ISO/IEC 42001 - Aligned Governance Evidence for High-Stakes AI Decision Systems: A Systematic Literature Review

Bharathi Vivekanandan, J. Jeyabharathi

School of Computing, Kalasalingam Academy of Research and Education, Tamil Nadu, India

vivbharathi@gmail.com, j.jeyabharathi@klu.ac.in

*Corresponding author: Dr J. Jeyabharathi (j.jeyabharathi@klu.ac.in)

Abstract

Artificial intelligence (AI) is increasingly deployed in high-stakes sectors such as healthcare, finance, and autonomous systems, creating a need for governance approaches that address risks related to bias, opacity, accountability, and compliance. This systematic literature review (SLR) examines how the literature aligns AI governance for high-stakes decision systems with ISO/IEC 42001:2023, with particular attention to evidence generation, auditability, and the operationalization of compliance. Using a structured search and multi-stage screening process, 39 studies were selected for analysis from an initial pool of 2,918 records. The review finds growing convergence toward risk-based, lifecycle-oriented AI governance, supported by mechanisms such as traceability, continuous monitoring, and machine-readable evidence artefacts. The literature also indicates increasing interest in automation-enabled compliance, including blockchain-based logging and agentic support for governance processes. However, evidence remains fragmented, with limited large-scale validation, inconsistent evidence formats, and uneven treatment of regional regulatory requirements. The review concludes that ISO/IEC 42001-aligned governance for high-stakes AI requires evidence-centric, auditable, and context-sensitive oversight that combines automation with human judgment.

Keywords: ISO/IEC 42001, AI Governance, Governance Evidence Layer (GEL), High-Stakes Decision Systems, Automated Compliance.

1 Introduction

Artificial intelligence (AI) is increasingly used in high-stakes sectors such as autonomous systems, healthcare, and finance [1–12]. In these contexts, failures in design, deployment, or oversight can lead to bias, opacity, discrimination, and breakdowns in accountability. Public repositories reporting more than 3,000 AI-related incidents indicate that these risks are no longer theoretical and require systematic governance responses [13]. Traditional technical safeguards alone are insufficient for high-stakes AI decision systems. Effective oversight requires governance mechanisms that align AI development and use with legal, ethical, and operational objectives. [13–16]. ISO/IEC 42001:2023 is especially important in this context because it is the first international management system standard developed specifically for AI [16–18]. Prior studies also emphasise AI-specific governance challenges, including limited explainability, lifecycle variability, and changing system behaviour over time. [18–20].

Recent literature has increasingly discussed auditable evidence, traceability, decision logging, and automation-assisted compliance as important elements of AI governance. [19, 21–25]. For analytical convenience, this review uses the term Governance Evidence Layer (GEL) to describe this emerging set of mechanisms for generating and verifying evidence. GEL is used here as a synthesis construct derived from the literature, not as a universally established architectural standard. Despite growing interest in AI governance, the literature remains fragmented regarding how evidence-centric governance is operationalised in high-stakes AI systems. [13, 16, 25]. This systematic literature review (SLR) synthesises 39 primary studies to examine four questions: who is accountable for oversight, what system elements are governed, when governance interventions occur across the AI lifecycle, and how governance is implemented through controls, tools, and evidence mechanisms in relation to ISO/IEC 42001 and related regulatory expectations. [13, 16, 17, 26]. The review provides a structured synthesis

Proceedings of ICICCS-2026 (Poster Papers), DMPedia Lecture Notes in Computer Science & Engineering.
ISBN: 978-81-993813-1-5

of current knowledge and identifies key gaps in verifiable, risk-based AI governance for high-stakes environments.

2 Background and Related Work

AI governance has become a practical requirement in high-stakes domains because technical performance alone does not address broader concerns such as safety, accountability, transparency, privacy, and liability. [13, 19, 27]. Prior literature generally defines AI governance as the set of organisational, regulatory, and technical mechanisms used to align AI systems with legal, ethical, and strategic objectives. [13, 15, 28]. In this context, governance is not limited to model behaviour, but extends to data, lifecycle controls, human oversight, and assurance processes.

2.1 ISO/IEC 42001 and AI governance frameworks

Recent literature has increasingly positioned ISO/IEC 42001:2023 as an important reference point for structured, risk-based AI governance. [15, 26, 28]. As the first international management system standard specifically developed for AI, it provides a governance-oriented framework for addressing issues such as lifecycle management, accountability, and the distinctive risks posed by evolving, opaque AI systems. [17, 19, 20]. At the same time, several studies note that ISO/IEC 42001 should be understood as a foundational baseline rather than a complete regulatory solution, since additional interpretation or controls may be needed to address domain-specific and regional requirements, including those associated with the EU AI Act. [14, 19, 26].

2.2 Evidence-centric operationalisation and remaining gaps

Related work also shows growing interest in translating governance principles into operational, auditable practice through traceability, decision logging, monitoring, and automated compliance mechanisms. [19, 21–25] For analytical convenience, this review uses the term Governance Evidence Layer (GEL) to describe the emerging class of mechanisms for generating, verifying, and tracing evidence discussed in the literature. This term is used here as a synthesis construct rather than as a universally established architectural standard. Despite this progress, prior work indicates that the field remains constrained by fragmented implementation approaches, uneven empirical validation, and the absence of widely adopted evidence standards for interoperable, scalable governance. [13, 16, 19, 24, 25, 29, 30] These limitations motivate the present review's focus on how governance is operationalised across accountability, governed system elements, lifecycle intervention points, and implementation mechanisms in high-stakes AI settings.

3 Systematic Review Methodology

This systematic literature review (SLR) followed a structured review process aligned with PRISMA 2020 to support transparent study identification, screening, and synthesis. [4, 13, 27]. The review process was organised into three phases: planning, conducting, and reporting. [13].

3.1 Search and study selection

The literature search was conducted in January 2026 across Google Scholar and OpenAlex and covered studies published between 2020 and 2025 [4, 13]. An initial set of 2,918 records was retrieved using search terms centred on artificial intelligence and governance. [4, 7, 13]. Duplicate records were removed before screening. [13]. The remaining studies were screened through a multi-stage process that included title and abstract review, full-text assessment, and forward and backwards snowballing to identify additional relevant studies. [4, 13]. Earlier sources cited elsewhere in the paper were used selectively for context on standards, conceptual framing, or domain background.

Studies were evaluated against predefined inclusion and exclusion criteria designed to retain work relevant to high-stakes AI governance. These criteria considered governance relevance, high-stakes application context, methodological substance, and organisational or operational applicability.[13]. The screening process resulted in a final sample of 39 studies for review. From this set, 28 studies formed the core basis for detailed thematic synthesis because they provided the strongest coverage of the review questions and the most direct treatment of governance implementation in high-stakes settings. [13].

3.2 Quality Assessment

The selected studies were further assessed using a five-question quality framework adapted from prior review work. [13, 34]. The assessment considered research design, clarity of reporting, relevance to the review scope, and contribution to understanding AI governance in high-stakes contexts. Studies assessed as poor quality were excluded from the final synthesis, resulting in an evidence base composed of good- and fair-quality studies only. [13].

3.3 Data Extraction and Synthesis

Data were extracted into a structured review repository and synthesised using four analytical dimensions derived from the review questions. [13]. These dimensions were: **who**, referring to accountability and oversight roles; **what**, referring to governed system elements such as data and models. [9, 13]; **when** referring to governance intervention points across the AI lifecycle [13, 14]; and **how**, referring to implementation mechanisms, including standards, evidence artefacts, and automation-supported governance approaches [16, 22, 24–26, 30]. This structure supported consistent comparison of studies and guided the thematic synthesis reported in the next section.

4 Data Synthesis and Analysis

The 39 selected studies show increasing attention to operational AI governance in high-stakes settings, moving beyond general ethical principles toward more structured assurance and compliance mechanisms. [13, 15, 20, 26, 33, 34]. The findings are organised around four themes: accountability, governed system elements, lifecycle intervention points, and implementation mechanisms. [13].

4.1 Accountability (WHO)

The literature treats accountability as distributed across organisational, technical, and regulatory levels. [13, 35]. Governance approaches often operate through internal risk and policy structures. [9, 36], while high-stakes domains also rely on multidisciplinary oversight involving safety, ethics, compliance, and domain expertise [4–12, 37, 38]. Several studies also highlight the interaction between regulatory oversight and standards-based assurance, including ISO/IEC 42001-related compliance approaches. [14, 19, 26].

4.2 Governed system elements (WHAT)

The reviewed studies distinguish between data-oriented and model-oriented governance. [13]. Data governance focuses on stewardship, quality, provenance, and bias mitigation. [9, 13, 15], whereas model governance emphasises interpretability, robustness, safety assurance, and post-deployment monitoring [19, 20, 25].

4.3 Lifecycle intervention points (WHEN)

Governance is consistently presented as a lifecycle concern rather than a one-time control. [13]. Before development, studies emphasise impact assessment and policy definition. [24, 35]. During development, they discuss testing, traceability, and compliance-oriented controls. [19,

22]. After deployment, the emphasis shifts to monitoring, logging, telemetry, and auditability [21, 23, 25].

4.4 Implementation mechanisms (HOW)

A recurring theme is the move toward more continuous, evidence-based governance. [13, 19, 21–25]. Reported mechanisms include traceability artefacts, policy-linked logs, blockchain-supported audit records, automation-assisted compliance checks, and privacy-preserving attestations. [19, 21–25]. In this review, these are grouped analytically under the term **Governance Evidence Layer (GEL)**, used as a synthesis construct rather than a standardised architecture. [24, 25].

5 Discussion

The reviewed studies suggest a broader shift in AI governance from principle-level responsible AI discourse toward more operational, evidence-based management approaches. [13, 15, 20, 26, 34, 35]. In high-stakes settings, this shift is reflected in growing attention to lifecycle governance, auditable controls, and mechanisms that connect organisational oversight with technical implementation. [13, 19, 21–25].

5.1 ISO/IEC 42001 as a governance baseline

Across the reviewed literature, ISO/IEC 42001 is increasingly treated as an important baseline for AI governance in high-stakes environments. [13, 26, 34, 35]. However, the studies also indicate that the standard is not fully sufficient on its own, since additional controls or interpretations may be needed to address regional and sector-specific requirements, including those associated with the EU AI Act. [14, 26]. This is especially relevant in domains such as healthcare and autonomous systems, where governance is closely tied to safety and accountability. [19, 20, 37].

5.2 Evidence-centric governance and GEL

The findings also indicate growing interest in evidence-centric governance mechanisms that make compliance and oversight more auditable in practice. [19, 21–25]. In this review, these mechanisms are grouped analytically under the term **Governance Evidence Layer (GEL)**, used as a synthesis construct rather than a standardised architecture. [24, 25]. The literature discusses artefacts such as traceable logs, policy-linked records, and runtime evidence as ways to strengthen assurance. However, implementation remains uneven, and no common evidence standard has yet emerged. [19, 24, 25].

5.3 Automation and human oversight

Another recurring theme is the shift from periodic manual review to more continuous, automated governance. [13, 19, 21, 22, 25]. The literature suggests that automation can improve scalability and timeliness, but it does not remove the need for human oversight. In high-stakes settings, human judgment remains necessary for escalation, exception handling, and accountability. [9, 13, 17, 27–29].

5.4 Implementation implications

The reviewed studies also point to practical challenges, including resource constraints, integration with legacy systems, and the lack of widely adopted interoperable evidence standards. [15, 19, 34, 35, 39, 40]. Overall, the findings suggest that implementation is likely to be incremental, with organisations combining standards-based governance with modular evidence, monitoring, and oversight mechanisms.

6 Conclusion and Future Work

This systematic literature review synthesised 39 studies on governance for high-stakes AI decision systems and found increasing attention to operational, risk-based, and auditable governance approaches. [13, 15, 20, 26, 34, 35]. The reviewed literature positions ISO/IEC 42001 as an important governance baseline, while also highlighting growing interest in evidence-centric mechanisms such as traceability, logging, monitoring, and policy-linked artefacts [19, 21–26]. In this review, these mechanisms were examined under the analytic concept of a Governance Evidence Layer (GEL) [24, 25].

The findings suggest that governance in high-stakes domains is increasingly treated as an operational requirement rather than a peripheral compliance activity. [4–12, 37, 38]. However, the evidence base remains fragmented, with limited large-scale validation and no widely adopted standard for interoperable governance evidence. [13, 19, 24, 25, 29]. Future work should therefore focus on longitudinal evaluation, standardisation of machine-readable evidence artefacts, stronger human-automation calibration, and governance for increasingly agentic AI systems. [13, 15, 24, 26–28, 34, 37].

Appendix A – Terminology

For clarity, this appendix defines key terms used throughout the review.

- **AIMS:** AI management system defined in ISO/IEC 42001:2023.
- **GEL:** In this review, a synthesis term for evidence, traceability, logging, and verification mechanisms supporting auditable AI governance.
- **High-stakes AI system:** An AI system whose failure or bias may cause significant harm to safety, rights, health, or financial outcomes.
- **HITL:** Human review or intervention within an AI-supported decision process.
- **Policy cards:** Structured artefacts linking system behaviour or controls to policy or regulatory requirements.

References

1. Jody Nelson, Lin, C.: Responsible AI System Development in Automotive Applications: A Framework. Presented at the WCX SAE World Congress Experience, April 1 (2025)

2. Suelen Daiene De Oliveira Bastos, Kalinka Castelo Branco, André Luiz De Oliveira: Bridging the Gaps: A Comparative Analysis of ISO 21434, ISO 26262 and Machine Learning in Autonomous Vehicles. In: 2025 Brazilian Symposium on Robotics (SBR) and 2025 Workshop on Robotics in Education (WRE). pp. 25–30. IEEE, Vitoria, Brazil (2025)

3. Shanza Ali Zafar, Jessica Kelly, Lena Heidemann, Mata, N.: Leveraging Existing Road-Vehicle Standards to Address EU AI Act Compliance. In: 2025 IEEE/ACM International Workshop on Responsible AI Engineering (RAIE). pp. 74–81. IEEE, Ottawa, ON, Canada (2025)

4. Marçal Mora-Cantallops, Elena García-Barriocanal, Miguel-Ángel Sicilia: Trustworthy AI Guidelines in Biomedical Decision-Making Applications: A Scoping Review. Big Data Cogn. Comput. 8, 73 (2024). https://doi.org/10.3390/bdcc8070073

5. Jee Young Kim, William Boag, Freya Gulamali, Alifia Hasan, Henry David Jeffry Hogg, Lifson, M., Mulligan, D., Patel, M., Raji, I.D., Sehgal, A., Shaw, K., Tobey, D., Valladares, A., Vidal, D., Balu, S., Sendak, M.: Organizational Governance of Emerging Technologies: AI Adoption in Healthcare. In: 2023 ACM Conference on Fairness, Accountability, and Transparency. pp. 1396–1417. ACM, Chicago, IL, USA (2023)

6. Maryam, B., Rob, B., Siobhan, C., Yalemisew, A.: A risk governance framework for healthcare decision support systems based on socio-technical analysis. In: Critical Perspectives on Artificial Intelligence Ethics 2020. , Edinburgh, Scotland (2020)

7. Arian Ranjbar, Eilin Mork, Jesper Ravn, Helga Brøgger, Per Myrseth, Hans Peter Østrem, Harry Hallock: Managing Risk and Quality of AI in Healthcare: Are Hospitals Ready for Implementation? Risk Manag. Healthc. Policy. Volume 17, 877–882 (2024). https://doi.org/10.2147/RMHP.S452337

8. Sanjay Nakharu Prasad Kumar: Ethical Frameworks for AI-Driven Decision Systems: A Comprehensive Analysis. Glob. J. Comput. Sci. Technol. 53–60 (2025). https://doi.org/10.34257/GJCSTDVOL25IS1PG53

9. Razak Abdul: AI-Driven Ethical Governance: Balancing Innovation and Accountability In Data-Intensive Enterprises. Int. J. Eng. Sci. Adv. Technol. 25, 17–21 (2025). https://doi.org/10.64771/ijesat.2025.v25.i10.pp17-21

10. Henrik Axelsen, Valdemar Licht, Jan Damsgaard: Agentic AI for Financial Crime Compliance, https://arxiv.org/abs/2509.13137, (2025)

11. Priscilla Samuel Nwachukwu, Onyeka Kelvin Chima, Chinelo Harriet Okolo: The artificial intelligence governance framework for finance: A control-by-design approach to algorithmic decision-making in accounting. Finance Account. Res. J. 7, 350–379 (2025). https://doi.org/10.51594/farj.v7i8.2016

12. Pratik Chawande: Model Risk Governance for AI-Based Compliance Systems in Investment Banking. Int. J. Multidiscip. Res. Growth Eval. 6, 2027–2035 (2025). https://doi.org/10.54660/.IJMRGE.2025.6.3.2027-2035

13. Amna Batool, Didar Zowghi, Muneera Bano: AI governance: a systematic literature review. AI Ethics. 5, 3265–3279 (2025). https://doi.org/10.1007/s43681-024-00653-w

14. Gregory Falco, Ben Shneiderman, Julia Badger, Ryan Carrier, Anton Dahbura, Danks, D., Eling, M., Goodloe, A., Gupta, J., Hart, C., Jirotka, M., Johnson, H., LaPointe, C., Llorens, A.J., Mackworth, A.K., Maple, C., Pálsson, S.E., Pasquale, F., Winfield, A., Yeong, Z.K.: Governing AI safety through independent audits. Nat. Mach. Intell. 3, 566–571 (2021). https://doi.org/10.1038/s42256-021-00370-7

15. Avinash Agarwal, Manisha J. Nene: A five-layer framework for AI governance: integrating regulation, standards, and certification. Transform. Gov. People Process Policy. 19, 535–555 (2025). https://doi.org/10.1108/TG-03-2025-0065

16. Delaram Golpayegani, Harshvardhan J. Pandit, Dave Lewis: Comparison and Analysis of 3 Key AI Documents: EU's Proposed AI Act, Assessment List for Trustworthy AI (ALTAI), and ISO/IEC 42001 AI Management System. In: Luca Longo and Ruairi O'Reilly (eds.) Artificial Intelligence and Cognitive Science. pp. 189–200. Springer Nature Switzerland, Cham (2023)

17. ISO/IEC: Information technology — Artificial intelligence — Management system, https://www.iso.org/standard/42001.html, (2023)

18. Jordan Pötsch: Interplay of ISMS and AIMS in context of the EU AI Act, http://arxiv.org/abs/2412.18670, (2024)

19. Zareen Hossain, Ritam Rajak, Indra Vijay Singh, Tansuhree Das, Abhinandan Pal: Self-Attesting Intelligence: A Framework for Inherently Verifiable AI Systems. Adv.

Knowl.-Based Syst. Data Sci. Cybersecurity. 02, 291–309 (2025). https://doi.org/10.54364/cybersecurityjournal.2025.1115

20. Dr Rambabu Kalathoti: Explainable AI in High-Stakes Decision Making: Beyond Accuracy. Sci. J. Artif. Intell. Blockchain Technol. 2, (2025). https://doi.org/10.63345/sjaibt.v2.i3.103

21. Suleiman S. Abba, Oluwadayo Mafolasere Olaniyi, Oluseun Babatunde Oladoyinbo, Olalekan Jamiu Okunleye, Valerie Ojinika Ejiofor: AI-Driven Automation of Cybersecurity Certification Processes: Evaluating Efficiency, Transparency and Risk Mitigation in Digital Governance Systems. J. Eng. Res. Rep. 27, 70–91 (2025). https://doi.org/10.9734/jerr/2025/v27i121728

22. Aman Sardana, Swaminathan Sethuraman, Priya Dharshini Kalyanasundaram: Compliance-as-Code 2.0: Orchestrating Regulatory Operations with Agentic AI. J. Artif. Intell. Gen. Sci. JAIGS ISSN3006-4023. 5, 546–563 (2024). https://doi.org/10.60087/jaigs.v5i1.366

23. Panda, D., Padhy, N. and Sharma, K: February. Strengthening IoT Resilience: A Study on Backdoor Malware and DNS Spoofing Detection Methods. In *2025 International Conference on Emerging Systems and Intelligent Computing (ESIC),* pp. 795-800, (2025).

24. Juraj Mavracic: Policy Cards: Machine-Readable Runtime Governance for Autonomous AI Agents, https://zenodo.org/doi/10.5281/zenodo.17464706, (2025)

25. Prince Enyiorji: Blockchain-enforced data lineage architectures with formal verification workflows enabling auditable AI decision chains across regulated fintech compliance regimes and supervisory reporting. Int. J. Sci. Res. Arch. 9, 1201–1217 (2023). https://doi.org/10.30574/ijsra.2023.9.2.0559

26. Martin Ebers: Standardising AI - The Case of the European Commission's Proposal for an Artificial Intelligence Act. SSRN Electron. J. (2021). https://doi.org/10.2139/ssrn.3900378

27. Donglin Wang, Weiyun Liang, Chunyuan Chen, Jing Xu, Yulong Fu: Governable AI: Provable Safety Under Extreme Threat Models, https://arxiv.org/abs/2508.20411, (2025)

28. Antonio Aloisi, Valerio De Stefano: Between risk mitigation and labour rights enforcement: Assessing the transatlantic race to govern AI-driven decision-making through a comparative lens. Eur. Labour Law J. 14, 283–307 (2023). https://doi.org/10.1177/20319525231167982

29. Sharma, H., Kumar, P., Shrivastava, G., Sharma, K. and Bhola, A: Using Machine Learning for Protecting the Security and Privacy of Internet of Medical Things (IoMT) Systems. In *Integrating Cloud, Fog, and Edge Computing in Healthcare: Federated Learning and Blockchain Approaches: Harnessing Distributed Technologies for Enhanced Healthcare Delivery,* pp. 123-138 (2026).

30. Adam Swidan: Human-in-the-Loop Reinforcement Learning for Ethical and Compliant Risk Decisions. Int. J. Sci. Res. Sci. Eng. Technol. 12, 121–131 (2025). https://doi.org/10.32628/IJSRSET2513857

31. Alex Dantart: Gobernanza y trazabilidad "a prueba de AI Act" para casos de uso legales: un marco técnico-jurídico, métricas forenses y evidencias auditables, https://arxiv.org/abs/2510.12830, (2025)

32. Panda D, Padhy N, Sharma K. DDOS Attack Detection and Performance Analysis in IOT Network using Machine Learning Approaches. Scalable Computing: Practice and Experience. 2025 Feb 10;26(2):950-63.

33. Anil Kumar Pakina: AI Governance via Explainable Reinforcement Learning (XRL) for Adaptive Cyber Deception in Zero-Trust Networks. J. Inf. Syst. Eng. Manag. 10, 98–115 (2025). https://doi.org/10.52783/jisem.v10i43s.8308

34. Ahmed Bahgat: AI governance and data privacy in cross-border contexts: A comparative analysis of regulatory frameworks. J. Data Prot. Priv. 8, 8 (2025). https://doi.org/10.69554/BOJL3033

35. David Leslie, Christopher Burr, Mhairi Aitken, Michael Katell, Morgan Briggs, Cami Rincon: Human Rights, Democracy, and the Rule of Law Assurance Framework for AI Systems: A Proposal. SSRN Electron. J. (2021). https://doi.org/10.2139/ssrn.4027875

36. Jain N, Jaiswal P, Sharma S, Sharma K, Sharma V: A machine learning based approach to detect phishing attack. In2023 5th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N) pp. 305-309). IEEE. (2023).

37. Aoun E Muhammad, Kin Choong Yow, Jamel Baili, Yongwon Cho, Yunyoung Nam: CORTEX: Composite Overlay for Risk Tiering and Exposure in Operational AI Systems, https://arxiv.org/abs/2508.19281, (2025)

38. Sharma P, Nagpal T, Shrivastava G, Kumar JD. Machine Learning for Fake News Detection A Survey. In2023 5th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), pp. 850-855. (2023).

39. Annette Kott, Andreas Rössler, Rainer Groß, Harald Kosch, Florian Ries: AI Regulates AI: An Artefact for Automated Risk Assessment. (2025). https://doi.org/10.18420/INF2025_117

40. Jessica Kelly, Shanza Ali Zafar, Lena Heidemann, João-Vitor Zacchi, Delfina Espinoza, Núria Mata: Navigating the EU AI Act: A Methodological Approach to Compliance for Safety-critical Products. In: 2024 IEEE Conference on Artificial Intelligence (CAI). pp. 979–984. IEEE, Singapore, Singapore (2024)