

BARRIERS TO ARTIFICIAL INTELLIGENCE IN ORGANISATIONAL INTERNAL CONTROL: BEHAVIOURAL AND TECHNOLOGICAL PERSPECTIVES

Seweryn CHRUSCIEL¹, Piotr SENKUS^{2*}

¹ Institute of Management Sciences and Quality, Doctoral School, University of Kalisz;
szd.1.2025@uniwersytetkaliski.edu.pl, ORCID: 0009-0006-3438-4112

² Institute of Management Sciences and Quality, University of Kalisz; p.senkus@uniwersytetkaliski.edu.pl,
ORCID: 0000-0001-9033-6437

* Correspondence author

Purpose: The aim of this paper is to identify and systematise the behavioural and technological barriers that limit the effectiveness of artificial intelligence (AI) implementations in organisational internal control systems, and to evaluate the implications of these barriers for regulatory compliance under the European Artificial Intelligence Act (EU AI Act).

Design/methodology/approach: A systematising critical literature review informed by PRISMA methodology was conducted. The databases Scopus, Web of Science, and Google Scholar were searched using keyword combinations spanning AI technology and behavioural psychology. Foundational works from 1995-2015 were included via snowballing, while the primary query targeted publications up to 2025. Following the application of inclusion and exclusion criteria, 22 publications were selected for in-depth qualitative analysis.

Findings: Two principal and mutually opposing psychological mechanisms were identified as the dominant barriers: automation bias — the tendency of auditors to uncritically accept algorithmic recommendations while suspending professional scepticism — and algorithm aversion — the irrational rejection of AI systems following the observation of even minor errors. Organisational culture and auditors' behavioural capital serve as key moderators. The technological barrier of model opacity compounds these behavioural dysfunctions by undermining conditions for informed human oversight as mandated by EU AI Act Article 14.

Research limitations/implications: The study relies on a systematising review of 22 secondary sources without primary empirical data. Future research should empirically verify de-biasing training effectiveness under controlled audit conditions.

Practical implications: Organisations deploying AI in internal control should prioritise Explainable AI (XAI) systems and develop enabling organisational cultures that support professional scepticism. Parallel investment in technology and cognitive competences of audit personnel is essential.

Social implications: Inadequate management of human–AI interaction may lead to a compliance illusion, wherein human oversight is formally satisfied but substantively hollow, exposing organisations and the public to systemic risk.

Originality/value: The study integrates dispersed findings from automation psychology, accounting information systems, and EU regulatory analysis into a unified analytical

framework. It demonstrates that the EU AI Act's human oversight mandate cannot be meaningfully fulfilled without addressing cognitive barriers of system users.

Keywords: automation bias, algorithm aversion, artificial intelligence, internal control, EU AI Act.

Category of the paper: Conceptual paper / Literature review.

1. Introduction

The implementation of artificial intelligence (AI) in risk management, auditing, and internal control processes currently represents one of the most dynamically developing areas of digital transformation in contemporary organisations. Traditional control methods, relying largely on statistical sampling and retrospective analysis of documentation, are becoming insufficient in the face of the growing volume, volatility, and velocity of financial and operational data generation. In response to these challenges, organisations are increasingly implementing advanced analytical systems based on machine learning (ML), natural language processing (NLP), and robotic process automation (RPA). These systems offer the potential to transition from sampling-based control to continuous auditing, analysing the entire population of transactions in real time and enabling proactive identification of anomalies (Boritz, No, 2015; Kokina, Davenport, 2017).

Nevertheless, a growing body of research documents what may be termed the "paradox of algorithmic control": increasing capital expenditures on sophisticated analytical technology do not translate proportionally into an actual improvement in the effectiveness of detecting irregularities and preventing fraud (Streatfield, 2001; Kokina et al., 2025). On closer examination, the weakest link in the modern architecture of control systems is not the technology itself, but the human–algorithm interface, burdened by primary behavioural and cognitive barriers.

At the root of this paradox lie two opposing, though often coexisting, psychological phenomena. On the one hand, automation bias — the propensity of decision-makers to uncritically accept algorithmic recommendations while reducing professional scepticism (Parasuraman, Manzey, 2010). On the other hand, algorithm aversion — the complete rejection of supporting technology following the observation of even a single, marginal error, despite the system's aggregate predictive accuracy significantly exceeding that of human judgement (Dietvorst, Simmons, Massey, 2015).

Managing these mechanisms becomes critical given the regulatory environment. The EU AI Act (Regulation 2024/1689) categorises systems used for financial and operational risk management as "high-risk", imposing an obligation to ensure effective human oversight. This requirement, however, appears difficult to fulfil without first diagnosing and neutralising the cognitive barriers of system users: automation bias renders human oversight illusory,

reducing it to mechanical approval of algorithmic outputs — a phenomenon that may be termed compliance illusion.

Despite extensive separate research streams on automation psychology, accounting information systems, and AI regulation, the literature lacks an integrative framework that simultaneously addresses all three dimensions in the context of internal control. This gap is particularly acute in the Central European context, where organisations face the dual challenge of rapid AI adoption and compliance with the EU AI Act without established traditions of enabling organisational culture.

The aim of this article is to identify and systematise the behavioural and technological barriers limiting the effectiveness of AI implementations in internal control systems. Specifically, the study addresses the following research questions:

RQ1: What are the principal behavioural mechanisms that impair the effectiveness of AI-supported internal control, and how do automation bias and algorithm aversion manifest in audit environments?

RQ2: How does organisational culture — specifically the distinction between enabling and coercive bureaucratic models — moderate the intensity of behavioural barriers to AI adoption in control processes?

RQ3: To what extent do existing regulatory frameworks, particularly the EU AI Act's human oversight requirement, address the identified behavioural and technological barriers, and what are the conditions for their effective implementation?

The contribution of this paper lies in integrating dispersed findings from automation psychology with the latest regulatory challenges and the specificity of internal controllers' work, thereby formulating a coherent analytical framework for managing technological change in audit units (Raisch, Krakowski, 2021).

2. Methods

The paper was prepared on the basis of a systematising critical literature review informed by the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) methodology. Given the interdisciplinary nature of the investigation — spanning behavioural psychology, accounting information systems, organisational management, and EU regulatory analysis — a multi-source search strategy was adopted to ensure comprehensive coverage of the relevant scholarly discourse.

The primary databases searched were Scopus and Web of Science Core Collection, which were selected for their extensive coverage of international peer-reviewed literature in the fields of management, accounting, and information systems. In addition, Google Scholar was

searched to capture Polish-language monographs and regulatory sources from the domain of management sciences that are not fully indexed in the aforementioned international databases.

The search strategy employed keyword combinations designed to bridge the technological and behavioural dimensions of the research problem. The following terms were used in various Boolean combinations: "internal control" OR "audit" combined with "artificial intelligence" OR "automation" combined with "behavioral" OR "trust" OR "algorithm aversion" OR "automation bias" OR "moral disengagement". The primary temporal scope of the query encompassed recent literature up to 2025, reflecting the rapidly evolving nature of AI technology and its regulatory environment. However, foundational works from the period 1995-2015 — particularly seminal contributions establishing the theoretical underpinnings of automation bias (Parasuraman, Manzey, 2010), algorithm aversion (Dietvorst, Simmons, Massey, 2015), organisational trust (McAllister, 1995), and bureaucratic typologies (Adler, Borys, 1996) — were incorporated through a snowballing procedure to ensure theoretical completeness.

The inclusion criteria encompassed: (a) peer-reviewed scientific articles published in English or Polish; (b) academic monographs with established citation reputation; (c) official reports and standards issued by internationally recognised professional bodies (specifically the IIA and ICAEW); and (d) key legislative instruments directly relevant to AI governance in the European context (EU AI Act). Exclusion criteria were applied to filter out purely computational or engineering-oriented publications focused exclusively on algorithmic optimisation, neural network architecture, or hyperparameter tuning, which lacked any substantive engagement with the managerial, behavioural, or organisational dimensions of AI implementation.

Following the systematic application of these criteria, 22 publications were selected for in-depth qualitative analysis. The analytical approach was based on deductive content coding, which enabled the grouping of identified barriers into two overarching categories: technological barriers (centred on the problem of model explainability) and behavioural barriers (encompassing dysfunctional trust patterns in human–algorithm interaction). An additional analytical layer involved triangulating findings from the scholarly literature with the normative requirements defined in Regulation (EU) 2024/1689. The complete list of 22 sources constituting the empirical foundation of this review is provided in the References section of this paper.

3. Results

3.1. Technological barriers: model opacity and the explainability imperative

The analysis of the selected sources indicates that artificial intelligence brings unprecedented transformational potential to the internal control environment. Traditional oversight systems relied on sampling methods where random data selection prevailed. In organisations processing millions of transactions daily, reliance on a sample of several dozen documents carries substantial risk of failing to detect systemic financial anomalies. AI-class systems and advanced machine learning algorithms enable the implementation of a continuous assurance model, which monitors the entire population of phenomena in the enterprise in real time, automatically categorising them according to risk level (Boritz, No, 2015).

Nevertheless, the implementation of this potential immediately encounters a critical technological barrier. The literature on accounting information systems consistently emphasises that AI adoption in audit units is impeded by the problem of model opacity, commonly known as the "black box" problem. Complex algorithms — in particular deep learning neural networks — generate recommendations based on multidimensional correlation analyses that remain entirely illegible to the human cognitive apparatus (Rudin, 2019). In the domain of financial audit and risk assessment, the inability to reconstruct the logic underlying an algorithmic decision disqualifies such tools in the eyes of professionals who bear ultimate criminal and disciplinary responsibility for control conclusions (Zhang, Cho, Vasarhelyi, 2022).

This limitation has prompted the formulation in the literature of a postulate to develop and deploy only algorithms from the Explainable AI (XAI) category. Such systems, in addition to generating the outcome itself (e.g., "transaction flagged as potential fraud"), provide the auditor with substantive justification and a transparent list of weighting factors that led to the conclusion. Without adherence to the principle of output transparency, technological tools risk becoming a source of new, poorly understood risks in the compliance environment (Kokina et al., 2025).

3.2. Automation bias: the erosion of professional scepticism

The results of the literature analysis demonstrate that behavioural factors constitute the principal predictor of AI system failure in risk management. The phenomenon of automation bias consists in the unreflective tendency of decision-makers to approve suggestions originating from automated systems while simultaneously lowering criticism towards these recommendations (Parasuraman, Manzey, 2010). In audit and compliance units, this mechanism assumes a particularly dangerous form: auditors, operating under chronic time pressure, treat positive algorithmic evaluations as definitive confirmation of the absence of irregularities, ultimately refraining from independent verification activities. Routine interaction

with the system leads to what the literature terms complacency error — the progressive atrophy of active monitoring skills.

The ICAEW (2024) has identified over-reliance on automation as one of the greatest contemporary sources of audit risk. Employees subjected to intensive algorithmic supervision lose a sense of individual agency, which directly diminishes their initiative and capacity to detect unconventional anomalies that the model has not been trained to recognise (Kellogg, Valentine, Christin, 2020). In environments characterised by excessive algorithmic signal generation — a condition known as alert fatigue — the mechanical dismissal of alerts becomes commonplace, entirely defeating the purpose of the automated system (Munoko, Brown-Liburd, Vasarhelyi, 2020).

3.3. Algorithm aversion: the asymmetry of error tolerance

At the opposite pole of the behavioural spectrum, the literature identifies algorithm aversion. Experimental research demonstrates that system users tend to disproportionately penalise machines for errors. Dietvorst, Simmons, and Massey (2015) established that individuals lose trust in AI-based systems with extreme rapidity upon observing even a single obvious error. Paradoxically, the same level of error produced by a human expert is tolerated as a "human factor", whereas absolute infallibility is subconsciously demanded of algorithms.

In audit departments, algorithm aversion manifests as an effective nullification of system utility and its silent boycott by working teams. Even when a machine learning model achieves a 95% accuracy rate — far exceeding traditional human capabilities in large-scale data analysis — a single false positive suffices to discourage professionals from further engagement with the system (Benbya, Pachidi, Jarvenpaa, 2021). This dynamic is compounded by progressive deskilling: as auditors less frequently perform independent data evaluation, they become increasingly inclined to reject algorithmic conclusions in ambiguous situations.

The phenomenon stands in fundamental contradiction to the postulates of Human-Centered AI design (Shneiderman, 2022). A properly designed control system should stimulate active trust calibration: auditors must be trained to understand the boundaries of algorithmic utility and to rationally assess when human judgement should override machine recommendations, rather than wholesale rejecting the platform following isolated errors.

3.4. Organisational culture as a moderator of behavioural barriers

The intensity of both mechanisms described above is critically modulated by the organisational context. In the classic bureaucratic typology formulated by Adler and Borys (1996), organisations grounded in coercion (coercive bureaucracy) enforce unquestioning submission to technological standards, thereby amplifying automation bias. Employees in such systems use algorithms as a protective shield that relieves them of personal responsibility for control outcomes. This contrasts with enabling models (enabling bureaucracy), in which employee knowledge and human capital are prioritised, and the algorithm functions as

a supporting tool that does not supplant the obligation to exercise independent professional judgement (Grzesiak, 2021).

This distinction presupposes the presence of cognition-based trust — a relationship in which the auditor possesses sufficient competence and confidence to challenge decisions produced by the machine (McAllister, 1995). The internal control standards issued by the IIA (2024) explicitly state that professional scepticism remains an inalienable foundation of the auditor's work, and that the responsibility for formulating an opinion can never be transferred to a software system, regardless of its sophistication. A direct implication follows: building organisational resilience in the era of AI depends on trust and the continuous upgrading of competences, not merely on delegating duties to IT systems (Kozmiński, 2004; Bugdol, 2010).

4. Discussion

4.1. Synthesis: the automation–augmentation paradox in internal control

The results presented above converge to illuminate what Raisch and Krakowski (2021) have conceptualised as the automation–augmentation paradox: technology simultaneously possesses the capacity to enhance and to undermine human decision-making, depending critically on the implementation context. In the domain of internal control, this paradox manifests with particular acuity. The very systems designed to augment the analytical power of auditors may, through the mechanisms of automation bias, erode the professional scepticism they were intended to support. Conversely, algorithm aversion may cause organisations to forfeit the genuine analytical advantages that AI systems demonstrably offer.

The evidence gathered in this review indicates that resolving this paradox is not merely a matter of managerial preference or optional organisational improvement but constitutes a pressing regulatory necessity. The EU AI Act (Regulation 2024/1689) categorises software used in broadly understood control infrastructure as high-risk systems. Article 14 of this regulation introduces the concept of mandatory human oversight, requiring entities to guarantee employees' real ability to identify system dysfunctions, override recommendations, or halt the system entirely in emergency situations. An analysis of the literature juxtaposed with this legal requirement reveals a significant pragmatic conflict: human oversight is legally mandated, yet from the perspective of behavioural science, it becomes entirely fictitious in organisations with immature management cultures. An employee who has succumbed to automation bias will approve any system-generated report, formally satisfying the "human oversight" postulate while substantively constituting a circumvention of legislative intent (Szpringer, 2024). This phenomenon — which may be termed compliance illusion — represents a critical gap between regulatory aspiration and organisational reality.

4.2. Organisational context: differential barriers across institutional types

An important dimension insufficiently addressed in the existing literature concerns the variation in AI adoption barriers across different organisational contexts. Three distinct environments merit consideration.

Large financial corporations typically possess substantial resources for AI implementation, including dedicated data science teams and established compliance infrastructure. However, their rigid hierarchical structures and extensive regulatory obligations may foster coercive bureaucratic tendencies (Adler, Borys, 1996). In such settings, automation bias poses the principal risk: auditors operating within highly proceduralised environments may perceive algorithmic recommendations as an extension of institutional authority and thus apply minimal critical scrutiny. The scale and complexity of operations — often spanning multiple jurisdictions — compounds the challenge, as human oversight must be maintained across thousands of automated decisions daily.

Public institutions, including government audit offices and regulatory agencies, face a distinct constellation of barriers. These organisations frequently operate under severe resource constraints, characterised by legacy IT infrastructure and limited access to specialist AI competences. The dominant bureaucratic culture in many public institutions tends strongly toward coercion (Kellogg, Valentine, Christin, 2020), creating an environment in which algorithmic outputs may be uncritically adopted as an efficiency measure rather than a genuine analytical enhancement. Furthermore, the public accountability framework introduces additional complexity: errors attributable to AI systems may generate disproportionate reputational damage, potentially amplifying algorithm aversion among risk-averse public servants.

Small and medium-sized enterprises (SMEs) confront the most acute version of the competence gap identified by Szpringer (2024). Many SME audit departments lack both the tools and the specialised personnel necessary to evaluate AI system compliance with regulatory requirements. For these organisations, the first practical step toward meaningful AI integration is likely to consist of basic XAI literacy training — enabling controllers to pose the right questions about algorithmic recommendations before accepting or rejecting them (Grzesiak, 2021). The informal organisational cultures characteristic of many SMEs may, paradoxically, offer advantages: the absence of deeply entrenched bureaucratic rigidity may facilitate the development of enabling cultures supportive of critical thinking and trust calibration (Bugdol, 2010).

4.3. Explainable AI in audit practice: from postulate to implementation

The consistent call in the literature for Explainable AI (XAI) as a prerequisite for responsible AI deployment in control systems (Rudin, 2019; Zhang, Cho, Vasarhelyi, 2022) raises a natural question regarding concrete implementation mechanisms. Two technical approaches merit particular attention in the context of audit applications.

First, **SHAP (SHapley Additive exPlanations) values (Lundberg, Lee, 2017)** represent a model-agnostic method that enables auditors to understand the contribution of each individual input feature to a specific algorithmic decision. In the context of fraud detection, for instance, a machine learning model that flags a particular transaction as suspicious can, via SHAP analysis, present the auditor with a transparent decomposition of the factors driving the classification — such as transaction amount, counterparty novelty, temporal deviation from established patterns, or geographic anomaly. This granular attribution directly addresses the "black box" problem identified by Rudin (2019), transforming an opaque algorithmic verdict into a structured basis for professional judgement. Crucially, SHAP values enable auditors to exercise the informed scepticism mandated by professional standards (IIA, 2024), as they can evaluate whether the system's reasoning aligns with their domain knowledge of risk indicators.

Second, **locally interpretable model-agnostic explanations (LIME)** offer a complementary approach that generates simplified, human-readable approximations of complex model behaviour for individual predictions. In continuous auditing environments — where AI systems process entire transaction populations in real time (Boritz, No, 2015) — LIME can provide auditors with case-by-case explanations of why specific transactions were escalated for manual review. This mechanism directly mitigates both of the identified behavioural barriers: it reduces automation bias by compelling the auditor to engage with the reasoning behind the recommendation rather than accepting it passively, and it reduces algorithm aversion by making the system's decision logic transparent, thereby allowing users to calibrate their trust based on substantive understanding rather than superficial impressions of system performance (Zhang, Cho, Vasarhelyi, 2022; Shneiderman, 2022).

The deployment of such XAI mechanisms addresses the core requirement of Article 14 of the EU AI Act: effective human oversight presupposes that users "fully understand the capabilities and limitations of the system" (Regulation EU 2024/1689). Without interpretable outputs, this understanding — and consequently genuine human oversight — remains unattainable.

4.4. Regulatory implications: bridging the gap between legal mandate and organisational reality

The regulatory framework established by the EU AI Act provides a normative foundation that, if implemented with due regard for behavioural science, can serve as a catalyst for organisational transformation. The internal control standards issued by the IIA (2024) reinforce

this position by affirming that professional scepticism remains non-negotiable even in highly automated environments. A direct implication for management follows: security management in the AI era is fundamentally an exercise in building organisational resilience grounded in trust and the continuous development of human competences, rather than in the mere delegation of cognitive duties to software systems (Koźmiński, 2004; Bugdol, 2010).

However, the analysis reveals that regulatory compliance alone is insufficient without corresponding investment in the behavioural dimension. The effective transition from automation to augmentation (Raisch, Krakowski, 2021) requires the simultaneous development of three pillars: (a) a technological pillar, mandating the deployment of XAI-compliant models that provide auditors with actionable explanations; (b) a regulatory pillar, enforcing active and genuine human oversight consistent with EU directives; and (c) a behavioural pillar, developing enabling organisational cultures in which professional doubt is treated not as disloyalty but as a manifestation of mature professionalism. Asymmetric investment — heavy expenditure on AI licenses without corresponding investment in the cognitive capital of the human team — leads to the façade nature of internal control processes and exposes the organisation to significant legal and reputational risk.

5. Conclusions

5.1. Knowledge contribution

This literature review makes three principal contributions to the scholarly understanding of AI implementation in internal control. First, it integrates previously dispersed findings from automation psychology, accounting information systems, and EU regulatory analysis into a unified analytical framework, demonstrating that behavioural, technological, and regulatory barriers are not independent phenomena but mutually reinforcing dimensions of a single systemic challenge. Second, it identifies the compliance illusion — the condition in which the EU AI Act's human oversight requirement is formally satisfied but substantively hollow due to unaddressed cognitive biases — as a critical and under-recognised risk in the digitalisation of control processes. Third, it demonstrates that the enabling-coercive bureaucratic typology (Adler, Borys, 1996) provides a theoretically grounded and practically actionable lens for predicting the differential susceptibility of organisations to automation bias and algorithm aversion.

5.2. Theoretical and practical implications

From a theoretical standpoint, the study advances the automation–augmentation paradox framework (Raisch, Krakowski, 2021) by demonstrating its specific manifestation in the internal control domain, where the stakes of cognitive failure are amplified by legal liability and public accountability. The integration of behavioural mechanisms with regulatory requirements contributes to the emerging literature on human-centred AI governance.

From a practical perspective, the findings yield three recommendations for management and audit leadership. First, organisations deploying AI in control processes should mandate the use of Explainable AI systems — specifically techniques such as SHAP or LIME — that enable auditors to exercise informed professional scepticism rather than passive acceptance or uninformed rejection of algorithmic outputs. Second, management should invest in structured de-biasing training programmes designed to calibrate auditors' trust in AI systems, incorporating simulated error scenarios that develop the capacity to distinguish between meaningful system failures and statistically expected imprecision. Third, organisational leaders should actively cultivate enabling bureaucratic cultures that reward critical questioning and professional doubt, as the review demonstrates that coercive cultures systematically amplify both automation bias and algorithm aversion.

5.3. Limitations

Several limitations of this study must be acknowledged. First, the analysis is based on a systematising review of 22 secondary sources and does not incorporate primary empirical data; consequently, the conclusions, while theoretically robust, require empirical verification in specific organisational settings. Second, the rapid pace of technological change in artificial intelligence means that some of the reviewed sources may not fully capture the most recent developments in model interpretability, real-time auditing capabilities, or regulatory implementation practices. Third, the study relies predominantly on conceptual analysis and theoretical extrapolation from existing literature; the proposed framework has not been tested through large-scale empirical investigation, and the organisational context analysis (Section 4.2) draws on logical inference from established theories rather than direct comparative evidence.

5.4. Future research directions

The direction for further research postulated on the basis of this review should prioritise the experimental verification of the effectiveness of specific de-biasing training interventions aimed at calibrating employees' trust in decision support systems under rigorously controlled conditions of auditors' work. Particularly valuable would be comparative studies examining the differential effectiveness of XAI interventions across organisational types — large financial corporations, public institutions, and SMEs — to validate the contextual framework proposed

in Section 4.2. Additionally, longitudinal studies tracking the evolution of automation bias and algorithm aversion over extended periods of human–AI collaboration would contribute significantly to understanding the temporal dynamics of trust calibration in audit environments.

References

1. Adler, P.S., Borys, B. (1996). Two Types of Bureaucracy: Enabling and Coercive. *Administrative Science Quarterly*, Vol. 41, Iss. 1, pp. 61-89, doi: <https://doi.org/10.2307/2393986>
2. Benbya, H., Pachidi, S., Jarvenpaa, S.L. (2021). Artificial Intelligence in Organizations: Implications for Information Systems Research. *Journal of the Association for Information Systems*, Vol. 22, Iss. 2, pp. 281-303, doi: <https://doi.org/10.17705/1jais.00662>
3. Boritz, J. E., No, W.G. (2015). Continuous Assurance: The External Auditors' Perspective. *Current Issues in Auditing*, Vol. 9, Iss. 1, pp. P1-P14, doi: <https://doi.org/10.2308/ciaa-51022>
4. Bugdol, M. (2010). *Wymiary i problemy zarządzania organizacją opartą na zaufaniu*. Kraków: Wydawnictwo UJ, pp. 34-56.
5. Dietvorst, B.J., Simmons, J.P., Massey, C. (2015). Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err. *Journal of Experimental Psychology: General*, Vol. 144, Iss. 1, pp. 114-126, doi: <https://doi.org/10.1037/xge0000033>
6. Grzesiak, L. (2021). *Kapitał ludzki jako czynnik skuteczności audytu wewnętrznego*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego, pp. 45-70.
7. ICAEW (2024). *Over-reliance on Automation* (Report). London: Institute of Chartered Accountants in England and Wales.
8. IIA (2024). *Global Internal Audit Standards* (Report). Lake Mary: The Institute of Internal Auditors.
9. Kellogg, K.C., Valentine, M.A., Christin, A. (2020). Algorithms at Work: The New Contested Terrain of Control. *Academy of Management Annals*, Vol. 14, Iss. 1, pp. 366-410, doi: <https://doi.org/10.5465/annals.2018.0174>
10. Kokina, J., Blanchette, S., Davenport, T.H., Pachamanova, D. (2025). Challenges and Opportunities for Artificial Intelligence in Auditing: Evidence from the Field. *International Journal of Accounting Information Systems*, Vol. 56, No. 100734, doi: <https://doi.org/10.1016/j.accinf.2024.100734>
11. Kokina, J., Davenport, T.H. (2017). The Emergence of Artificial Intelligence: How Automation Is Changing Auditing. *Journal of Emerging Technologies in Accounting*, Vol. 14, Iss. 1, pp. 115-122, doi: <https://doi.org/10.2308/jeta-51730>

12. Koźmiński, A.K. (2004). *Zarządzanie w warunkach niepewności: podręcznik dla zaawansowanych*. Warszawa: PWN, pp. 78-95.
13. Lundberg, S.M., Lee, S.I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*, 30, 4765-4774.
14. McAllister, D.J. (1995). Affect- and Cognition-Based Trust as Foundations for Interpersonal Cooperation in Organizations. *Academy of Management Journal*, Vol. 38, Iss. 1, pp. 24-59, doi: <https://doi.org/10.2307/256727>
15. Munoko, I., Brown-Liburd, H.L., Vasarhelyi, M.A. (2020). The Ethical Implications of Using Artificial Intelligence in Auditing. *Journal of Business Ethics*, Vol. 167, Iss. 2, pp. 209-234, doi: <https://doi.org/10.1007/s10551-019-04411-4>
16. Parasuraman, R., Manzey, D.H. (2010). Complacency and Bias in Human Use of Automation: An Attentional Integration. *Human Factors*, Vol. 52, Iss. 3, pp. 381-410, doi: <https://doi.org/10.1177/0018720810376055>
17. Raisch, S., Krakowski, S. (2021). Artificial Intelligence and Management: The Automation–Augmentation Paradox. *Academy of Management Review*, Vol. 46, Iss. 1, pp. 192-210, doi: <https://doi.org/10.5465/amr.2018.0072>
18. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (EU AI Act), OJ EU L, 2024/1689 (2024).
19. Ribeiro, M.T., Singh, S., Guestrin, C. (2016). “Why Should I Trust You?”: Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135-1144. doi: <https://doi.org/10.1145/2939672.2939778>
20. Rudin, C. (2019). Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nature Machine Intelligence*, Vol. 1, Iss. 5, pp. 206-215, doi: <https://doi.org/10.1038/s42256-019-0048-x>
21. Shneiderman, B. (2022). *Human-Centered AI*. Oxford: Oxford University Press, pp. 45-78.
22. Streatfield, P.J. (2001). *The Paradox of Control in Organizations*. London: Routledge, pp. 23-45.
23. Szpringer, W. (2024). *Sztuczna inteligencja w zarządzaniu. Regulacja a konkurencja*. Warszawa: Poltext, pp. 78-110.
24. Zhang, C.A., Cho, S., Vasarhelyi, M.A. (2022). Explainable Artificial Intelligence (XAI) in Auditing. *International Journal of Accounting Information Systems*, Vol. 46, No. 100572, doi: <https://doi.org/10.1016/j.accinf.2022.100572>