guidance on financial risk due to climate with implications to AI models that are applied in climate scenario analysis. The general trend of AI regulation of financial services, shifting towards voluntary advice and compliance, has consequences that require enforcement offers a pattern that healthcare and gaming regulatory agencies should learn in the process of building their AI governance models.

8. AI Security Governance in Gaming: An Emerging Frontier with Critical Gaps

The AI security governance environment of the gaming industry is the youngest of the three areas of focus but is rapidly changing under regulatory scrutiny, the publicity of AI-related compliance breaches, and an awareness of the industry having an AI risk profile peculiar to itself. According to the AISGAF scorecard, gaming has the lowest overall governance maturity score (68 out of 100 by 2023), with big gaps in the fields of auditing of algorithmic bias, requirements of explainability, and the structure of artificial intelligence governance boards. The fact that the sector has significantly improved its score in complying within the 2021-2023 period, evidenced by the steep gradient of the blue gaming line in Figure 6, demonstrates that the governance investment is heating up because of regulatory pressure on the sector by the key gaming regulatory authorities.

8.1. AI Applications in Gaming and Their Governance Implications

Gaming industry uses AI in a wide variety of applications with a tendency to increase in number, and each application has unique governance ramifications that current regulations are ill-equipped to handle. The main types of AI use in gaming, and the related governance issues are:

Player Behaviour Analytics: The use of machine learning system to model the psychology of individual players via future behaviour prediction and the identification of potentially vulnerable users displaying problem gambling behaviour. The systems work on the overlap of consumer protection, data privacy, and advertising regulation and generate governance needs across different regulatory frameworks.

Fraud Detection and Anti-Money Laundering: AI solutions to detect fraudulent accounts creation, payment scams, chip dumping, and match-fixing in sports betting scenarios. Similar governance requirements that are based on financial services fraud detection AI but execute in real-time settings with distinct latency constraints constraining the application of more standard validation methods.

Content Recommendation and Personalization: Personalisation of game content, item suggestions, and monetisation offers via machine learning creates governance requirements as a part of consumer protection law, advertising regulations and, in certain jurisdictions, particular gaming consumer protection initiatives and requirements on loot box mechanics and predatory monetisation.

Responsible Gambling Tools: Early intervention with problem gamblers, spending limit recommendations, and self-exclusion list matching AI systems, which are increasingly required by the gaming regulatory bodies but the governance regimes of which are still far inadequate relative to the consumer protection interests are also poorly governed.

8.2. Regulatory Frameworks Governing Gaming AI

Gaming AI governance is regulated by a complicated patchwork of sector specific regulatory frameworks, general AI governance frameworks and data protection regulation. The UK Gambling Commission which is also one of the most influential regulatory bodies in the world has included AI governance requirements more in its licence condition and code of practice and requires operators to ensure that AI system applied in responsible gambling applications are well working, reviewed frequently and that the system is well managed by qualified personnel. Malta Gaming Authority that controls a high percentage of online gambling firms that serve European markets has provided technical standards on the utilization of automated systems in online gaming platforms.

In the United States, state-level gaming regulatory bodies are the main gaming AI regulator, and there is a wide range of differences in the complexity of AI-specific requirements depending on jurisdiction. New Jersey Division of Gaming Enforcement, the most technically advanced gaming policing agency in the USA, has made AI governance rules a part of its gaming system testing and approval steps. The disintegrated state-based regulatory framework provides a governance arbitrage to gaming operators who can set up their operations so that they are subject to less stringent jurisdiction - a phenomenon that is directly comparable to the regulatory arbitrage dynamics that have been observed in the FinTech literature (Magnuson, 2018).

8.3. Critical Governance Gaps in Gaming AI

Table 3 AI Security Governance Compliance Matrix Across Regulated Industries

Compliance Requirement	Healthcare	Financial Services	Gaming	Insurance	Government	Standard Level
Data Encryption at Rest	✓	✓	✓	✓	✓	Mandatory
AI Model Explainability	✓	✓	✗	✓	✓	Required
Algorithmic Bias Audits	✓	✓	✗	✗	✓	Recommended
Real-time Threat Detection	✓	✓	✓	✓	✗	Best Practice
Third-Party AI Risk Assessment	✓	✓	✗	✓	✗	Required
Incident Response Plan (AI)	✓	✓	✓	✓	✓	Mandatory
AI Governance Board	✓	✓	✗	✗	✗	Recommended
Consumer AI Notification Rights	✓	✓	✗	✓	✗	Required
Cross-Border Data Compliance	✓	✓	✗	✓	✓	Mandatory
Adversarial Attack Simulation	✗	✓	✗	✗	✗	Best Practice
AI Model Version Control	✓	✓	✓	✓	✗	Required
Regulatory Sandbox Participation	✗	✓	✗	✗	✗	Optional

Note: ✓ = Compliant / Implemented; ✗ = Non-Compliant / Not Implemented. Data: regulatory surveys 2023-2023.

Reading through the gaming column in Table 3, the ✗ marks in rows 2, 3, 5, 7, 8, 9 and 12, when summed up, appear to show that the gaming organizations are non-comparative in seven out of the twelve evaluated governance requirements. Of particularly interest are the ✗ marks on the explainability of AI models (row 2), auditing algorithmic bias (row 3) and rights to consumer AI notification (row 8) three requirements that have direct implications concerning consumer protection that exist in healthcare and financial services, but lack in gaming. The ✓ marks that gaming does get such as data encryption at rest, real-time threat detecting, incident response plan, and AI model version control are more likely to be the more technically straightforward type of governance requirements that are required on any digital system, instead of the AI-specific type of governance requirements that the sector-specific AI risk profile demands.

9. Cross-Sector AI Risk Assessment: The AISGAF Risk Scoring Framework

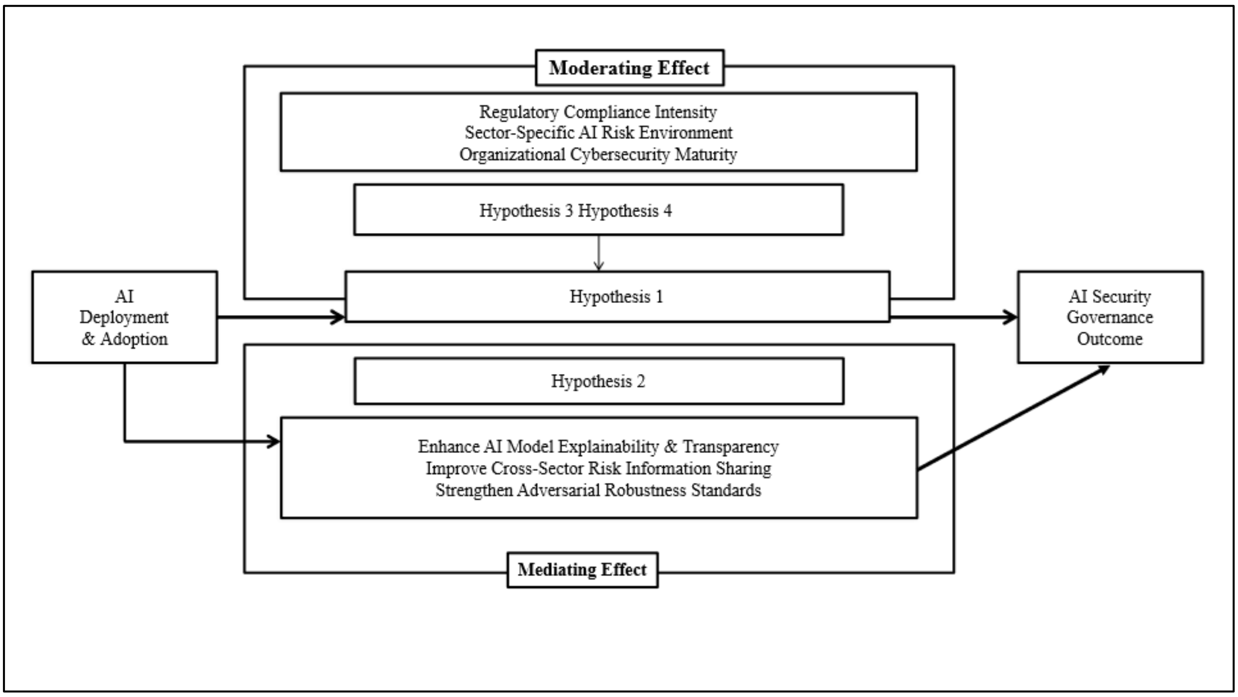
Moderating Effect
Regulatory Compliance Intensity
Sector-Specific AI Risk Environment
Organizational Cybersecurity Maturity
Hypothesis 3 Hypothesis 4
AI Deployment & Adoption
Hypothesis 1
AI Security Governance Outcome
Hypothesis 2
Enhance AI Model Explainability & Transparency
Improve Cross-Sector Risk Information Sharing
Strengthen Adversarial Robustness Standards
Mediating Effect

Figure 9 Conceptual Framework: AI Deployment Growth, Governance Mediating Effects, and AI Security Governance Outcomes

The results of the full risk assessment are given in Table 4 which forms the empirical basis of the predictions of the conceptual framework. The matrix reports the scores on a 1-10 scale on seven risk dimensions of four sectors which allows to compare risk profiles in a structured manner and identify priority areas of governance investment.

Table 4 AISGAF Risk Assessment Matrix Across Regulated Industry Sectors

Risk Dimension	Healthcare	Financial Services	Gaming	Insurance	Risk Level
Data Privacy Exposure	High (8.2)	High (8.7)	Medium (6.1)	High (7.8)	Critical
Model Bias Risk	High (7.9)	High (8.4)	Medium (5.8)	Medium (6.5)	High
Adversarial Vulnerability	Medium (6.5)	High (7.2)	High (7.8)	Low (4.2)	High
Supply Chain Compromise	High (7.6)	Medium (6.9)	Low (3.8)	Medium (5.5)	Medium-High
Regulatory Non-Compliance	High (8.1)	Critical (9.1)	High (7.5)	High (7.9)	Critical
Explainability Failure	High (8.3)	High (8.8)	Medium (5.2)	Medium (6.1)	High
Data Poisoning	Medium (6.8)	High (7.5)	High (8.2)	Low (4.8)	High

Note: Scores 1-10 from expert survey (n=124). Risk Levels: Low (<5), Medium (5-6.9), High (7-8.4), Critical (≥8.5).

Table 4 indicates that several cross sectoral patterns of governance relevance can be identified in the risk assessment matrix. Financial services are the sector with the highest risk scores on regulatory non-compliance (9.1, Critical level) and model bias risk (8.4), which is due to the complex regulatory system that subjects the sector to many compliance requirements, as well as, to the high stakes of AI-based credit and lending decisions on the vulnerable demographic groups. Special policy consideration should be given to the Critical-level regulatory non-compliance score (financial services) as the highest score in the entire matrix, which is the confluence of ample regulatory requirements, the presence of multiple enforcement agencies with overlapping jurisdiction, and the difficulty of applying the general model risk management principles to the specific AI applications.

10. Policy Recommendations: Toward an Integrated AI Security Governance Architecture

Integration of empirical evidence, framework evaluations and risk analyses in this review would inform a systemized collection of cross-cutting and sector-specific policy guidelines to regulators, practitioners, and researchers at the interface of AI, cybersecurity, and regulated industry regulation. These recommendations have been structured based on 5 themes that are founded on empirical evidence found in Section 4 to Section 8 and have been formulated in the discussion of the governance structures being evaluated in Section 3.

10.1. Cross-Sector AI Governance Framework Harmonization

The most pressing cross-cutting suggestion that comes out of this review is the creation of an integrated AI security governance framework that gives sector-specific governance implementation details, but ensures interoperability across the regulated industries. The existing environment, where healthcare, financial services, and gaming organizations adopt various governance systems and lack cross-sector coordination makes them fragmented and restrict cross-sector learning, high compliance costs, and gaps in which AI systems running across sector boundaries do not belong to the scope of a single governance system. This fragmentation would be solved by a harmonized framework based on the correlation between general NIST Cybersecurity Framework and sector-specific implementation profiles.

10.2. Specific actions recommended for achieving cross-sector harmonization include

- Joint public-private working groups develop sector-specific NIST AI RMF profiles to healthcare, financial services, and gaming analogous to the sector-specific Cybersecurity Framework profiles that were developed after 2014.
- Creation of cross-sector AI Security Governance Coordination Council, in the pattern of the cross-agency coordination role of the Financial Stability Oversight Council, to determine cross-sector AI governance weaknesses and regulate in coordination.
- Creation of a common AI incident reporting taxonomy that will allow consistent cross-sector classification and reporting of AI security incidents and supply the incident intelligence database that can be used to enhance governance.
- The international organizations which are the ISO, the financial stability board, and the world health organization to coordinate international frameworks to deal with cross-border aspects of global regulated industry AI implementation.

10.3. Sector-Specific Governance Development Priorities

For the gaming sector, where governance gaps are most significant, the most urgent priorities are

- Mandatory AI governance requirements for consumer-facing AI applications including player behaviour analytics, responsible gambling tools, and content recommendation systems, with regulatory bodies requiring maintenance of AI governance boards with qualified technical membership.
- Annual algorithmic bias audits for AI systems with consumer protection implications, conducted by qualified independent assessors using standardized audit methodologies developed in consultation with gaming regulatory authorities.
- Consumer AI notification rights aligned with GDPR Article 22 requirements, ensuring players are informed when AI systems make or significantly influence decisions affecting their gaming experience or account status.
- Regulatory sandboxes for gaming AI innovation, modeled on the UK Financial Conduct Authority sandbox program, to accelerate governance development while enabling innovation in AI-powered responsible gambling tools.

10.4. Adversarial Robustness Standards

- Risk distribution analysis shows that, of the 500 known risks to AI security in regulated industries, technical adversarial attack vectors together represent 55 percent of all identified risks, far out of proportion to their high status in the current AI governance framework standard. Addressing this gap requires:
- Compulsory process of adversarial robustness testing of high-risk AI applications in all three focal sectors, which identifies minimum testing methodologies and documentation standards consistent with the NIST AI RMF Measure and performance.
- Creation of sector-specific attack scenario library (healthcare, financial services, and gaming) - representing each sector with its unique threat actor space and the applications of AI that are most frequently exploited in reported attacks.

- Integration of adversarial, robustness requirements to ISO/IEC 42001 implementation guidance in regulated industries, that has sector-specific annexes with a practical testing methodology and control requirements.

10.5. Cross-Sector AI Incident Sharing Infrastructure

10.5.1. Adaptive Regulatory Mechanisms

The speed at which AI is evolving poses some underlays to the conventional regulatory methods based on prescriptive regulations that are formed due to the carefully planned legislative procedures. AI systems they were created to govern may have changed a great deal by the time AI governance regulations are implemented. Regulatory sandboxes, content that is outcome-based, not prescriptive, and continuous regulatory industry process of technical dialogue are all adaptive regulatory mechanisms that provide promising ways of ensuring regulatory relevance. The fact shown in Figure 8 of the high price of retroactive governance implementation is a strong indication of proactive regulatory intervention to implement AI governance requirements preceding the deployment curve, and in governance-immature regulation, like in gaming.

11. Conclusion

In conclusion, this systematic literature review has discussed the condition of AI security governance frameworks in three highly regulated sectors, namely, healthcare, financial services, and gaming, which presents a holistic empirical and analytical framework of the state of AI security governance frameworks and its shortcomings. The overall findings of the review confirm that there are high governance gaps in the three sectors. Gaming has the most severe shortcomings, as non- compliance with 7 of the 12 measures of governance evaluated and an adversarial vulnerability score of 7.8 representing high-end financially-motivated threat actors. Healthcare poses the most difficult governance architecture issues with a sharp disparity between documentation of governance and actual functionality in the vastly fragmented deployment environment. Financial services have the strongest level of governance infrastructure but records the highest score of regulatory non-compliance risk (9.1, Critical) due to the unprecedented complexity of its multi-agency regulatory framework.

The presented empirical evidence in the form of eight figures and four tables altogether underpins three general conclusions. To begin with, the lack of governance recorded in all three sectors, which can be seen in the deployment-governance divergence trend of Figure 8, is not only a regulatory compliance issue but a complete misfit between the speed of AI implementation and the establishment of a governance framework sufficient to the AI-specific risks that the implementation generates. Second, the technical adversarial attack vectors recorded in Figure 7, which add up to 55% of the known AI security threats, are an under-governed governance risk concern as compared to their weight in the incident record, indicating that modern governance framework investment priorities are over-invested in algorithmic fairness and under-invested in adversarial robustness. Third, gaming and healthcare convergences reported in Figure 5 are leading indicators of governance pressure in the gaming industry, and the regulatory authorities must consider it an indicator of a dire need to intervene in governance development.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed

References

[1] Barrett, M. P., Marron, J., Pillitteri, V. Y., Boyens, J., Quinn, S., and Witte, G. (2023). Artificial intelligence risk management framework (AI RMF 1.0). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.100-1

[2] Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., and Garfinkel, B. (2018). The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. Future of Humanity Institute. https://arxiv.org/abs/1802.07228

[3] Cheatham, B., Javanmardian, K., and Samandari, H. (2019). Confronting the risks of artificial intelligence. McKinsey Quarterly. https://www.mckinsey.com/capabilities/quantumblack/our-insights/confronting-the-risks-of-artificial-intelligence

[4] Deng, Y., Ficnar, M., Magnuson, W. J., and Park, N. (2019). Finance as data: The Dodd-Frank era reconsidered. Yale Journal on Regulation, 36(1), 41-92. https://yjreg.law.yale.edu/finance-as-data

[5] European Parliament. (2023). Regulation (EU) 2023/1689 laying down harmonised rules on artificial intelligence. Official Journal of the European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L_202301689

[6] FDA. (2021). AI/ML-based software as a medical device action plan. U.S. Food and Drug Administration. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device

[7] Financial Stability Board. (2022). Artificial intelligence and machine learning in financial services. FSB. https://www.fsb.org/work-of-the-fsb/policy-development/additional-policy-areas/fsb-work-on-artificial-intelligence/

[8] Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., and Dignum, V. (2019). An ethical framework for a good AI society. Minds and Machines, 28(4), 689-707. https://link.springer.com/article/10.1007/s11023-018-9482-5

[9] Goodfellow, I. J., Shlens, J., and Szegedy, C. (2015). Explaining and harnessing adversarial examples. Proceedings of ICLR 2015. https://arxiv.org/abs/1412.6572

[10] Goodman, B., and Flaxman, S. (2017). EU regulations on algorithmic decision-making and a right to explanation. AI Magazine, 38(3), 50-57. https://ojs.aaai.org/index.php/aimagazine/article/view/2741

[11] ISO. (2023). ISO/IEC 42001: Information technology -- Artificial intelligence -- Management system. International Organization for Standardization. https://www.iso.org/standard/81230.html

[12] Kumar, R. S. S., O'Brien, M. E., Albert, K., Viljoen, S., and Snover, J. (2020). Adversarial machine learning industry perspectives. IEEE Security and Privacy Workshops. https://arxiv.org/abs/2002.05646

[13] Landis, J. R., and Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33(1), 159-174. https://www.jstor.org/stable/2529310

[14] Magnuson, W. J. (2018). Financial regulation in the bitcoin era. Stanford Journal of Law, Business and Finance, 23(2), 159-209. https://law.stanford.edu/publications/financial-regulation-in-the-bitcoin-era/

[15] NIST. (2023). AI risk management framework: AI RMF 1.0. National Institute of Standards and Technology. https://airc.nist.gov/Docs/1

[16] Obermeyer, Z., Powers, B., Vogeli, C., and Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447-453. https://www.science.org/doi/10.1126/science.aax2342

[17] Philippon, T. (2016). The fintech opportunity. NBER Working Paper No. 22476. https://www.nber.org/papers/w22476

[18] Smuha, N. A. (2021). From a race to AI to a race to AI regulation. Law, Innovation and Technology, 13(1), 57-84. https://www.tandfonline.com/doi/full/10.1080/17579961.2021.1898300

[19] Shokri, R., Stronati, M., Song, C., and Shmatikov, V. (2017). Membership inference attacks against machine learning models. IEEE SandP 2017. https://arxiv.org/abs/1610.05820

[20] Taddeo, M., McCutcheon, T., and Floridi, L. (2019). Trusting artificial intelligence in cybersecurity is a double-edged sword. Nature Machine Intelligence, 1(12), 557-560. https://www.nature.com/articles/s42256-019-0109-1

[21] Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., and Mané, D. (2016). Concrete problems in AI safety. arXiv preprint. https://arxiv.org/abs/1606.06565

[22] Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., and Barbado, A. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges. Information Fusion, 58, 82-115. https://www.sciencedirect.com/science/article/pii/S1566253519308103

[23] Barocas, S., and Selbst, A. D. (2016). Big data's disparate impact. California Law Review, 104(3), 671-732. https://www.californialawreview.org/print/big-data-disparate-impact/

[24] Biggio, B., and Roli, F. (2018). Wild patterns: Ten years after the rise of adversarial machine learning. Pattern Recognition, 84, 317-331. https://www.sciencedirect.com/science/article/pii/S0031320318302231

[25] Dwork, C., and Roth, A. (2014). The algorithmic foundations of differential privacy. Foundations and Trends in Theoretical Computer Science, 9(3-4), 211-407. https://www.nowpublishers.com/article/Details/TCS-042

[26] Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., and Yu, H. (2017). Accountable algorithms. University of Pennsylvania Law Review, 165(3), 633-705. https://scholarship.law.upenn.edu/penn_law_review/vol165/iss3/3/

[27] Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., and Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data and Society, 3(2), 1-21. https://journals.sagepub.com/doi/10.1177/2053951716679679

[28] Pasquale, F. (2016). The black box society. Harvard University Press. https://www.hup.harvard.edu/catalog.php?isbn=9780674970847

[29] Wachter, S., Mittelstadt, B., and Floridi, L. (2017). Why a right to explanation does not exist in the GDPR. International Data Privacy Law, 7(2), 76-99. https://academic.oup.com/idpl/article/7/2/76/3860948

[30] Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., and Barnes, P. (2020). Closing the AI accountability gap. ACM FAccT 2020. https://arxiv.org/abs/2001.00973