IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, November 2025
Impact Factor: 7.67

Towards a Comprehensive ML Auditing Framework: Extending the Core Criteria Catalog

Padalkar Juie Shailesh[1], Khedkar Rushikesh Dyandeo[2], Londhe Neha Prafulla[3], Prof. Palve P. B[4]
Students, Department of Computer Engineering[1 2 3]
Professor, Department of Computer Engineering[4]
Adsul Technical Campus, Chas, Ahilyanagar, Maharashtra, India

Abstract: *As more ML technologies enter crucial sectors like medicine and governance, maintaining reliability, responsibility, and openness via audits is increasingly important. The document introduces an enhanced machine learning audit system constructed upon the fundamental standards outlined in Schwarz's research group. The statement has been restated in another manner without altering its core message: [2]. A preliminary investigation into "auditable artificial intelligence" was undertaken through systematic review of existing research, followed by in-depth interviews with machine learning practitioners, subject matter specialists, and audit professionals for input on this topic. The system enhances its database through inclusion of new aspects such as control over information rights, transparency in decision-making processes, reliability and safety measures, moral considerations, and ongoing surveillance capabilities. Additionally, we create an effective auditing form containing over forty specific inquiries tailored to meet those requirements. Next, our focus is on integrating an expanded audit process within the machine learning development cycle of enterprises, including its advantages, obstacles, and ongoing research efforts.*

Keywords*:* ML auditing, AI governance, auditing framework, explainability, robustness, ethics.

I. INTRODUCTION

Modern machine learning technologies have become prevalent in various industries including medicine, banking, and governmental services. These technologies offer potential benefits in terms of innovation and productivity but come with significant threats such as bias, lack of transparency, improper use of data, and system instability which erode confidence. Implementing robust audit procedures is crucial in order to manage potential threats associated with machine learning model evaluations.

The authors Schwarz et al. Their newly developed ML Auditing Core Criteria Catalog serves as a crucial foundational element based on insights gathered through a thorough qualitative research process. The catalogue categorizes audit-related matters under these main sections: Fundamental Concepts, Information Processing Techniques, and Evaluation Criteria. Nevertheless, although it excels in certain areas, the catalogue fails to adequately cover issues like data protection, continuous surveillance, and harmonizing multiple stakeholders' interests.

The document introduces an enhanced machine learning audit system built upon Schwarz et al. 's work. Her efforts have been recognized. We suggest refining audit standards through an integration of scholarly reviews, input from stakeholders, and textual examination by analyzing data systematically. Our goal is to assist machine learning experts, evaluators, and institutions in guaranteeing that their applications adhere to ethical standards, maintain transparency, and safeguard security.

II. RELATED WORK

1. The authors Schwarz et al. In 2024: A preliminary investigation into "auditable artificial intelligence" was carried out through thematic data analysis employing methods outlined by Mayringet al. , supplemented by additional insights provided in Bortz and Döring's work; this led to the creation of a comprehensive set of thirty questions for an audit

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 5, November 2025 Impact Factor: 7.67

framework focused on ensuring transparency within AI systems. The product directory comprises sections labeled as Basic Concepts, Information Theory, and Evaluation Criteria. Mendeley.

2. The topic of discussion revolves around establishing standardized for conducting AI audit procedures. To illustrate, consider Manheim et al. Suggest creating an AI Auditing Standards Council to guarantee uniform and dynamic auditing procedures. The arXiv repository is an online platform for sharing scientific papers in various fields of study.

3. Value-driven assessments focus on stakeholders' input for implementing ethical within machine learning models. The arXiv is an online repository for scientific papers in various fields of study.

4. Analyzing Data Usage: Current research focuses on auditing models for their use of personal data through member identification techniques, tackling issues related to data sovereignty and user privacy. The arXiv is an online repository for scientific papers.

5. Accessibility of audits: Casper & Co. Assert that complete reliance on opaque methods falls short in thorough evaluations; advocate instead for both transparent and innovative approaches such as examining internal systems and reviewing documents. The arXiv repository is an online platform for sharing scientific papers in various fields of study.

III. METHODOLOGY

Our methodological approach consists of the following steps:

A. Scoping Study

- Conducting thorough research across various academic journals by using keywords like "Auditable Artificial Intelligence", "Machine Learning Audit Frameworks", and "Governance Audits of AI".
- Inclusion of peer-reviewed papers, white papers, standards documents

B. Qualitative Content Analysis

- Employing an inductive approach, I analyzed specific texts through schemas derived from Mayring's 2000 work and Miles & Huberman’s 1994 theory.
- Identification of themes, categories, and sub-dimensions

C. Stakeholder Consultation

- Engage in organized discussions or training sessions with: software engineers specializing in machine learning algorithms, subject matter specialists such as those working within health care industries, internal compliance officers responsible for auditing operations internally, and independent verification professionals tasked with overseeing financial audits externally.
- Employ the Delphi approach or an analogous process for achieving criterion refinement through collaboration.

D. Drafting the Extended Criteria Catalog

- Suggest revised standards grounded in research findings combined with feedback from stakeholders.
- For each criterion, design audit questions
- Validate draft with stakeholders and refine

E. Future Pilot / Validation

- Implement the use of this framework across at least two practical machine learning applications.
- Collect opinions, assess functionality, and continuously refine the collection system.

IV. EXTENDED ML AUDITING CRITERIA FRAMEWORK

Our proposed framework encompasses eight key aspects:

1. Conceptual Basics

- Governance & Accountability
- Stakeholder Values & Ethics

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 5, November 2025 Impact Factor: 7.67

- Legal & Regulatory Compliance

2. Data Ownership & Privacy

- Data Consent & Ownership
- Data Lineage & Provenance
- Anonymization / Pseudonymization

3. Model Design & Data Quality

- Data Representativeness & Bias
- Feature Engineering & Selection
- Fairness Constraints & Techniques

4. Assessment Metrics

- Performance (Accuracy, Calibration)
- Fairness (Group / Individual)
- Interpretability & Explainability

5. Robustness & Security

- Adversarial Robustness
- Resilience to Data Drift
- Model Fallback / Recovery Mechanisms

6. Transparency & Explainability

- Post-hoc Explanations (SHAP, LIME, Counterfactual)
- Decision Documentation & Traceability
- User-facing Explanation Interfaces

7. Continuous Monitoring & Certification

- Audit Logging & Trails
- Retraining / Retriggering Audits
- External Certification or Internal Compliance

V. AUDIT QUESTIONNAIRE (SAMPLE)

Here is an example of auditable inquiries corresponding to those specified in the preceding categories:

Data Ownership & Privacy:

1.Whose legal rights pertain to the information utilized within your machine learning framework?
2.Are there any records of individual users' permission granted for collecting and utilizing their information?
3.Are pseudonymization or anonymization mechanisms implemented?

Model Robustness:

1.Is it true that the model has undergone testing in relation to deceptive inputs or exceptional scenarios?
2.What methods evaluate appropriate conduct in situations where changes occur due to alterations in data inputs?
3.Wouldn't it be helpful to have an alternative option available in case something goes wrong with the system?

Continuous Monitoring:

1.Is every interaction tracked securely within an encrypted record of activities?
2.When there is an instance where the model's performance declines or shows signs of deviation over time?
3.Is there regular scrutiny of audit records conducted either internally through an auditing group or externally via another entity?

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, November 2025
Impact Factor: 7.67

VI. DISCUSSION

A. Benefits

- Offers an effective, organized resource consisting of a questionnaire along with evaluation standards which organizations may employ throughout their project phases.
- Frequently overlooked by solely technical evaluations is an emphasis on stakeholders' perspectives regarding ethics and data sovereignty.
- Promotes ongoing scrutiny beyond single audits, thereby enhancing credibility and security.

B. Limitations

- Utilizing the entire system might require substantial effort and specialized manpower, particularly in cases of small-scale entities.
- Certain standards like ethics compatibility and clarity vary widely in their acceptance among individuals due to inherent subjectivity.
- Regular modifications of the system's efficacy hinge upon technological advancements and regulatory shifts over time.

C. Challenges

- Getting buy-in from diverse stakeholders (engineers, auditors, leadership)
- Embedding audit processes smoothly within the machine learning development cycle while maintaining agility in technological advancement.
- Balancing depth of audit with cost and scalability

VII. FUTURE WORK

1. Implementing & Testing: Integrate the system into practical machine learning applications such as those found in health care and finance sectors, gather information, and optimize performance.
2. Developing tool enhancements—such as online monitoring interfaces, automatic verification forms, and log recording mechanisms—is crucial for ensuring robust auditing capabilities.
3. Proposing an Audit Execution Governance Model: Define who conducts audits, their frequency, and ensure certifications.
4. Continuous refinement through iterative processes utilizes initial assessments for improvement based on user insights and project stakeholders' contributions.

VIII. CONCLUSION

This study expands upon the existing core machine learning audit framework introduced by Schwarz et al. Transform it into an enriched, layered perspective. Integrating features such as transparency, control over data assets, resilience against attacks, ethical considerations, and ongoing surveillance will create an actionable audit solution tailored to user requirements. The extensive survey comprising more than forty inquiries facilitates ML team navigation throughout the development, testing, implementation, and oversight stages. Our objective is for this blueprint to foster dependable, open, and responsible AI technologies—encouraging scholars and professionals to implement, evaluate, and continually improve upon its design.

REFERENCES

[1]. M. Schwarz, L. C. Hinske, U. Mansmann, F. Albashiti, “Designing an ML Auditing Criteria Catalog as Starting Point for the Development of a Framework,” *IEEE Access*, vol. 12, pp. 39953–39967, 2024.

[2]. D. Manheim, S. Martin, M. Bailey, M. Samin, R. Greutzmacher, “The Necessity of AI Audit Standards Boards,” arXiv, 2024.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, November 2025
Impact Factor: 7.67

[3]. M. Yurrita, D. Murray-Rust, A. Balayn, A. Bozzon, "Towards a multi-stakeholder value-based assessment framework for algorithmic systems," arXiv, 2022.

[4]. Z. Huang, N. Z. Gong, M. K. Reiter, "A General Framework for Data-Use Auditing of ML Models," arXiv, 2024.

[5]. S. Casper, C. Ezell, C. Siegmann, …, M. Von Hagen, "Black-Box Access is Insufficient for Rigorous AI Audits," arXiv, 2024.

[6]. D. Leocádio, L. Malheiro, J. Reis, "Artificial Intelligence in Auditing: A Conceptual Framework for Auditing Practices," *Administrative Sciences*, vol. 14, no. 10, 2024.