From Principles to Controls: A Design-Science Justification of an AI Auditing Meta-Framework

Derya S. Esen[1]

[1] Goethe University Frankfurt, Germany
derya@dsozen.com

Abstract. Due to the proliferation of AI governance frameworks since 2023, organizations need auditing approaches that are interoperable, evidence based, and applicable across the full AI lifecycle. Building on a previously presented PRISMA-guided scoping review that proposed a 25-control AI auditing meta-framework, this paper examines the design rationale underlying that artefact. It addresses a key weakness in emerging AI auditing meta-frameworks by making their design logic explicit through design-science research and auditing theory. The contribution is not a second framework, but a justification of how high-level governance principles can be translated into concrete, verifiable audit controls through a structured control-derivation schema. The paper positions AI auditing meta-frameworks not as universal standards, but as translation layers that help organizations interpret regulatory regimes in a consistent and defensible way.

Keywords: AI Auditing, Design Science Research, AI Governance, Meta-Frameworks, Compliance, Audit Control.

1 Introduction

AI governance has shifted from a predominantly principles-based discourse toward enforceable duties, management system standards, and assurance expectations. Instruments such as the EU Artificial Intelligence (AI) Act [1], ISO/IEC 42001 [2], and the NIST AI Risk Management Framework [3] illustrate this transition: they differ in legal force, terminology, scope, and risk model, yet converge on the expectation that organizations can demonstrate trustworthy AI through documentation, monitoring, testing, oversight, and accountability mechanisms.

This creates a design problem rather than a merely classificatory one. Organizations do not need another list of values; they need a defensible way to translate heterogeneous governance claims into controls that auditors can inspect, since principles alone cannot guarantee ethical AI [4]. Recent AI auditing meta-frameworks respond to this need by consolidating requirements across regimes. However, consolidation is vulnerable to a legitimacy critique: if the rationale for selecting and structuring controls is implicit, a meta-framework may appear arbitrary, ad hoc, or purely technocratic.

This paper builds directly on prior work in which the author conducted a PRISMA-guided scoping review of 28 post-2023 AI governance frameworks and proposed a unified 25-control meta-framework for AI auditing [5]. That earlier contribution

synthesized common requirements across regulatory, standards-based, and policy frameworks, while identifying three blind spots: post-deployment drift monitoring, foundation-model assurance, and sustainability. The present paper takes the next step: it asks how such a meta-framework can be justified as a design artefact rather than treated as an arbitrary consolidation of controls.

The guiding research question is therefore: how can the translation from AI governance principles to auditable controls be justified through Design Science Research (DSR)? The answer proposed here is that the artefact must be evaluated not only by coverage of source frameworks, but by the explicitness of its design principles, the traceability of its control derivation, and the quality of evidence it makes available for audit judgement.

2 Research Continuity and Design Science Lens

Table 1 clarifies the relationship between the earlier study and the present paper.

Table 1. Research positioning of the prior and present papers.

Aspect	Prior paper [5]	Present paper
Main aim	Map the post-2023 AI governance landscape and propose a unified audit meta-framework.	Justify the meta-framework through design-science research and auditing theory [6].
Method	PRISMA-guided scoping review and framework mapping.	Ex post DSR reconstruction of the artefact, design principles, and control derivation.
Output	A 25-control AI auditing meta-framework.	A principled schema linking governance principles, controls, evidence, and verification.

DSR is appropriate because the object of inquiry is an artefact intended to solve a situated organizational problem [7]. An AI auditing meta-framework is not a neutral taxonomy. It defines categories, prioritizes risks, and specifies the evidence through which compliance becomes visible. In this sense, it is a normative socio-technical instrument, meaning it mediates between legal and governance principles on one side and organizational practice, engineering artefacts, and audit judgement on the other.

Methodologically, this paper adopts DSR as an ex post justificatory lens [8] for the artefact introduced in prior work. Rather than repeating the full scoping review, it reconstructs the design logic of that artefact by identifying the design problem, deriving normative design principles, mapping those principles to control requirements, and specifying the evidentiary outputs that make the controls auditable [9].

3 Design Principles and Control-Derivation Schema

Four design principles guide the construction and defense of the meta-framework. Interoperability requires that controls map to several governance regimes without collapsing their differences. Evidentiary auditability requires each control to produce inspectable evidence rather than rely on unsupported declarations. Lifecycle completeness requires coverage from system conception and data collection through deployment, monitoring, incident response, and retirement. Risk proportionality requires control intensity to scale with the AI system's potential harm, context, and legal classification.

These principles support a structured control-derivation schema: (1) identify a governance claim or obligation; (2) infer the audit objective implied by that claim; (3) formulate an organizational or technical control; (4) define the evidence artefact generated by the control; and (5) specify the auditor verification point. A control is justified only when its link to a governance principle and its evidentiary function are both explicit. Table 2 illustrates this logic using selected controls from the prior framework.

Table 2. Design-science derivation from principles to auditable controls.

Design principle	Example control from prior framework	Evidence expected	Auditor verification point
Interoperability	Domain-regulation alignment	Mapping matrix to EU AI Act, ISO/IEC 42001, NIST AI RMF	Confirm simultaneous compliance claims and unresolved gaps.
Evidentiary auditability	Traceable decision logs	Tamper-evident input-output and processing records	Reconstruct high-impact decisions from retained logs.
Lifecycle completeness	Continuous monitoring and drift triggers	Baseline metrics, thresholds, alerts, retraining records	Check whether drift events trigger investigation or revalidation.
Risk proportionality	Risk-tiering and third-party review	Risk class, rationale, escalation decision	Assess whether assurance depth matches system risk.

4 Discussion: Meta-Frameworks as Translation Layers

The added value of this paper lies in making explicit what remained partly implicit in the prior framework proposal: why these controls are legitimate design choices. The prior study established the fragmented regulatory problem and proposed a practical control corpus. This paper argues that such a corpus gains legitimacy only when its derivation is transparent, traceable, and grounded in design principles that preserve regulatory intent.

This framing is particularly important for current blind spots in AI assurance. Post-deployment drift, foundation-model assurance, and sustainability metrics often remain

weakly specified across governance frameworks. A design-led meta-framework can make those gaps explicit by adding controls for continuous monitoring, foundation-model risk assessment, red-teaming, documentation, compute-hour logging, and carbon or energy disclosure. These additions are not arbitrary extensions; they are justified because they translate lifecycle, auditability, and proportionality principles into evidence-generating controls.

The translation-layer view also clarifies the role of expertise. Control derivation cannot be fully automated, because legal concepts such as transparency, oversight, fairness, and appropriate risk management depend on context. Expert-guided synthesis is therefore not a weakness of the method, but a necessary feature of socio-technical auditing.

5 Conclusion and Future Work

This paper has argued that the legitimacy of AI auditing meta-frameworks depends on making their design logic explicit. A DSR perspective shows that the key contribution is not only the aggregation of AI governance instruments, but the justified transformation of principles into controls, evidence artefacts, and verification points. The resulting artefact functions as a translation layer between fragmented governance regimes and auditable organizational practice. Future work should empirically validate the artefact through ex-pert review, Delphi studies, or pilot audits with auditors, regulators, compliance officers, and developers.

References

1. European Parliament and Council: Regulation (EU) 2024/169 laying down harmonized rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union (2024)
2. ISO/IEC: ISO/IEC 42001:2023. Information technology - Artificial intelligence - Management system. International Organization for Standardization, Geneva (2023)
3. National Institute of Standards and Technology: Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1, Gaithersburg (2023)
4. Mittelstadt, B.: Principles alone cannot guarantee ethical AI. Nature Machine Intelligence 1, 501-507 (2019)
5. Esen, D.S.: Toward a Unified Meta-Framework for AI Auditing: A PRISMA Scoping Review of Post-2023 Governance Frameworks and Blind Spots. Conference paper presented at IFIP2025, Copenhagen (2025)
6. ISO: ISO 19011:2018. Guidelines for auditing management systems. International Organization for Standardization, Geneva (2018)
7. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design science in information systems research. MIS Quarterly 28(1), 75-105 (2004)
8. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A design science research methodology for information systems research. Journal of Management Information Systems 24(3), 45-77 (2007)
9. Gregor, S., Hevner, A.R.: Positioning and presenting design science research for maximum impact. MIS Quarterly 37(2), 337-355 (2013)