International Journal of Sustainable Studies, Technologies, and Assessments

IJSSTA

Review

Ethical AI Framework and Auditing: A Comprehensive Overview

Narendra Kumar
Department of Computer Science and Engineering
Amity University Jharkhand, Ranchi, Jharkhand, India
E-mail: narendra298@gmail.com

Anurag Aeron
MIET, Meerut, Uttar Pradesh India
E-mail: anuragaeroniitr@gmail.com

Received: 23 February 2025 / Revised: 12 April 2025 / Accepted: 28 May 2025 / Published online: 06 June 2025

Abstract: With the rapid integration of Artificial Intelligence (AI) technologies into various sectors of society, it has become paramount to ensure their ethical use. All this is to maintain the principles of an ethical structural framework. Ethical AI frameworks consist of principles and guidelines that govern AI technologies' development, deployment, and use. These principles typically include transparency, accountability, fairness, confidentiality, and social benefit. Such a framework serves as a guideline for developers, policymakers, and users, guiding their decisions and actions to minimize potential ethical risks associated with AI. Auditing of AI systems involves a systematic assessment process to assess their compliance with ethical principles and regulatory requirements. This process examines various aspects of AI systems, including their algorithms, decision-making processes, data inputs, and impacts on individuals and society. By conducting thorough audits, organizations can identify bias, discrimination, privacy violations, and other ethical concerns within AI systems, thus being able to take corrective action. Additionally, the article discusses the challenges and opportunities associated with implementing an ethical AI framework and conducting AI audits. Challenges include the complexity of AI systems, the dynamic nature of ethical considerations, and the need for interdisciplinary collaboration. However, the abstract also highlights opportunities to promote trust, innovation, and social well-being through ethical AI practices. Finally, the abstract emphasizes integrating ethical AI frameworks and audit mechanisms into AI development and deployment processes. By doing so, stakeholders can maintain ethical standards, minimize risks, and maximize the positive impact of AI on individuals and society.

Keywords: Auditing, AI Framework, Artificial Intelligence, Ethically Driven Robotics, AI governance.

1. Introduction

Developing an ethical AI framework ensures that artificial intelligence is used responsibly and ethically. By establishing guidelines and standards, Organizations may manage the intricacies of AI technology, focusing on fairness, accountability, transparency, and privacy. The ISO/IEC JTC 1 Guidelines Committee on the Development of Artificial Intelligence (SC 42) and the working groups of IEEE SA's AI standards are two organizations involved in the development and establishment of standards related to artificial intelligence (AI) are developing

ethical frameworks for AI and Machine Learning. ISO/IEC 22989 includes over 100 commonly used AI terminologies to increase information-sharing efficiency and create terminology-consistent standards. ISO/IEC 23053:2022 for Machine Learning provides a lexicon of AI and ML. The IEEE Standard for Transparency of Autonomous Systems (7001-2021) defines transparency levels for autonomous systems, including safety-critical systems. IEEE P7006 outlines the technical requirements for developing and providing access to personalized AI systems. The IEEE Draft Ontological Standard for Ethically Driven Robotics and Automation Systems (P7007/D1) defines definitions and connections to enable ethically driven robotics and automation systems (R&A) development. Well-being metrics for ethical AI (IEEE P7010) allow for better consideration of how services and products can improve human well-being based on a broader range of metrics than just economic growth. Ethical issues for intelligent and autonomous systems could stifle innovation by introducing unwelcome regulations.

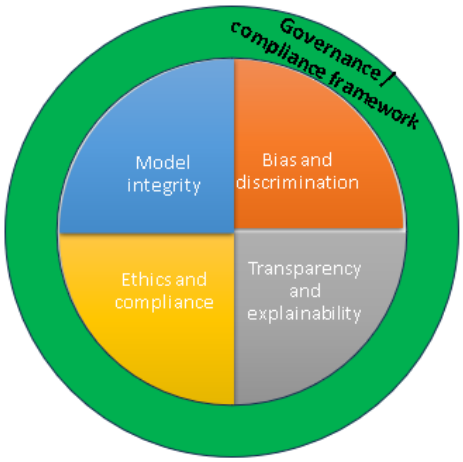

Fig1. Governance /compliance framework

IEEE P7000 and P7002 are standards for Addressing Ethics Concerns in System Design and software development. P7000 focuses on detecting and analyzing potential ethical problems, while P7002 provides specifications for privacy-oriented considerations for goods, services, and systems using personal information. Both standards cover the entire life cycle, from policy to development, quality control, and value realization. P7003 emphasizes accountability and clarity regarding algorithms' targeting, evaluation, and influence on users and stakeholders. It aims to guide individuals and organizations in developing algorithms and communicate best practices to regulatory authorities and users to prevent unwarranted differential impacts on users. The ethical principles of transparency, privacy, accountability, and fairness are essential in evaluating the fairness of AI frameworks. Transparency explains how AI makes decisions, privacy controls information, accountability assigns responsibility, and fairness promotes social justice by avoiding unfair AI biases. These principles form a broad framework for evaluating AI frameworks, with noncompliance resulting in liability.

Auditing AI systems is a critical component of upholding ethical standards. It involves assessing algorithms, data inputs, decision-making processes, and outcomes to identify biases, errors, or unintended consequences. Through rigorous auditing practices, organizations can mitigate risks and ensure that their AI systems align with ethical principles.

Developing an ethical AI framework ensures that artificial intelligence is used responsibly and ethically. By establishing guidelines and standards, organizations can navigate the complexities of AI technology, focusing on fairness, transparency, accountability, and privacy.

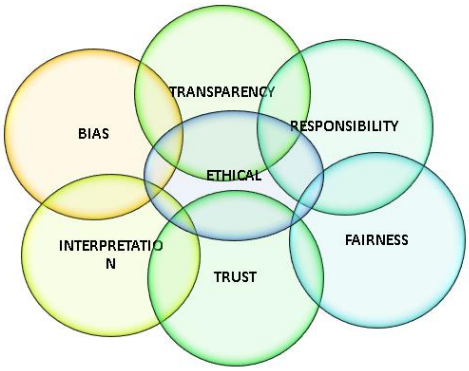

Fig 2. AI Ethics component

Auditing AI systems is a critical component of maintaining ethical standards. It includes algorithms, data inputs, decision-making processes, and assessing outcomes to identify biases, errors, or unintended consequences. Through rigorous audit practices, organizations can reduce risks and ensure that their AI systems align with ethical principles. Developing an ethical AI framework ensures that artificial intelligence is used responsibly and ethically. By establishing guidelines and standards, organizations can navigate the complexities of AI technology, focusing on fairness, accountability, transparency, and privacy.

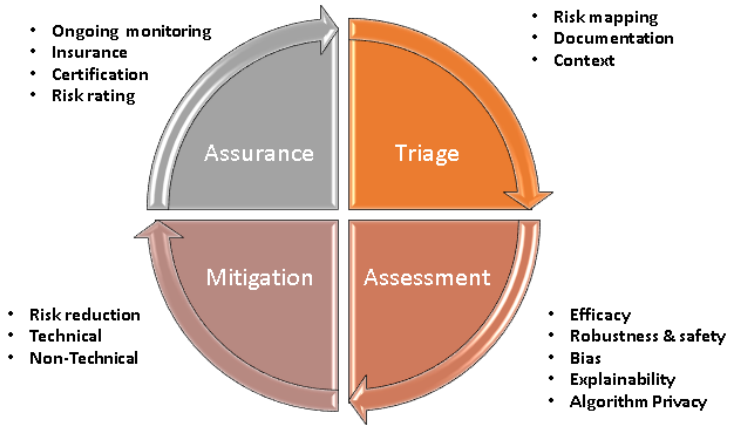

Fig 3. AI auditing process

Auditing AI systems is a critical component of maintaining ethical standards. It includes algorithms, data inputs, decision-making processes, and assessing outcomes to identify biases, errors, or unintended consequences. Through rigorous audit practices, organizations can reduce risks and ensure that their AI systems align with ethical principles. Developing an ethical AI framework ensures that artificial intelligence is used responsibly and ethically. By establishing guidelines and standards, organizations can navigate the complexities of AI technology, focusing on fairness, accountability, transparency, and privacy. Auditing AI systems is a critical component of upholding ethical standards. It involves assessing algorithms, data inputs, decision-making processes, and outcomes to identify biases, errors, or unintended consequences. Through rigorous auditing practices, organizations can mitigate risks and ensure that their AI systems align with ethical principles.

A comprehensive overview of ethical AI framework and auditing underscores the importance of integrating ethics into AI technologies' design, development, deployment, and monitoring. By prioritizing ethical considerations throughout the AI lifecycle, we can harness the Power of artificial intelligence for positive impact while safeguarding against potential harm.

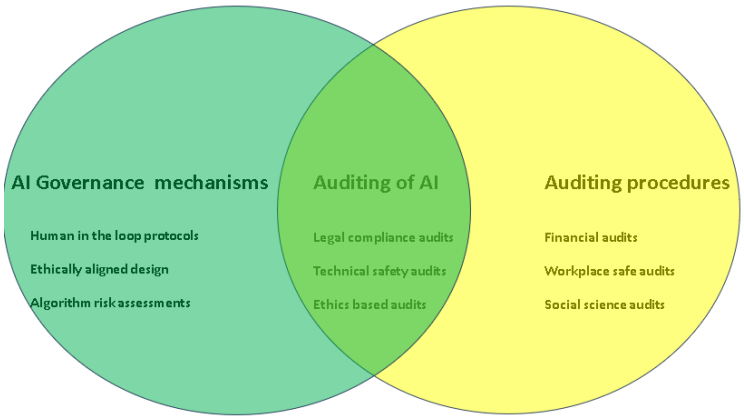

Fig 4. A conceptual outline of how AI auditing ties to earlier work on AI governance.

The subsequent sections of this article are organized as follows. Section 2 reviews the development of auditing as a means of governance. It explores how auditing has enhanced transparency and accountability in several domains, such as safety engineering and financial accounting. In Section 3, we utilize current societal advancements to demonstrate that the requirement to examine AI systems arises from a combination of influences from both higher authorities and grassroots movements. Section 4 reviews prior scholarly material about AI auditing. When conducting AI audits, we differentiate between specific and comprehensive understandings of the process and the legal, ethical, and technical methods used. Section 5 provides the articles that are part of this thematic compilation. In Section 6, we demonstrate how these papers contribute to the existing variety of auditing methods established to identify and reduce the risks presented by AI systems.

2. Literature review

This section examines several Ethical AI Framework and Auditing techniques depending on through rigorous auditing practices, organizations can mitigate risks and ensure that their AI systems align with ethical principles. In [1], we provide a comprehensive framework for developing and deploying trusted AI, emphasizing transparency, accountability, and respect for human autonomy. It outlines critical ethical principles and proposes concrete implementation requirements in AI systems. In [2], the book explores the ethical issues associated with AI development and deployment. It discusses various ethical frameworks and principles applicable to AI, accountability, including transparency, fairness, and security. The authors also address the challenges of applying these principles into practice. In [3], the paper introduces a method for auditing black-box machine learning models to identify and mitigate indirect forms of bias and discrimination. It highlights the importance of auditing AI systems to ensure fairness and transparency in decision-making. In [4], this paper proposes using model cards as a standardized format for reporting the performance and characteristics of machine learning models, including information on bias, fairness, and ethical considerations. It aims to improve transparency and facilitate model auditing processes. In [5], this paper presents an approach to auditing AI systems using automated techniques and human

judgment. It proposes a hybrid human-machine analysis framework to flag system failures, identify ethical concerns, and increase AI development and deployment accountability. In [6], this article examines the challenges of achieving fairness and accountability in sociotechnical systems, including AI systems. It argues the importance of considering the broader socio-political context and power dynamics in designing and auditing AI systems to address ethical concerns effectively. In [7], this paper proposes mechanisms to support verifiable claims about the behavior and properties of AI systems, including ways of auditing and certifying an AI system's compliance with ethical and regulatory requirements. It addresses the need for transparent and accountable AI development processes. In [8], this paper discusses ethical considerations and challenges in achieving trustworthy AI. It examines various aspects of AI development, including fairness, transparency, accountability, and privacy, and proposes strategies to address ethical concerns through auditing and regulation. In [9], this paper argues the importance of contextual integrity in ethical AI design and auditing. It emphasizes the need to consider the broader socio-cultural context and power dynamics in assessing the ethical implications of AI systems beyond mere interpretation. In[10], this article examines the ethical challenges organizations such as Google DeepMind face in AI research and development. It discusses the need for ethical solid frameworks and auditing mechanisms to ensure responsible AI innovation. This paper explores the promise and dangers of auditing AI systems for fairness, accountability, and transparency [11]. It highlights the challenges of biased data, algorithmic ambiguity, and unintended consequences and discusses the implications of developing an ethical AI framework. This paper examines the risks of relying on machine learning models for security applications [12]. It emphasizes the importance of rigorous auditing and validation processes to ensure AI's reliability and ethical use in security-critical domains.

3. The Development of Auditing as a System of Governance:

Auditing is a promising AI governance mechanism based on three key ideas: procedural regularity, transparency, proactivity in AI system design, and operational independence. However, auditing has become ambiguous due to its rapid growth and multidisciplinary nature. This underdetermination can make it difficult to verify claims of auditing and potentially exacerbate bias and harm. To effectively address this issue, it is imperative to conduct a thorough examination of the historical background of auditing in the fields of financial accounting, safety engineering, and social science research. The application of principles in practice. The methods and best practices in these areas have influenced current efforts to audit AI systems, emphasizing the importance of a collective understanding of auditing and its ability to address bias and mitigate harm. The advancements in these areas have significantly influenced the current endeavors to scrutinize AI systems.

A. The Social Science Audit Studies:

Research methods known as audit studies are commonly employed in the social sciences to analyze the behavior of individuals or the dynamics of social processes. The field experiments conducted in this study aim to replicate natural science experiments by employing a randomized research design within a real-world context. Researchers can study individuals and groups in their natural environment, differentiating it from surveys or interviews. Since the 1950s, audit studies have been used to investigate behaviors that are challenging to detect, such as racial and gender discrimination.

Audit studies can differ in two ways: the specific area of study and the independent variable selection. Some examples that can be considered are race, gender, age, and religion. Additionally, social science audit studies can

be utilized to collect data on the discriminatory effects of AI systems. Research conducted by Buolamwini and Gebru (2018) demonstrates that AI systems when employed to categorize images of individuals based on gender, exhibit higher levels of accuracy when applied to males with lighter skin tones compared to females with darker skin tones. There exists a tension in research methods between those focused on explaining social phenomena through empirical evidence and those driven by an activist agenda to bring about normative change or improve the conditions of the subjects under study. Both approaches have their strengths and often intersection in real-world scenarios. Nevertheless, how researchers engage with their investigation subject is significant, and audit studies in the social sciences have traditionally been linked to research with an activist orientation.

B. The Audits for safety

Safety audits are integral to auditing, intending to mitigate financial risks and identify potential health and safety hazards. They possess a rich historical background that traces its origins to the Industrial Revolution in 19th-century Britain. During this period, workers were confronted with unfavorable working conditions and significant hazards to their safety and well-being. Employers are held accountable through institutionalizing safety audits to reduce significant accident risks. Safety audits utilize various methods and tools to evaluate the effectiveness of safety management systems within organizations. AI audits need to consider the organizational culture of the entities involved in designing or deploying these systems. Nevertheless, safety audits possess certain limitations as governance mechanisms. While these measures may decrease the likelihood of incidents, they cannot completely eradicate the risk. Additionally, they may inadvertently give individuals a misguided sense of safety and encounter difficulties obtaining necessary evidence. This issue may interest AI auditors, as their access is often restricted due to privacy legislation and intellectual property rights. Financial and safety audits have distinct differences in substance, but they have commonalities regarding functions and procedures. Their main objectives are to validate the auditee's claims, aiming to mitigate risks and ensure management's responsibility. Nevertheless, term auditing has been employed in various ways in other contexts, as evidenced by social science audit studies.

C. Audits of Financial

The term audit has its roots in the Latin word audits, which translates to 'a hearing'. Over time, it has developed to encompass the verification of written records and enforce accountability. Auditing and financial accounting have a long history in the Middle Ages. During this time, audits were conducted to ensure the integrity of fiscal obligations. Auditing was initiated in 1844 with the enactment of the Joint Stock Companies Act by the British Parliament. The directors were required by this act to furnish audited financial statements to investors. The UK was the birthplace of the initial public accountancy organizations, and the 1980s witnessed the emergence of risk-based auditing. In his work, Power (1997) asserts that financial auditing serves as a verification ritual, investigating possible fraud and providing a sense of reassurance. Structured audits can foster trust among stakeholders with conflicting interests by ensuring transparency and procedure consistency. The relationship between auditors and auditees is multifaceted, characterized by operational independence and a shared objective of collaboration. This tension has led to the development of a model known as the three lines of defense, which can mitigate the risks associated with AI systems.

Financial auditing and accounting have become a significant industry on a global scale, with a market size exceeding $110 billion. Numerous organizations have leveraged their expertise and market positions to broaden their scope by providing additional auditing services, including AI auditing.

4. The Requirement for AI System Audits

Auditing procedures are implemented to address the perceived requirements of individuals and groups who are seeking information or reassurance regarding the behavior of others. The emergence of financial and safety audits can be attributed to the need of investors and the social and political pressures to enhance working conditions. The practice of AI auditing has become increasingly prevalent, as it aims to meet the demands of various stakeholders who are seeking information or reassurance. The necessity to conduct audits on AI systems is driven by both top-down and bottom-up pressures. Technology providers are subject to regulatory mandates and normative expectations from external stakeholders, including policymakers and advocacy groups. On the other hand, technology providers ensure competitiveness by implementing voluntary measures that involve ongoing software development and testing.

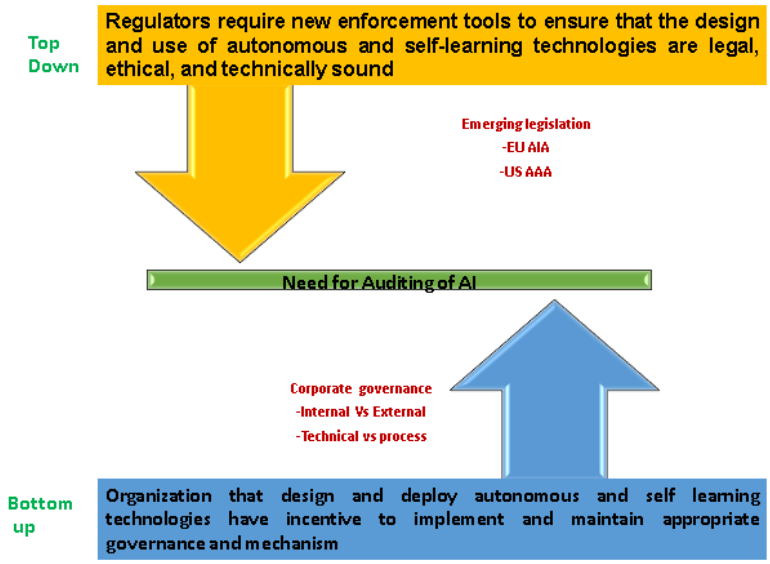

Fig 5: Both top-down and bottom-up forces support the need for AI system audits.

A. The significance of AI auditing in corporate governance

The involvement of private companies is essential in the development and deployment of AI systems, as these systems have significant impacts on social justice, economic growth, and public safety. The company possesses a robust presence in commercial applications and fundamental research pertaining to computational techniques that serve as the underpinning of artificial intelligence (AI) systems. In 2018, the quantity of research papers on machine learning that were published by private companies and laboratories in the United States surpassed the number of papers published by academics by a margin of more than 50%. The policies and governance mechanisms implemented by private companies to regulate the design and utilization of AI systems have substantial societal importance.

Policymakers are driven by the motivation to mandate audits of AI systems, whereas technology providers are similarly motivated to willingly subject their AI systems to independent audits. The concept of corporate AI governance encompasses a comprehensive framework that includes a set of rules, practices, processes, and technological tools. The purpose of this system is to ensure that an organization effectively utilizes artificial intelligence (AI) systems in alignment with its strategies, objectives, and values. However, the corporate governance landscape is dynamic, and private companies face ongoing demands to innovate and improve their products.

Technology providers have implemented various measures to ensure that their products and services conform to established quality standards and effectively meet the needs and expectations of consumers. Audits can provide valuable insights into the governance of corporate AI systems, addressing both technical and normative inquiries. Technology providers can evaluate the capabilities and limitations of AI systems before deployment, identifying and addressing potential risks to prevent harm. In addition, audits encourage investments in effective risk management. Organizations undergo independent audits to evaluate and enhance their software development processes and Quality Management Systems (QMS).

There has been significant public criticism regarding the intentions and methods behind the development of AI systems rather than focusing solely on their technical shortcomings. In 2020, Clearview AI received significant criticism following investigations that exposed its practice of collecting billions of images from social media platforms without obtaining users' consent. These images were then used to create its training dataset. Clearview AI experienced substantial damage to its reputation and was subject to legal actions, resulting in a settlement prohibiting the company from selling its technologies to private entities in the United States. Examining the technical and ethical dimensions of AI systems in audits can assist technology providers in mitigating financial and reputational risks. Efficiently communicating audit findings can give companies a competitive edge, such as showcasing the measures taken to ensure ethical AI systems can positively impact marketing and public relations. Thorough and scholarly audits of AI systems can assist organizations in enhancing various business metrics, such as regulatory readiness, data protection, talent recruitment, reputation management, and streamlining processes.

Given these grassroots influences, numerous technology providers have chosen to proactively assess their AI systems to ensure compliance with various ethical principles. Nevertheless, it is essential to approach this development with caution. Sloane (2021) has raised concerns about needing more independence in audits commissioned by technology providers. Bandy (2021) has highlighted the difficulty in verifying the claims made by technology providers regarding auditing their AI systems, especially in the absence of agreed standards. This section aims to assess the benefits of utilizing AI systems auditing as a governance mechanism. The objective is to highlight the mutual interest of policymakers and technology providers in the development and promotion of procedures for auditing these systems. Academic researchers are the most suitable individuals to assess the feasibility and efficacy of these auditing procedures.

B. The use of auditing as a mechanism to enforce legislation

AI auditing procedures are motivated by the growing influence of government regulations. AI systems can enhance economic growth and human well-being by enhancing information processing speed and accuracy and fostering the creation of novel solutions. Nevertheless, ethical, social, and legal challenges are associated with them, including the possibility of harm due to bias, discrimination, privacy breaches, human misconduct, and the erosion of self-determination. Policymakers must navigate the delicate balance between safeguarding against harm and fostering innovation. Significant advancements in large language models (LLMs) have garnered widespread interest, exemplified by the emergence of ChatGPT. LLMs can produce text closely resembling human writing, but they have faced significant criticism and opposition due to their design and usage. Several studies have demonstrated that LLMs have the potential to generate language that is deemed unethical, such as making racist and sexist remarks, and their responses frequently contain factual inaccuracies. Open-source business models enable the utilization of LLMs for purposes beyond their initial scope, resulting in a significant public backlash.

There is a growing demand for policymakers to establish regulations governing the design and utilization of AI systems. The European AIA stands as the pioneering regulatory framework of a major global economy, with subsequent proposals by governments focusing on more specific legislation. Canada's Directive on Automated Decision-Making, Singapore's guidelines on responsible AI system design, and the Algorithmic Accountability Act of 2022 (AAA) are under consideration by the US Congress. The draft regulations vary in scope and content, yet they all establish rules and requirements that organizations must adhere to when designing or deploying AI systems.

For regulations to be successfully implemented and enforced, they must be accompanied by effective governance mechanisms. As an illustration, technology providers who do not meet AIA's requirements may face significant fines. In order to assess compliance, it is necessary to evaluate the available mechanisms for determining a provider's actions. Auditing plays a crucial role in this context. Similar to auditing financial transactions for accuracy, comprehensiveness, and legality, the design and utilization of AI systems can also undergo audits to ensure technical reliability and adherence to legal requirements.

The EU AIA requires high-risk AI systems to undergo conformity assessments before deployment, while the UK Information Commissioner's Office has guided auditing AI systems. In 2021, New York City implemented the AI Audit Law. This law mandates that AI systems utilized to influence employment-related judgments must undergo independent audits. Regulators can utilize AI audits to evaluate the legal compliance of an AI system. Nonetheless, a significant distinction exists between financial audits and audits of AI systems that are required by law. Managers face pressure from investors to effectively manage financial risk, while technology providers are under pressure from policymakers to maintain political legitimacy. As automation continues to advance, policymakers' political legitimacy will rely heavily on their capacity to effectively address AI systems' ethical and social implications. As a result, there is a growing need to establish standardized protocols for auditing AI systems.

C. Audit of AI's Multidisciplinary Foundations

The literature review on AI systems auditing primarily examines the procedures used to assess the consistency of AI systems with relevant specifications, regulations, or ethical principles. It is crucial to revisit and further elaborate on the definition provided in the introduction.

D. Broad Vs. Narrow Conceptions of Auditing AI Systems

There are two main types of AI auditing: narrow and broad conceptions. One perspective emphasizes examining and evaluating the results of AI systems with various input data, while another evaluates the effectiveness of software development processes and QMS technology providers. An auditing process can be defined as the systematic and repetitive querying of an algorithm with inputs while observing its outputs. This allows for drawing inferences about the algorithm's internal operations, which may need to be more readily apparent. These types are ideal for collecting evidence of unlawful discrimination and are commonly supported by experimental designs. Alternatively, broader perspectives center on the governance structures of organizations involved in designing and implementing AI systems. This practice uses traditional IT audits and procedures for managing technology risks. An auditor's responsibilities extend beyond simply reviewing the algorithm and management measures. They must also consider the data, development methods, and algorithm optimization carefully.

Extensive understandings of auditing enable researchers to identify and examine illegal, erroneous, or unethical behaviors exhibited by AI systems and delve into the origins of these behaviors. For instance, discriminatory behavior can arise from training datasets that are incomplete or unrepresentative or from inadequate testing and validation procedures. Researchers have established methods for auditing the complete AI system development and deployment process. This includes evaluating the governance structures of technology providers to ensure proper staff training, assembling training datasets, assessing AI system limitations before deployment, and monitoring their behavior on an ongoing basis.

E. AI Auditing Literature Review

AI auditing is a process employed by different societal stakeholders to evaluate the legality, risk management, and involvement of AI systems. The process involves evaluating an entity's behavior to determine if it aligns with established standards, regulations, or norms. Auditing procedures vary depending on the audited subject, including individuals, organizations, or technical systems. Functionality audits assess the logical basis for decisions made, code audits evaluate the source code of an AI system, and impact audits investigate the different types, severity, and frequency of effects that arise from the output of an AI system. In order to ensure proper auditing procedures, it is of utmost importance to maintain operational independence between the auditor and the auditee. In addition, a predefined baseline is utilized as a reference point for the purpose of evaluation. The literature on auditing AI systems is characterized by its diversity. The scope of this includes a wide range of sources, including academic articles, books, auditing tools and procedures created by private companies, industry standards, and draft legislation and guidance documents issued by policymakers. Literature on AI systems can be categorized into different approaches, including auditing, technical, legal, and ethical perspectives. Additionally, various strands of research focus on proposing, developing, employing, or critiquing auditing procedures for AI systems.

F. Legal, Technical, and Ethics-Based Approaches

Researchers employ various methodologies to conduct audits on AI systems. These audits typically involve establishing a predetermined baseline that encompasses technical specifications, legal obligations, and voluntary ethical principles. The categorization of the AI systems auditing literature includes technical, legal, and ethical approaches. The assessment of technical approaches involves evaluating the technical characteristics of AI systems, such as their accuracy, robustness, and safety. The evaluation of AI systems prior to market deployment involves the use of ex-ante and ex-post audits. The concept of auditing software has undergone advancements, with certain researchers creating open-source toolkits to assess the performance of AI systems across different tasks and datasets. The purpose of these procedures is to conduct audits on AI systems in order to assess their compliance with predefined technical, functionality, and reliability standards. The legal approaches encompass auditing procedures that evaluate the compliance of AI systems with applicable regulations, including data privacy regulations, discrimination laws, sector-specific certification mandates, and general transparency obligations.

Legal scholars have debated the application of these regulations to AI systems. Many procedures have been proposed and implemented, such as auditing AI systems for unlawful price discrimination and Facebook's ad delivery algorithm. Alternatively, ethics-based approaches rely on voluntary ethics principles as the standard baseline. The audits can be categorized as either collaborative or adversarial. Collaborative audits are conducted with

the goal of providing assurance, while adversarial audits are intended to uncover and expose harm. Ethics-based audits prioritize key factors such as transparency, explainability, bias mitigation, fairness, and accountability. Several private companies, including AstraZeneca, have voluntarily undergone ethics-based audits. The demarcation lines between technical, legal, and ethics-based audits can often be indistinct, as auditors commonly depend on technical methodologies to collect evidence pertaining to the properties and effects of AI systems. The ethical consideration of an AI system often requires technical robustness and legal compliance as prerequisites. The three audit types can be understood as a continuum of complementary approaches that each have distinct focal points. The utilization of technical, legal, and ethical approaches proves advantageous for two primary reasons: firstly, it aligns with the terminology employed by policymakers, and secondly, it facilitates the differentiation of various audits that serve distinct objectives.

G. Who Audits the Auditors?

A substantial amount of literature supports the need for auditing AI systems, emphasizing the potential social, ethical, and legal risks they present. Studies indicate that audits are crucial in promoting good governance by ensuring adherence to procedures and transparency. Additionally, they help prevent potential harm by encouraging the proactive design of AI systems. Many of these contributions consist of commentary or viewpoint articles advocating for using structured and independent audits as a practical solution to address governance challenges. Academic researchers have created procedures and tools to audit tangible AI systems. These can be categorized into two main groups: high-level procedures that provide a framework for audits and tools that auditors can use for specific tasks. Additionally, empirical studies are conducted to evaluate the design and impact of AI systems on users and societies.

A small yet expanding group of researchers is focused on studying the practicality and efficacy of auditing as a governance mechanism for AI systems. Nevertheless, the prevailing focus in auditing procedures has revolved around theoretical criticisms. For instance, Sloane (2021) contends that the current procedures lack effectiveness and might even have adverse consequences. In a recent study, Engler (2021) highlighted the challenges independent auditors face in holding technology providers accountable. One major obstacle is the refusal of these providers to grant access to their data and models. This problem can only be effectively addressed through the implementation of sector-specific legislation. However, empirical research is still needed to support the claims regarding the limitations of AI systems auditing as a governance mechanism.

The field of AI auditing is experiencing significant growth in both research and practice. However, there is a noticeable disparity between the level of attention it has received and the limited amount of empirical academic research on the efficacy and practicality of various auditing methods. Auditing of AI: Legal, Ethical, and Technical Approaches is a topical collection recently published by the Digital Society. It aims to address the gaps in our understanding of AI through concise and academic discussions. The collection includes six articles that discuss the challenges and best practices related to designing and conducting AI audits. Lessons from Practice, authored by Hasan et al. (2022), provides a comprehensive analysis of the challenges [17] faced by auditors and industry practitioners in the process of designing and conducting AI audits. The study offers valuable insights into the practical aspects of addressing algorithmic bias and risk assessments. The authors highlight the importance of taking into account the broader context of AI systems as integral components of larger sociotechnical systems

when designing audits. This emphasis is based on their extensive experience providing advice and conducting AI audits for clients across different industries.

5. Conclusion

This article briefly summarizes prior research on AI auditing, emphasizing three main aspects: (1) It is essential for current efforts to audit AI systems to draw lessons from past audits in fields such as financial accounting, safety engineering, and social sciences. (2) Academic researchers are vital in examining the practicality and efficacy of various AI auditing methods. (3) Auditing is a multidisciplinary endeavor that benefits from integrating different approaches that support and strengthen one another. The article emphasizes the significance of applying established practices from financial and IT auditing to AI auditing. It underscores the necessity for further research to establish solid methodologies and draw on accumulated experiences to develop practical AI auditing procedures. The text also emphasizes the current demand for auditing AI systems due to various pressures from different directions. AI auditing requires a multidisciplinary approach involving researchers from various fields who examine the subject's legal, ethical, and technical aspects. Legal compliance audits frequently employ technical methodologies to gather evidence pertaining to the characteristics and effects of artificial intelligence (AI) systems. Prior to considering an AI system as ethical, it is typically required to ensure its technical robustness and legal compliance. The careful consideration of AI auditing procedures' design and implementation is of utmost importance. In order for AI auditing procedures to be both feasible and practical, it is essential that they are well-organized and transparent. The assessment procedures should include a well-defined material scope and incorporate elements from both technology-oriented and process-oriented assessments. In order to ensure comprehensive oversight, it is recommended to incorporate continuous monitoring into the process. Furthermore, it is advisable to engage independent third-party auditors to carry out the auditing procedures.

References:

[1] M. Cannarsa, “Ethics guidelines for Trustworthy Ai,” The Cambridge Handbook of Lawyering in the Digital Age, pp. 283–297, Nov. 2021. doi:10.1017/9781108936040.022

[2] J. Lamuth, “Ethical Artificial Intelligence,” SciVee, Sep. 2010. doi:10.4016/22623.01

[3] P. Adler et al., “Auditing black-box models for indirect influence,” 2016 IEEE 16th International Conference on Data Mining (ICDM), Dec. 2016. doi:10.1109/icdm.2016.0011

[4] M. Mitchell et al., “Model cards for Model Reporting,” Proceedings of the Conference on Fairness, Accountability, and Transparency, Jan. 2019. doi:10.1145/3287560.3287596

[5] B. Nushi, E. Kamar, and E. Horvitz, “Towards accountable AI: Hybrid human-machine analyses for characterizing system failure,” Proceedings of the AAAI Conference on Human Computation and Crowdsourcing, vol. 6, pp. 126–135, Jun. 2018. doi:10.1609/hcomp.v6i1.13337

[6] A. D. Selbst, D. Boyd, S. A. Friedler, S. Venkatasubramanian, and J. Vertesi, “Fairness and abstraction in Sociotechnical Systems,” Proceedings of the Conference on Fairness, Accountability, and Transparency, Jan. 2019. doi:10.1145/3287560.3287598

[7] E. Hine, C. Novelli, M. Taddeo, and L. Floridi, “Supporting trustworthy AI through machine unlearning,” SSRN Electronic Journal, 2023. doi:10.2139/ssrn.4643518

[8] J. Wang and L. Fu, “Review of ‘Ai assurance: Towards trustworthy, explainable, safe, and ethical AI’ by Feras A. Batarseh and Laura J. Freeman, Academic Press, 2023,” AI & SOCIETY, Nov. 2023. doi:10.1007/s00146-023-01802-1

[9] C. Henin and D. Le Métayer, “Beyond explainability: Justifiability and contestability of Algorithmic Decision Systems,” AI & SOCIETY, vol. 37, no. 4, pp. 1397–1410, Jul. 2021. doi:10.1007/s00146-021-01251-8

[10] A. Ananthaswamy, “Deepmind AI topples experts at Complex Game Stratego,” Nature, Dec. 2022. doi:10.1038/d41586-022-04246-7

[11] C. T. Okolo, “The promise and perils of Generative AI: Case Studies in an African context,” Proceedings of the 4th African Human Computer Interaction Conference, Nov. 2023. doi:10.1145/3628096.3629066

[12] S. Shiva, “Testing and trusting machine learning systems,” Online Journal of Robotics & Automation Technology, vol. 1, no. 1, Feb. 2021. doi:10.33552/ojrat.2021.01.000503

[13] N. Schöppl, M. Taddeo, and L. Floridi, “Ethics auditing: Lessons from business ethics for ethics auditing of ai,” The 2021 Yearbook of the Digital Ethics Lab, pp. 209–227, 2022. doi:10.1007/978-3-031-09846-8_13

[14] S. Basnet, “Navigating the ethical landscape: Considerations in implementing AI-ML Systems in human resources,” International Journal of Research Publication and Reviews, vol. 5, no. 3, pp. 3436–3447, Mar. 2024. doi:10.55248/gengpi.5.0324.0755

[15] Merging risk-based auditing and integrated auditing with Agile Auditing,” Agile Auditing, pp. 225–230, Jun. 2021. doi:10.1002/9781119693529.ch15

[16] M. J. Walter, A. Barrett, D. J. Walker, and K. Tam, "Adversarial AI test cases for Maritime Autonomous Systems," AI, Computer Science, and Robotics Technology, vol. 2, Apr. 2023. doi:10.5772/acrt.15

[17] Hasan, Md.M. et al. (2023) Review on the Evaluation and development of Artificial Intelligence for covid-19 containment’, Sensors, 23(1), p. 527. doi:10.3390/s23010527.