Keeping AI in Check: The SCARPI Process Model for Continuous Ethical Auditing of AI-Enabled IT Systems

Clemens Brackmann[1][0009-0002-6667-0165]

[1] University of Duisburg-Essen, Essen, Germany
clemens.brackmann@uni-due.de

Abstract. As artificial intelligence (AI) systems become deeply embedded in organizational IT infrastructures, the need for structured mechanisms to ensure ethical compliance has become urgent. Existing governance frameworks such as COBIT and ISO 19011:2018 provide foundation but lack procedures for continuously auditing AI Systems throughout their development. Through a systematic literature review across three major databases, this paper synthesizes existing guidance to address this gap proposing the SCARPI process model comprising Define Scope, Collect and Assess, Reflect and Plan, and Implement. Grounded in established IT governance and audit structures, SCARPI enables organizations to systematically identify and mitigate bias and fairness issues in AI systems through an iterative, stakeholder-informed approach. By integrating continuous auditing with transparency artifacts such as a Public Transparency Statement, SCARPI offers a concrete, governance-aligned path toward responsible and trustworthy AI deployment. This work represents a first step toward bridging the gap between AI accountability mechanisms and existing IT governance structures.

Keywords: AI auditing, Ethical AI, IT governance, Process model.

1 Introduction

Despite decades of artificial intelligence (AI) research, the widespread deployment of AI systems in real-world organizational contexts has introduced challenges that existing governance structures are ill-equipped to handle. As organizations increasingly adopt AI to drive growth and innovation, the risk of deploying unfair, biased, or non-compliant systems has grown in parallel [38]. This makes accountability and ethical oversight not merely desirable, but essential. In this context, audits play a crucial role in assessing AI systems in terms of their fairness and compliance with the law [25]. Implementing AI audits therefore not only contributes to ethical compliance; organizations can also benefit significantly in several other ways. By improving performance, gaining customer insights, and mitigating risks, organizations can harness the full potential of AI while safeguarding their interests and maintaining public trust [28].

The recent adoption of the EU AI Act underscores the growing demand for regulation and ethical standards in AI [8]. While AI technologies have existed for decades, the rise of generative AI has dramatically accelerated both adoption rates and public

awareness of the technology's potential. This rapid diffusion further intensifies the need for robust regulatory and ethical frameworks [36].

Regardless of the pace of adoption, ensuring trustworthy and responsible use of AI is critical to minimizing the risk of harm [42]. The current lack of process guidelines and best practices increases the risk of developing and deploying unfair and biased AI systems, since potential hazards can hardly be identified before deployment [11]. AI audits are well-suited instruments to address these issues, as one of their main objectives is to promote transparency, traceability, and accountability [29, 42]. However, existing AI auditing frameworks remain scarce. Much of the prior work has focused either on audit and governance structures within information systems without a clear focus on AI [23, 24], or on high-level AI principles [14, 18, 32] and technical solutions such as fairness metrics [11, 20].

A critical gap emerges from this review: most existing frameworks fail to address the integration of current IT governance and audit structures into the AI auditing process [38, 39, 49]. This is a significant shortcoming, as researchers argue that drawing on established governance solutions is fundamental when constructing audit structures for AI systems [49]. Given the complex and far-reaching implications of AI [38], any viable AI audit framework must address integration with existing governance structures while also considering multiple stakeholder groups to assess bias and fairness [3, 30]. Such an approach enables organizations to maintain public trust while promoting transparency and traceability throughout their product development lifecycles [19, 42]. Current governance models, however, do not support the continuous integration of audit systems into AI development; a gap that remains unaddressed in literature. This motivates the following research question:

How can IT systems utilizing AI continuously be audited regarding ethical considerations?

To answer this question, we construct a structured process model that bridges the missing alignment between existing IT governance structures and the specific demands of AI auditing. The process model form was chosen deliberately: unlike a set of static principles or a high-level framework, a process model provides actionable, step-by-step guidance that can be integrated into existing organizational workflows. The model is grounded in established frameworks and a systematic literature analysis following vom Brocke et al. [44].

2 Theoretical Foundation

A defining characteristic of AI systems, and the primary source of their audit complexity, is their degree of autonomy and self-learning agency. Unlike conventional IT systems, which execute predefined instructions without adaptation, AI systems analyse their environment and take actions autonomously to achieve specific goals [18]. This distinction is critical: it is precisely this autonomy that introduces the risk of undesirable, unpredictable, and potentially harmful outcomes that static IT governance structures are not designed to handle [18, 34]. While AI is applied across a wide range of industries, including pharmaceuticals, finance, and aerospace, no single universal

definition encompasses the full breadth of its applications [38]. For the purposes of this paper, we follow the European Commission's definition, which describes AI systems that display intelligent behaviour by analysing their environment and taking actions with some degree of autonomy to achieve specific goals [18].

In response to the risks associated with this autonomy, a set of AI principles has emerged from ethical AI discourse. These become apparent as best practices and rules intended to guide the responsible and trustworthy development and deployment of AI systems [12]. Regulatory bodies and leading organizations have each formulated their own versions of these principles as guidelines [14, 43, 48]. The ongoing legislative discussions around the EU AI Act make clear that principles alone are insufficient to guarantee accountability and ensure trust in AI systems [8]. A more comprehensive and enforceable approach is therefore needed.

Novelli et al. [33] provides a useful framework for understanding what full accountability in AI requires, identifying four interconnected mechanisms. First, achieving compliance requires that specific standards for design, development, and deployment are met [24]. Second, reporting demands that practices are in place to ensure explainability and justifiability in AI systems [29, 33, 36]. Third, oversight focuses on evaluating AI system performance across its lifecycle by identifying relevant facts and evidence [33, 45]. Fourth, enforcement addresses the consequences drawn from such evaluations, typically in the form of assessments or reviews highlighting accountability as a combined results [33]. Audits are particularly well-suited to fulfilling the oversight and enforcement dimensions of this accountability framework. Unlike voluntary principles, audits are systematic, independent, and documented processes for determining whether predefined audit criteria have been met, based on the collection and evaluation of relevant objects or evidence [21, 24]. Over recent years, multiple audit types have emerged, each with a distinct focus and domain [24, 39, 49], thereby expanding the range of functions audits can serve. Audits have consequently established themselves as highly effective accountability instruments [3].

When auditing AI systems, a broad range of components can be examined. In the IT domain, these include change management practices, data, service management, IT budget and costs, and incident response, among others [31]. Within the AI domain specifically, audit-relevant components extend to model features, model inputs and outputs, as well as contextual factors such as the cultural environment, first-party interpretation, and third-party understanding [28]. Since AI systems frequently operate as subsidiaries of larger IT systems, auditing an IT component will often simultaneously surface AI-related issues [9, 18]. Zinda [49] further argues that governance, assurance, and risk are central considerations during an AI audit, given that the audit process helps identify potential ethical issues, biases, and unintended consequences arising from data collection, algorithmic decision-making, and societal impact [3].

Conducting such audits is not without challenges. In some cases, access to AI system components is limited to the model output retrieved via an API, making deeper inspection difficult [4]. Components such as training data or model development processes are frequently classified as trade secrets and therefore inaccessible to external auditors [4]. These structural barriers underscore a fundamental problem: without a defined, continuous process for integrating audits into the AI development lifecycle,

accountability remains fragmented and reactive. The absence of a structured, continuous audit process that aligns with existing IT governance frameworks motivates the process model proposed in this paper.

3 Methodology

This paper follows the Design Science Research (DSR) paradigm as its overarching methodological foundation. DSR is concerned with the creation and evaluation of artifacts that solve identified organizational or sociotechnical problems [17]. Following the artifact taxonomy of Hevner et al. [17], SCARPI is classified as a *method artifact*: a process model that provides a concrete sequence of steps for achieving a specified goal. The design and development of SCARPI follows the DSR process model proposed by Peffers et al. [37], specifically the phases of problem identification and motivation, definition of objectives, design and development, and communication of results. Artifact evaluation, identified by Hevner et al. [17] as a core DSR activity, is acknowledged as a necessary next step and is discussed in the limitations of this paper.

To inform the design of the artifact, we applied a structured literature review as the primary knowledge base, following the approach presented by Webster & Watson [46] in combination with vom Brocke et al. [44]. The combination of both approaches ensures qualitative, traceable, and reproducible results. We conducted a systematic query across the three literature databases, ACM Digital Library, Scopus, and Web of Science, searching across title, abstract, and keywords. The initial search results can be seen in Fig. 1.

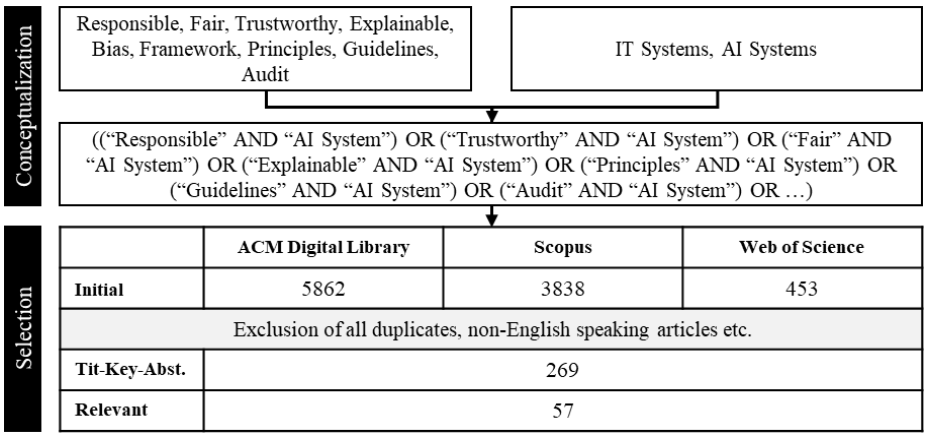

Fig. 1. Literature Review Overview.

The screening process proceeded in two stages. First, we removed duplicates as well as articles that were non-English, not peer-reviewed (including panels and commentaries), or unavailable in full text. Second, we applied quality criteria to the remaining set

following Bandara et al. [1], evaluating each article based on its title, abstract, and full content. The quality criteria were:

- Papers proposing guidance for auditing IT or AI systems,
- Papers presenting guidelines for developing fair IT or AI systems, and
- Papers comparing or evaluating existing frameworks related to auditing.

Following this filtering process, 57 papers were identified as relevant for further analysis. The synthesis of these papers revealed that existing literature predominantly offers high-level recommendations and general guidelines, but stops short of providing a precise, actionable procedure for continuously auditing AI systems. This finding directly motivated the design of SCARPI as a method artifact that operationalizes the accumulated guidance from the literature into a governance-integrated process model, thereby responding to both the practical gap identified in the problem statement and the DSR imperative to produce artifacts of demonstrable utility [17, 37].

4 Review of Existing Governance and Audit Frameworks

Existing IT governance and audit frameworks offer a valuable starting point for developing structured approaches to AI auditing. By examining these frameworks critically, organizations can identify transferable principles while also recognizing where adaptation is necessary to account for AI's unique characteristics.

4.1 Transferable Principles from IT Governance

Several core principles from established IT governance practice translate directly to the AI domain. Data integrity, for instance, is a central concern in frameworks such as COBIT [23] and becomes even more critical in AI contexts, where the quality and bias of training data directly affect model performance and fairness [40]. Similarly, transparency and traceability, which are foundational principles in both COBIT and ISO 19011:2018, are equally essential when governing AI systems, which are often complex and difficult to interpret [15]. The concept of explainable AI [7] has emerged as the AI-specific counterpart to these principles, reflecting the growing recognition that AI decision-making processes must be communicated in a clear and understandable manner. At the same time, not all IT governance principles extend seamlessly to AI. Certain aspects require tailored consideration due to AI's dynamic and evolving nature [6]. In particular, the explainability and interpretability of AI decision-making present challenges not typically encountered in traditional IT systems, where algorithms are comparatively transparent [20].

4.2 Strengths and Limitations of Existing Frameworks

Given that auditing frameworks for AI systems remain largely immature [26], adopting and adapting established frameworks such as ISO 19011:2018 and COBIT is widely recommended as a foundation for implementing effective AI auditing systems. Both

frameworks are widely recognized, have undergone multiple revisions since their initial development, and embody accumulated best practices for governing and auditing IT systems. ISO standards are formally reviewed every five years, while COBIT has evolved through six versions since its introduction in 1996, most recently as COBIT 2019 [23].

ISO 19011:2018 provides a broad auditing solution centred on a continuous auditing approach and a structured audit program, which aligns well with the iterative nature of AI development. However, its process flow remains highly general and requires further refinement to be applicable to the AI domain. COBIT, on the other hand, offers valuable governance and risk assessment objectives that complement ISO 19011:2018's processes well [47]. Yet COBIT does not adequately address the software development lifecycle [49] and provides limited guidance on execution. Moreover, fully complying with COBIT in agile software engineering environments introduces additional organizational challenges [35], as AI development is inherently iterative rather than waterfall-based [38]. In sum, neither framework alone is sufficient; both require supplementation with additional tools, procedures, and processes to be applicable to AI system auditing.

4.3 The Role of Stakeholders

An effective audit of AI systems demands the involvement of diverse stakeholder groups, both internal and external. Internal stakeholders typically include executive management, IT managers, business managers, and board members [23], while external stakeholders encompass state actors, external auditors, domain experts, and members of civil society [10]. The inclusion of external stakeholders is particularly important: as they are not directly involved in the development or operation of the system under review, they are positioned to conduct impartial assessments and provide objective insights into potential harms or unintended consequences [10, 47].

4.4 Limitations of Existing AI-Specific Process Models

Beyond IT governance frameworks, several AI-specific approaches have emerged, ranging from comprehensive governance frameworks and ethical principles [14, 18, 32] to algorithmic and statistical audit mechanisms [11]. However, existing literature has focused predominantly on rules, guidelines, and principles rather than structured process models [18, 39, 49]. Process models, by contrast, offer a concrete and implementable means of embedding audit and governance structures into organizational workflows [12, 49].

The few existing process models in this space exhibit notable limitations. The SMACTR framework proposed by Raji et al. [38] defines development phases and corresponding documentation artifacts but focuses primarily on internal algorithmic auditing, failing to account for the influence of diverse stakeholder groups or the integration of existing governance structures throughout the product development lifecycle. Other process models address the development and deployment of fair AI systems through statistical optimization techniques [5, 13, 20, 27], but similarly do not provide a comprehensive, governance-integrated auditing procedure.

Taken together, these findings point to a consistent gap: no existing framework combines a structured, continuous process model with integration of established IT governance principles and meaningful stakeholder involvement. It is precisely this gap that the process model proposed in this paper seeks to address.

5 The SCARPI Process Model

SCARPI, an acronym derived from its four core phases: *Define Scope*, *Collect and Assess*, *Reflect and Plan*, and *Implement*, was developed by aligning insights from existing audit and governance structures with the product development lifecycle of AI applications [38]. The model adopts a continuous audit program approach, reflecting that the review and monitoring of an AI system's performance is an ongoing, iterative task rather than a discrete event [24].

The core processes of SCARPI (see Fig. 2) are connected sequentially, with each phase feeding into the next through a set of documentation artifacts. These artifacts serve a dual purpose: they are stored in a shared repository to enhance transparency and traceability, and selected artifacts are communicated to relevant stakeholder groups to inform and guide the development and deployment of the AI application. Consistent with best practice in AI accountability, stakeholder involvement is not limited to internal parties, thus external stakeholders play an equally important role in ensuring impartial assessment [3, 30]. While conflicts may arise between internal and external stakeholders regarding the transparency of documents, the model allows for internal and external versions of the documented artifacts. The external versions (e) should be within the shared repository, while internal versions (i) are kept confidently due to trade secret protections.

5.1 Define Scope

This phase aligns with the audit program establishment process described in ISO 19011:2018 [24] and serves as the foundation for the entire audit cycle. Its primary purpose is to define the boundaries and objectives of the audit before any evidence is collected. Key elements to be documented at this stage include:

- **Objectives:** what the audit aims to achieve and the ethical or fairness concerns it seeks to address.
- **Criteria:** the standards, principles, or regulatory requirements against which the AI system will be evaluated.
- **Methods:** how the audit will be conducted, including planned techniques such as interviews, data analysis, or automated testing.
- **Team Composition:** who will be involved, including the roles and responsibilities of both internal and external participants.
- **Timeframe:** the planned duration and key milestones of the audit process.

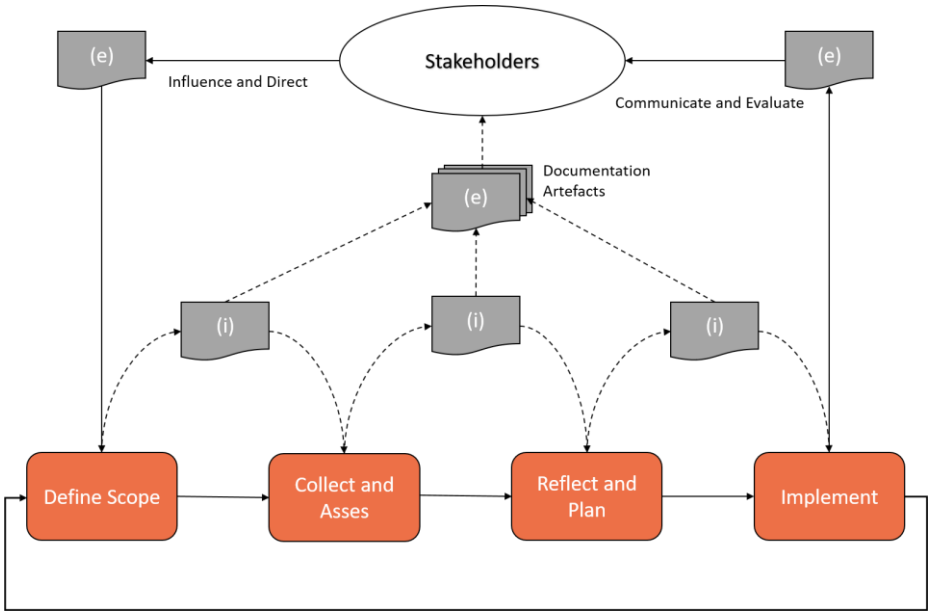

Fig. 2. The SCARPI Process Model with internal (i) and external (e) versions of shared documentation artefacts.

To establish meaningful objectives, it is essential to monitor industry trends and regulatory developments. Frameworks such as COBIT and its proposed expansions [23, 47] provide a useful starting point for identifying relevant risks, legal obligations, and governance concerns. Beyond current trends, this phase also requires anticipating potential harm and proactively investigating the AI system's behaviour. External stakeholders, including journalists, researchers, and end users, can provide valuable input in identifying such risks [3, 30].

It is important to note that this phase focuses on indications and observations rather than concrete evidence of harm. Ethical principles, such as those set out in the European Commission's guidelines for trustworthy AI [18], should guide not only the formulation of audit objectives but also the broader development process, ensuring ongoing alignment with established values. Supplementary governance mechanisms, such as an internal ethics review board, can further safeguard human rights and well-being throughout this phase [38].

5.2 Collect and Assess

This phase focuses on gathering and evaluating audit evidence: the information required to assess the AI system against the criteria established in the Define Scope phase. The appropriate scope and sample size of this evidence will vary depending on the specific audit context. The outcome of this phase is a set of audit findings, which serve as the primary input to the subsequent Reflect and Plan phase.

Assessing bias and fairness requires drawing on a diverse range of information sources. Internal stakeholders may be interviewed to surface past decision-making processes that shaped the AI system, as even seemingly minor design choices during development can have significant, difficult-to-measure consequences [38, 45]. This is referred to as first-party interpretation [28]. However, relying solely on internal perspectives or statistical methods is insufficient for evaluating complex AI systems, given the inherently political and socially embedded nature of AI [41, 49]. Third-party understanding, which includes perspectives from those external to the development process, must therefore also be incorporated.

Landers & Behrend [28] identify six key components of AI models that can serve as information sources during this phase: *Model Input, Model Output, Model Design, Model Development, Model Features*, and *Model Processes*.

When assessing bias, it is important to identify its specific type and origin, as bias can arise from multiple sources including training data, feature selection, and model design [16]. Modifying outputs or applying counteracting techniques should be approached with caution: each proposed change must be carefully evaluated, as interventions targeting one form of bias can inadvertently introduce or amplify others [16, 45]. Fairness and performance metrics provide structured means of evaluating the model's behaviour against the audit's predefined objectives and criteria.

5.3 Reflect and Plan

Text This phase focuses on analysing and interpreting the audit findings produced in the Collect and Assess phase. Following the approach set out in ISO 19011:2018, audit conclusions are drawn by systematically reviewing these findings against the objectives and criteria established during Define Scope. These conclusions address how well the audit objectives were met, the comprehensiveness of the audit scope, and the degree to which predefined criteria were fulfilled.

The conclusions then serve as the foundation for a mitigation plan, which prioritizes identified risks and proposes concrete design recommendations to improve the AI system's fairness and ethical compliance. COBIT's objective APO02 (Managed Strategy) provides a useful reference for structuring this planning process [23]. Determining the threshold at which an AI system presents an unacceptable ethical or fairness risk is a non-trivial challenge; guidance can be drawn from resources such as the European Commission's guidelines for trustworthy AI [18] or the ICO's Draft Guidance on AI and Data Protection [22].

A key transparency artifact produced during this phase is the algorithmic design history file (ADHF), a concept adapted from the SMACTR framework [38]. This file serves as an auditable record of the audit process, capturing the evaluation of findings, the rationale behind major decisions, and the progression of the mitigation plan. Within SCARPI, this artifact supports the principles of explainable AI and aligns with the BAI12 Managed Black Box objective proposed in the COBIT expansion by Witt et al. [47].

5.4 Implement

This phase, aligned with the act stage of the Plan-Do-Check-Act (PDCA) cycle and the follow-up concept in ISO 19011:2018 [24], focuses on executing the corrective actions identified in the mitigation plan. It is functionally analogous to error and exception management in COBIT's DSS06 objective (Manage Service Delivery and Support) [23]. The goal is to directly improve the AI system's perceived fairness and to reduce identified bias.

Optimizing the bias-variance trade-off is central to achieving statistical fairness. Commonly employed techniques include pre-processing, in-processing, and post-processing methods, each targeting different stages of the model development pipeline [2, 40]. The range of available options will vary depending on the specific model architecture and the degree of access the engineering team has to the system's internals. Throughout implementation, it is essential that all changes are guided by the prior phases' objectives, findings, and mitigation plan. Where further issues or unexpected side effects are discovered during implementation, these should be thoroughly documented and deferred to a subsequent audit iteration, as their full risk profile and impact cannot yet be adequately assessed [38]. If a critical failure breaching the risk threshold established during Reflect and Plan is identified, such as evidence of unlawful discrimination or a safety-relevant harm, the affected deployment must be halted or rolled back, with the authority to do so residing with a designated governance body such as the internal ethics review board or executive management. A new audit cycle is then initiated immediately rather than deferred, ensuring the system is reassessed before redeployment.

5.5 Post-Deployment: The Public Transparency Statement

Following the successful implementation of the mitigation plan, SCARPI prescribes the production of a Public Transparency Statement (PTS). This artifact is derived from the ADHF compiled during Reflect and Plan and is designed to communicate the AI system's audit progress to external audiences in a structured and accessible format, while excluding commercially sensitive information and trade secrets. That means, the PTS is the external (e) counterpart of the ADHF. It serves two core functions: enhancing public trust and demonstrating organizational transparency. As the EU AI Act increasingly mandates transparency from organizations deploying AI systems [8], the PTS provides a concrete mechanism for these obligations. Key components may include first-party interpretations of system behaviour, validation methods employed, fairness metrics in use and their improvement over time, and information on known residual biases currently under investigation. It is important to note that meaningful organizational transparency requires sustained commitment across multiple audit cycles; a single PTS is a starting point, not a destination [50].

6 Discussion

This paper advances IS governance research by demonstrating that the accountability gap in AI-enabled IT systems cannot be addressed by technical solutions or ethical

principles alone. By positioning AI systems as sociotechnical extensions of existing IT infrastructure rather than standalone entities, the paper builds a case for why IS governance frameworks such as COBIT and ISO 19011:2018 are not merely analogous to AI governance but are foundational to it. This reframing has implications beyond SCARPI itself: it suggests that the IS governance literature holds underexplored resources for the broader AI accountability discourse, and that future research at this intersection is both warranted and overdue.

Compared to the most closely related artifact in the literature, the SMACTR framework [38], SCARPI advances on three dimensions. First, it explicitly integrates with established IT governance objectives from COBIT and ISO 19011:2018, rather than proposing a standalone auditing procedure. Second, it incorporates both internal and external stakeholder groups as active participants throughout the audit process, rather than treating auditing as a purely internal engineering activity. Third, its continuous audit program structure reflects the iterative nature of AI development, whereas SMACTR is primarily oriented toward discrete pre-deployment review. Other process models in this space address fairness through statistical optimization techniques [5, 20] but do not engage with organizational governance structures at all, leaving precisely the gap that SCARPI is designed to fill. Taken together, these distinctions position SCARPI not as an incremental refinement of existing work but as a structurally different type of contribution. This treats the organizational and governance context of AI systems as a first-order concern rather than an afterthought.

Several limitations of this work must be acknowledged. Most critically, SCARPI has not yet been subjected to empirical evaluation. As a DSR method artifact [17], its utility, feasibility, and organizational fit remain to be demonstrated through real-world application. The model was derived from a synthesis of existing literature and governance frameworks, which ensures its theoretical grounding but does not substitute for practical validation. Further, SCARPI's current integration is primarily anchored in COBIT and ISO 19011:2018; organizations operating under different governance frameworks may require adaptation before the model can be applied effectively.

7 Conclusion

This paper has proposed SCARPI, a structured process model for the continuous ethical auditing of AI-enabled IT systems. The central objective of the framework is to foster a productive alignment between AI accountability mechanisms and established IT governance structures. Existing frameworks have thus far failed to formalize this relationship. By integrating AI auditing with the governance principles and audit processes embodied in COBIT and ISO 19011:2018, SCARPI enables organizations to proactively identify, evaluate, and mitigate the ethical risks of AI deployment, supporting a more secure, responsible, and trustworthy integration of AI into workflows.

The key contribution of SCARPI lies in its dual integration: it aligns with existing IT governance structures, ensuring organizational compatibility, while simultaneously extending these structures to address the specific ethical challenges posed by AI systems. The continuous audit program approach at the model's core reflects the iterative

nature of AI development and ensures that oversight is embedded throughout the product lifecycle rather than applied as a post-hoc measure. The inclusion of structured transparency artifacts, most notably the PST, addresses emerging regulatory demands under the EU AI Act and contributes to building broader societal trust in AI systems.

Future research should prioritize empirical validation through case studies within organizations operating under COBIT-aligned governance structures, as these would provide the most direct basis for assessing SCARPI's real-world utility. Beyond validation, the integration of additional governance frameworks, the examination of SCARPI's applicability across EU AI Act risk categories, and the standardization of the Public Transparency Statement as a regulatory reporting instrument each represent promising avenues for further inquiry. Ultimately, the convergence of IS governance principles and AI accountability is not an optional enhancement; it is a structural prerequisite for trustworthy AI at organizational scale.

References

1. Bandara W, Furtmueller E, Gorbacheva E et al. (2015) Achieving Rigor in Literature Reviews: Insights from Qualitative Data Analysis and Tool-Support. Communications of the Association for Information Systems 37. doi: 10.17705/1CAIS.03708
2. Briscoe E, Feldman J (2011) Conceptual complexity and the bias/variance tradeoff. Cognition 118:2–16. doi: 10.1016/j.cognition.2010.10.004}
3. Buolamwini J, Constanza-Chock S, Rajo ID (2023) Change from the outside: towards credible third-party audits of AI systems. In: Prud'homme B, Régis C, Farnadi G (eds) Missing links in AI governance. United Nations Educational Scientific and Cultural Organization; Mila - Québec Artificial Intelligence Institute, Paris, Montréal, pp 6–26
4. Burrell J (2016) How the machine 'thinks': Understanding opacity in machine learning algorithms. Big Data & Society 3. doi: 10.1177/2053951715622512
5. Deng, W. H., Nagireddy, M., Lee, M. S. A., Singh, J., Wu, Z. S., Holstein, K., & Zhu, H. (2022). *Exploring How Machine Learning Practitioners (Try To) Use Fairness Toolkits*. https://doi.org/10.1145/3531146.3533113
6. Dennehy D, Griva A, Pouloudi N et al. (eds) (2021) Responsible AI and analytics for an ethical and inclusive digitized society. 20th IFIP WG 6.11 Conference on e-Business, e-Services and e-Society, I3E 2021, Galway, Ireland, september 1–3, 2021, proceedings. Springer eBook Collection, vol 12896. Springer International Publishing; Springer, Cham
7. Doran, D., Schulz, S., & Besold, T. R. (2017). *What Does Explainable AI Really Mean? A New Conceptualization of Perspectives*.
8. European Parliament, & Council of the European Union. (2024, June 13). *Laying down harmonised rules on artificial intelligence (Artificial Intelligence Act).* European Parliament, Council of the European Union. https://eur-lex.europa.eu/eli/reg/2024/1689/oj
9. Falkenberg ED (1998) An Integrated Discipline Emerging. IFIP TC8/WG8.1 International Conference on Information System Concepts, 1st ed. Springer, Boston
10. Farnadi G (2023) The AI industry through the lens of ethics and fairness. In: Prud'homme B, Régis C, Farnadi G (eds) Missing links in AI governance. United Nations Educational Scientific and Cultural Organization; Mila - Québec Artificial Intelligence Institute, Paris, Montréal, pp 27–50
11. Feldman, T., & Peake, A. (2021). *End-To-End Bias Mitigation: Removing Gender Bias in Deep Learning*.

12. Floridi L, Cowls J (2019) A Unified Framework of Five Principles for AI in Society. Harvard Data Science Review. doi: 10.1162/99608f92.8cd550d1
13. Friedler SA, Scheidegger C, Venkatasubramanian S (2021) The (Im)possibility of fairness. Commun ACM 64:136–143. doi: 10.1145/3433949
14. Google (2025) AI Principles. https://ai.google/principles/
15. Gunning D, Stefik M, Choi J et al. (2019) XAI—Explainable artificial intelligence. American Association for the Advancement of Science
16. Hellström, T., Dignum, V., & Bensch, S. (2020). *Bias in Machine Learning -- What is it Good for?*
17. Hevner AR, March ST, Park J et al. (2004) Design Science in Information Systems Research1. MIS Quarterly 28:75–106. doi: 10.2307/25148625
18. High-Level Expert Group on Artificial Intelligence. (2019). *Ethics Guidelines for Trustworthy AI.* European Commission. https://www.europarl.europa.eu/cmsdata/196377/AI%20HLEG_Ethics%20Guidelines%20for%20Trustworthy%20AI.pdf
19. Hutchinson, B., & Mitchell, M. (2018). *50 Years of Test (Un)fairness*. https://doi.org/10.1145/3287560.3287600
20. Hutchinson B, Smart A, Hanna A et al. (2021) Towards Accountability for Machine Learning Datasets: Practices from Software Engineering and Infrastructure. In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. ACM, New York, NY, USA, pp 560–575
21. IEEE Computer Society (2008) IEEE Standard for Software Reviews and Audits. IEEE Std 1028-2008:1–53. doi: 10.1109/IEEESTD.2008.4601584
22. Information Commissioner's Office. (2023, March 15). *Guidance on AI and Data Protection.* Information Commissioner's Office. https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/
23. ISACA. (2018). COBIT 2019 Framework: Introduction and Methodology.
24. ISO (2018) Guidelines for auditing management systems 03.120.20(19011:2018). https://www.iso.org/standard/70017.html. Accessed 18 Mar 2026
25. Kazim E, Denny DMT, Koshiyama A (2021) AI auditing and impact assessment: according to the UK information commissioner's office. AI Ethics 1:301–310. doi: 10.1007/s43681-021-00039-2
26. Kokina J, Blanchette S, Davenport TH et al. (2025) Challenges and opportunities for artificial intelligence in auditing: Evidence from the field. International Journal of Accounting Information Systems 56:100734. doi: 10.1016/j.accinf.2025.100734
27. Kroll JA, Huey J, Barocas S et al. (2017) ACCOUNTABLE ALGORITHMS. University of Pennsylvania Law Review 165:633–705
28. Landers RN, Behrend TS (2023) Auditing the AI auditors: A framework for evaluating fairness and bias in high stakes AI predictive models. Am Psychol 78:36–49. doi: 10.1037/amp0000972}
29. Li B, Qi P, Liu B et al. (2023) Trustworthy AI: From Principles to Practices. ACM Comput Surv 55:1–46. doi: 10.1145/3555803
30. Madaio, M., Egede, L., Subramonyam, H., Vaughan, J. W., & Wallach, H. (2021). *Assessing the Fairness of AI Systems: AI Practitioners' Processes, Challenges, and Needs for Support*.
31. Majdalawieh M, Zaghloul I (2009) Paradigm shift in information systems auditing. Managerial Auditing Journal 24:352–367. doi: 10.1108/02686900910948198
32. Microsoft. (2022). *Microsoft Responsible AI Standard: General Requirements for External Release*. https://cdn-dynmedia-

1.microsoft.com/is/content/microsoftcorp/microsoft/final/en-us/microsoft-brand/documents/Microsoft-Responsible-AI-Standard-General-Requirements.pdf?culture=en-us&country=us
33. Novelli C, Taddeo M, Floridi L (2024) Accountability in artificial intelligence: what it is and how it works. AI & Soc 39:1871–1882. doi: 10.1007/s00146-023-01635-y
34. Nurkamid M, Jazuli A (2018) Information Systems for Accreditation Assessment Department base on The Web (Case Study: Informatic Engineering Department). In:
35. Ozkan N, Tarhan A, Gören B et al. (2020) Harmonizing IT Frameworks and Agile Methods: Challenges and Solutions for the case of COBIT and Scrum. In: Proceedings of the 2020 Federated Conference on Computer Science and Information Systems. IEEE, pp 709–719
36. Pant A, Hoda R, Spiegler SV et al. (2024) Ethics in the Age of AI: An Analysis of AI Practitioners' Awareness and Challenges. ACM Trans Softw Eng Methodol 33:1–35. doi: 10.1145/3635715
37. Peffers K, Tuunanen T, Rothenberger MA et al. (2007) A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems 24:45–77. doi: 10.2753/MIS0742-1222240302
38. Raji ID, Smart A, White RN et al. (2020) Closing the AI accountability gap. In: Hildebrandt M, Castillo C, Celis E et al. (eds) Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. ACM, New York, NY, USA, pp 33–44
39. Schöppl N, Taddeo M, Floridi L (2022) Ethics Auditing: Lessons from Business Ethics for Ethics Auditing of AI. The 2021 Yearbook of the Digital Ethics Lab:209–227. doi: 10.1007/978-3-031-09846-8_13
40. Sehra S, Flores D, Montanez GD (2021) Undecidability of Underfitting in Learning Algorithms. In: 2021 2nd International Conference on Computing and Data Science (CDS). IEEE, pp 591–594
41. Steinacker L (2022) Code Capital: A Sociotechnical Framework to Understand the Implications of Artificially Intelligent Systems from Design to Deployment. Code Capital: A Sociotechnical Framework to Understand the Implications of Artificially Intelligent Systems from Design to Deployment
42. Thiebes S, Lins S, Sunyaev A (2021) Trustworthy artificial intelligence. Electron Markets 31:447–464. doi: 10.1007/s12525-020-00441-4
43. Université de Montréal (2018) The Montréal Declaration for a Responsible Development of Artificial Intelligence
44. vom Brocke J, Winter R, Hevner A et al. (2020) Special Issue Editorial –Accumulation and Evolution of Design Knowledge in Design Science Research: A Journey Through Time and Space. Journal of the Association for Information Systems 21:520–544. doi: 10.17705/1jais.00611
45. Wang C, Wang K, Bian AY et al. (2023) When Biased Humans Meet Debiased AI: A Case Study in College Major Recommendation. ACM Trans Interact Intell Syst 13:1–28. doi: 10.1145/3611313
46. Webster J, Watson RT (2002) Analyzing the Past to Prepare for the Future: Writing a Literature Review. MIS Quarterly 26:xiii–xxiii
47. Witt BA, Wolanske SJ, Merhout JW (2021) I&T Governance Framework for Artificial Intelligence in Marketing: A COBIT 2019 Expansion. ISACA JOURNAL
48. Zhang L (2019) China: AI Governance Principles Released. Global Legal Monitor
49. Zinda N (2022) Ethics Auditing Framework for Trustworthy AI: Lessons from the IT Audit Literature. The 2021 Yearbook of the Digital Ethics Lab:183–207. doi: 10.1007/978-3-031-09846-8_12
50. Züger T, Pothmann D (2023) The AI Transparency Cycle. doi: 10.5281/ZENODO.8273113