Scientific African 25 (2024) e02281

Contents lists available at ScienceDirect

Scientific African

journal homepage: www.elsevier.com/locate/sciaf

Bias and ethics of AI systems applied in auditing - A systematic review

Wilberforce Murikah [a,*], Jeff Kimanga Nthenge [b], Faith Mueni Musyoka [b]

[a] Department of Computing and Informatics, United States International University - Africa, Nairobi, Kenya
[b] Department of Computing and Information Technology, University of Embu, Embu, Kenya

ARTICLE INFO

Editor: DR B Gyampoh

Keywords:
Artificial intelligence
Audit
Bias
Ethics
Machine learning
AI

ABSTRACT

The integration of artificial intelligence into auditing shows great potential in enhancing automation and gaining insights from complex data. However, it also presents significant ethical challenges, including algorithmic biases, transparency, accountability, and fairness. This study aimed to investigate the sources of bias and risks posed by AI systems applied in auditing and the complex downstream interactions and effects they have. The study also explored the technical and ethical guardrails proposed and recommendations for translating principles into auditing practice. A systematic methodology was employed to acquire relevant studies across scientific databases. This involved a three-step process, including a targeted search query using Boolean operators and snowballing to yield 310 preliminary publications. A systematic review process was then conducted to identify 123 relevant articles focused on AI's implications for auditing, accounting, finance, or assurance contexts. Finally, screening and filtering on research quality distilled 83 high-quality publications from the year 2018 to 2023 spanning computer science, accounting, management science, and ethics disciplines. The analysis revealed five primary sources driving technical and human biases: data deficiencies, demographic homogeneity, spurious correlations, improper comparators, and cognitive biases. It also highlighted wider issues, such as trade-offs between efficiency and diligence, erosion of human skills and judgement, data dependence risks, and privacy violations from uncontrolled personal data exploitation. The study found promising remedies, including causal modeling to enable auditors to uncover subtle biases, representative algorithmic testing to evaluate fairness, periodic auditing of AI systems, human oversight alongside automation, and embedding ethical values like fairness and accountability into system design. The study concludes that auditors play a crucial role in assessing and ensuring AI's reliable and socially beneficial integration. It recommends governance, risk assessment before deployment, ongoing performance monitoring, and policies fostering trust and collaboration to responsibly translate principles into auditing practice.

Introduction

Auditing is crucial for ensuring the accuracy of financial information used in investment decisions. However, auditors face challenges in analysing large datasets efficiently, which can impact audit quality, risk identification, and predictive capabilities [1]. As a result, mistakes or irregularities may go undetected during standardized testing processes as companies embrace digital

* Corresponding authors.
E-mail address: wmurikah@gmail.com (W. Murikah).

https://doi.org/10.1016/j.sciaf.2024.e02281

Available online 13 June 2024

transformation. Artificial intelligence (AI) can significantly enhance auditing by automating repetitive tasks, such as data entry and account reconciliation. It can also identify hidden discrepancies in reports, such as unusual patterns in transactions or inconsistencies between a company's reported revenue and supplier payments. Additionally, AI can analyze various sources of evidence using its powerful data processing capabilities [2,3]. Techniques such as robotic process automation, anomaly detection algorithms, and automated reconciliation engines, enhanced with natural language processing, can quickly review vast amounts of structured and unstructured data to uncover potential inaccuracies.

With the rapid advancement of AI capabilities, audit firms have enthusiastically adopted this technology, with major firms making substantial investments to integrate AI [4,5]. AI is being integrated into specific use cases to support auditors in their workflows. For instance, AI tools are used to automatically reconcile large volumes of transactions by analysing all transactions, eliminating the need for auditors to manually sample and review subsets of data. The AI reconciliation system identifies any discrepancies or anomalous transactions, which the auditor can then investigate further. This allows auditors to focus on identified risks and exceptions while taking advantage of the computational power of AI to process the entire population of transactions. The integration of this intelligent process automation improves the efficiency and effectiveness of the traditional reconciliation task. Robotic process automation and anomaly detection algorithms can analyze now 100 % of a client's transactions, as opposed to the traditional random sampling method [6]. Machine learning models identify transactions with characteristics that differ from organizational baselines by considering factors such as timing, frequency, location, or parties involved [7]. Alerts are generated for unusual payments, accelerated billing, excessive modifications, or suspicious vendors. This allows human efforts to focus on investigating the subset of entries with the highest risk, which could otherwise be easily concealed within large volumes of data.

Additionally, automated reconciliation engines quickly compare financial reports to corresponding ledger postings, backup documents, and related external news [8]. Natural language processing parses thousands of unstructured pages to extract key figures, statistics, commentary, dates, and other fields for matching [9]. Any inconsistencies between a company's public statements and internal records could indicate faulty reporting or attempts to mislead. For example, CEO remarks praising overseas growth while quarterly earnings mainly declined abroad would require examination. Thus, AI strengthens multifaceted confirmation procedures through holistic machine reading at scale.

However, ethical concerns are growing regarding risks of unfair biases and inadequate transparency given the black box nature of AI systems, where even internal data scientists cannot always fully explain why systems behave a certain way as flexibility and predictive accuracy become prioritized over simplicity or causality [10,11]. Recent cases like Apple's credit algorithm and UnitedHealth's medical algorithm systematically discriminating against minorities exemplify potential pitfalls of AI [12,13,14]. Critics argue that productivity gains should not override emerging hazards or erode public trust [15,16,17,18]. There are widespread calls among experts to proactively address risks related to unfair biases, causal misattributions, deception, destabilizing, and lack of interpretability effects that could undermine reliability and fairness [19,20]. Specifically, reliance on training datasets that reflect existing societal biases can further propagate discrimination, even if unintentionally ingrained [21]. This may lead to financial

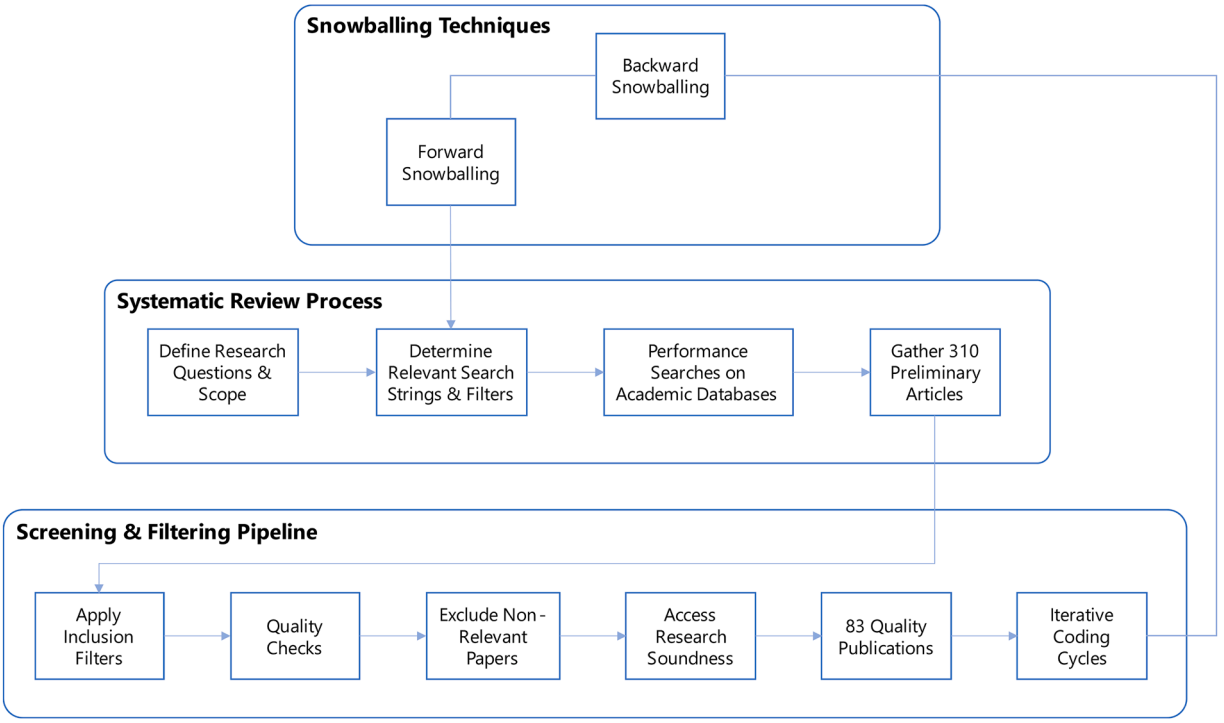

Fig. 1. Systematic literature review methodology.

algorithms penalizing minorities, predictive policing systems targeting marginalized communities to a greater extent, and facial recognition misidentifying people of color at significantly higher rates. For instance, an anomaly detection algorithm intended to identify suspicious transactions could end up flagging a greater number of transactions from minority communities if the historical data shows racial disparities in financial audits. Alternatively, an automated risk scoring system could assign higher risk scores to companies with more leadership diversity if the algorithm associates majority demographics with lower risk, thus incorporating problematic assumptions.

Scholars argue that solutions require combining technical rigor and ethics to align AI with auditing's public interest duty [22,23,24]. Integrating such diverse considerations when creating dependable sociotechnical systems is underexplored within auditing contexts [25]. However, insights can be drawn from pioneering fields. For example, the Asilomar AI Principles developed by leaders across technology, law, and philosophy provide actionable guidance on topics like transparency, accountability, non-maleficence, and social benefit [26,27]. Additionally, Google's "People + AI Guidebook," based on six years of human-cantered machine learning research, details best practices for mindful, inclusive design.

This study conducts a systematic literature review synthesizing current knowledge surrounding: 1) sources of bias and risks posed by AI systems, 2) their complex downstream interactions and effects, 3) evolving technical and ethical guardrails proposed, and 4) recommendations for responsibly translating principles into auditing practice. By consolidating dispersed findings across computer science, social science, regulatory, and practitioner communities, this review combines technical precision and conscience to outline pathways for reliable and ethical AI augmentation that upholds auditing's truth-seeking mission.

Review methodology

A systematic approach that adheres to established best practices guided the search strategy, as depicted in Fig. 1. This approach involved following a comprehensive and well-structured plan to ensure that the search process was thorough, efficient, and effective [26,27]. The search strategy utilized scientific databases, such as Web of Science and Google Scholar, to identify relevant peer-reviewed academic publications and preprints on artificial intelligence, machine learning, bias, transparency, ethics, and auditing. The search yielded a total of 310 preliminary results, which were sourced from a variety of platforms. To refine the search and ensure that the results were more specific, Boolean operators were employed to combine terminology related to accounting, auditing, finance, or assurance areas.

As illustrated in Fig. 2, database searches on artificial intelligence and auditing underwent phased relevance and quality checks, resulting in 123 publications. During full-text screening, studies that briefly mentioned artificial intelligence without detailed technical or ethical auditing analysis were manually excluded. Articles directly analysing algorithmic opacity, accountability, and implications for audits, control assessments, and fraud analytics were prioritized. As shown in the Multi-Stage Study Screening Protocol (Fig. 2), strict inclusion protocols further refined the 123 studies. To assess the quality of the studies, the ROBINS-I tools, a standardized critical

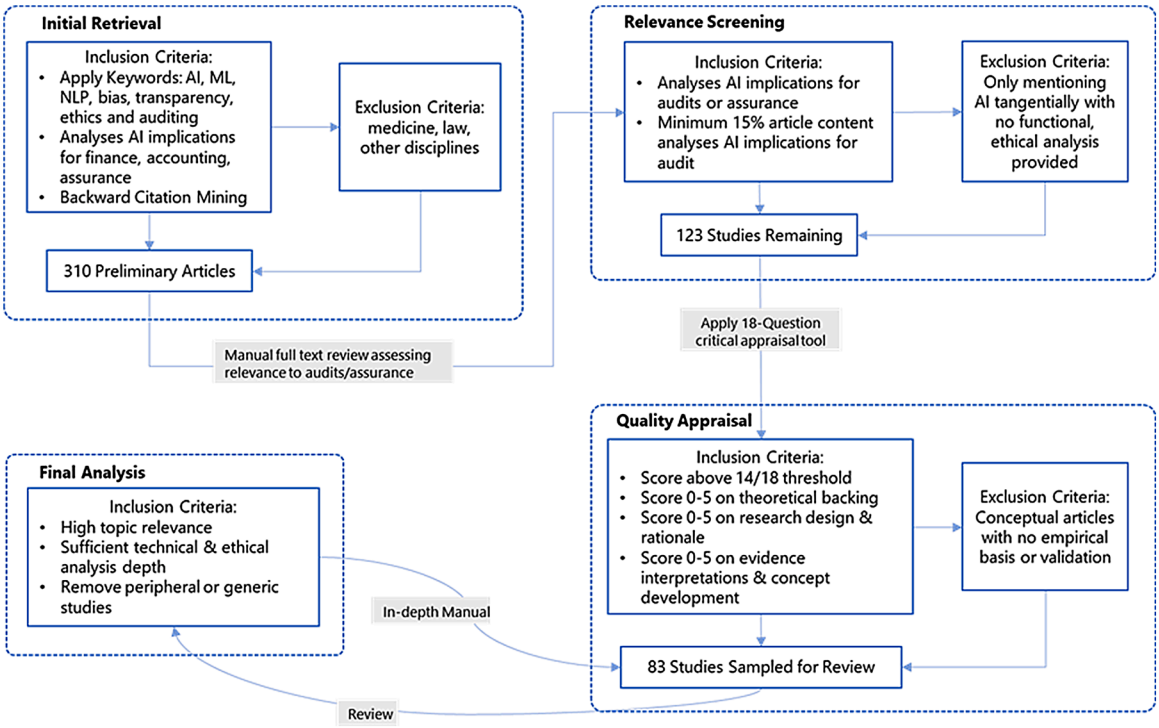

Fig. 2. Multi-stage study screening protocol.

appraisal tool, was utilized to eliminate any papers with weaknesses in theoretical foundations, methods, analytical validity, reliability, validation, or generalizability [21]. This tool also helped identify and remove studies lacking methodological rigor or generalizability.

Subsequently, a total of 83 high-quality publications from the years 2018 to 2023 in the fields of computer science, accounting, management, and ethics were selected and analysed through iterative coding cycles (as illustrated in Fig. 1). These final empirical publications encompassed topics such as external financial audits, fraud detection, quality assurance audits, and risk reviews [28,29,30]. Specifically, these publications focused on technical aspects, such as algorithmic opacity, as well as ethical dimensions, such as accountability.

The final 83 publications provided an overview across computer science, accounting, management, and ethics, selected based on relevance, complexity, and depth of analysis. Iterative coding cycles provided insights into artificial intelligence innovations in auditing. Formal modeling techniques [31] were used to connect conceptual relationships, categorizing codes into themes based on shared attributes to reveal challenges and opportunities presented by artificial intelligence.

Results

Sources of algorithmic biases and unfairness

Analysis of the literature corpus revealed five primary sources feeding biases into AI systems applied in auditing as depicted in Table 1. These sources intricately interplay throughout the AI system development and application lifecycles. Design choices, subjective judgments, and social biases become ingrained in algorithmic systems and data-driven inferences, contributing to unfairness [32].

The literature analysis highlighted how seemingly benign technical factors lead to discriminatory and unethical outcomes. Issues begin with biased data collection methods, such as convenience samples and misrepresentations. Historical discrimination embeds itself in training datasets, resulting in AI models that neglect contextual factors and fail minority groups due to demographic homogeneity. Deployed models perpetuate stereotypes and societal prejudices, presenting biased outcomes under the guise of impartial algorithmic determinations [56].

In a study conducted to explore the use of machine learning for anomaly detection in financial data, several potential sources of bias and ethical concerns emerged during the development, training, and evaluation of the models [7]. Imbalanced datasets, where normal transactions significantly outnumbered anomalies (e.g. 99.5 % vs 0.5 %), were found to bias models towards the majority class and hinder the detection of rare anomalies. Additionally, the use of synthetic anomalies posed risks as they may not accurately represent real-world anomalies. This could introduce subjective biases based on the auditor's opinions of anomaly criteria. Another study supported this argument, suggesting that assuming training data as non-anomalous overlooks inherent anomalies in financial transactions and poses ethical risks if valid anomalies are wrongly classified as normal [57]. Predefined anomaly categories may also fail to capture the full range of anomalies, thereby limiting detection and the evaluation metrics could distort perceptions of model performance due to class imbalance. These risks highlight ongoing debates in auditing concerning subjective biases, analytical assumptions, and ethical sampling risks. While anomaly detection models aim to minimize human biases, biases stemming from data and evaluation choices raise similar concerns. Therefore, it is crucial to adhere to ethical and impartial practices when applying machine learning to real-world financial anomaly detection.

The paper by Gao and Han examines the impact of artificial intelligence (AI) on the goals and processes of auditing financial statements [39]. They argue that relying on manual inspection of accounting records for auditing can introduce biases and ethical risks, whereas an AI-enabled approach can address these issues. Auditing solely based on verifying compliance with accounting principles may lead to audits being biased towards procedural legitimacy, rather than ensuring fairness and reliability of financial information. Moreover, using predefined anomaly categories can limit the scope of what is considered an audit red flag. However, the author asserts that using synthetic anomalies to train AI systems assumes the normality of the original data and reflects the auditors' subjective opinions on what is considered an anomaly. This approach may fail to capture the complexities of the real world and introduce bias. The study argues that AI offers techniques to identify a broader, interconnected set of audit propositions to capture

Table 1
Sources of biases in ai-based auditing systems.

S/N	Source	Description	Example Manifestations	References	% Studies Identifying
1	Data Deficiencies	Errors, uncertainty, or lack of diversity in training data	Gender bias, demographic skew	[7,8,33,34,35,36,37,38,39,40]	42 %
2	Demographic Homogeneity	Models trained on limited population diversity	Discrimination against minorities	[23,41,42,43,44,45,46,47,48,49]	24 %
3	Spurious Correlations	Proxy variables correlating with protected attributes	Racial discrimination through zip codes	[50,51]	9 %
4	Improper Comparators	Unfair benchmarking groups reinforcing disparities	Evaluating only on high-income groups	[23,52,53]	11 %
5	Cognitive Biases	Designers' skewed assumptions and thinking embedded in systems	Confirmation bias, selective perception	[23,32,54,55]	14 %

complex risks. By leveraging more comprehensive data, biases from sampling and over-reliance on internal records can be reduced. Furthermore, AI-based anomaly detection that utilizes multiple data sources and validation methods can help mitigate bias stemming from subjective definitions of abnormality.

In their research on the application of big data and artificial intelligence to audit, Xing et al. argue that AI can improve audit efficiency but also introduces new technical and systemic risks [40]. Biases in data collection, such as limited samples or subjective definitions of anomalies, can affect the models. Legacy auditing sampling approaches that overlook rare but significant anomalies persist in AI systems trained on biased data. The authors conclude that the assumptions embedded in AI algorithms and training data also raise ethical risks. Assuming anomaly-free data fails to account for legitimate irregularities that are dismissed as normal. They state that automated anomaly detection focused on predefined categories limits the scope of identified issues and risks overlooking risks not captured by the specified features. While the researchers acknowledge that AI provides productivity gains, they suggest that auditors must evaluate systems for bias, inform clients about standardization needs, and continually assess solutions against audit principles. No model can fully automate subjective domain knowledge and ethical judgment. Similar to sampling, AI remains vulnerable to bad input data and biases that overlook contextual risks. Fair and representative data curation, as well as ongoing human oversight, are crucial. The risks of bias and ethical lapses from AI reflect and amplify existing debates on balancing efficiency, analysis, and professional judgement. However, informed usage of AI's analytical power can counterbalance human limitations if aligned with core audit values. This requires considering its capabilities and limitations within the overall audit process.

A recent examination of the ethical challenges posed by the increasing use of AI-based decision-making in accounting identifies five main areas of concern: objectivity, privacy, transparency, accountability, and trustworthiness [54]. Analysing these through Rest's well-known four-component model of ethical decision-making antecedents, Smith et al. find that AI systems currently lack essential attributes related to moral sensitivity, judgment, motivation, and character to make ethical automated choices on their own. Consequently, the authors argue that appropriate governance processes and updated auditing mechanisms must reinforce shared accountability between humans and AI to adjust to this emerging technology. Since AI cannot meet crucial requirements for ethical decision-making independently Lehner, O. M., et al. caution that risks related to embedded biases, opacity, and diffuse responsibilities require careful oversight and collaboration [58]. While AI holds the promise of significant gains in efficiency and insight from large datasets, unchecked adoption of AI in accounting also poses a threat of replicating existing challenges in the profession regarding analytical subjectivity, transparency, and public trust. By outlining these risks, Smith et al. offer a timely analysis of how governance and assurance reforms can support ethical AI-assisted decision-making as automation becomes more prevalent in the accounting field.

The increasing integration of artificial intelligence into the auditing process raises ethical and social concerns alongside efficiency gains from automating repetitive tasks, as outlined in recent research [51]. Core auditing principles such as professional scepticism, competence, care, and judgment could be compromised by unexplainable AI systems that hinder human oversight and understanding. Additionally, reliance on AI in continuous monitoring and evaluation increases privacy and cybersecurity risks. Lack of trust in opaque AI could create accountability gaps, impacting public responsibility perceptions. More broadly, accelerated job displacement also poses challenges in the absence of governance guardrails. analysing these issues affecting duty ethics, utilitarian outcomes, and social contracts, Munoko et al. emphasize the importance of transparency, independence, and human-AI collaboration in auditing functions, even with automation [59]. From judgment gaps to displaced careers, supposedly neutral AI carries ethical and social risks linked to users, clients, and the public, which updated policies and professional codes must address as adoption accelerates.

The reviewed studies demonstrate that some risks associated with AI highlighted or intensified existing auditing challenges related to bias, assumptions, sampling, and professional judgment. For instance, biases in AI models perpetuated historical discrimination issues, while reliance on predefined anomalies risked overlooking risks similar to limitations of traditional sampling methods. However, although the aforementioned studies examined potential sources of bias in AI applied to auditing, it is crucial to recognize the limitations and gaps. Table 2 provides an overview of the key gaps identified in multiple studies. Addressing these gaps would have significantly enhanced AI auditing methods, which is essential as reliance on algorithmic systems increases. From these findings, it is clear that AI in auditing emphasizes real-world testing, interdisciplinary collaboration, exploration of human involvement, standardization of audit methods, resolution of data quality issues, and consideration of external influencing factors more broadly. By

Table 2
Gaps in studies on sources of biases in AI auditing systems.

S/No	Gaps/Limitations	Reference
1	Studies did not consider real-world application and comparison with existing models	[7,8,35,43,44,46,49]
2	Lack of Multidisciplinary approach	[23]
3	Overfitting, difficulty determining neuron counts, and dependency on quality/quantity of data.	[36]
4	Lacks exploration of other variables and some variables used may not apply well across different economic and regional contexts	[37,50]
5	Challenges posed by diverse dataset licensing and attribution in AI	[38]
6	Further investigation needed for human control in algorithmic systems and its impact on risk levels	[41]
7	No consensus on bias metrics and fairness definitions to standardize audits of AI systems	[42]
8	Data preparation, cleaning, labeling and minimization challenges	[47,52,55]
9	Imbalanced Datasets	[48]
10	Lack of significant evidence on influences of audit quality	[34,51]
11	Need for comprehensive specifications developed collaboratively with LLM application creator	[54]
12	Limited scope of audit's impact and potential overemphasis on technical improvements	[32]

taking measures to address these knowledge gaps, researchers could have further developed best practices for auditing AI that tackled both longstanding auditing concerns and new ethical risks introduced by AI systems. This would have ensured that AI auditing approaches continued to progress responsibly and accountably as automation and algorithmic decision-making became more prevalent in the auditing profession.

Wider ethical risks of AI augmentation

Apart from technical biases, the integration of AI poses several ethical risks and normative tensions around goals, values and accountability illuminated through the reviewed studies as conveyed in Table 3. Experts from regulatory bodies, standard-setting organizations, academic research, and senior practitioners in the field audit and compliance caution against prioritizing productivity over the core principles of professional judgment, impartiality, and truthfulness that form the foundation of auditing's ethical code. The fundamental principles of independence and professional scepticism that underpin audit quality are at risk of being compromised by an improper balancing of automation and human discretion. Leading researchers warn that the rapid decline in human capabilities and the lack of transparency regarding AI system limitations could jointly erode the conscience and social accountability of auditors over time [60]. Some argue that the principles of professional scepticism and impartiality, which guide auditors toward the public interest, should take precedence over the commercial goals of efficiency and cost-savings that AI promises [61]. Others contend that the opaque nature of AI systems conceals the actual level of human oversight and discretion involved. In the absence of transparency on appropriate checks, the rapid advancements in AI could quietly erode the diligence and social responsibility of auditors.

Complex downstream interactions amplifying harm

The analysis found AI biases can propagate through multiple complex socio-technical interactions during design, training, deployment, and monitoring stages as shown in Fig. 3.

Unchecked technical biases accumulate into discriminatory decisions and behaviours towards vulnerable groups. Cascading network effects then magnify harm as marginalization inhibits access to opportunities for redress or skills development. Vicious cycles reinforce structural barriers and widen capability divides [32]. Without deliberate controls, reliance on flawed historical data, narrow demographics, inappropriate evaluation metrics and proxy variables becomes deeply ingrained.

Biases accumulate through feedback loops between interacting model versions, benchmarking tests, performance metrics and pipelines automated at scale. Real-world impacts remain invisible until substantial later stage harms manifest.

Evolving technical and ethical guardrails

To mitigate these risks, researchers suggested implementing technical and ethical safeguards to ensure trustworthy and socially beneficial AI integration, as outlined in Table 4. The studies highlight that technical interventions should be supported by strong governance mechanisms for responsible AI use [87]. As shown in Fig. 4, there is an interconnected relationship among auditing, transparency, stakeholder prioritization, and proactive algorithm thinking, which are essential to establish ethical guidelines. From the studies, it was observed that collaboration among engineers, auditors, and ethicists during system lifecycles can integrate ethical values into sociotechnical architectures that prioritize the public interest. Continual evolution of the framework across the areas depicted in Fig. 4 is crucial to managing stakeholder trade-offs as new risks emerge with AI advancements. Ensuring responsible oversight involves establishing clear organizational structures and competencies in policy, industry, and society, going beyond empty words.

Table 4 provides an overview of guardrails proposed to enable trustworthy and socially aligned auditing innovation with AI, including preventive and detective controls. The data and ethical guardrails take a proactive approach by enhancing input diversity and incorporating values such as fairness into system design using techniques like synthetic data and impact assessments [88,113]. On the other hand, algorithmic and evaluation guardrails take a reactive approach, focusing on interpretability, behavioural monitoring, and extending assessment metrics beyond just accuracy [41,110]. Regardless of the category, maintaining human discretion and

Table 3
Ethical risks from AI integration in auditing.

S/N	Risk Category	Key Concerns	Example Harms	References
1	Goal Alignment	Conflict between efficiency gains and audit rigor or public duties	Lower diligence, diluted scepticism	[62,63,64,65,66]
2	Value Alignment	Lack of transparency erodes accountability	Public trust deficits, culpability ambiguities	[67,68,69]
3	Accountability	Overreliance on algorithms sans human judgment	Deskilling effects, conscience gaps	[69,70,71]
4	Skill Substitution	Rapid automation destabilizing work domains	Audit profession hollowed out	[59,72,73,74,75]
5	Impartiality	Client data dependence risks auditor independence	Conflicts of interest	[76,77,78,79,80]
6	Privacy	Personal data exploitation without consent	Confidentiality breaches	[59,78,79,80,81,82,83,84,85,86]

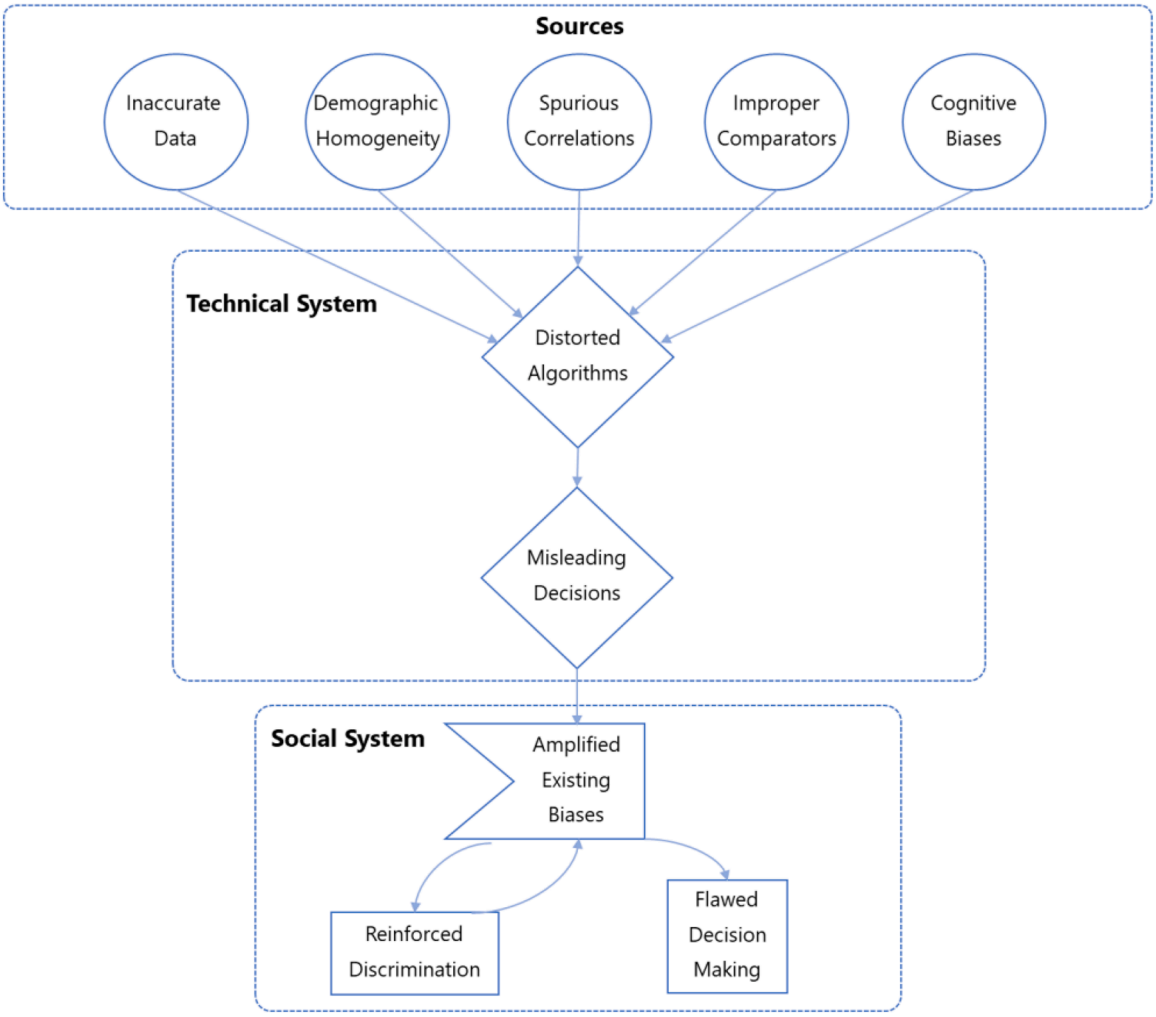

Fig. 3. AI Bias propagation pathways across sociotechnical dimensions.

Table 4
Emergent guardrails for trustworthy AI-based auditing.

S/N	Guardrail	Category	Objective	Example Techniques	References
1	Data Guardrails	Preventive	Enhance data diversity; Prevent distortion roots	Targeted collection, reweighting, synthetic data	[88,89,90,91,92,93]
2	Algorithmic Guardrails	Detective	Build interpretability; Assess model behaviours	Explainability interfaces, sandbox testing	[93,94,95,96,97,98,99,100,101]
3	Evaluation Guardrails	Detective	Broaden assessment metrics; Test on representative groups	Statistical parity measures, subgroup validation	[41,42,102,103,104,105,106,107,108,109,110]
4	Ethical Guardrails	Preventive	Embed values like fairness into system design	Impact assessments, ethics boards, redress pathways	[110,111,112,113]
5	Human Guardrails	Responsible stakeholder	Maintain reasonable discretion over automation	Judgment-based authority boundaries	[86,97,110,114,115,116,117,118,119,120]

judgment is crucial for prioritizing the public interest [86,97,110,114,120]. The multifaceted guardrails aim to preserve benefits while managing risks as AI plays a more significant role in audits. However, effective oversight involves coordinating interventions across policy, organizational, and social dimensions rather than relying solely on technical controls.

Consolidation and analysis of the findings from the auditing context revealed that effective oversight of AI auditing innovation requires integrating information across various dimensions. As shown in Fig. 4, the iterative process of independently auditing AI systems allows for crucial feedback loops to shape auditing principles and meet stakeholder needs. Transparent AI policies and risk

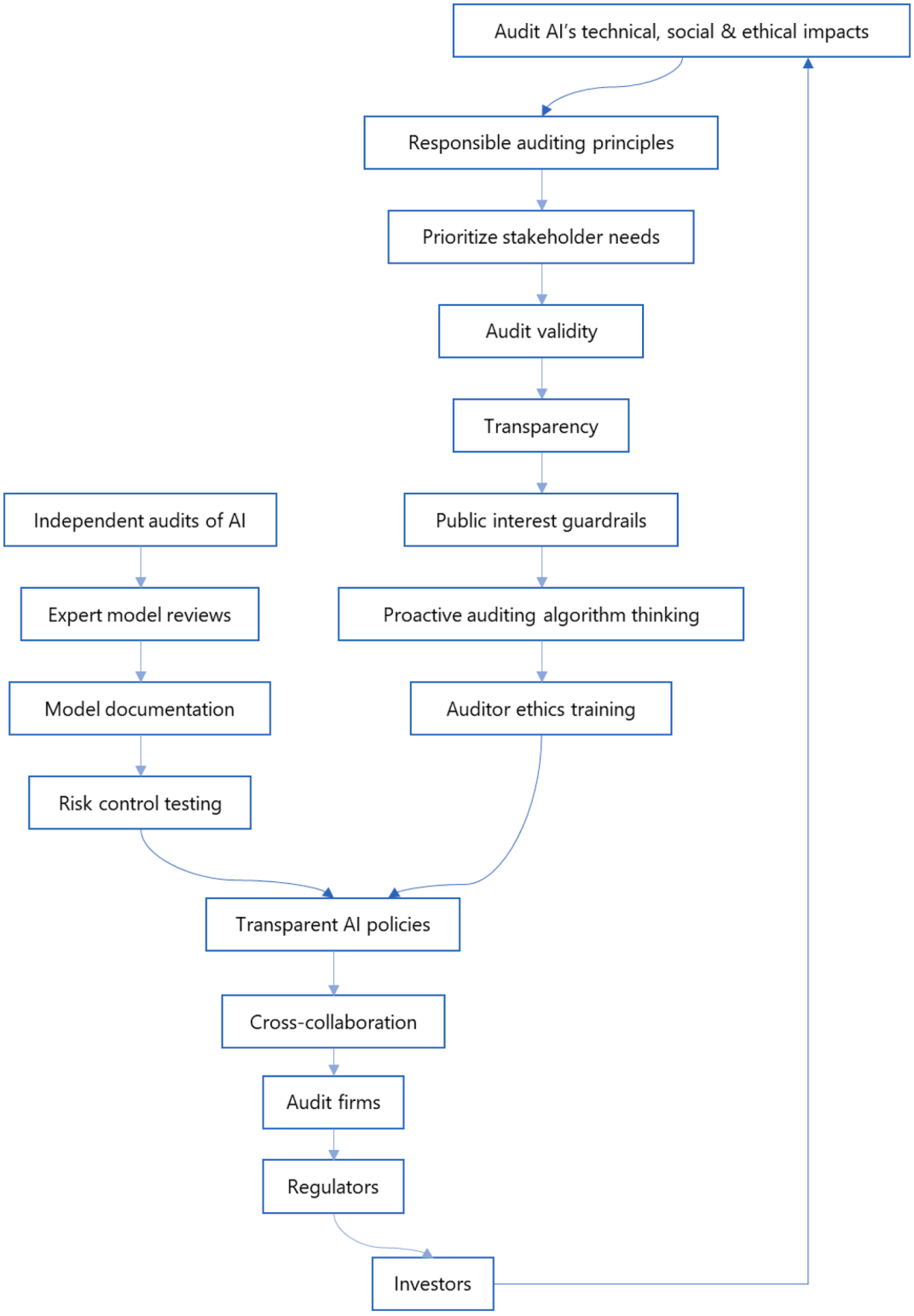

Fig. 4. Cascading flows for instilling guardrails in AI-Based auditing.

control testing connect field deployments to public interest guidelines. The interconnected processes illustrate that technical auditing cannot function in isolation – cooperation among audit firms, regulators, investors, and civil society is essential. The interconnected relationships also highlight the connections from detailed algorithmic evaluations by expert review boards to reassessing broader impacts. In totality, Fig. 4 highlights the studies emphasize on an intricate connection that call for a socio-technical systems perspective to achieve comprehensive governance objectives. Isolated auditing procedures fail to address emerging risks as automated decision systems become embedded in organizational workflows. However, shared responsibility entails collaborative actions to proactively establish guidelines that can support ethical innovation.

Discussion

Significance of prevalent bias sources

The study uncovered significant issues stemming from data deficiencies, which have significant implications for auditors' reliance on AI. Legacy financial systems with poor data integrity undermine credibility regarding algorithmic dependability [121]. Additionally, there are constraints regarding access to high-quality, representative, and diverse training data, often collected opportunistically or lacking diligence [122]. Furthermore, historical discrimination is ingrained in datasets, necessitating extensive transparency and cleaning interventions [123]. The prevailing demographic homogeneity and lack of diverse perspectives is an under-examined but primary source of harm. Homogeneity perpetuates structural inequities and inhibits impartiality essential for truth-seeking [124]. Tools based solely on technical features often exacerbate prejudice through unexamined feedback loops across interacting algorithmic systems over time [122].

Auditors have a crucial oversight role in scrutinizing the integrity of training data, questioning assumptions encoded in models, and proactively assessing risks of unfairness towards minorities. However, relying solely on interpretability mechanisms to explain model behaviours also has limitations. The rationales provided describe correlations rather than causations, still allowing the possibility of erroneous inferences [125]. To safeguard against the erosion of human conscience and biases, it is necessary to maintain judicious boundaries on automation scope guided by risk-benefit analyses weighing both technical accuracy and social impacts [55]. Regaining lost public trust also requires transparency on the extent of human discretion retained despite advances in artificial autonomy [67].

Wider risk horizons beyond accuracy

The study uncovered ethical tensions, including diluting diligence, overreliance on efficiency, and threats to impartiality from market incentives, that warranted governance. Predictive models focused narrowly on mining insights risked compromising principles of professional scepticism vital for assurance quality [86]. Rapid automation also posed challenges to auditing's professional ethos by deskilling human capabilities like critical thinking, ethical discernment, and impartial judgement [64]. This made symbiotic and transparent approaches that responsibly balanced machine and human proficiencies necessary.

Unrestricted access to sensitive data granted to external AI providers may have necessitated oversight mechanisms to preserve independence and prevent irreversible industry concentration [78,80]. Impact evaluation through multistakeholder participation, continuous improvement cycles and public deliberation around trade-offs also fostered trust-building essential for stable adoption. Beyond technical accuracy, solutions promoting holistic prosperity demanded evaluating well-being advancements collectively through an empathetic lens prioritizing marginalized groups.

Cascading impacts of algorithmic biases

The review highlighted the significant and long-lasting consequences of even small technical biases in AI systems. Over time, these biases can accumulate and lead to discriminatory decisions, affecting access to important opportunities such as credit, housing, education, and political participation [44]. This lack of empathy allowed for the development of reinforcing feedback loops that perpetuated existing barriers and widened the gap between communities. As a result, vulnerable populations experienced limited development opportunities and increased instability [19].

To address these issues, it was clear that solutions based solely on technical metrics such as accuracy or profit were insufficient. A multidimensional equity lens was necessary to evaluate the holistic social impacts of AI systems, and guardrails that prioritized affected communities were needed alongside technical assurances for trustworthy AI [22,112]. Urgent action was required to prevent the negative consequences of flawed AI systems from impacting universal prosperity.

Balance through collaborative guardrails

Researchers highly recommend collaborative governance frameworks that incorporate technical, social, and ethical aspects among various stakeholders, based on the "SUCCESS" principles [126]. Solitary technical interventions are usually inadequate. Instead, it is essential to have cross-functional oversight committees, formal review procedures, and executive leadership supervision to guarantee responsible AI development, deployment, and monitoring [34]. This helps avoid unintended harm and fosters ethical consistency. As shown in Fig. 5, setting up ethical boundaries along innovation pathways facilitates progress without hindering it [67]. These boundaries are enforced through updated regulations, committee evaluations, training initiatives, and participatory design processes.

Specifically, Fig. 5 provides an overview of the proposed ethical guardrails flow, covering policy, organizational, and competency

layers. By implementing multidisciplinary guardrails through updated regulations, expert oversight boards, training programs, and participative development workflows, firms can proactively assess and mitigate ethical risks. This upholds both the innovative promise and public trust imperatives surrounding the advancement of AI systems [34,67]. Economic goals should not compromise principles of impartiality or professional scepticism. In fact, a virtue-driven brand reputation founded on collective conscience can provide the most sustainable competitive advantage. Moving forward requires a continuous, participative synthesis of insights across society to balance technical promise and ethical responsibility.

While principles provide useful guidelines, concrete oversight structures are crucial for implementing trustworthy AI innovation in accounting and auditing [34]. Recent United States of America governments initiatives have shown that involving multiple parties from industry, government, and civil society promotes accountability. For example, the Biden Administration's AI executive order mandates independent auditing and the establishment of an expert safety board to regularly assess the risks of financial and accounting AI systems [127]. In addition, major accounting and audit firms stress the importance of maintaining human-in-the-loop decision--making despite the growing automation. Regulations alone are not sufficient without developing competencies at the intersection of ethics and technology. Proactive design methodologies can strengthen development processes that balance guidelines and innovations [67]. In general, by taking collaborative action across policy, organizational, and competency domains, firms can establish multi-disciplinary guidelines to ensure that systems are aligned with ethical principles. This necessitates continuously synthesizing insights from stakeholders to strike a balance between innovation and public trust.

Fig. 5. Proposed framework for instilling ethical guardrails.

Conclusion and recommendations

Conclusion

The research identified five primary sources of unfair bias originating from data inadequacies, demographic homogeneity, spurious correlations, improper comparators, and embedded cognitive biases. It also revealed associated ethical tensions, such as conflicting priorities, threats to sound judgment, impartiality erosion, privacy violations, and marginalizing impacts. Furthermore, the study found that even isolated technical biases can propagate through socioeconomic structures, limiting opportunities for vulnerable communities.

The research consolidated findings from multiple disciplines to propose pathways towards reliable and ethical AI integration into assurance processes. It emphasized the need for guardrails that encompass data quality enhancement, model transparency, organizational accountability, and the preservation of human discretion. The conceptual cascade model of bias accumulation illustrates how unfairness intensifies over time.

However, the review methods had limitations, including the lack of specificity in model evaluation procedures, unclear metrics definitions, insufficient contextual validation across economic settings, and inadequate assessment of societal impacts. Numerous research gaps remain, such as the need for clearer boundaries around human involvement, standardized risk assessment criteria, data licensing remedies, bias audit selectivity, and real-world comparisons.

Recommendations

Considering the increasing permeation of artificial intelligence (AI) and machine learning (ML) in business contexts, it is essential that auditors remain vigilant in ensuring that predictive models do not perpetuate historical biases and make fair, accountable, and transparent decisions. However, there is still work to be done to mature the field of algorithm auditing and promote responsible AI practices across industries. Future research should focus on strengthening legal oversight for independent ethics boards and mandating diverse development teams while formulating practitioner codes of ethics. Other open questions warrant exploration, including technical solutions for improving data veracity and model interpretability, integrating tailored bias testing protocols across audit workflows, and developing risk frameworks for autonomous systems based on empirical evidence and cost-benefit analyses.

Several promising avenues for further research exist to expand the body of knowledge in this emerging discipline. For academics, these include exploring technical solutions to improve model interpretability and data diversity, developing tailored bias testing toolkits for audit workflows, examining the impact of explainability measures on auditor judgments, and proposing risk frameworks to address concerns around autonomous systems.

Policymakers also have a crucial role to play in investigating regulations around independent audits and ethics boards for AI, requiring diverse development teams, incorporating re-validation needs as models drift over time, and instituting practitioner oaths that cement commitments to ethical AI principles.

Audit firms can build competency at the intersection of technology and ethics through impact assessments of AI systems and formalizing debiasing protocols and techniques like data perturbation techniques to preserve privacy. Technology vendors must also prioritize transparency through interfaces and testing reports, while collaborating with diverse domain experts to uncover blind spots in AI assurance tools. To ensure AI safety and fairness and scale trustworthy algorithm auditing practices globally, stakeholders across ecosystems must work collectively. Pursuing open research questions can significantly mature the discipline.

Discipline

Information Technology and Engineering

CRediT authorship contribution statement

Wilberforce Murikah: Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Investigation, Resources, Project administration. **Jeff Kimanga Nthenge:** Validation, Resources, Writing – review & editing, Visualization. **Faith Mueni Musyoka:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] C. Balfe, P. Button, M. Penn, D.J. Schwegman, Infrequent identity signals, multiple correspondence, and detection risks in audit correspondence studies, Field Methods 35 (1) (2023) 3–17, https://doi.org/10.1177/1525822X211057623/SUPPL_FILE/SJ-PDF-1-FMX-10.1177_1525822X211057623.PDF.

[2] D.L. Marino, et al., AI augmentation for trustworthy AI: augmented robot teleoperation, in: International Conference on Human System Interaction, HSI 2020, 2020, pp. 155–161, https://doi.org/10.1109/HSI49210.2020.9142659.

[3] S. Latifi, AI Or IA (Intelligence Augmentation)- Future Trends, 2023, p. 5, https://doi.org/10.1109/CMVIT57620.2023.00010. –5.

[4] T.J. DeStefano, T. Teodorovicz, J. Cho, H. Kim, J. Paik, What, in: Determines AI Adoption?, 2022, 2022, https://doi.org/10.5465/AMBPP.2022.14791abstract.
[5] L. Cao, T-Shaped Teams: Organizing to Adopt AI and Big Data At Investment Firms, 2021, https://doi.org/10.56227/22.1.13.
[6] H.T. Alattas, et al., Extract Compliance-Related Evidence Using Machine Learning, in: Proceedings - 2022 14th IEEE International Conference on Computational Intelligence and Communication Networks, CICN 2022, 2022, pp. 537–542, https://doi.org/10.1109/CICN56167.2022.10008324.
[7] A. Bakumenko, A. Elragal, Detecting anomalies in financial data using machine learning algorithms, Systems 2022 10 (5) (2022) 130, https://doi.org/10.3390/SYSTEMS10050130. *Vol. 10, Page 130*.
[8] S.V. Ivakhnenkov, Application of artificial intelligence in auditing, Scientific notes of NaUKMA. Econ. Sci. 8 (1) (2023) 54–60, https://doi.org/10.18523/2519-4739.2023.8.1.54-60.
[9] C. Zhang (Abigail), Predict audit quality using machine learning algorithms, SSRN Elect. J. (2018), https://doi.org/10.2139/SSRN.3449848.
[10] F. Giannotti, AI3SD Video: Explainable machine Learning For Trustworthy AI, 2021, https://doi.org/10.5258/SOTON/AI3SD0157.
[11] J.-M. John-Mathews, Critical empirical study on black-box explanations in AI, in: International Conference on Interaction Sciences, 2021.
[12] S. Shrestha, S. Das, Exploring gender biases in ML and AI academic research through systematic literature review, Front Artif Intell 5 (2022) 976838, https://doi.org/10.3389/FRAI.2022.976838/BIBTEX.
[13] N. Schmidt and B.E. Stephens, "An introduction to artificial intelligence and solutions to the problems of algorithmic discrimination," *arXiv.org*, 2019.
[14] M. DeCamp, C. Lindvall, Mitigating bias in AI at the point of care, Science (1979) 381 (6654) (2023) 150–151, https://doi.org/10.1126/SCIENCE.ADH2713.
[15] C.X. Kerasidou, A. Kerasidou, M. Buscher, S. Wilkinson, Before and beyond trust: reliance in medical AI, J. Med. Ethics 48 (11) (2022) 852–856, https://doi.org/10.1136/MEDETHICS-2020-107095.
[16] A.A. Nichol, M.C. Halley, C.A. Federico, M.K. Cho, P.L. Sankar, Not in my AI: moral engagement and disengagement in health care AI development, in: Pacific Symposium on Biocomputing, 2023, pp. 496–506, https://doi.org/10.1142/9789811270611_0045.
[17] M. Khosravy, K. Nakamura, A. Pasquali, O. Witkowski, N. Nitta, N. Babaguchi, When AI facilitates trust violation: an ethical report on deep model inversion privacy attack, in: Proceedings - 2022 International Conference on Computational Science and Computational Intelligence, CSCI 2022, 2022, pp. 929–935, https://doi.org/10.1109/CSCI58124.2022.00166.
[18] T.Y. Zhuo, Y. Huang, C. Chen, Z. Xing, Red Teaming ChatGPT Via Jailbreaking: Bias, Robustness, Reliability and Toxicity, 2023. Accessed: Nov. 29, 2023. [Online]. Available: https://arxiv.org/abs/2301.12867v4.
[19] P. Gupta, Addressing Bias and Fairness Issues in Artificial Intelligence, 2023.
[20] C.G. Harris, Mitigating age biases in resume screening AI models, in: The International FLAIRS Conference Proceedings 36, 2023, https://doi.org/10.32473/FLAIRS.36.133236.
[21] J. Menke, et al., Establishing institutional scores with the rigor and transparency index: large-scale analysis of scientific reporting quality, J. Med. Internet Res. 24 (6) (Jun. 2022) e37324, https://doi.org/10.2196/37324.
[22] M. Del Carmen Fernández Martínez, A. Fernández, AI in recruiting. Multi-agent systems architecture for ethical and legal auditing, in: IJCAI International Joint Conference on Artificial Intelligence 2019, 2019, pp. 6428–6429, https://doi.org/10.24963/IJCAI.2019/903.
[23] R.N. Landers, T.S. Behrend, Auditing the AI auditors: a framework for evaluating fairness and bias in high stakes ai predictive models, Am. Psychol. 78 (1) (2022) 36–49, https://doi.org/10.1037/AMP0000972.
[24] J. Koreff, L. Baudot, S.G. Sutton, Exploring the impact of technology dominance on audit professionalism through data analytic-driven healthcare audits, J. Inf. Syst. 37 (3) (2023) 59–80, https://doi.org/10.2308/ISYS-2022-023.
[25] A. Krakowski, E. Greenwald, T. Hurt, B. Nonnecke, M. Cannady, Authentic integration of ethics and AI through sociotechnical, problem-based learning, in: Proceedings of the AAAI Conference on Artificial Intelligence 36, Jun. 2022, pp. 12774–12782, https://doi.org/10.1609/AAAI.V36I11.21556.
[26] A. Matthews, Sociotechnical imaginaries in the present and future university: a corpus-assisted discourse analysis of UK higher education texts, Learn Media Technol 46 (2) (Apr. 2021) 204–217, https://doi.org/10.1080/17439884.2021.1864398.
[27] G. Stefani, M. Biggeri, L. Ferrone, Sustainable transitions narratives: an analysis of the literature through topic modelling, Sustainability 2022 14 (4) (2022) 2085, https://doi.org/10.3390/SU14042085. *Vol. 14, Page 2085*.
[28] J.A. McDermid, Y. Jia, Z. Porter, I. Habli, Artificial intelligence explainability: the technical and ethical dimensions, Philosoph. Transact. Royal Society A 379 (2207) (2021), https://doi.org/10.1098/RSTA.2020.0363.
[29] J. Goldenfein, Algorithmic Transparency and Decision-Making Accountability: Thoughts for Buying Machine Learning Algorithms, 2019.
[30] A.Z. Huriye, The ethics of artificial intelligence: examining the ethical considerations surrounding the development and use of AI, Am. J. Technol. 2 (1) (Apr. 2023) 37–44, https://doi.org/10.58425/ajt.v2i1.142.
[31] W. Mansell, V. Huddy, The Assessment and Modeling of Perceptual Control: A Transformation in Research Methodology to Address the Replication Crisis, 22, 2018, pp. 305–320, https://doi.org/10.1037/gpr0000147.
[32] I.D. Raji, J. Buolamwini, Actionable auditing: investigating the impact of publicly naming biased performance results of commercial AI products, in: AIES 2019 - Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, 2019, pp. 429–435, https://doi.org/10.1145/3306618.3314244.
[33] O. Adamyk, V. Benson, B. Adamyk, H. Al-Khateeb, A. Chinnaswamy, Does artificial intelligence help reduce audit risks?, in: 2023 13th International Conference on Advanced Computer Information Technologies, ACIT 2023 - Proceedings, 2023, pp. 294–298, https://doi.org/10.1109/ACIT58437.2023.10275661.
[34] Z. Wang, et al., An Exploratory Study of AI System Risk Assessment from the Lens of Data Distribution and Uncertainty, Cornell University, Dec. 2022. Accessed: Nov. 29, 2023. [Online]. Available: https://arxiv.org/abs/2212.06828v1.
[35] W. Xiao, Research on applied strategies of business financial audit in the age of artificial intelligence, in: Proceedings - 2022 18th International Conference on Computational Intelligence and Security, CIS 2022, 2022, pp. 436–439, https://doi.org/10.1109/CIS58238.2022.00098.
[36] Q.-J. Smith, R. Valverde, J. Molson, A perceptron based neural network data analytics architecture for the detection of fraud in credit card transactions in financial legacy systems, WSEAS Transactions on Systems and Control 16 (2021) 358–374, https://doi.org/10.37394/23203.2021.16.31.
[37] D.J. Chi, Z. De Shen, Using hybrid artificial intelligence and machine learning technologies for sustainability in going-concern prediction, Sustainability 14 (3) (2022) 1810, https://doi.org/10.3390/SU14031810. *2022Vol. 14, Page 1810*.
[38] S. Longpre, et al., The Data Provenance Initiative: A Large Scale Audit of Dataset Licensing & Attribution in AI, Cornell University, 2023. Accessed: Nov. 29, 2023. [Online]. Available: http://arxiv.org/abs/2310.16787.
[39] Y. Gao, L. Han, Implications of artificial intelligence on the objectives of auditing financial statements and ways to achieve them, Microprocess. Microsyst. (2021) 104036, https://doi.org/10.1016/J.MICPRO.2021.104036.
[40] Z. Xing, L. Zhu, Z. Lijun, A study on the application of the technology of big data and artificial intelligence to audit, in: Proceedings - 2020 International Conference on Computer Engineering and Application, ICCEA 2020, Mar. 2020, pp. 797–800, https://doi.org/10.1109/ICCEA50009.2020.00174.
[41] E. Kazim, A.S. Koshiyama, A. Hilliard, R. Polle, Systematizing audit in algorithmic recruitment, J. Intellig. 2021 9 (3) (2021) 46, https://doi.org/10.3390/JINTELLIGENCE9030046. *Vol. 9, Page 46*.
[42] P. Saleiro, et al., Aequitas: A Bias and Fairness Audit Toolkit, Cornell University, 2019. Accessed: Nov. 29, 2023. [Online]. Available: https://arxiv.org/abs/1811.05577v2.
[43] S. Pentyala, D. Melanson, M. De Cock, G. Farnadi, PrivFair: a Library for Privacy-Preserving Fairness Auditing, Cornell University, 2022. Accessed: Nov. 29, 2023. [Online]. Available: http://arxiv.org/abs/2202.04058.
[44] S. Muhammad, Auditing Algorithmic Fairness With Unsupervised Bias Discovery, Amsterdam Intelligence (2021). https://amsterdamintelligence.com/posts/bias-discovery.
[45] N.I. Raddatz, P.A. Raddatz, K. Sorensen, K. Ogunade, The adverse effects of the 'Anticipation of Racial Discrimination' on auditors who are black, indigenous, or people of color (BIPOC): an exploratory study with research propositions, Account. Horizons (2023) 1–9, https://doi.org/10.2308/HORIZONS-2022-098.
[46] S.D. Jaiswal, A.K. Verma, A. Mukherjee, Auditing Gender Analyzers On Text Data, Cornell University, 2023. Accessed: Nov. 29, 2023. [Online]. Available: https://arxiv.org/abs/2310.06061v1.

[47] G.G. Clavell, M.M. n. Zamorano, C. Castillo, O. Smith, A. Matic, Auditing algorithms: on lessons learned and the risks of data minimization, in: AIES 2020 - Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, 2020, pp. 265–271, https://doi.org/10.1145/3375627.3375852.

[48] Y. Xu, P. Terhörst, K. Raja, M. Pedersen, A Comprehensive Analysis of AI Biases in DeepFake Detection With Massively Annotated Databases, Cornell University, 2022. Accessed: Nov. 29, 2023. [Online]. Available: http://arxiv.org/abs/2208.05845.

[49] W. Zekun, S. Bulathwela, A.S. Koshiyama, Towards Auditing Large Language Models: Improving Text-Based Stereotype Detection, Cornell University, 2023. Accessed: Nov. 29, 2023. [Online]. Available: http://arxiv.org/abs/2311.14126.

[50] M.J. Rahman, H. Zhu, M.M. Hossain, Auditor choice and audit fees through the lens of agency theory: evidence from Chinese family firms, J. Family Business Manag. 13 (4) (2023) 1248–1276, https://doi.org/10.1108/JFBM-02-2023-0027/FULL/PDF.

[51] A. Puspitasari, Z. Baridwan, A.F. Rahman, The effect of audit competence, independence, and professional skepticism on audit quality with auditor's ethics as moderation variables, Int. J. Business, Econ. Law 18 (2019) 135–144.

[52] A. Mishra, Y. Gorana, Who Decides if AI is Fair? The Labels Problem in Algorithmic Auditing, Cornell University, 2021. Accessed: Nov. 29, 2023. [Online]. Available: https://arxiv.org/abs/2111.08723v1.

[53] P. Dimaio, Knowledge representation for algorithmic auding to detangle systemic bias, in: International Semantic Intelligence Conference, 2021, pp. 149–157.

[54] C. Rastogi, M. Tulio Ribeiro, N. King, H. Nori, S. Amershi, Supporting human-AI collaboration in auditing LLMs with LLMs, in: AIES 2023 - Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society, 2023, pp. 913–926, https://doi.org/10.1145/3600211.3604712.

[55] M. Soleimani, A. Intezari, N. Taskin, D. Pauleen, Cognitive biases in developing biased Artificial Intelligence recruitment system, in: Proceedings of the Annual Hawaii International Conference on System Sciences 2020, 2021, pp. 5091–5099, https://doi.org/10.24251/HICSS.2021.620.

[56] H. Suresh, K.M. Lewis, J. Guttag, A. Satyanarayan, Intuitively assessing ML model reliability through example-based explanations and editing model inputs, in: International Conference on Intelligent User Interfaces, Proceedings IUI 22, 2022, pp. 767–781, https://doi.org/10.1145/3490099.3511160.

[57] M. Lokanan, V. Tran, N.H. Vuong, Detecting anomalies in financial statements using machine learning algorithm: the case of Vietnamese listed firms, Asian J. Account. Res. 4 (2) (2019) 181–201, https://doi.org/10.1108/AJAR-09-2018-0032/FULL/PDF.

[58] O.M. Lehner, K. Ittonen, H. Silvola, E. Ström, A. Wührleitner, Artificial intelligence based decision-making in accounting and auditing: ethical challenges and normative thinking, Account., Audit. Accountabil. J. 35 (9) (2022) 109–135, https://doi.org/10.1108/AAAJ-09-2020-4934/FULL/PDF.

[59] I. Munoko, H.L. Brown-Liburd, M. Vasarhelyi, The ethical implications of using artificial intelligence in auditing, J. Bus. Ethics 167 (2) (2020) 209–234, https://doi.org/10.1007/S10551-019-04407-1.

[60] M. Blösser, A. Weihrauch, A consumer perspective of AI certification – the current certification landscape, consumer approval and directions for future research, Eur. J. Mark 58 (2) (2023) 441–470, https://doi.org/10.1108/EJM-01-2023-0009/FULL/PDF.

[61] M.T. Fülöp, D.I. Topor, C.A. Ionescu, J. Cifuentes-Faura, N. Mägdaş, Ethical concerns associated with artificial intelligence in the accounting profession: a curse or a blessing? J. Business Econ. Manag. 24 (2) (2023) 387, https://doi.org/10.3846/JBEM.2023.19251. -404–387–404.

[62] R. Libby, P. Witz, Can artificial intelligence reduce the effect of independence conflicts on audit firm liability? SSRN Elect. J. (2020) https://doi.org/10.2139/SSRN.3734629.

[63] A. Fedyk, J. Hodson, N. Khimich, T. Fedyk, Is artificial intelligence improving the audit process? Rev. Account. Stud. 27 (3) (2022) 938–985, https://doi.org/10.1007/S11142-022-09697-X/TABLES/11.

[64] C.P.H. Peters, Auditor automation usage and professional skepticism, SSRN Elect. J. (2022), https://doi.org/10.2139/SSRN.4309348.

[65] G. Puthukulam, A. Ravikumar, R.V.K. Sharma, K.M. Meesaala, Auditors' perception on the impact of artificial intelligence on professional skepticism and judgment in oman, Univer. J. Account. Finance 9 (5) (2021) 1184–1190, https://doi.org/10.13189/UJAF.2021.090527.

[66] N.A. Noordin, K. Hussainey, A.F. Hayek, The use of artificial intelligence and audit quality: an analysis from the perspectives of external auditors in the UAE, J. Risk and Financ. Manag. 15 (8) (2022) 339, https://doi.org/10.3390/JRFM15080339. *2022, Vol. 15, Page 339*.

[67] B. Knowles, J.T. Richards, The sanction of authority: promoting public trust in AI, in: FAccT 2021 - Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, Mar. 2021, pp. 262–271, https://doi.org/10.1145/3442188.3445890.

[68] C. Dwork, M. Minow, Distrust of artificial intelligence: sources & responses from computer science & Law, Daedalus 151 (2) (2022) 309–321, https://doi.org/10.1162/DAED_A_01918.

[69] E.P. Goodman, J. Trehu, AI audit washing and accountability, SSRN Elect. J. (2022), https://doi.org/10.2139/SSRN.4227350.

[70] J. Rafner, et al., Deskilling, upskilling, and reskilling: a case for hybrid intelligence, Morals & Mach. 1 (2) (2021) 24–39, https://doi.org/10.5771/2747-5174-2021-2-24.

[71] O.Ozmen Garibay, et al., Six human-centered artificial intelligence grand challenges, Int. J. Hum. Comput. Interact. 39 (3) (2023) 391–437, https://doi.org/10.1080/10447318.2022.2153320.

[72] C. Ho, "Can Artificial Intelligence Transform Auditing and Our Fear of That Transformation," Public Company Accounting Oversight Board (PCAOB), 2023 https://pcaobus.org/news-events/speeches/speech-detail/can-artificial-intelligence-transform-auditing-and-our-fear-of-that-transformation.

[73] A.R. Hasan, Artificial Intelligence (AI) in accounting & auditing: a literature review, Open J. Business Manag. 10 (01) (2022) 440–465, https://doi.org/10.4236/OJBM.2022.101026.

[74] R. Seethamraju, A. Hecimovic, Adoption of artificial intelligence in auditing: an exploratory study, Australian J. Manag. 48 (4) (2022) 780–800, https://doi.org/10.1177/03128962221108440.

[75] G. Almufadda, N.A. Almezeini, Artificial intelligence applications in the auditing profession: a literature review, J. Emerg. Technol. Account. 19 (2) (2022) 29–42, https://doi.org/10.2308/JETA-2020-083.

[76] A. Qayyum, A. Watson, A. Buchanan, M. Paterson, Y. Hakimpour, The data-driven audit: how automation and AI are changing the audit and the role of the auditor, Chartered Professional Accountants of Canada (2020).

[77] K.M. Stein, "Algorithms, Audits, and the Auditor," Public Company Accounting Oversight Board (PCAOB), 2023 https://pcaobus.org/news-events/speeches/speech-detail/algorithms-audits-and-the-auditor.

[78] B.C. Stahl, D. Wright, Ethics and privacy in AI and big data: implementing responsible research and innovation, IEEE Secur. Priv. 16 (3) (2018) 26–33, https://doi.org/10.1109/MSP.2018.2701164.

[79] J. Ayling, A. Chapman, Putting AI ethics to work: are the tools fit for purpose? AI and Ethics 2 (3) (2021) 405–429, https://doi.org/10.1007/S43681-021-00084-X. *2021 2:3*,.

[80] R.S.T. Lee, AI ethics, security and privacy, Artificial Intelligence in Daily Life (2020) 369–384, https://doi.org/10.1007/978-981-15-7695-9_14.

[81] H.S.A. Ahmed, "Challenges of AI and Data Privacy—And How to Solve them," Information Systems Audit and Control Association (ISACA), 2023 https://www.isaca.org/resources/news-and-trends/newsletters/atisaca/2021/volume-32/challenges-of-ai-and-data-privacy-and-how-to-solve-them.

[82] J. Banja, How might artificial intelligence applications impact risk management? AMA J. Ethics 22 (11) (2020) 945–951, https://doi.org/10.1001/AMAJETHICS.2020.945.

[83] K. Manheim, L. Kaplan, Artificial intelligence: risks to privacy and democracy, Yale J. Law and Technol. 21 (2019). Accessed: Dec. 14, 2023. [Online] Available, https://heinonline.org/HOL/Page?handle=hein.journals/yjolt21&id=106&div=&collection=.

[84] L. Mitrou, Data protection, artificial intelligence and cognitive services: is the general data protection regulation (GDPR) 'Artificial Intelligence-Proof'?. Tilburg: TILT Law & Technology Working Paper Series (Topic), 2018, https://doi.org/10.2139/SSRN.3386914.

[85] T. Pasquier, J. Singh, J. Powles, D. Eyers, M. Seltzer, J. Bacon, Data provenance to audit compliance with privacy policy in the Internet of Things, Pers Ubiquitous Comput 22 (2) (2018) 333–344, https://doi.org/10.1007/S00779-017-1067-4.

[86] R. Akula, I. Garibay, Audit and Assurance of AI Algorithms: A framework to Ensure Ethical Algorithmic Practices in Artificial Intelligence, Cornell University, 2021. Accessed: Dec. 14, 2023. [Online]. Available: https://arxiv.org/abs/2107.14046v1.

[87] G. Finocchiaro, "The regulation of artificial intelligence," AI Soc, vol. 1, pp. 1–8, Apr. 2023, doi: 10.1007/S00146-023-01650-Z/METRICS.

[88] H. Yu, Z. Yang, R.O. Sinnott, Decentralized big data auditing for smart city environments leveraging blockchain technology, IEEE Access 7 (2019) 6288–6296, https://doi.org/10.1109/ACCESS.2018.2888940.

[89] B. Soleymanian, R. Solgi, Application of artificial intelligence model to identify the distorted financial statements, Preprints (Business, Computer Science) (2021), https://doi.org/10.20944/PREPRINTS202109.0223.V1.

[90] B. Ghai, K. Mueller, D-BIAS: a causality-based human-in-the-loop system for tackling algorithmic bias, IEEE Trans Vis Comput. Graph. 29 (1) (2022) 473–482, https://doi.org/10.1109/TVCG.2022.3209484.

[91] E.P. Goodman, J. Trehu, Algorithmic auditing: chasing ai accountability, Santa Clara High Technol. Law J. 39 (2022). Accessed: Dec. 14, 2023. [Online]. Available: https://heinonline.org/HOL/Page?handle=hein.journals/sccj39&id=289&div=&collection=.

[92] C. Schweimer, S. Scher, Real-world-robustness of Tree-Based Classifiers, Cornell University, 2022, https://doi.org/10.48550/ARXIV.2208.10354.

[93] M. Langer, K. Baum, K. Hartmann, S. Hessel, T. Speith, J. Wahl, Explainability auditing for intelligent systems: a rationale for multi-disciplinary perspectives, in: 2021 IEEE 29th International Requirements Engineering Conference Workshops (REW) 2021-September, 2021, pp. 164–168, https://doi.org/10.1109/REW53955.2021.00030.

[94] E. Pitoura, Social-minded measures of data quality, ACM J Data Inf Qual 12 (3) (2020), https://doi.org/10.1145/3404193.

[95] E. Toreini, et al., Technologies For Trustworthy Machine Learning: A Survey in a Socio-Technical Context, Cornell University, 2020.

[96] Q. Liao, M. Research, C. Kush, R. Varshney, K.R. Varshney, Human-Centered Explainable AI (XAI): From Algorithms to User Experiences, Cornell University, 2021.

[97] R. Confalonieri, L. Coba, B. Wagner, T.R. Besold, A historical perspective of explainable Artificial Intelligence, Wiley Interdiscip Rev Data Min Knowl Discov 11 (1) (2020), https://doi.org/10.1002/WIDM.1391.

[98] A.F. Markus, J.A. Kors, P.R. Rijnbeek, The role of explainability in creating trustworthy artificial intelligence for health care: a comprehensive survey of the terminology, design choices, and evaluation strategies, J. Biomed. Inform. 113 (2020), https://doi.org/10.1016/J.JBI.2020.103655.

[99] K. Siddiqui, T.E. Doyle, Trust metrics for medical deep learning using explainable-AI ensemble for time series classification, Canadian Conference on Electrical and Computer Engineering 2022-September (2022) 370–377, https://doi.org/10.1109/CCECE49351.2022.9918458.

[100] U. Ehsan, Q.V. Liao, M. Muller, M.O. Riedl, J.D. Weisz, Expanding explainability: towards social transparency in AI systems, in: International Conference on Human Factors in Computing Systems, 2021, https://doi.org/10.1145/3411764.3445188.

[101] A. Shah, Frameworks for improving AI explainability using accountability through regulation and design, CompSciRN: Artificial Intelligence (Topic) (2020), https://doi.org/10.2139/SSRN.3617349.

[102] A. Colley, K. Väänänen, J. Häkkilä, Tangible explainable AI - an initial conceptual framework, in: International Conference on Mobile and Ubiquitous Multimedia, 2022, pp. 22–27, https://doi.org/10.1145/3568444.3568456.

[103] S. Genovesi, J.M. Mönig, Acknowledging sustainability in the framework of ethical certification for AI, Sustainability 14 (7) (Apr. 2022), https://doi.org/10.3390/SU14074157.

[104] S. Saralajew, et al., A Human-Centric Assessment Framework for AI, Cornell University, 2022, https://doi.org/10.48550/ARXIV.2205.12749.

[105] I. Naja, M. Markovic, P. Edwards, W. Pang, C. Cottrill, R. Williams, Using knowledge graphs to unlock practical collection, integration, and audit of AI accountability information, IEEE Access 10 (2022) 74383–74411, https://doi.org/10.1109/ACCESS.2022.3188967.

[106] K. Cachel, E. Rundensteiner, FINS auditing framework: group fairness for subset selections, in: AAAI/ACM Conference on AI, Ethics, and Society, 2022, pp. 144–155, https://doi.org/10.1145/3514094.3534160.

[107] M.K. Ahuja, et al., Opening the software engineering toolbox for the assessment of trustworthy AI, NeHuAI@ECAI (2020).

[108] S. Brown, J. Davidovic, A. Hasan, The algorithm audit: scoring the algorithms that score us, Big Data Soc 8 (1) (2021), https://doi.org/10.1177/2053951720983865.

[109] J. Mökander, J. Schuett, Hannah, R. Kirk, Luciano Floridi, Auditing large language models: a three-layered approach, AI and Ethics 2023 1 (2023) 1–31, https://doi.org/10.1007/S43681-023-00289-2.

[110] J. Mökander, L. Floridi, Ethics-based auditing to develop trustworthy AI, Minds Mach (Dordr) 31 (2) (2021) 323–327, https://doi.org/10.1007/S11023-021-09557-8/TABLES/1.

[111] O. Mek, HHS Trustworthy Artificial Intelligence (AI) Playbook, U.S. Department of Health & Human Services, 2021, pp. 1–109. Accessed: Dec. 14, 2023. [Online]. Available: https://www.hhs.gov/sites/default/files/hhs-trustworthy-ai-playbook.pdf.

[112] B. Ammanath, et al., Trustworthy AI in Practice, 2022 [Online]. Available: www.deloitte.com/us/AIInstitute.

[113] P. Ala-Pietilä, B. O'Sullivan, N. Boujemaa, Ethics guidelines for trustworthy AI, European Commission (2019) 1–41 [Online]. Available: https://ec.europa.eu/digital.

[114] M. Minkkinen, J. Laine, M. Mäntymäki, M.M. Fi, Continuous auditing of artificial intelligence: a conceptualization and assessment of tools and frameworks, Digital Society 1 (3) (2022) 1–27, https://doi.org/10.1007/S44206-022-00022-2. *2022 1:3*.

[115] G. Falco, et al., Governing AI safety through independent audits, Nat Mach Intell 3 (7) (2021) 566–571, https://doi.org/10.1038/S42256-021-00370-7.

[116] M.G. Alles, G.L. Gray, Will the medium become the message? A framework for understanding the coming automation of the audit process, J. Inform. Syst. 34 (2) (2020) 109–130, https://doi.org/10.2308/ISYS-52633.

[117] S. Cao, L.W. Cong, M. Han, Q. Hou, B. Yang, Blockchain architecture for auditing automation and trust building in public markets, Comput. (Long Beach Calif) 53 (7) (2020) 20–28, https://doi.org/10.1109/MC.2020.2989789.

[118] B. Couceiro, I. Pedrosa, A. Marini, State of the art of artificial intelligence in internal audit context, in: Iberian Conference on Information Systems and Technologies 2020, 2020, https://doi.org/10.23919/CISTI49556.2020.9140863.

[119] B.P. Commerford, S.A. Dennis, J.R. Joe, J.W. Ulla, Man versus machine: complex estimates and auditor reliance on artificial intelligence, Social Science Research Network (2020), https://doi.org/10.2139/SSRN.3422591.

[120] L. Rodrigues, J. Pereira, A.F. da Silva, H. Ribeiro, The impact of artificial intelligence on audit profession, J. Inform. Syst. Eng. Manag. 8 (1) (2023), https://doi.org/10.55267/IADT.07.12743.

[121] W. Li, Q. Zhou, J. Ren, S. Spector, Data mining optimization model for financial management information system based on improved genetic algorithm, Inform. Syst. E-Business Manag. 18 (4) (Dec. 2019) 747–765, https://doi.org/10.1007/S10257-018-00394-4.

[122] E. Ntoutsi, et al., Bias in data-driven artificial intelligence systems—An introductory survey, Wiley Interdiscip Rev Data Min Knowl Discov 10 (3) (2020), https://doi.org/10.1002/WIDM.1356.

[123] D. Madras, E. Creager, T. Pitassi, R. Zemel, Fairness through causal awareness: learning causal latent-variable models for biased data, in: Proceedings of the Conference on Fairness, Accountability, and Transparency, 2018, pp. 349–358, https://doi.org/10.1145/3287560.3287564.

[124] R. Bommasani, et al., On the Opportunities and Risks of Foundation Models, 2021 *arXiv.org*.

[125] H. Senetaire, D. Garreau, J. Frellsen, P.-A. Mattei, Explainability as statistical inference, in: International Conference on Machine Learning, 2022, https://doi.org/10.48550/ARXIV.2212.03131.

[126] J. Adams, Defending explicability as a principle for the ethics of artificial intelligence in medicine, Med. Health Care Philos. (2023) 1–9, https://doi.org/10.1007/S11019-023-10175-7/METRICS.

[127] Fact sheet: president biden issues executive order on safe, secure, and trustworthy artificial intelligence, The White House (2024). AccessedFeb. 22,[Online]. Available, https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/.