Conceptual Foundations for the Design and Implementation of a Comprehensive Framework for Auditing Artificial Intelligence Systems

M'hamedi Fattoum, Accounting and Auditing, University of Blida 2 - Lounici Ali ,Blida (Algeria) mfetoum @yahoo.fr

Received: 26/02/2026 *Accepted: 14/09/2026* *Published:30/09/2026*

Abstract:

This research aims to generate valuable insights for both academic scholarship and professional practice, contributing to the advancement of understanding in AI auditing by developing adaptive methodologies that integrate technical and ethical dimensions.

The study found that effective AI auditing requires clear and unified standards, active stakeholder collaboration, access to relevant information, systematic documentation, and audit teams with specialized competencies. Audits should be conducted independently to ensure objectivity, while audit priorities and procedures must be continuously updated to maintain reliability, strengthen controls, monitor risks systematically, and adapt to evolving operational conditions.

Keywords: Artificial Intelligence (AI); auditing; AI auditing methodologies.

Jel Classification Codes: O33, M42, M48

1. INTRODUCTION

Artificial intelligence has emerged as a principal driver of Industry 5.0, enabling innovative solutions to complex analytical and operational challenges. As organizations increasingly integrate AI technologies into core business processes, significant risks have arisen, including algorithmic bias, limited transparency, privacy erosion, and potential infringements of individual rights. These concerns underscore the critical importance of AI system auditing, which poses distinctive challenges due to the inherent complexity, opacity, rapid evolution, and deep organizational integration of AI technologies. Accordingly, AI system auditing represents an evolving field that requires a rigorous and structured methodology based on sequential steps integrating both technical and ethical dimensions. It entails comprehensive verification of the implementation of appropriate controls and safeguards throughout the entire system lifecycle to ensure compliance with applicable legal frameworks, established trust and safety standards, and evolving societal expectations.

The primary objective of this research is to generate valuable insights for both academic scholarship and professional practice, contributing to a deeper understanding of how AI systems can be effectively audited. It aims to develop adaptive methodologies that integrate both technical and ethical dimensions, thereby providing auditors with a flexible framework for conducting customized audits. Furthermore, the study proposes a methodology that combines best practices with a structured roadmap grounded in standardized and context-sensitive evaluation frameworks. These frameworks can then be operationalized into robust, validated, and open-source methodologies and tools, systematically developed and embedded within a coherent and standardized audit infrastructure, rather than being created on an ad hoc basis.

- **Methodology:** This study adopts an exploratory analytical approach grounded in a comprehensive review of recent literature to examine current AI auditing practices, identify key auditability metrics, and determine the competencies required to conduct effective audits of AI systems.

- Hypotheses:

- The availability of comprehensive auditability requirements enhances auditors' ability to conduct systematic and reliable audits of AI systems.
- Adopting clear AI auditing standards, supported by stakeholder collaboration and access to relevant information, enhances transparency, auditability, and accountability.
- The presence of specialized and integrated competencies within AI audit teams positively influences the effective implementation of audit procedures.
- Conducting audits through independent external parties or organizationally autonomous internal units enhances the impartiality and objectivity of audit outcomes.
- The continuous updating of AI auditing priorities and procedures improves audit reliability and aligns audit objectives with dynamic changes in the business environment.

2. The Conceptual Framework for Auditing Artificial Intelligence Systems

2.1 Definition of Artificial Intelligence:

2.1.1 Artificial Intelligence in Organizations: Artificial Intelligence (AI) has fundamentally transformed organizational decision-making and control structures. Ranging from predictive analytics to generative models such as ChatGPT, AI systems are now deployed across diverse domains, including customer service automation, fraud detection, financial forecasting, and compliance monitoring. While these technologies enhance operational efficiency, flexibility, and scalability, they also introduce systemic risks—such as algorithmic bias, model opacity, and limited explainability—that challenge conventional risk management and governance mechanisms. The use of AI in high-stakes decision contexts has raised significant ethical, legal, and technical concerns, particularly with respect to accountability.

Scholars caution that organizational adoption of AI often outpaces the development of sufficient internal oversight mechanisms. Consequently, assurance functions, including internal audit, must fundamentally reconsider traditional auditing practices, especially when evaluating systems whose behavior may be unpredictable and not fully interpretable (Jiun & Razli, 2025, p. 2286).

Artificial intelligence (AI) can be defined as computer-based systems capable of simulating human cognitive functions by analyzing large volumes of data, identifying patterns, and performing tasks such as reasoning, problem-solving, and decision-making with varying degrees of autonomy. Within organizations, AI is widely used to support planning, organizing, controlling and leadership functions. Early AI applications, such as expert systems and knowledge management systems, primarily supported human decision-making and exhibited limited autonomy. However, advances in real-time data collection and processing have significantly expanded both the scope of AI applications and their operational independence. Today, AI systems autonomously carry out tasks such as recruitment, task allocation, and employee performance evaluation, often with minimal human intervention.

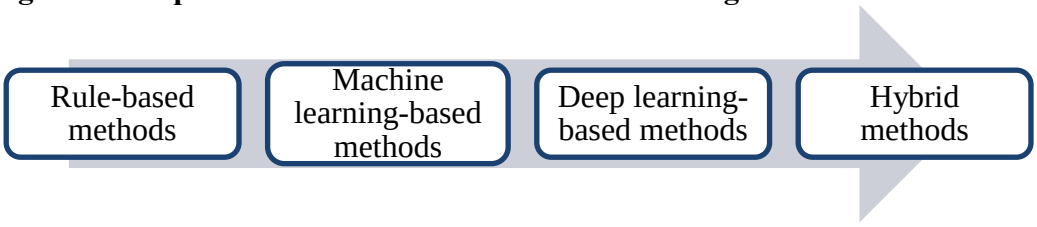
2.1.2 Types of Artificial Intelligence Techniques: AI techniques can be classified into the following categories:

- **Machine Learning (ML):** A branch of AI focused on developing algorithms capable of learning from data and progressively improving their performance. It includes supervised learning, unsupervised learning, and reinforcement learning.
- **Neural Networks & Deep Learning:** A branch of machine learning comprising algorithms that mimic the structure and function of the human brain, with deep learning as a subset focused on multi-layered artificial neuron architectures.

- **Data Mining, Knowledge Discovery and Advanced Analytics:** These techniques focus on extracting meaningful insights from large datasets to inform predictions and support decision-making.
- **Rule-Based Modeling and Decision-Making:** Systems that utilize a predefined set of rules to analyze input data and generate corresponding decisions are referred to as rule-based systems.
- **Fuzzy Logic-Based Approach:** A mathematical logic framework designed to handle uncertainty and imprecision in data, where reasoning is based on degrees of truth rather than binary values, characterizing the fuzzy nature of the information.
- **Knowledge Representation, Uncertainty Reasoning, and Expert System Modeling:** Techniques that structure knowledge for machine use and enable reasoning under uncertainty.
- **Case-Based Reasoning:** Solving new problems by comparing them with similar cases stored in a database of past experiences.
- **Text Mining and Natural Language Processing:** Used to analyze and understand linguistic data such as text and speech.
- **Visual Analytics, Computer Vision, and Pattern Recognition:** Techniques for analyzing visual data, including images and videos.
- **Hybridization, Searching and Optimization:** Methods that combine multiple AI techniques to solve complex problems (Kaur, 2024, p. 126).

2.1.3 **Techniques for Classification in Artificial Intelligence:** Figure 1 illustrates the main classification methods in **AI**:

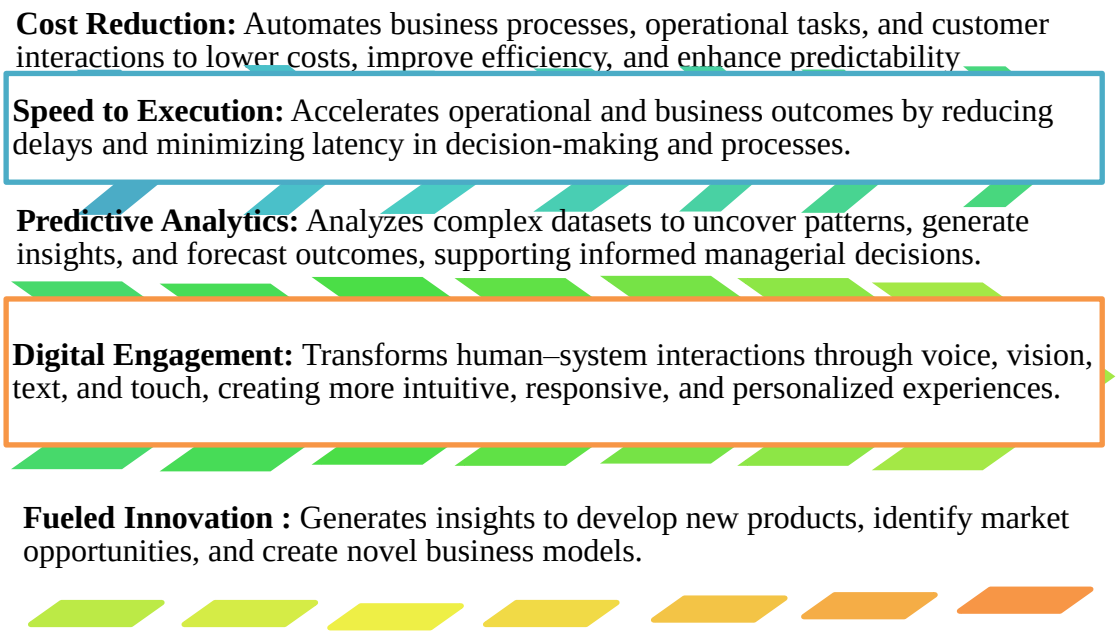
Fig.1. Techniques for Classification in Artificial Intelligence



Source: Built by the author based on: (Kaur, 2024, p. 126)

2.1.4 How Organizations Use Artificial Intelligence to Create Value: Organizations increasingly leverage artificial intelligence (AI) as a strategic resource to enhance value creation across multiple dimensions, as illustrated in the following figure:

Fig.1. How Organizations Use Artificial Intelligence to Create Value



Source: Built by the author based on: (Calagna, Cassidy, & Park, 2021 , p. 3)

2.2 Auditing Artificial Intelligence Systems:

Auditing artificial intelligence (AI) systems has gained increasing attention among researchers, professional service firms, and policymakers, particularly following the foundational study by Sandvig et al. (2014) on algorithm auditing. Since then, a growing body of academic research has emphasized the role of auditing in identifying and mitigating the ethical, legal, and technical risks associated with AI systems. For example, ProPublica's audit of the recidivism prediction algorithm used by the U.S. justice system—the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) system—revealed systemic racial bias against defendants of color. Similar instances have emerged across multiple

sectors, and when combined with the increasing ubiquity of AI in diverse applications—including safety-critical systems—these cases have intensified discussions on the necessity of auditing and ensuring the auditability of AI. Moreover, regulatory initiatives have been introduced globally to address growing concerns regarding the potential misuse and harm of AI, particularly in high-risk domains such as finance, insurance, education, employment, law enforcement, and healthcare (Himanshu, Padh, & Thelisson, 2025, p. 2). In parallel, a specialized AI auditing ecosystem has emerged in practice, with major professional firms—such as EY (2018), Deloitte (2020), KPMG (2020), and PwC (2020)—providing audit and assurance services aimed at ensuring the responsible design and deployment of AI systems. Overall, AI auditing is increasingly conceptualized as an independent mechanism for enhancing trust and transparency, analogous to financial auditing, by evaluating not only the technical performance of AI systems but also their compliance with organizational policies and regulatory requirements (Mökander, 2023, p. 2).

2.2.1 Artificial Intelligence Regulations and AI Auditing Frameworks: Regulatory bodies in the United States, Europe, and other regions are increasingly seeking to develop frameworks and policies to govern the use of artificial intelligence (AI) systems and ensure their reliability. In the United States, the National Institute of Standards and Technology (NIST) is developing the AI Risk Management Framework (AI RMF), while the Institute of Internal Auditors (IIA) has proposed a framework for auditing AI systems with a particular focus on governance and human factors. In addition, Executive Order 13960 establishes principles for the use of AI within the public sector. The Committee of Sponsoring Organizations of the Treadway Commission (COSO), in collaboration with Deloitte, has also developed guidance on AI governance and risk management. In Europe, the European Union's Artificial Intelligence Act (AI Act) stands out as the first comprehensive legislative initiative to classify AI systems according to risk levels, with the aim of ensuring the safety of high-risk systems. This is complemented by additional guidance issued by the Information

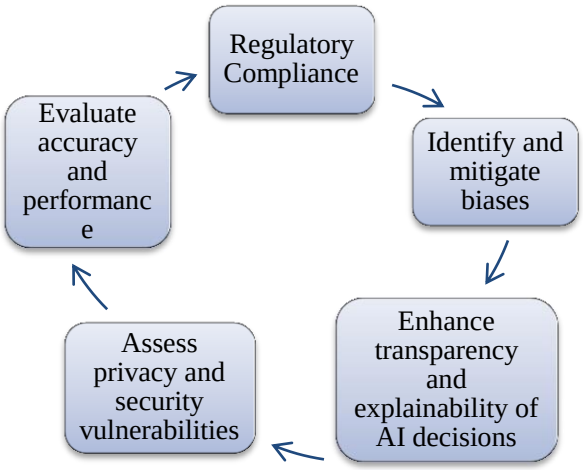
Commissioner's Office (ICO) and the European Commission's High-Level Expert Group on AI. Other countries, such as Canada, Singapore, and China, have likewise developed frameworks addressing AI governance, accountability, and ethical compliance. At the international level, the Organisation for Economic Co-operation and Development (OECD) has issued principles for trustworthy AI (Li & Goel, 2025, p. 5.6). Another recent example of AI regulation is Australia's proposed Mandatory Guardrails for High-Risk AI and Voluntary AI Ethics. Although differing in legal enforceability and scope, both frameworks emphasize transparency, accountability, and auditability through measures such as risk management, documentation and recordkeeping, data governance, human intervention and oversight, and conformity assessments (Himanshu, Padh, & Thelisson, 2025, p. 5). Most regulatory initiatives emphasize accountability, transparency, fairness, safety, and privacy, while comparatively less attention is devoted to data quality and traceability. Regulatory priorities also vary across jurisdictions; for example, the United States places greater emphasis on explainability, whereas this aspect receives less focus in Europe. Such divergence has contributed to the absence of international consensus on ethical principles and auditing standards. Recent literature increasingly highlights ethical AI auditing as a mechanism for enhancing accountability and bridging the gap between ethical principles and practical implementation, viewing trust in AI as the outcome of multi-level governance involving governmental oversight, external auditing, and the efforts of professional and civil society organizations (Li & Goel, 2025, p. 7).

2.2.2 Definition of AI Auditing: AI auditing involves the examination of evidence of a process or activity, like financial transactions or an engineering process, and then evaluation of the evidence against some standards or metrics, which could be a regulation or standards regime, or internal management metrics. Conducting an audit requires the availability of auditable artifacts that document decisions, systems, and processes. Auditing also relies on non-technical governance to ensure consistency with relevant principles or norms, and the independence of the assessment and

audit is used to ensure transparency and places liabilities on both the parties assessing and the assessed parties (Ayling & Chapman , 2022, p. 408). AI auditing is considered a subfield of information technology (IT) auditing that typically focuses on identifying risk factors and evaluating control mechanisms. According to the Institute for Electrical and Electronics Engineers (IEEE), an audit is defined as "an independent evaluation of conformance of software products and processes to applicable regulations, standards, guidelines, plans, specifications, and procedures." Accordingly, an IT audit typically involves the systematic collection of evidence related to risks and controls and assesses an organization's technological infrastructure at a specific point in time to ensure the accuracy, efficiency, security, and regulatory compliance of its systems and processes (Laine, Minkkinen, & Mäntymäki, 2024, p. 1). AI auditing can be defined as a systematic and comprehensive examination of the design, training, deployment, and operational processes of artificial intelligence systems, conducted to assess their compliance with applicable regulations, internal policies, governance frameworks, and ethical standards. This evaluation encompasses the examination of data quality, model architecture, result interpretability, and decision-making processes, with the objective of ensuring transparency, accountability, effective risk management, and the safe and responsible use of AI systems.

2.2.3 Objectives of Artificial Intelligence Auditing: The following figure illustrates the objectives of artificial intelligence auditing:

Fig.4. Objectives of Artificial Intelligence Auditing



Source: Built by the author based on: (Sharma, 2025, p. 3485)

2.2.4 Classifications of AI Systems Auditing: The results of current practices in auditing artificial intelligence systems indicate that AI auditing can be classified into four main types:

- **Ethics-based auditing:** This type focuses on issues such as model bias, fairness, explainability, transparency, privacy, security, and safety.
- **Policy and regulatory compliance auditing:** In this approach, auditors assess the extent to which AI systems comply with relevant regulations, standards, or industry norms.
- **Process and governance auditing:** This type examines AI systems across the stages of design, development, deployment, and operation.
- **Quality assurance auditing:** This focuses on evaluating the reliability, validity, and accuracy of AI system outputs.

In addition, within the context of AI systems auditing, a distinction can be made between **technology-oriented audits**, which concentrate on examining the technical characteristics and capabilities of AI systems, and **process-oriented audits**, which focus on governance structures and quality management systems adopted by technology providers.

2.2.5 Categories of AI Auditors: Three relevant categories of auditors emerge (Farley & Lansang, 2025, p. 9.13):

- **Internal Auditors:** Employed within the organizations they assess, internal auditors benefit from close access to AI systems and direct collaboration with other departments, which mitigates concerns over trade secret disclosure. However, internal audits face limitations, including the absence of standardized procedures, insufficient independence from the auditee, and variability in standards across companies. Consequently, audits may selectively evaluate only certain aspects of AI systems, results may go unpublished, and auditors may be incentivized to report favorable outcomes, potentially undermining the effectiveness and transparency of internal oversight.
- **Third-Party Auditors:** Third-party auditors operate independently of the organizations they assess, often functioning as researchers examining the societal impacts of AI. While their independence represents a key advantage over internal auditors, they face significant constraints, including limited access to data and algorithmic systems, potential obstruction by the audited entities, and a narrowly defined scope shaped by their own research objectives. Legal and technical barriers—such as restrictions imposed by Terms of Service and the ambiguous application of laws like the Computer Fraud and Abuse Act (CFAA)—further impede their efforts, potentially diminishing the effectiveness and overall impact of external AI audits.
- **External Auditors:** External auditors are independent third-party entities engaged to evaluate AI systems without being involved in the organization's daily operations. They typically have greater access to AI systems than independent auditors, while maintaining more detachment than internal auditors, allowing them to assess compliance with legal requirements, fairness standards, and organizational policies. External audits can enhance reputational

credibility and verify system accuracy. However, they face limitations, including potential conflicts of interest driven by reputational or financial considerations, the absence of mandated auditing standards, and the risk of "audit washing," where auditors may produce narrow or favorable findings to secure future engagements.

2.3 Challenges in Auditing Artificial Intelligence: Auditing AI systems faces several practical challenges, as highlighted in the literature. The most significant challenges include:

- Companies that hold intellectual property rights over their AI systems are often reluctant to permit external audits, primarily to safeguard proprietary assets and confidential information.
- Auditors frequently face difficulties in accessing AI systems for testing and evaluation purposes, as responsible entities tend to restrict transparent or easy access to their systems.
- High monetary and time costs associated with conducting external AI system audits represent a major barrier to their broader adoption.
- The limited availability of specialized tools hinders the effective execution of AI auditing processes (FERNANDES, 2021, p. 54).
- Challenges Inherent to Self-Learning Systems and Large Language Models: Self-learning AI systems, particularly those employing reinforcement learning, pose significant auditing challenges due to their continuous adaptation, dependence on uncurated or biased data, and vulnerability to model drift and adversarial attacks. Notable cases, such as Microsoft's Tay chatbot, demonstrate how these systems can rapidly reproduce harmful or biased behaviors. Proprietary Large Language Models (LLMs) further exacerbate auditability difficulties due to restricted access to training data and provider-imposed limitations. Without intentional design choices that enhance auditability—such as documentation, Model Cards, or audit APIs—these systems present elevated risks of undetected bias and harm, complicating efforts to ensure effective and safe auditing

(Kilian, Jäck, & Ebel, 2025, p. 3).

3. Foundational Principles for the Development of Comprehensive AI Systems Audit Procedures:

3.1 Key Requirements for Auditing Artificial Intelligence Systems:

3.1.1 AI Auditability: refers to the inherent characteristics, technical and organizational processes, and design features of an AI system that enable effective auditing. It encompasses the availability, accessibility, and traceability of data, decisions, and operational processes necessary for systematic evaluation. Auditability allows independent third parties to examine, monitor, and assess an AI system's behavior by ensuring adequate disclosure of relevant information, thereby facilitating the identification of risks, biases, and errors. A high level of auditability ensures that AI systems can be evaluated comprehensively and reliably, supporting external oversight, strengthening accountability, and fostering stakeholder trust. Unlike transparency and explainability, auditability concerns the system's capacity to be independently assessed for compliance with predefined standards, even when the system itself is not fully transparent or easily interpretable (Himanshu, Padh, & Thelisson, 2025, p. 3).

AI systems consist of five core components: training data, the model, processes, governance, and people, each requiring specific auditability measures. Data auditability entails ensuring data accessibility, documenting data characteristics, and maintaining quality assurance. Model auditability requires thorough documentation, effective monitoring, and enhanced explainability and transparency. Process auditability depends on comprehensive logging throughout the AI lifecycle. Governance auditability involves establishing clear internal and external rules that regulate AI operations and ensure quality. People-related auditability requires clearly defined roles and responsibilities of key stakeholders involved in the development, deployment, and monitoring of the AI system (Li & Goel, 2025, pp. 28,29). Each of these five AI components demands a distinct set of auditability characteristics to be implemented, as summarized in the following table:

Table 1. Auditability Measures of AI System Components

Component for Audits	Auditability Measures
1 Training data sets	
▪ Data accessibility	➢ Ensure availability and accessibility of training data.
▪ Data characteristics	➢ The description of the data, including the features and uses of the training data, should be well captured.
▪ Data quality	➢ Ensure the quality of training data, and provide quality metrics used to measure data quality.
▪ Security and privacy considerations	➢ Security and privacy considerations need to be given early during model design and development.
2 The underlying model	
▪ Model intended use	➢ Well-documented intended use including the AI system scope and the domain companies and contexts to which the AI system applies.
▪ Model explainability and transparency	➢ Implement explainability features in the model. ➢ As the black-box nature is not acceptable, well commented sections on AI algorithms, automated information tracking on model performance and user behavior within the system, or other methods that enhance explainability can be implemented to explain highly complicated models.
▪ Model selection and model values	➢ Decision-making on model selection and the value of the model needs to be documented. ➢ The audited party needs to describe the level of automation for the underlying model.
▪ Model operations	➢ Description of the model characteristics, model performance, and the performance evaluation metrics on an ongoing basis. ➢ Description of the causal relationships. ➢ Records of historical performance results and risk analyses results of the model throughout the development phases.
▪ Model deployment and monitoring	➢ Documentation of how the model is embedded into the overall product and process. ➢ Documentation of how the model is monitored over the lifetime of the AI system. ➢ Description of the measures expected to be used in terms of retraining the model and figuring out when the model fails.
3 4 5 Process, governance, and people	

➢ Data process and governance	➢ Documentation and logs should be maintained for how the data was sourced, evaluated, cleaned, used, and stored. ➢ Data privacy and security need to be well-governed with robust external privacy and security standards or internal rules and criteria. ➢ Quality measures for training data should be well developed and stated, especially when the data are purchased from external providers.
➢ Management governance structure	➢ Well-documented governance standards and continuous assurance and compliance model. ➢ Documented information about management governance structure. ➢ Documentation should include people who are responsible for each phase, people who authorized different deliverables, the decision escalation process from developers to C-suites, and the performance metrics and thresholds.
➢ Model governance	➢ Well-documented governance structures around model development, including how the model is being selected, fit, managed, restricted, and deployed.

Source: (Li & Goel, 2025, p. 29)

3.1.2 Knowledge and Skills Required for AI Audits: The audit team should possess a comprehensive set of core competencies, including (Li & Goel, 2025, pp. 34,35):

- **Knowledge of Compliance Requirements:** AI auditors must have a strong understanding of the legal, regulatory, and ethical requirements relevant to AI systems.
- **AI Knowledge:** AI auditors should possess both foundational and practical knowledge of artificial intelligence concepts and technologies.

- **Knowledge of the Underlying AI Model Being Audited:** AI auditors should thoroughly understand the specific AI model being audited, including its structure, functionality, and intended purpose.
- **Business Knowledge:** AI auditors should have a strong understanding of the client's business processes and the operational context in which the AI system operates.
- **Data Science Expertise:** AI auditors should possess expertise in data science, statistics, and mathematics to assess AI model performance and reliability.
- **Data Engineering Expertise:** AI auditors should have expertise in data engineering, including databases, system architecture, data modeling, sampling, and analytics.
- **Knowledge of Social Implications:** AI auditors should have a thorough understanding of the societal and ethical implications of AI technologies.
- **Auditing Skills:** AI auditors should be skilled in auditing principles and methodologies specific to AI audit projects.
- **Coding Competency:** AI auditors should possess strong programming skills to effectively analyze, test, and validate AI systems.
- **Competency in AI Audit Tools:** AI auditors should be proficient in using specialized tools and technologies for AI auditing.
- **Software Engineering Skills:** AI auditors should possess knowledge of software engineering or practical experience in software development.
- **Soft Skills:** AI auditors should demonstrate social responsibility, ethical awareness, and strong communication and storytelling abilities.

3.1.2 AI Audit Standards: Mandatory, independent external AI audits are increasingly discussed as a means to bridge the gap between high-level principles and practical implementation. While most initiatives are country-specific and risk regulatory overlap, efforts such as the U.S. NIST guidelines, the UK AI Safety Institute's evaluations, and proposed French initiatives provide a foundation for effective AI auditing. However, these efforts primarily focus on the evaluation of AI models themselves. Ensuring the continuous development of AI auditing standards requires addressing

three key needs (David , Sammy, Mark , Mikhail , & Ross, 2025, p. 6612.6620):

- **Audit the process, not just the product:** AI systems pose evolving risks that cannot be adequately assessed through final-product evaluation alone. Black-box models, imperceptible training data vulnerabilities, and ethical concerns related to data sourcing and reinforcement learning processes necessitate comprehensive oversight of inputs, development practices, and lifecycle management. Similar to highly regulated industries such as pharmaceuticals and nuclear power, AI safety requires process-based evaluation rather than isolated product assessment. Moreover, the growing interaction among multiple AI systems increases complexity and reinforces the need for broader, human-centered auditing. Effective audits should therefore combine model evaluation with structured development processes, including both pre- and post-training assessments, to ensure systems operate safely and as intended.
    - Process audits and stakeholder participation
    - Audits for process beyond development
    - Interim audits during AI training
    - Rapid progress requires evolving standards
    - Continued monitoring and ongoing analysis
    - Inadequate audits can cause failures
- **Governance for safety**: Effective AI governance requires fostering a strong culture of safety, as technical safeguards alone are insufficient. Drawing lessons from industries such as nuclear energy and aviation, safety must be embedded at multiple cultural levels artifacts, stated values, and underlying assumptions. Industry-wide commitment to shared standards, ideally coordinated through dedicated standards boards, is essential, as fragmented standards weaken accountability. However, cultivating a meaningful safety culture in the AI sector remains challenging due to persistent complacency, within this framework:

- Safety culture enables effective audits
- Internal auditing is insufficient
- Audits should provide transparency, not create requirements
- Audits must allow for failure
- Auditing for governance and culture

- **Auditing standards body, not standard audits**: Current software safety and ethical standards are insufficient and fragmented, particularly in the context of AI, where ad hoc and model-specific updates undermine accountability. Rather than relying on isolated or one-time standard-setting efforts, there is a need for a dedicated auditing standards body responsible for the continuous development, harmonization, and oversight of AI auditing standards. Such a body would also help define the appropriate scope of these standards. This approach would involve:
  - Boards providing guidance, not static rules or regulations
  - Balancing principles and rules
  - Scope and applications-specific audits
  - Safety-washing and industry dynamics
  - Implementation concerns

3.2 Phases of the AI Auditing Process:

AI system auditing procedures involve assessing the extent to which systems comply with established AI principles and recognized technical, ethical, and social standards. These procedures further include evaluating the accuracy, completeness, and reliability of system outputs, examining governance structures and processes related to the development, deployment, and operation of AI systems, and ensuring that quality assurance requirements are met throughout the system lifecycle. To preserve objectivity and impartiality, audits are typically conducted either by independent external parties or by functionally independent internal units within the organization (Li & Goel, 2025, p. 17.18). A formalized

approach to AI auditing can be defined through the following stages (Anne, 2024, p. 63.65):

3.2.1 Phase 1 Understanding: The first phase of AI auditing is iterative and focuses on developing a comprehensive understanding of the complex audit context, structured according to the CRISP-DM framework and its activities.

- **Step 1 Business Understanding:** Focuses on comprehending the business context and the AI system's role. It aims to align the audit objectives with organizational goals and relevant regulations through activities such as intake interviews, determining audit objectives, and assessing the business environment.
- **Step 2 – Data Understanding:** Aims to develop a comprehensive understanding of all datasets relevant to the audit objectives, including their collection, description, exploration, and verification. It requires continuous alignment with the business context, which may necessitate revisiting earlier stages to ensure accurate interpretation and consistency with audit goals.
- **Step 3 Model Understanding**: Involves analyzing the AI model's architecture, behavior, and performance within its business setting, including its complexity, autonomy, and influence on the business. It is conducted iteratively in conjunction with business and data understanding, enabling the continuous refinement of audit objectives, data requirements, and model analysis.
- **Step 4 Risk Assessment:** Evaluates the risks associated with the AI system, drawing on insights from the business context, utilized data, and model purpose, and estimates potential issues such as bias or other context-specific harms.

3.2.2 **Phase 2 Research Object:** This phase establishes the focus of the AI audit by referencing the preliminary model and considering the "Make Context" and/or "Use Context." It includes defining an ethical frame (Step 5) to identify the specific risks relevant to the audit, while other AI risks are considered out of scope. Based on the ethical frame, the audit plan (Step 6)

is developed and presented in consultation with the client. Audit planning is a common step in regular audits. The key distinction in AI audits lies in the initiation phase preceding audit planning, which requires more specialized and context-specific knowledge than conventional audits.

3.2.3 Phase 3 Environment and/or Model: This phase involves examining the AI system's environment (Use Context), the model itself (Make Context), or both, depending on the scope defined by the auditor:

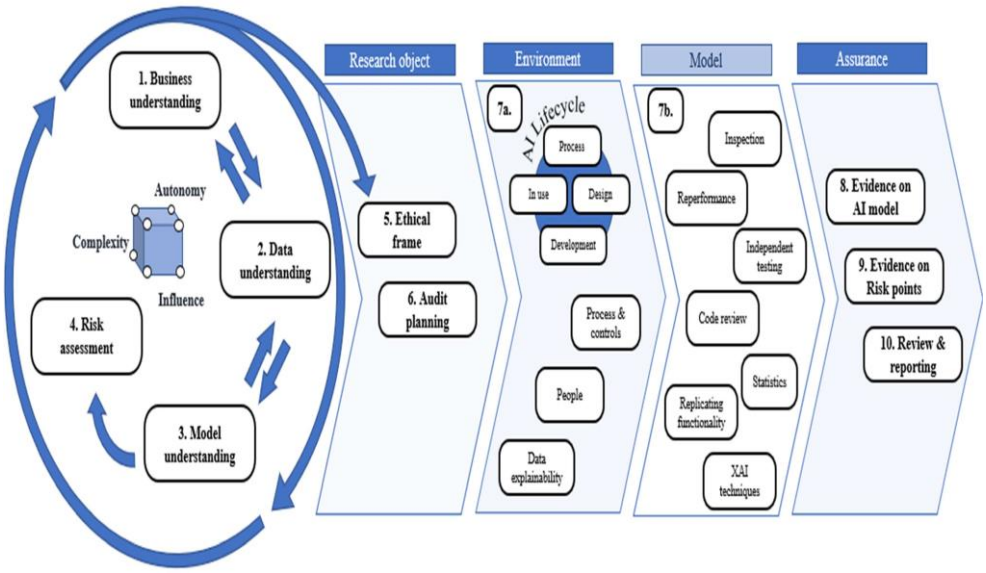
a. **Step 7 Environment:** assesses the AI system's environment, including controls, monitoring mechanisms, and lifecycle management. It highlights adherence to structured development processes, comprehensive documentation, IT- Like controls (e.g., change management, identity and access management), and data explainability, ensuring transparency, insight into model training, and data quality controls.
b. **Step 7 Step Model:** centers on auditing the AI model itself through methods as inspection or reperformance of AI lifecycle activities, independent testing with external datasets, and code review. Post-hoc techniques from Explainable AI (XAI) can offer insights into black-box decisions, while statistical measures, such as recall, provide relative evaluations of reliability. However, these approaches deliver only "reasonable assurance," as standardized quantitative benchmarks are lacking, and certain model functionalities may be inherently difficult to audit directly.

Generally, three main types of AI audit reports can be distinguished, as follows (CLAVELL, 2024, pp. 19,20):

- ✍ **Internal AI Audit Reports**: These reports document the audit process, identified issues, and applied or recommended mitigation measures. Unlike financial audits, algorithmic auditors may propose solutions, oversee their implementation, and report the outcomes. Publication of internal audit reports is not mandatory.

- ✍ **Public AI Audit Reports**: These reports present the final audit, detailing the AI system, auditing methodology, implemented mitigation measures, and any further recommendations. They also specify the frequency and methods or metrics to be used in follow-up audits.
- ✍ **Periodic AI Audit Reports**: Follow-up reports that reference the initial audit and confirm that developers continue to test for bias, implement mitigation measures, and monitor system impacts. Depending on the system's complexity, both internal and public versions may be produced.

Fig. The AI Auditing process



Source: (Anne, 2024, p. 66)

4. CONCLUSION

Auditing AI systems is a central component of an organization's AI strategy, whether as an end user or a provider of AI solutions. Its purpose is to verify system integrity, assess reliability, mitigate risks, strengthen controls, and provide assurance over AI-dependent processes. Effective AI

auditing requires approaches tailored to the system's characteristics, including comprehensive information on data and processes, quality assurance mechanisms, and governance frameworks. It also necessitates detailed documentation of model development, deployment, and monitoring, alongside a clear governance structure to ensure compliance with ethical and legal standards. Additionally, auditing must be guided by established standards, reference frameworks, or defined objectives, supported by auditors with specialized knowledge and skills in AI systems.

One of the key findings of this study is as follows:

- ✍ AI system auditing primarily focuses on assessing the extent to which developers and operators adhere to required procedures throughout the system's lifecycle. This ensures that the outcomes of these systems are consistent with applicable legal frameworks, established trust and safety standards, and societal expectations, while also strengthening accountability in the context of evolving technological and legal challenges.
- ✍ AI system auditability necessitates comprehensive measures at the model, data, and process/governance levels. This includes ensuring model clarity, transparency, and explainability; securing accessible, high-quality, and privacy-compliant data; and documenting standards, procedures, and stakeholder roles to enable systematic and reliable auditing.
- ✍ Effective AI system auditing requires clear and unified standards and requirements, stakeholder collaboration, and access to relevant information. Systematic documentation and detailed recordkeeping enhance transparency, auditability, and accountability, while audit teams must possess specialized and integrated competencies to perform audits reliably.
- ✍ AI audit standards are currently underregulated, creating risks of ethical "reputation washing," where organizations adopt abstract

ethical guidelines without enforcing mechanisms or verifiable accountability frameworks.

- ✍ Maintaining integrity and objectivity in AI audits requires that they be conducted by independent external auditors or internally by organizationally autonomous units, ensuring impartiality and independence the developers or operators of the AI system.
- ✍ Effective AI auditing requires continuously updating priorities and procedures to assess system reliability, ensure adequate assurance, strengthen controls, monitor risks systematically, and align audit objectives with dynamic operational changes.

Based on the above discussion, several recommendations can be proposed, summarized as follows:

- ☑ AI auditing should take a comprehensive, strategic approach covering data, models, and deployment, fostering a holistic understanding of risks and impacts. Viewed as a tool rather than a cost, it strengthens trust, mitigates risks, and supports AI development in line with its expected benefits.
- ☑ Organizations should implement robust AI governance frameworks, establish independent internal audit functions, and provide continuous training to auditors, ensuring effective oversight, risk management, and specialized assurance of AI systems.
- ☑ Auditors should continuously enhance their skills and knowledge to keep pace with technological advancements, supporting their scientific and practical qualifications and their collaboration with relevant authorities in establishing specialized AI auditing standards.
- ☑ Encouraging research and development through specialized conferences and scholarly dialogue enhances understanding of AI auditing and advances its theoretical and practical frameworks.

5. Bibliography List :

- Theses:

1 Anne, B. (2024). AI auditing: Incorporating eXplainable AI in an auditing framework. Master's thesis. Turku School of Economics, Turku, Finland: University of Turku.

2 FERNANDES, R. (2021). AUDIT AI: CO-DESIGNING WEB PLATFORM FOR EXTERNAL ALGORITHM AUDITING. MASTER IN COMPUTER SCIENCE AND ENGINEERING. DEPARTMENT OF COMPUTER SCIENCE , Lisbon, Portugal: NOVA University .

- Journal article :

1 Ayling, J., & Chapman , A. (2022). Putting AI ethics to work: are the tools fit for purpose? AI and Ethics, 2(3).

2 David , M., Sammy, M., Mark , B., Mikhail , S., & Ross, G. (2025). The necessity of AI audit standards boards. AI & SOCIETY, 40.

3 Dheyaa , Z., Firas, k., Ali, u., & Evan, M. (2024). The Impact of Artificial Intelligence Techniques on External Audit. Al-Qadisiyah Journal for Administrative and Economic Sciences(Special Issue).

4 Farley, E., & Lansang, C. (2025). AI AUDITING: FIRST STEPS TOWARDS THE EFFECTIVE REGULATION OF ARTIFICIAL INTELLIGENCE SYSTEMS. Harvard Journal of Law & Technology, 38.

5 Himanshu, V., Padh, K., & Thelisson, E. (2025). Can AI be Auditable?

6 Jiun, D., & Razli, R. (2025). Auditing Artificial Intelligence in Practice: Insights from Internal Auditors on Risk, Adoption, and Assurance. INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN BUSINESS AND SOCIAL SCIENCES, 15(12).

7 Kaur, M. (2024). A Comprehensive Overview of Artificial Intelligence-Based Classification Techniques. International Journal of Science and Research Archive, 11(02).

8 Kilian, R., Jäck, L., & Ebel, D. (2025). European AI Standards – Technical Standardisation and Implementation Challenges under the EU AI Act. European Journal of Risk Regulation , 16(3).

9 Laine, J., Minkkinen, M., & Mäntymäki, M. (2024). Ethics-based AI auditing: A systematic literature review on conceptualizations of ethical principles and knowledge contributions to stakeholders. Information & Management, 61(5).

10 Li, Y., & Goel, S. (2025). Artificial intelligence auditability and auditor readiness for auditing artificial intelligence systems. International Journal of Accounting Information Systems, 56.

11 Mökander, J. (2023). Auditing of AI: Legal, Ethical and Technical Approaches. Journal Digital Society, 49(2).

12 Robert, P., Pierce, C., Marquis, E., Kim, S., & Alahmad, R. (2020). Designing Fair AI for Managing Employees in Organizations: A Review, Critique, and Design Agenda. Human-Computer Interaction.

13 Sharma, S. (2025). AUDITING ARTIFICIAL INTELLIGENCE: A COMPREHENSIVE FRAMEWORK. International Research Journal of Modernization in Engineering Technology and Science, 07(02).

- Internet websites:

1 Calagna, K., Cassidy, B., & Park, A. (2021). Realize the Full Potential of Artificial Intelligence - Applying the COSO Framework and Principles to Help Iimplement and Scale Artificial Intelligence. (C. o. Commission, Éd.) Récupéré sur https://www.riskcue.id/uploads/ebook/20220327203805-2022-03-27ebook203757.pdf.

2 CLAVELL, G. (2024). AI Auditing Checklist for AI Auditing. Récupéré sur https://www.edpb.europa.eu/system/files/2024-06/ai-auditing_checklist-for-ai-auditing-scores_edpb-spe-programme_en.pdf.