

Article

From the EU AI Act to Audit Practice: A Governance-to-Controls Framework for Quality Management and Evidence

János Kálmán 

Faculty of Law and Political Sciences, Széchenyi István University, 9026 Győr, Hungary; kalman.janos@sze.hu

Abstract

Artificial intelligence (AI) tools—including audit data analytics, robotic process automation, machine-learning models, and generative AI—are changing how audit teams identify risks, select procedures, and evaluate evidence. At the same time, Regulation (EU) 2024/1689 (the EU AI Act) establishes a risk-based governance architecture built around risk management, data governance, technical documentation, logging, transparency, human oversight, robustness, cybersecurity, and post-market monitoring. The Act is not an auditing standard and does not directly regulate every tool used by audit firms. Nevertheless, its governance logic is relevant where audit firms develop, procure, or rely on AI-enabled systems that process sensitive client data, influence professional judgement, or become part of audit-relevant client systems. This conceptual study uses doctrinal requirements-to-controls mapping and design-oriented analysis to translate selected AI Act governance objectives into firm-level and engagement-level quality-management controls and into criteria for evaluating AI-enabled audit evidence. The paper specifies three modes of AI Act relevance: direct legal relevance where a regulated AI Act role is engaged; indirect relevance where AI compliance documentation becomes audit-relevant information; and benchmark relevance where the Act supplies governance objectives for quality management without creating an audit-law duty. The resulting artefacts are a traceable AI Act/IAASB standards crosswalk, an evidence-risk typology, a quality-management integration model, a documentation and review checklist, and a proportional maturity model. The framework clarifies when AI outputs remain triage or risk-assessment tools, when they provide directional or corroborative evidence, and the narrower conditions under which they may contribute to substantive evidence. It links reliance to data completeness, reconciliation, versioning, validation, false-positive and false-negative behaviour, explainability, logging, source-document corroboration, and reviewer challenge. The contribution is a scalable governance-to-controls framework that supports defensible reliance and inspection readiness without overstating the AI Act's direct legal applicability. Empirical validation in audit firms remains a priority for future research. It further explains how quantitative risk features and anomaly-detection outputs feed into qualitative audit judgement: models can route attention to unusual transactions or documents, but evidential weight still depends on base-rate-aware error analysis, source-document corroboration, and reviewer challenge.



Academic Editor: Lawrence Chui

Received: 23 April 2026

Revised: 10 July 2026

Accepted: 11 July 2026

Published: 15 July 2026

Copyright: © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: AI-assisted auditing; audit evidence; audit quality management; EU AI Act; AI governance; professional scepticism; explainability; auditability

1. Introduction

Audit practice is undergoing a significant technological transformation. Computer-assisted audit techniques have been used for decades, but recent advances in audit data

analytics, process mining, robotic process automation (RPA), machine learning, and generative AI permit population-level analysis, automate parts of evidence production, and support pattern recognition across structured and unstructured data [1–9]. These capabilities may broaden audit coverage and improve efficiency, but they also create new dependencies on data quality, model behaviour, system configuration, vendor architecture, and the way auditors interpret automated outputs.

The motivation for this inquiry is not merely compliance with a new statute. Recent studies show that generative and predictive AI can alter information processing, professional workflows, and firm-level innovation outcomes, while also creating adoption heterogeneity and risks of over-reliance [10–12]. In audit, this means that productivity-oriented AI adoption must be reconciled with documentation, professional scepticism, and evidential reliability. The legal question and the audit-quality question therefore converge: governance controls become valuable when they make AI-assisted work reviewable, reproducible, and proportionate to the intended evidential role.

In parallel, the European Union has enacted Regulation (EU) 2024/1689 (the EU AI Act), a horizontal and risk-based framework for AI systems [13–22]. Its governance architecture includes risk management, data governance, technical documentation, record-keeping, transparency, human oversight, accuracy, robustness, cybersecurity, and post-market monitoring. The AI Act is not an auditing standard, and its legal obligations do not automatically attach to every internal audit tool. Its control logic is nevertheless relevant to audit firms as deployers, developers, purchasers, and evaluators of AI-enabled systems.

The central audit problem is therefore not simply whether an AI tool is accurate in a technical sense. Audit quality also depends on whether the engagement team can establish the provenance and completeness of the data, understand the tool's intended use and limitations, reproduce or otherwise review the procedure, challenge the output with professional scepticism, and document the basis for reliance. Adoption research shows that organisational support, competence, responsibility allocation, and integration into audit methodology materially shape how analytical tools are used [23–30].

Recent research on digital transformation and legal innovation further demonstrates that technological benefits depend on the interaction between organisational capability and regulatory adaptation [31], while work on AI-enabled service innovation and legal compliance highlights the need to integrate technological deployment with governance structures [32]. These adjacent streams are valuable, but they do not resolve the audit-specific question of how contemporary AI-governance requirements should be translated into quality-management controls, workpaper evidence, reviewer responsibilities, and defensible reliance decisions.

This study addresses three connected gaps. First, the AI-assisted auditing literature identifies opportunities, adoption barriers, and emerging assurance needs, but offers limited traceable translation into firm-level and engagement-level quality-management controls. Second, the audit-evidence literature explains sufficiency and appropriateness but lacks an operational, AI-specific framework connecting data provenance, model validation, explainability, logging, human challenge, false-positive/false-negative behaviour, and source-document corroboration to the degree of evidential reliance. Third, EU AI-governance scholarship analyses legal duties and regulatory architecture, but rarely connects those duties to the practical design of audit procedures and workpaper documentation. The article therefore asks how selected AI Act governance objectives can be converted into auditable controls without presenting the Act as a substitute for audit standards.

The added value over existing audit-technology and AI-auditability scholarship is thus deliberately narrow but operational. The paper does not claim that documentation, validation, or professional scepticism are new. It clarifies what the AI Act contributes

as governance vocabulary, what existing audit standards continue to require, and what additional audit-specific design decisions are needed to move from a legal governance requirement to a reviewable audit-control response.

The paper is organised around three research questions, each addressing a distinct stage of the governance-to-evidence chain:

RQ1: How can the EU AI Act's governance requirements be mapped onto audit-firm quality-management systems and engagement-level quality management? This question concerns the institutional translation from legal governance objectives to responsibilities, policies, resources, monitoring, and remediation within risk-based systems of quality management [33–36].

RQ2: Which AI-governance requirements are most critical to the reliability and documentation of AI-enabled audit evidence, and how can they be operationalised in audit procedures? This question focuses on the evidential consequences of data integrity, model validation, traceability, explainability, and reproducibility, as well as the distinction between directional, corroborative, and potentially substantive uses of AI outputs [8,9,36–39].

RQ3: Which control mechanisms and record-keeping practices can mitigate AI-specific risks—especially data risk, model risk, human-oversight risk, and security risk—while preserving audit quality and professional scepticism? This question links technical and organisational controls to behavioural safeguards, including documented challenge, supervision, escalation, and alternative procedures [36–42]. Section 2 reviews the relevant scholarship; Section 3 sets out the EU legal and professional context; Section 4 explains the analytical design; Section 5 presents the framework; Section 6 evaluates its mechanisms, limits, and implications; and Section 7 concludes.

2. Literature Review

2.1. Technological Transformation of Audit Practice

Audit-technology scholarship has progressed from continuous auditing and continuous assurance to audit data analytics, RPA, machine learning, and generative AI. Early work on continuous auditing emphasised analytic monitoring and near-real-time exception reporting [1,2]. Later research describes audit data analytics as a family of techniques capable of transforming risk assessment and substantive testing through full-population analysis, richer anomaly detection, and the combination of structured and unstructured data [5–9].

RPA operates as a complementary layer by automating repetitive activities, such as extracting data from client systems, populating workpapers, and executing standardised queries. The principal benefit is the reallocation of professional effort towards higher-order judgement; the corresponding risk is that automation may reproduce flawed processes at scale when ownership, change control, and review are weak [3,4]. Generative AI further expands the technology stack by supporting natural-language interrogation, summarisation, drafting, and workflow coordination. Because its outputs can be non-deterministic, its appropriate role is generally assistance and triage rather than autonomous professional judgement [43].

Recent accounting information systems research has shifted attention from technical capability to implementation, auditability, and auditor readiness. The 2025 International Journal of Accounting Information Systems special issue illustrates this change by examining AI auditability, auditor adoption, field-level implementation challenges, machine-learning decision support, and adaptive audit processes [43–51]. This study builds on that agenda by examining how EU AI-governance requirements can be translated into controls that are usable within audit methodology.

2.2. *Evidential Reliability in Data-Driven Engagements*

The core auditing question is not whether AI can generate insights, but whether those insights can support audit conclusions as sufficient and appropriate evidence. Audit analytics produce risk scores, anomaly flags, predictions, classifications, and generated summaries. Their evidential value depends on the provenance and integrity of input data, the validity and stability of the model, the relevance of the output to the assertion tested, and the auditor's ability to understand and review the procedure [8,9].

AI outputs may play at least four roles in an engagement: planning evidence, directional evidence, corroborative evidence, or substantive evidence. The appropriate classification depends on reliance rather than on the technology label. An anomaly-detection model used to prioritise transactions ordinarily supports direction and selection; a model used as a substitute for an external confirmation would require materially stronger validation, traceability, and corroboration.

Research on analytical-tool adoption shows that perceived value does not automatically translate into audit quality. Effective use depends on organisational support, user competence, methodological integration, and the ability to explain the relationship between the output and the audit objective [23–25]. These findings are consistent with broader technology-adoption research on perceived usefulness, ease of use, self-efficacy, and organisational conditions [27–30].

A recent quantitative audit-analytics stream gives this evidence question a stronger empirical footing. Bertomeu et al. use machine-learning features from accounting, market, governance and audit data to detect material misstatements, showing how quantitative predictors can identify patterns that are difficult to specify *ex ante* [52]. Cheynel et al. show that detected fraud exhibits heavy-tailed, power-law behaviour, which matters for audit design because a small number of extreme cases can dominate expected loss and detection priorities [53]. Computer-science work on AutoAudit and on minimum-description-length anomaly detection in bookkeeping data similarly treats accounting records as temporal, graph-structured or pattern-based systems in which unusual nodes, periods, account combinations or description-length deviations can route attention to audit-relevant risk points [54,55]. These studies do not make AI outputs self-authenticating audit evidence; rather, they show where quantitative features can enter the audit process as risk indicators, triage signals and inputs to targeted substantive procedures.

2.3. *Quality Management, Professional Scepticism, and Organisational Adoption*

Audit quality is a multidimensional construct shaped by engagement inputs, process integrity, reporting outcomes, and the institutional environment. A recurring insight is that audit quality must be managed as a system rather than inferred from a single proxy [33–35]. Contemporary quality-management standards consequently require firms to establish quality objectives, identify and assess quality risks, design responses, and monitor and remediate deficiencies [36].

AI introduces a related family of quality risks: data risk, model risk, vendor risk, cybersecurity risk, confidentiality risk, and human-oversight risk. These risks cut across governance and leadership, resources, engagement performance, information and communication, and monitoring. Where responsibility for validation, version control, and incident response is unclear, the same tool may be used inconsistently across engagements, weakening both scalability and reviewability [23–25].

Professional scepticism is a critical behavioural component of this system. Automated recommendations can direct attention to anomalies, but they can also create automation bias and reduce the motivation to challenge a plausible-looking output. Models and measures of professional scepticism show that evidence characteristics, incentives, individual traits,

and the judgement environment interact to shape sceptical action [40,41]. AI governance in auditing must therefore combine technical validation with explicit human accountability, supervision, and documented challenge.

2.4. Explainability, Accountability, and System Auditability

Explainability is a recurring governance and evidence problem. Research distinguishes between post-hoc explanations for opaque models and inherently interpretable models, and cautions against relying on superficial explanations in high-stakes decisions [37–39]. In auditing, explanation has a practical function: engagement teams must justify their conclusions to reviewers, oversight bodies, and—where relevant—courts or other external evaluators. A procedure that cannot be traced, reproduced, or independently assessed may have limited evidential value even when its predictive performance appears strong.

AI ethics and governance scholarship documents a proliferation of principles, including fairness, accountability, transparency, and human control, while also showing how difficult it is to convert principles into operational requirements [42]. The EU AI Act is significant because it moves from broad principles towards enforceable lifecycle obligations for systems within its scope [13–22]. The present study uses a functional translation: selected requirements are treated as control objectives that can inform audit-quality management and evidence evaluation, while their direct legal applicability remains a separate question.

2.5. Emerging Assurance Services and the Unresolved Governance Gap

A wider market for assurance over AI is emerging alongside the use of AI within financial-statement audits. Organisations may seek independent assurance that AI systems are appropriately governed, robust, non-discriminatory, and compliant. Audit firms may also need to evaluate clients' AI controls during risk assessment or internal-control work, and firms that develop or substantially modify tools may assume responsibilities beyond those of an ordinary user [45,46,48].

This development creates an important two-way relationship. Governance documentation produced for an AI system—such as technical documentation, validation reports, logs, and monitoring records—may become audit-relevant information. At the same time, the audit firm's own use of AI requires governance artefacts that support supervision, re-performance, confidentiality, and inspection. The capabilities, responsibilities, and independence implications differ between "AI in audit" and "audit of AI", but both depend on system auditability.

The literature therefore leaves an unresolved translation problem. AI-assisted auditing research describes technologies and adoption conditions; audit-quality research describes systemic drivers of quality; audit-evidence research defines sufficiency and appropriateness; and AI-governance scholarship specifies lifecycle obligations. What is missing is a structured mechanism that connects those domains at the level of control objectives, responsible actors, evidence artefacts, and reliance decisions.

The contribution of this paper is to supply that mechanism. It does not claim that AI Act compliance proves audit quality, nor that every audit tool is regulated as high-risk. Instead, it develops a governance-to-controls framework in which legal and professional requirements are translated into testable organisational controls and documentation expectations. This distinction between direct legal obligation and functional governance benchmark is central to the analysis.

3. EU Regulatory and Standard-Setting Context

3.1. The EU AI Act's Governance Architecture and Relevance for Auditing

Regulation (EU) 2024/1689 establishes a horizontal framework intended to ensure that AI systems placed on the EU market, put into service, or used in the Union are governed in proportion to their risk [22]. For high-risk AI systems, Articles 9–15 address risk management, data and data governance, technical documentation, record-keeping, transparency and information for deployers, human oversight, and accuracy, robustness, and cybersecurity. Article 17 further requires providers of high-risk systems to establish a documented quality-management system. The Act combines product-safety architecture with fundamental-rights protection [13,15,18,20].

For audit firms, the legal relevance depends on their role and use case. A firm may act as a deployer when it uses a third-party AI system under its authority; it may become a provider where it develops and puts a system into service under its own name, or substantially modifies an existing system in circumstances covered by the Act. Classification and obligations are context-specific. Accordingly, this paper does not assume that an internal analytical tool is automatically high-risk, but it recognises that procurement, development, modification, and deployment can produce different compliance positions.

The analysis distinguishes three forms of AI Act relevance. Direct legal relevance arises where the audit firm, vendor, or client occupies an AI Act role and the relevant system falls within the Act's scope. Indirect relevance arises where AI Act documentation, conformity assessment material, incident reports, or monitoring records become audit-relevant information for a client's financial reporting risks. Benchmark relevance arises where the Act does not impose duties on the audit procedure itself, but its governance architecture supplies a structured model for documenting reliable, human-supervised, and reviewable AI use.

This distinction also separates AI-in-audit from audit-of-AI. AI-in-audit concerns the audit firm's own use of analytics, machine learning, RPA, process mining, or generative AI to perform audit work. Audit-of-AI concerns the auditor's evaluation of a client's AI systems, controls, compliance records, incidents, provisions, contingent liabilities, disclosures, and potential financial reporting exposure. The two settings involve different evidence sources, legal roles, independence considerations, competence requirements, and documentation expectations.

The Act is also relevant as a governance benchmark. Risk management, data lineage, technical documentation, logging, human oversight, and lifecycle monitoring are directly connected to questions an audit reviewer must ask: what was the tool intended to do, what data did it process, which version was used, what limitations were known, who challenged the output, and how were incidents or drift addressed? Standardisation and common specifications may further shape the technical artefacts available for review [18,22].

3.2. From Regulatory Risk to Audit-Quality Risk

Risk provides the conceptual bridge between AI governance and audit-quality management. The AI Act's risk categories are legal and institutional constructs, not merely statistical measurements [15,16]. Audit-quality management likewise requires professional judgement when identifying quality risks, assessing their likelihood and effect, and designing proportionate responses [36]. The objects of regulation differ, but both regimes rely on structured risk identification, documented controls, assigned responsibility, and monitoring.

This study therefore treats eight governance domains as control objectives: risk management; data governance; technical documentation and traceability; logging and record-keeping; transparency and communication; human oversight and accountability; accuracy,

robustness, and cybersecurity; and monitoring and remediation. These objectives map onto firm-level governance and leadership, ethical requirements, resources and competence, engagement performance, information and communication, and monitoring and remediation.

The bridge is functional rather than equivalential. An AI Act artefact does not automatically constitute sufficient appropriate audit evidence, and a well-designed audit control does not by itself establish legal compliance. The analytical value lies in identifying where the two systems ask structurally similar questions and where additional audit-specific judgement—especially on relevance, reliability, corroboration, and professional scepticism—is required.

3.3. Interaction with EU Audit Law and Professional Standards

Auditing in the EU is governed by a layered framework that includes Directive 2006/43/EC, Regulation (EU) No 537/2014 for statutory audits of public-interest entities, national implementation and oversight, and international standards adopted or applied in the relevant jurisdiction [36,56,57]. These sources impose requirements relating to independence, professional scepticism, engagement performance, documentation, quality management, and oversight. The EU AI Act does not amend those audit obligations, but it can affect the systems, information, and compliance risks that auditors encounter.

At firm level, ISQM 1 requires a risk-based system of quality management; ISA 220 (Revised) addresses engagement-level quality management; and ISA 500 frames the sufficiency and appropriateness of audit evidence [36]. AI-enabled procedures raise concrete questions under each of these standards. Firms require governance over tool approval, competence, vendor relationships, security, and monitoring, while engagement teams must decide whether a particular output is relevant and reliable for the audit purpose and document the basis for that decision.

The “audit of AI” dimension is equally important. A client’s use of prohibited or high-risk AI may create compliance costs, provisions, contingent liabilities, disclosure issues, operational disruption, and reputational consequences. Technical documentation, conformity-assessment materials, incident records, and post-market monitoring reports may therefore become audit-relevant information. Their existence does not establish operating effectiveness; auditors must still assess completeness, authenticity, and the possibility of paper compliance.

3.4. Standardisation, Conformity Assessment, and the Assurance Interface

The EU AI Act relies heavily on standardisation and conformity-assessment infrastructures. Harmonised standards may provide a presumption of conformity for specified requirements, while technical documentation and post-market records can translate governance duties into inspectable artefacts [18,22]. This architecture creates an assurance interface between AI governance and auditing, although the purpose and level of assurance remain distinct.

Upstream, audit firms that develop, customise, or procure AI tools need documentation, logging, change control, and monitoring sufficient to govern their own reliance. Downstream, auditors of AI-intensive clients may encounter conformity-assessment results, provider documentation, validation materials, and incident logs as potential evidence. In both directions, the audit task remains the same in principle: evaluate relevance and reliability, identify contradictory information, and determine whether corroboration or alternative procedures are required.

The interaction also creates independence and competence questions. Firms expanding into AI-governance assurance must separate advisory and assurance roles where required, ensure access to appropriate technical expertise, and avoid treating conformity documenta-

tion as a substitute for professional judgement. The framework developed below addresses these issues through responsibility allocation, review, and proportionality rather than through technology-specific mandates.

The framework should also be distinguished from implementation frameworks such as the NIST AI Risk Management Framework and ISO/IEC 42001. NIST AI RMF 1.0 offers a voluntary lifecycle risk-management structure organised around Govern, Map, Measure and Manage functions, while ISO/IEC 42001 specifies requirements for establishing, implementing, maintaining and continually improving an AI management system [58,59]. Those instruments are likely to guide how firms operationalise AI controls in practice. The EU AI Act adds a different layer: binding EU classifications, role-specific duties, conformity and logging artefacts for in-scope systems, and a legal vocabulary that may become audit-relevant information. Accordingly, this paper does not replace NIST- or ISO-based implementation systems; it translates the Act's governance vocabulary into audit-specific workpaper evidence, reliance decisions and reviewer tests.

4. Materials and Methods

4.1. Research Design and Analytical Framework

This study adopts a conceptual, doctrinal, and design-oriented research strategy. It does not estimate the causal effect of AI on audit quality. Instead, it examines how governance requirements can be translated into organisational controls and evidence practices that are intelligible within auditing. The unit of analysis is the governance requirement or control objective, and the principal analytical operation is a structured requirements-to-controls mapping.

The framework connects the three research questions to four source domains: (i) Regulation (EU) 2024/1689 and the legal scholarship interpreting its governance architecture [13–22]; (ii) EU statutory-audit law [56,57]; (iii) IAASB pronouncements on quality management, engagement quality, and audit evidence [36]; and (iv) peer-reviewed research on audit technologies, adoption, professional scepticism, explainability, and AI governance [1–9,23–30,33–35,37–51,60]. Adjacent work on AI-enabled organisational innovation and legal compliance was used to contextualise implementation conditions [31,32]. Figure 1 shows how the questions, sources, mapping steps, outputs, and evaluation criteria are connected.

4.2. Requirements-to-Controls Mapping, Scope, and Design Criteria

The mapping procedure comprised five sequential steps:

1. Requirement identification. The analysis extracted governance requirements most relevant to audit-system reliability and reviewability: risk management, data governance, technical documentation, logging, transparency, human oversight, robustness and cybersecurity, and lifecycle monitoring. The primary legal text was used to anchor the categories, while scholarship supported interpretation and boundary identification [13–22].
2. Decomposition into control objectives. Each legal or governance requirement was reformulated as a technology-neutral control objective—for example, verifying data provenance and quality, maintaining versioned documentation, preserving run logs, allocating human responsibility, and defining incident and remediation procedures.
3. Mapping to quality management. The control objectives were mapped to firm-level and engagement-level responsibilities, including governance and leadership, resources and competence, tool approval, engagement performance, communication, monitoring, and remediation. This step addresses RQ1 and is grounded in risk-based quality-management logic [33–36].

- 4. Mapping to the evidence lifecycle. The same objectives were mapped to data acquisition, model execution, output interpretation, documentation, review, retention, and the final reliance decision. This step addresses RQ2 and RQ3 by identifying where data, model, human-oversight, security, and explainability risks can impair evidential reliability [8,9,36–41].
- 5. Consolidation and analytical checking. The mappings were consolidated into five artefacts: the control matrix (Table 1), evidence-risk typology (Table 2), quality-management integration model (Table 3), documentation and review checklist (Table 4), and maturity model (Table 5). Cross-references among the artefacts were then reviewed to ensure that each recommended control had an identifiable governance source, responsible actor, evidence implication, and implementation record.

Research design and analytical workflow

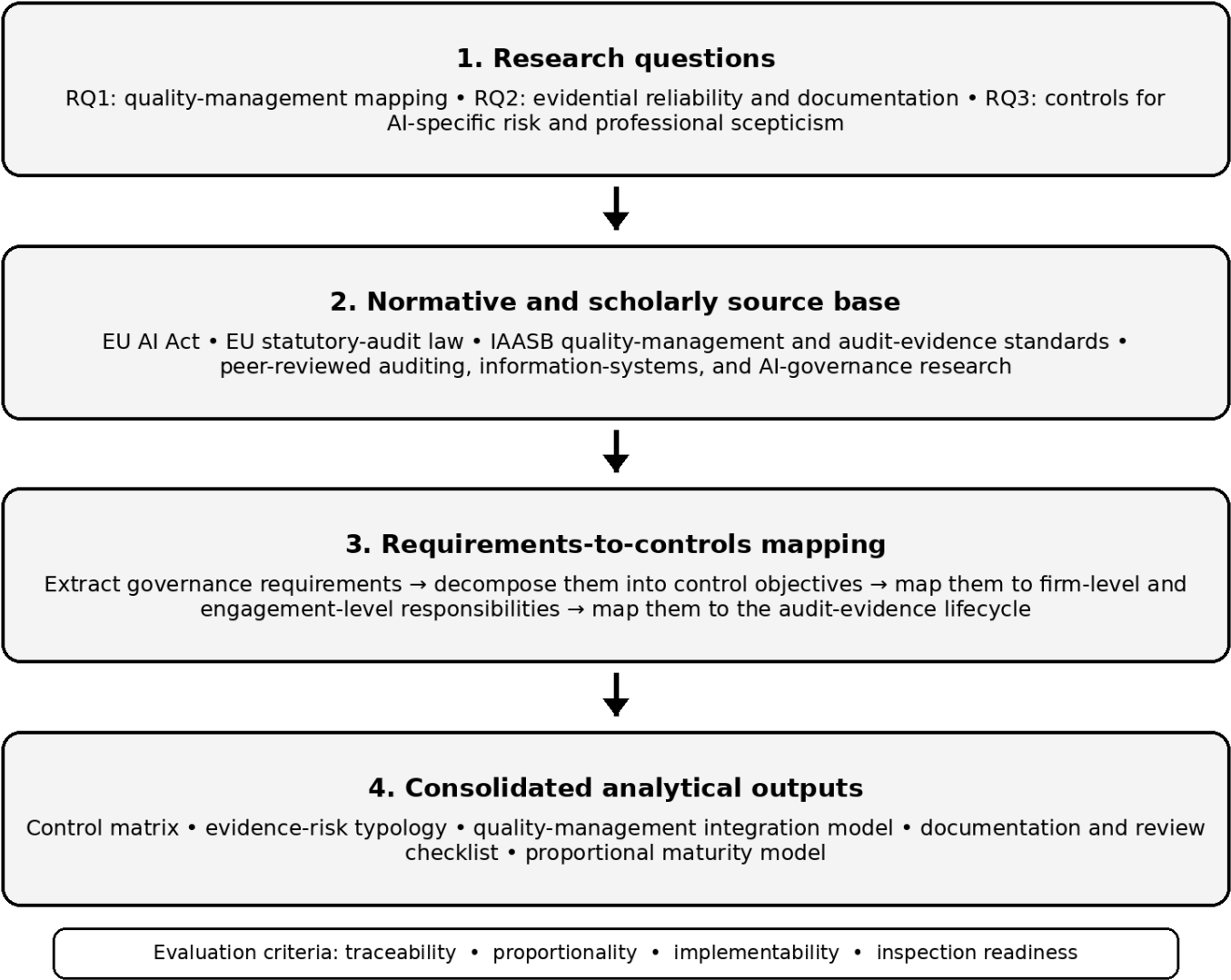


Figure 1. Research design and analytical workflow.

Table 1. AI Act-to-audit standards crosswalk for quality management and audit evidence.

AI Act/Audit-Standard Anchor	AI-Related Risk	Control Owner and Trigger	Expected Workpaper Evidence	Reviewer Test and Residual Risk
Risk management and QMS: AI Act Arts. 9, 17; ISQM 1 paras. 23–26, 34; ISA 220 (Revised) paras. 25–30, 32, 35–37; ISA 315 (Revised 2019) paras. 13, 27–31	AI use not captured as a quality risk; reliance not linked to assertions or risks of material misstatement.	Quality leader / AI governance lead and engagement partner. Trigger: QMS update, new tool, material model change, or engagement reliance.	AI use register; engagement AI risk memo; reliance class; escalation and consultation records.	Reviewer traces each AI use to objective, risk, owner and response. Residual risk: uncontrolled AI reliance.
Data governance: AI Act Art. 10; ISA 315 (Revised 2019) paras. 13, 25–26; ISA 500 paras. 7, 9; ISA 540 (Revised) paras. 13, 18, 23–26	Input population incomplete, biased, unreconciled, unrepresentative, or transformed without control.	Data owner, engagement team and methodology specialist. Trigger: extract, transformation, run, or population change.	Data lineage; extraction script; reconciliation; key / duplicate checks; feature-engineering note.	Reviewer tests completeness, accuracy and lineage. Residual risk: unreliable data impairs evidence.
Technical documentation and versioning: AI Act Art. 11; ISA 230 paras. 8–10; ISA 500 paras. 7, 9, 11	Purpose, limits, parameters, version or configuration cannot be re-performed or inspected.	Tool owner and engagement manager. Trigger: qualification, version update, parameter change, or high-reliance use.	Tool dossier / model card; validation summary; version ID; configuration snapshot; reproducibility package.	Reviewer asks whether another competent auditor could understand and re-perform. Residual risk: black-box conclusion.
Logging and record-keeping: AI Act Art. 12; ISA 230 paras. 7–10, 14–16; ISA 500 paras. 7, 9, 11	Inputs, prompts, runs, outputs, exceptions or user interactions missing or not retained.	Tool owner, IT/security and engagement team. Trigger: each AI-assisted run or prompt sequence used in the file.	Run logs; prompt logs where relevant; output repository; timestamped exception lists; retention confirmation.	Reviewer matches logs to conclusion and retained evidence. Residual risk: unsupported reliance and weak inspection trail.
Transparency and human oversight: AI Act Arts. 13–14; ISA 220 (Revised) paras. 25–30, 32, 35–37; ISA 230 paras. 8–10; ISA 500 paras. 7, 9, 11	Team over-relies on opaque outputs, fails to challenge exceptions, or cannot explain relevance.	Engagement partner, senior reviewer and specialist. Trigger: medium/high-reliance procedures and final review.	Procedure narrative; limitation statement; challenge log; alternative explanations; reviewer sign-off; source corroboration.	Reviewer evaluates whether challenge is substantive. Residual risk: automation bias and weak scepticism.
Accuracy, robustness and cybersecurity: AI Act Art. 15; ISQM 1 paras. 32(f)–(g), 37–42; ISA 220 (Revised) paras. 25–30, 32; ISA 500 paras. 7, 9, 11	Performance unstable, outside intended domain, vulnerable, or affected by vendor/platform incident.	Model owner, security lead and quality monitoring function. Trigger: approval, periodic test, vendor update, incident, drift signal.	Holdout/back-testing; sensitivity tests; threshold rationale; security assessment; vendor assurance; incident log.	Reviewer checks validation evidence and incident response. Residual risk: compromised or unstable outputs.
Monitoring, incidents and remediation: AI Act lifecycle logic, including Arts. 72–73 for regulated systems; ISQM 1 paras. 35–47; ISA 220 (Revised) paras. 35–37	Drift, recurring errors, incidents or remediation do not feed into QMS and engagement planning.	Quality monitoring function and AI governance lead. Trigger: monitoring, incident, failed reviewer test, or complaint.	Drift dashboard; root-cause analysis; remediation plan; re-validation memo; team communication.	Reviewer verifies remediation and need for additional procedures. Residual risk: repeated deficiencies and inconsistent quality.

Table 2. AI-enabled audit evidence risk typology and reliance conditions.

Evidence Risk Source	Typical Manifestations	Red Flags/Indicators	Recommended Audit Responses
Data risk	Incomplete or biased datasets; inconsistent joins; privacy limits affecting completeness; external data of uncertain provenance.	Unreconciled extract totals; missing primary keys; duplicate records without explanation; undocumented filters; sudden volume changes; unverified third-party data.	Reconcile to source systems and ledger; document lineage; test completeness and accuracy; retain extraction scripts; perform sensitivity analysis; obtain alternative evidence if data quality is weak.
Model-validation risk	Overfitting; inappropriate model choice; model used outside intended population; class imbalance; thresholds chosen without audit rationale.	No holdout/back-testing; no false-positive/false-negative analysis; no threshold rationale; model version mismatch; feature engineering undocumented; no independent review.	Validate against intended audit population; document performance metrics and threshold selection; assess precision/recall and base rate; prefer interpretable models for high reliance; restrict weak models to directional use.
Base-rate/error-cost risk	Material misstatements are rare; high accuracy hides low recall; false positives and false negatives have unequal audit costs.	Accuracy reported without prevalence, confusion matrix, threshold rationale or follow-up capacity analysis.	Report precision, recall/sensitivity, specificity and negative predictive value; set thresholds by materiality, expected follow-up effort and reliance role; sample or otherwise challenge unflagged items when false negatives are costly.
Automation-bias/oversight risk	Auditors accept scores, summaries or classifications without challenge; reviewer involvement is late or superficial.	No challenge log; identical conclusions across engagements; no alternative explanations; exceptions cleared by reference to AI output alone.	Mandate human oversight plan; require challenge and corroboration steps; involve qualified reviewer for medium/high reliance; document contrary evidence and professional scepticism.
Process-mining risk	Event logs, case identifiers or timestamps do not faithfully represent the process; excluded pathways distort conclusions.	Missing timestamps; duplicate events; inconsistent case IDs; system-interface gaps; unreconciled event counts; undocumented pathway exclusions.	Test event-log completeness; reconcile process data to financial records; review excluded populations; document process-model assumptions; corroborate unusual pathway findings.
Generative-AI/source-trace risk	Generated summaries or extracted clauses are inaccurate, incomplete, hallucinated, or unsupported by source documents.	Missing prompt log; no retrieval configuration; unsupported generated claim; no source cross-reference; confidential data entered into unmanaged tool.	Treat outputs as triage unless claims are traced to source records; retain prompts and source links; verify extracted claims manually; prohibit unsupported substitution of generated text for audit evidence.
Security and integrity risk	Data, model, prompt logs or outputs are altered, exposed, or accessed without authorisation; vendor updates change behaviour.	Missing access log; unexplained output changes; vendor update not reflected in workpaper; absent security assessment; incident not escalated.	Enforce access control and secure logging; perform vendor due diligence; document incidents and re-performance; treat compromised outputs as unreliable until remediated.
Explainability and reviewability risk	Outputs cannot be explained, reproduced, or linked to assertions; reviewers cannot inspect the basis of reliance.	No reproducibility package; proprietary tool with no explanation; missing model card; no configuration snapshot; conclusions not tied to assertion tested.	Require tool dossier and configuration snapshot; document limitations; restrict opaque tools for high-reliance use; add corroborative procedures and reviewer challenge.

Table 3. Practical integration of AI governance into audit quality management components.

Quality Management Component	AI Governance Integration	Key Responsibilities	Indicative Documentation
Governance and leadership	Establish AI governance policy; set risk appetite for AI reliance; ensure accountability.	Managing partner/quality leader; AI governance lead.	AI governance policy; AI use register; escalation pathways.
Ethics and independence	Assess independence threats from tool vendors, consulting relationships, and model development.	Ethics partner; engagement partner; compliance.	Vendor independence assessment; conflict checks; safeguards documentation.
Acceptance and continuance	Evaluate client AI environment; determine whether AI-related risks affect engagement acceptance or resource needs.	Engagement partner; risk committee.	Client AI risk profile; engagement acceptance memo; competence plan.
Resources and competence	Define competency requirements (data, AI, cybersecurity); training and staffing; use of specialists.	HR/learning; engagement leaders.	Training records; specialist involvement logs; competence assessment.
Engagement performance	Standardize AI-assisted procedures; oversight plan; templates for documenting AI outputs and limitations.	Engagement team; reviewer.	AI procedure workpapers; reliance rationale; reviewer sign-off.
Information and communication	Communicate tool updates and limitations; ensure consistent messaging to teams and reviewers.	AI tool owner; methodology team.	Release notes; user guidance; FAQs; limitation statements.
Monitoring and remediation	Track tool performance, drift, incidents and engagement outcomes; remediate deficiencies.	Quality monitoring function; AI tool owner.	Monitoring reports; incident log; remediation plan; follow-up testing.

Table 4. Checklist for documenting and reviewing AI-enabled audit evidence.

Checklist Domain	Core Questions for the Engagement Team	Minimum Workpaper Evidence/Artefacts
Use-case definition and relevance	What audit objective does the AI tool support? Which assertions and risks does it address? Is the use case directional or substantive?	Procedure description; linkage to risk assessment; mapping to assertions; reliance classification.
Tool identity and versioning	What tool/model/version was used? Were parameters changed? Is the configuration captured for reproducibility?	Tool dossier reference; version ID; configuration snapshot; change log.
Data provenance and integrity	What data were used? How were they extracted, cleaned and reconciled? Are completeness and accuracy checks documented?	Data lineage record; extraction scripts; reconciliations; sampling of source data; access permissions.
Model validation and limitations	What validation evidence supports the tool's use for this purpose? What are known limitations and assumptions? Is interpretability adequate?	Validation summary; performance metrics; limitation statement; sensitivity/plausibility checks.
Output interpretation and scepticism	How were outputs interpreted? What alternative explanations were considered? Were exceptions followed up with corroborating evidence?	Exception follow-up workpapers; challenge log; corroboration procedures; supervisory notes.
Human oversight and review	Who reviewed the AI-assisted procedure and why are they competent? Was an independent reviewer involved for high-reliance use cases?	Reviewer sign-off; competence record; consultation documentation; second-review evidence.
Logging and audit trail	Are run logs, inputs and outputs retained? Is there a secure audit trail that supports inspection and re-performance?	Run logs; output repository; timestamped artefacts; retention confirmation.
Security and confidentiality	Were cybersecurity and confidentiality controls applied? Are vendor/cloud risks addressed? Is access controlled?	Security assessment summary; access rights report; vendor assurance; incident check.
Conclusion and impact on audit opinion	How did the AI-assisted procedure affect the audit conclusion? Were additional procedures required due to uncertainty or drift?	Conclusion paragraph; linkage to overall audit conclusion; cross-references to additional procedures.

Table 5. Proportional maturity model for AI governance implementation in audit firms.

Maturity Level	Typical AI Use in Audits	Minimum Governance Controls	Quality Management and Evidence Implications
Level 1—Basic	Rule-based analytics; simple RPA; spreadsheet-driven exception testing.	Inventory of tools; basic documentation of data sources; manual review of outputs; access controls for client data.	AI outputs treated as directional evidence; limited reliance; emphasize reconciliations and clear workpaper narratives.
Level 2—Managed	Standardized audit data analytics; process mining; supervised ML models for risk scoring; limited generative AI for drafting/triage.	Tool dossiers and versioning; documented validation; engagement AI risk memo; reviewer sign-off for high-reliance procedures; secure logging.	Some outputs may contribute substantive evidence when validation is strong; stronger re-performance and inspection readiness.
Level 3—Advanced	Integrated AI platforms; continuous monitoring; extensive ML and generative AI; proprietary models; AI-enabled client controls testing.	Lifecycle monitoring (drift, incidents); formal change management; explainability standards; independent model risk function; periodic third-party assurance; quality dashboards.	Higher reliance possible but requires robust evidence governance; inspection readiness becomes a core design constraint; strong accountability structures.

The mapping is functional rather than equivalence-based. An AI Act requirement is not treated as an audit standard, and an audit workpaper is not treated as AI Act compliance documentation. Instead, the procedure identifies where AI governance concepts can support existing audit requirements concerning quality risk, engagement performance, documentation, sufficiency and appropriateness of evidence, risk assessment, estimates, supervision, consultation, review, and retention.

To make the mapping reproducible, each mapped item was processed through seven coding fields: (1) legal or governance anchor; (2) AI Act role and audit use case; (3) audit-standard anchor; (4) risk to quality management or evidential reliability; (5) control owner and trigger; (6) expected workpaper evidence and reviewer test; and (7) residual risk if the control fails. Items were included where they affected data integrity, model behaviour, human oversight, documentation, security, monitoring, or inspection readiness. Items were excluded where they related solely to market surveillance, administrative enforcement, or institutional obligations without a plausible audit-quality or evidence consequence.

Worked example under the seven-field protocol: AI Act Article 12 on logging. (1) Legal or governance anchor: Article 12's log-generation and record-keeping objective for high-risk systems, used as a direct obligation only where the system is in scope and otherwise as a governance benchmark. (2) AI Act role and audit use case: an audit firm deploys an anomaly-detection model for journal-entry risk ranking. (3) Audit-standard anchor: ISA 230 paragraphs 8–10, ISA 500 paragraphs 7, 9 and 11, and ISA 220 (Revised) paragraphs 29–32. (4) Risk: missing run logs, prompts, inputs, outputs or exceptions prevent re-performance and reviewer challenge. (5) Control owner and trigger: the tool owner and engagement manager operate the control at each model run or prompt sequence used in the engagement file. (6) Expected workpaper evidence and reviewer test: a timestamped run log, input-file hash or version, model/configuration ID, retained output list, exception disposition and reviewer cross-match from log to conclusion. (7) Residual risk if the control fails: unsupported reliance, inability to investigate drift or contradictory evidence, and a weak inspection trail.

Scope and limitations. The paper does not provide legal advice on the classification of a specific audit tool and does not treat AI Act requirements as universally applicable audit duties. It distinguishes direct legal applicability from functional use as a governance benchmark. It also does not claim empirical validation: the framework is an analytical design artefact intended for subsequent testing through field studies, experiments, inspections, and implementation research.

Because audit engagements vary in complexity, AI use, and degree of reliance, the framework is deliberately scalable. It covers low-reliance automation, such as rule-based reconciliations, as well as higher-reliance uses, such as machine-learning risk scores and generative-AI document review. The intensity of validation, documentation, review, and monitoring should increase with reliance, opacity, changeability, and potential consequence.

Design quality criteria. The resulting framework was evaluated against four practical criteria:

(i) Traceability: Each recommendation should be traceable to an identifiable governance objective and to an audit-quality or audit-evidence concern.

(ii) Proportionality: Control intensity should respond to reliance and risk; low-reliance uses should not be burdened with the same requirements as high-reliance, opaque, or rapidly changing systems.

(iii) Implementability: Recommendations should be expressible as realistic policies, tool dossiers, risk memoranda, workpaper templates, review steps, and monitoring procedures.

(iv) Inspection readiness: The framework should support supervision, re-performance, regulatory inspection, and post-incident analysis through logging, versioning, retained evidence, and documented reliance decisions.

5. Results

5.1. Governance-to-Controls Translation (RQ1)

Table 1 addresses RQ1 by replacing a general governance matrix with a traceable standards crosswalk. It links selected AI Act governance anchors to IAASB quality-management and audit-evidence requirements, identifies the audit-quality or evidence risk created by AI use, allocates the control to an owner, states the trigger for operation, and identifies the expected workpaper evidence, reviewer test, and residual risk. The table is not intended to collapse legal and auditing obligations into a single regime. Its purpose is to show how AI Act governance logic can be converted into auditable control design under existing professional requirements.

5.2. Evidence-Risk Typology and Reliance Conditions (RQ2)

Table 2 addresses RQ2 by specifying how different AI-enabled outputs carry different evidential weight. The categories of risk assessment, directional, corroborative, and potentially substantive use should not be treated as labels applied after the event. They are reliance decisions that depend on the audit objective, assertion tested, population completeness, data reconciliation, transformation controls, model validation, threshold choice, false-positive and false-negative profile, reproducibility, interpretability, reviewer challenge, and source-document corroboration.

The reliance boundary is especially important for the distinction between directional and substantive evidence. An AI output remains a triage or navigation tool when it merely helps the auditor find documents, transactions, or clauses for human inspection. It supports risk assessment when it reveals patterns or concentrations that shape further procedures. It becomes directional evidence when it prioritises items for testing, but the audit conclusion still depends on follow-up evidence. It can contribute to substantive evidence only in narrower circumstances: the input population must be complete and reconciled, transformations and feature engineering must be documented, validation must address the intended population and assertion, threshold selection must be justified, false-positive and false-negative implications must be considered, the run must be reproducible, and conclusions must be corroborated or challenged by retained source evidence.

Base-rate and error-cost considerations matter because audit procedures are often designed to detect rare but material exceptions. In low-prevalence settings, headline

accuracy can be misleading: a model may appear accurate because most transactions are non-problematic while still missing the small set of misstatements that matter most. The validation file should therefore report, where feasible, the confusion matrix and reliance-relevant measures such as precision, recall or sensitivity, specificity, negative predictive value and the expected audit cost of false positives and false negatives [52,53].

False positives and false negatives should be evaluated differently according to use. A high false-positive rate may be acceptable for screening or directional use if follow-up testing is efficient and transparent; it is much less acceptable where the model is used to support substantive conclusions. False negatives are especially consequential when unflagged items receive little or no further attention, so threshold selection should be justified by materiality, tolerable misstatement, expected follow-up capacity and the residual procedures applied to unflagged populations. Graph-based, temporal and MDL-style anomaly-detection approaches can help auditors prioritise unusual accounts, periods or transaction paths, but the audit conclusion still requires qualitative evaluation of business rationale, authorisation, source documents, management explanations and contradictory evidence [54,55].

Figure 2 summarises this staged reliance logic by showing how AI-enabled outputs move from triage and risk-assessment support toward directional, corroborative and, only under stricter validation and corroboration conditions, potentially substantive evidential use.

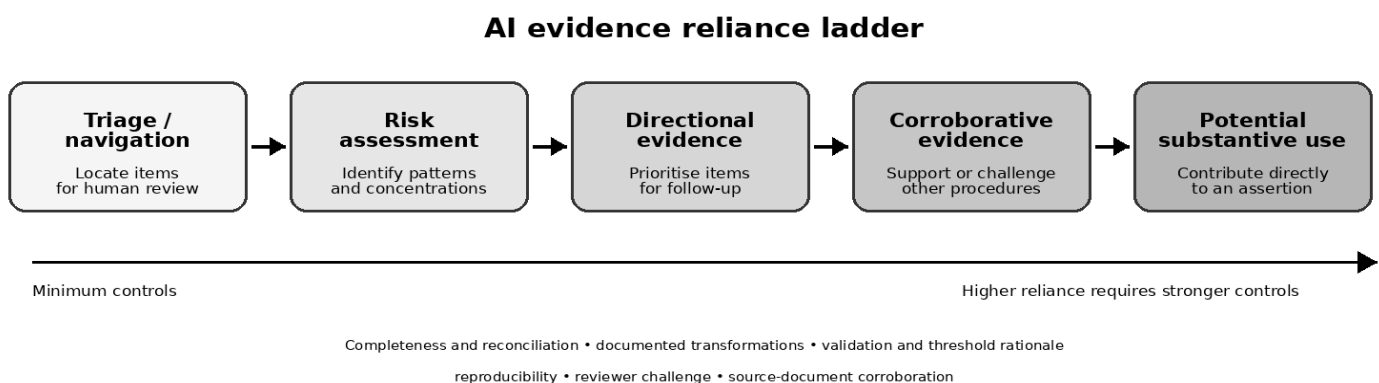


Figure 2. Evidence reliance ladder for AI-enabled audit outputs.

5.3. Quality-Management Implementation Mechanisms (RQ1 and RQ3)

The control matrix can be embedded in an audit firm's system of quality management through three linked implementation layers: governance and responsibility allocation; tool qualification and change management; and monitoring and remediation. Together, these layers operationalise RQ1 and RQ3 by connecting AI-specific risks to named owners, engagement responsibilities, review requirements, and retained documentation.

First, governance and responsibility allocation. Firms should define roles for AI tool ownership (including accountability for validation and documentation), engagement-level responsibility (engagement partner/manager), and independent quality review responsibility. For high-reliance use cases, separating development/configuration from review functions can support independence and scepticism. Training should not only cover tool operation but also interpretability limits and common cognitive biases [27–29,40,41].

Second, tool qualification and change management. Before an AI tool is used in engagements, the firm should maintain a 'tool dossier' covering the tool's purpose, limitations, validation results, data requirements, cybersecurity controls, and versioning. Changes to model versions, parameters or data pipelines should trigger re-qualification and communication to engagement teams. This aligns with the AI Act's emphasis on

technical documentation and lifecycle governance and addresses the practical need for reproducibility and inspection readiness.

Validation expectations should also vary by system type. Supervised models used for risk scoring or transaction classification require holdout testing, back-testing against known outcomes, threshold justification, class-imbalance assessment, feature-importance review, version control, and independent review when reliance is high. Anomaly-detection tools require population definition, outlier-logic documentation, treatment of rare but valid transactions, review of excluded populations, and follow-up protocols for flagged and unflagged items. Process-mining tools require event-log completeness, timestamp reliability, case-identifier consistency, duplicate-event handling, and reconciliation to financial records. Generative-AI tools require prompt retention, retrieval configuration, source tracing, hallucination testing, confidentiality restrictions, access controls, and human verification of every extracted claim used in the audit file.

Third, monitoring and remediation. Audit firms should monitor AI tool performance (e.g., drift indicators and error rates), usage patterns across engagements, and incident reports (e.g., data integrity issues and security events). Monitoring results should feed into root-cause analysis and remediation. From a quality perspective, monitoring also supports consistent application across engagements, which is essential for scalable audit methodologies [33–35].

5.4. Documentation and Review Controls (RQ2 and RQ3)

Table 4 converts traceability, transparency, oversight, and reviewability into workpaper questions and minimum artefacts. It therefore supports both the reliability analysis in RQ2 and the control and record-keeping analysis in RQ3. The checklist can be applied during engagement execution and during engagement-quality review [8,36–41].

The checklist is intended to be proportionate rather than identical for every tool. Low-reliance administrative or navigational use normally requires a procedure description, data-source identification, confidentiality control, and manual verification of any claim used in the file. Medium-reliance directional or corroborative use additionally requires data lineage, tool/version identification, validation summary, run logs, and documented follow-up of exceptions. High-reliance or potentially substantive use requires the full tool dossier, reconciliation of the input population, version and configuration snapshots, threshold rationale, independent reviewer challenge, alternative procedures where uncertainty remains, and explicit linkage to the audit conclusion.

5.5. Illustrative Application Scenarios

Scenario A: AI-assisted journal entry testing as a worked mapping example. Suppose an engagement team uses an anomaly-detection model to prioritise journal entries for follow-up testing. The AI Act governance anchor is risk management, data governance, documentation, logging, transparency, oversight, and robustness; the audit-standard anchor is quality risk assessment, engagement performance, audit documentation, and sufficiency/appropriateness of evidence. The audit team first defines the assertion and population, reconciles the general-ledger extract to the client system, documents filters and feature engineering, and captures the model version and configuration. The model owner documents validation, including threshold selection, back-testing where possible, and the expected false-positive and false-negative profile. The engagement team then treats the output as directional evidence unless validation, population completeness, reproducibility, and corroboration are strong enough to support a narrower substantive contribution. The reviewer tests whether flagged items were followed up, whether unflagged populations remain plausible sources of material misstatement, whether alternative explanations were

considered, and whether the final conclusion rests on retained source evidence rather than on the score alone.

More specifically, quantitative methods feed into this scenario at three points: feature construction, scoring and follow-up allocation. Features may include unusual posting times, rare account combinations, round amounts, manual overrides, user-pattern concentration, graph centrality or deviations from ordinary bookkeeping paths; scoring and thresholding then route attention to items that warrant human investigation. The qualitative side of the audit remains complementary: the auditor evaluates business rationale, authorisation, source documentation, management explanations, and whether the unflagged population still contains plausible material misstatement risk. This is why the framework treats quantitative analytics as a disciplined routing and corroboration mechanism unless the stronger validation, completeness and source-document conditions for substantive contribution are satisfied [52–55].

Scenario B: Generative AI for contract and narrative-disclosure review. Generative AI tools can help locate clauses, summarise board minutes, compare covenant language, or identify inconsistent disclosure narratives. Because outputs can be non-deterministic and unsupported, the framework treats them as triage unless every material claim is traced to a source document retained in the engagement file. The workpaper should retain the prompt or query, tool/version, retrieval configuration, source-document links, human verification steps, and confidentiality safeguards. A generated summary of a contract, loan covenant, board minute, or disclosure note should not substitute for inspection of the underlying record unless the relevant extracted claim is verified, cross-referenced, and evaluated under the applicable assertion.

Scenario C: Auditing clients' AI compliance signals. Where a client operates AI systems that are material to operations, pricing, lending, provisioning, customer treatment, or disclosures, AI Act documentation may have indirect audit relevance. Conformity assessment materials, technical documentation, incident reports, monitoring records, and remediation plans may inform the auditor's assessment of provisions, contingent liabilities, control deficiencies, regulatory exposure, or narrative disclosures. However, such documentation is not automatically reliable audit evidence. The auditor must assess its completeness, source, independence, consistency with operational records, and whether the client's documented compliance corresponds to actual control operation.

Across these scenarios, AI tools may support coverage and efficiency, but only where governance artefacts—documentation, logging, validation, oversight, monitoring, and source-document corroboration—are treated as part of the evidence chain rather than as decorative compliance material. The framework therefore supports more disciplined reliance decisions; it does not demonstrate that AI use itself improves audit outcomes.

5.6. A Proportional Maturity Model for Implementation

Audit firms vary widely in technological capability and degree of reliance. Some use scoped analytics and RPA; others deploy proprietary machine-learning platforms and generative-AI assistants. Table 5 responds to this variation through a proportional maturity model in which control expectations increase with reliance, complexity, opacity, changeability, and potential consequence.

At the basic level, the emphasis is on transparency and minimum traceability for low-reliance automation. At the managed level, tool qualification, data-lineage documentation, and reviewer oversight are formalised. At the advanced level, lifecycle monitoring, drift detection, explainability standards, independent model-risk capability, and integrated quality dashboards are added. Maturity does not justify automatic reliance; it increases the firm's capacity to evaluate and defend reliance.

6. Discussion

6.1. The Governance-to-Evidence Mechanism

The framework's principal analytical claim is that AI governance affects audit quality through a three-stage mechanism. First, a legal or governance requirement is translated into a control objective. Second, that objective is assigned to a responsible actor and embedded in firm-level or engagement-level quality management. Third, the resulting control produces or preserves information—such as a data-lineage record, model-version snapshot, run log, challenge record, or incident report—that can support the auditor's evaluation of evidential reliability. The mechanism therefore connects RQ1, RQ2, and RQ3 rather than treating the matrices and checklists as separate outputs.

The AI Act adds value to this mechanism in three limited ways. First, it supplies a lifecycle vocabulary—risk management, data governance, documentation, logging, transparency, human oversight, robustness, and monitoring—that can be converted into audit-control objectives. Second, it forces role clarity by distinguishing direct legal relevance from indirect audit relevance and benchmark relevance. Third, it exposes what the Act does not answer for auditors: how much evidential weight an AI output should carry, which workpapers prove sufficiency and appropriateness, how reviewers should test reliance, and when source-document corroboration remains indispensable. These are audit-specific additions supplied by the framework rather than by the Act itself.

This mechanism does not mean that governance documentation automatically becomes sufficient, appropriate audit evidence. Documentation is evidentially useful only when it permits the engagement team to test the origin, operation, limitations, and use of the system. A complete model card cannot cure incomplete source data; a detailed log cannot make an invalid model appropriate for the audit purpose; and a reviewer sign-off cannot replace meaningful challenge. Governance artefacts are necessary conditions for defensible reliance in many AI-enabled procedures, but they are not substitutes for substantive validation and professional judgement.

The role assigned to an AI output is consequently decisive. The same anomaly score may serve as planning or directional evidence when used to target follow-up work, but it may approach substantive evidence only when the underlying data are reconciled, the model is validated for the relevant population and assertion, limitations are understood, outputs are reproducible, and contradictory information is investigated. The framework thus shifts attention from the label "AI" to the degree and consequence of reliance.

The same point clarifies the complementarity between quantitative analytics and qualitative audit procedures. Quantitative methods expand the searchable population and make rare, complex or non-linear patterns more visible, but they do not determine materiality, explain business purpose, prove authorisation, or resolve contradictory evidence. Qualitative procedures supply context and corroboration, while quantitative outputs discipline where sceptical attention should be directed and how much residual testing remains necessary. The workpaper should therefore document not only the model result, but also how the result changed the audit plan, which flagged and unflagged items were tested, and why the remaining risk was acceptable.

6.2. Boundary Conditions and Implementation Trade-Offs

The first boundary condition is legal applicability. The AI Act's governance vocabulary can inform audit controls even where a particular tool is not legally classified as high-risk. However, the distinction between binding obligation and voluntary benchmark must remain explicit. Overstating direct applicability would create unnecessary compliance claims; ignoring the governance logic would leave audit firms without a coherent response to data, model, oversight, and monitoring risks.

A second trade-off concerns explainability and predictive performance. Highly complex models may identify patterns that simpler models miss, but opacity raises the cost of review and challenge. For high-reliance audit uses, interpretability, reproducibility, or strong independent validation may therefore justify choosing a less complex model. Where a black-box tool is retained, the response should be stronger corroboration, tighter scope, enhanced reviewer involvement, and explicit documentation of residual uncertainty—not a generic explanation statement.

A third trade-off lies between centralised governance and engagement-level judgement. Central functions are better placed to approve tools, manage vendors, validate versions, monitor incidents, and maintain common documentation. Engagement teams remain responsible for client-specific data integrity, the fit between the tool and the assertion, interpretation of exceptions, and the final reliance decision. The framework allocates these responsibilities separately to prevent both uncontrolled local experimentation and mechanical reliance on centrally approved tools.

Proportionality creates a further tension with inspection readiness. Excessive controls can discourage useful low-risk automation, while insufficient documentation makes high-reliance procedures impossible to defend. The maturity model addresses this by scaling control intensity with reliance, complexity, changeability, and potential harm. Vendor opacity, confidentiality constraints, and restricted access to training or validation information may nevertheless cap the degree of permissible reliance even in otherwise mature firms.

6.3. Implications for Audit Firms, Regulators, and Assurance Markets

For audit firms, the immediate implication is to manage AI as a portfolio of governed methods rather than as a collection of isolated applications. A central inventory, role allocation, tool qualification, version control, change management, incident handling, and minimum workpaper requirements create the infrastructure within which engagement-level judgement can operate. Training should address not only tool use, but also data lineage, validation limits, automation bias, confidentiality, and the circumstances requiring consultation or alternative procedures.

For audit regulators and oversight bodies, the framework supports an outcome-oriented approach. Supervisory attention can focus on whether the firm can demonstrate reliable data, appropriate validation, reviewable outputs, accountable human decisions, and effective remediation, rather than prescribing a particular technology. Coordination with AI and data-protection authorities may become increasingly important where the same documentation serves multiple regulatory purposes, but audit evidence must still be assessed under audit-specific criteria.

Audit committees and those charged with governance can use the checklist to question how AI-assisted procedures affected coverage, judgement, confidentiality, and the audit conclusion. At the same time, expanding assurance markets for AI governance create opportunities and independence risks. Firms offering both implementation advice and assurance must safeguard objectivity, clarify the subject matter and criteria, and ensure that technical specialists support rather than displace the accountable assurance practitioner.

6.4. Limitations and Research Agenda

The study is conceptual, and its outputs have not yet been validated in live audit engagements. The mapping is structured and traceable, but it is not a statistical meta-analysis and does not measure whether a particular control improves inspection outcomes, fraud detection, efficiency, professional scepticism, or judgement quality. The framework should therefore be read as a proposed design artefact for audit-quality management and evidence documentation, not as an empirically verified model. Implementation will also vary across

firm size, national oversight arrangements, client complexity, tool architecture, AI maturity, vendor access, and the evolving interpretation of the AI Act and related standards.

Future research should test the framework at several levels. Field studies can examine how firms allocate responsibility and which controls are feasible for small and medium-sized practices. Experiments can assess how explanation formats, confidence indicators, mandatory challenge steps, and reviewer prompts affect automation bias and professional scepticism. Archival or inspection-based studies can test whether tool dossiers, logs, and monitoring reduce deficiencies. Case studies can evaluate generative-AI workpaper documentation, while comparative research can examine how AI Act compliance material functions as audit evidence for clients operating regulated AI systems.

The framework also generates testable propositions. Greater traceability should increase reviewability but may not improve judgement unless auditors possess sufficient competence; stronger human-oversight requirements should reduce automation bias but may add delay and cost; and mature central governance should improve consistency while creating a risk of over-reliance on centrally approved tools. Empirical examination of these tensions would convert the present conceptual design into evidence-based implementation guidance.

7. Conclusions

This study examined how selected EU AI Act governance requirements can be translated into audit-quality management and audit-evidence practices. Regarding RQ1, the analysis shows that eight governance domains can be mapped to firm-level and engagement-level responsibilities through risk-based quality management. The resulting control matrix and integration model allocate responsibility for tool approval, competence, data governance, documentation, oversight, monitoring, and remediation.

Regarding RQ2, the analysis identifies data integrity, model validity, traceability, explainability, reproducibility, and human challenge as the principal conditions affecting the reliability of AI-enabled audit evidence. The appropriate evidential role depends on reliance: outputs may support planning, direction, corroboration, or—where validation and reviewability are sufficiently strong—substantive procedures. The evidence-risk typology and documentation checklist make these conditions explicit.

Regarding RQ3, the framework identifies scalable controls including AI-use registers, tool dossiers, engagement risk memoranda, version and run logs, prompt logs, challenge records, reviewer sign-offs, incident procedures, drift monitoring, and source-document corroboration. These controls can support professional scepticism, evidence documentation, reviewer challenge, and inspection readiness, but they do not convert AI Act compliance into audit evidence, eliminate the need for professional judgement, or prove that AI adoption improves audit quality in practice. The contribution is therefore a governance-to-controls mechanism that links EU AI regulation with auditing while preserving the legal and professional boundaries between them. Empirical validation in audit firms, including firms of different sizes and AI maturity, is the next step.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data and materials discussed in this study are available within the article and in the publicly available sources cited in the References. No separate empirical dataset was generated or deposited for this study.

Acknowledgments: The author thanks the anonymous reviewers for their constructive comments.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Chan, D.Y.; Vasarhelyi, M.A. Innovation and Practice of Continuous Auditing. *Int. J. Account. Inf. Syst.* **2011**, *12*, 152–160. [\[CrossRef\]](#)
2. Vasarhelyi, M.A.; Alles, M.; Kogan, A. Principles of Analytic Monitoring for Continuous Assurance. *Int. J. Account. Inf. Syst.* **2012**, *13*, 267–281. [\[CrossRef\]](#)
3. Cooper, L.A.; Holderness, D.K.; Sorensen, T.L.; Wood, D.A. Robotic Process Automation in Public Accounting. *Account. Horiz.* **2019**, *33*, 15–35. [\[CrossRef\]](#)
4. Bakarich, K.M.; O'Brien, P.E. The Robots Are Coming. . . but Aren't Here Yet: The Use of Artificial Intelligence Technologies in the Public Accounting Profession. *J. Emerg. Technol. Account.* **2021**, *18*, 27–43. [\[CrossRef\]](#)
5. Cao, M.; Chychyla, R.; Stewart, T. Big Data Analytics in Financial Statement Audits. *Account. Horiz.* **2015**, *29*, 423–429. [\[CrossRef\]](#)
6. Brown-Liburd, H.; Issa, H.; Lombardi, D. Behavioral Implications of Big Data's Impact on Audit Judgment and Decision Making and Future Research Directions. *Account. Horiz.* **2015**, *29*, 451–468. [\[CrossRef\]](#)
7. Schneider, G.P.; Dai, J.; Janvrin, D.J.; Ajayi, K.; Raschke, R. Infer, Predict, and Assure: Accounting Opportunities in Data Analytics. *Account. Horiz.* **2015**, *29*, 719–742. [\[CrossRef\]](#)
8. Appelbaum, D.; Kogan, A.; Vasarhelyi, M.A. Big Data and Analytics in the Modern Audit Engagement: Research Needs. *Audit. A J. Pract. Theory* **2017**, *36*, 1–27. [\[CrossRef\]](#)
9. Earley, C.E. Data Analytics in Auditing: Opportunities and Challenges. *Bus. Horiz.* **2015**, *58*, 493–500. [\[CrossRef\]](#)
10. Bertomeu, J.; Lin, Y.; Liu, Y.; Ni, Z. The Impact of Generative AI on Information Processing: Evidence from the Ban of ChatGPT in Italy. *J. Account. Econ.* **2025**, *80*, 101782. [\[CrossRef\]](#)
11. Choi, J.H.; Xie, C.L. Human + AI in Accounting: Early Evidence from the Field. *J. Account. Res.* **2026**, *64*, 1333–1373. [\[CrossRef\]](#)
12. Babina, T.; Fedyk, A.; He, A.X.; Hodson, J. Artificial Intelligence, Firm Growth, and Product Innovation. *J. Financ. Econ.* **2024**, *151*, 103745. [\[CrossRef\]](#)
13. Almada, M.; Petit, N. The EU AI Act: Between the Rock of Product Safety and the Hard Place of Fundamental Rights. *Common Mark. Law Rev.* **2025**, *62*, 85–120. [\[CrossRef\]](#)
14. Hacker, P. Sustainable AI Regulation. *Common Mark. Law Rev.* **2024**, *61*, 345–386. [\[CrossRef\]](#)
15. De Gregorio, G.; Dunn, P. The European Risk-Based Approach: Regulating AI by (De)Constructing the Notion of Risk. *Common Mark. Law Rev.* **2022**, *59*, 473–500. [\[CrossRef\]](#)
16. Schuett, J. Risk Management in the Artificial Intelligence Act. *Eur. J. Risk Regul.* **2024**, *15*, 367–385. [\[CrossRef\]](#)
17. Ebers, M. The EU AI Act: Local and Global Risk of Algorithmic Harm. *Eur. J. Risk Regul.* **2025**, *16*, 684–703. [\[CrossRef\]](#)
18. Laux, J.; Wachter, S.; Mittelstadt, B. Three Pathways for Standardisation and the AI Act: Standards for High-Risk AI and a Route to Systemic Bias. *Comput. Law Secur. Rev.* **2024**, *53*, 105957. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Neuwirth, R.J. Prohibited Artificial Intelligence Practices: A Critical Analysis of the AI Act. *Comput. Law Secur. Rev.* **2023**, *48*, 105798. [\[CrossRef\]](#)
20. Almada, M.; Radu, A. The Brussels Side-Effect: How the AI Act Can Reduce the Global Reach of EU Policy. *Ger. Law J.* **2024**, *25*, 646–663. [\[CrossRef\]](#)
21. Bakiner, O. Pluralistic Sociotechnical Imaginaries in Artificial Intelligence (AI) Law: The Case of the European Union's AI Act. *Law Innov. Technol.* **2023**, *15*, 558–582. [\[CrossRef\]](#)
22. European Parliament and Council of the European Union. Regulation (EU) 2024/1689 of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence and Amending Regulations and Directives (Artificial Intelligence Act). *Off. J. Eur. Union* **2024**, *L*, 2024/1689.
23. Perdana, A.; Lee, W.E.; Kim, C.M. Prototyping and Implementing Robotic Process Automation in Accounting Firms: Benefits, Challenges and Opportunities to Audit Automation. *Int. J. Account. Inf. Syst.* **2023**, *51*, 100641. [\[CrossRef\]](#)
24. Krieger, F.; Drews, P.; Velte, P. Explaining the (Non-)Adoption of Advanced Data Analytics in Auditing: A Process Theory. *Int. J. Account. Inf. Syst.* **2021**, *41*, 100511. [\[CrossRef\]](#)
25. Li, X.; Dai, J.; Gershberg, T.; Vasarhelyi, M.A. Understanding Usage and Value of Audit Analytics for Internal Auditors: An Organizational Approach. *Int. J. Account. Inf. Syst.* **2018**, *28*, 59–76. [\[CrossRef\]](#)
26. Kim, H.-J.; Mannino, M.; Nieschwietz, R.J. Information Technology Acceptance in the Internal Audit Profession: Impact of Technology Features and Complexity. *Int. J. Account. Inf. Syst.* **2009**, *10*, 214–228. [\[CrossRef\]](#)
27. Davis, F.D. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Q.* **1989**, *13*, 319–340. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Ajzen, I. The Theory of Planned Behavior. *Organ. Behav. Hum. Decis. Processes* **1991**, *50*, 179–211. [\[CrossRef\]](#)
29. Compeau, D.R.; Higgins, C.A. Computer Self-Efficacy: Development of a Measure and Initial Test. *MIS Q.* **1995**, *19*, 189–211. [\[CrossRef\]](#) [\[PubMed\]](#)

30. Damanpour, F. Organizational Complexity and Innovation: Developing and Testing Multiple Contingency Models. *Manag. Sci.* **1996**, *42*, 693–716. [[CrossRef](#)]
31. Deng, Z.; Sumbal; Li, L.; Khaskheli, M.B. Inclusive and Sustainable Digital Technological Transformation, Legal Innovation in Marketing and Business Performance of China and the European Union. *Sustainability* **2026**, *18*, 2523. [[CrossRef](#)]
32. Li, Y.; Wang, S.; Khaskheli, M.B. Integrating Artificial Intelligence into Service Innovation, Business Development, and Legal Compliance: Insights from the Hainan Free Trade Port Era. *Systems* **2024**, *12*, 463. [[CrossRef](#)]
33. Francis, J.R. A Framework for Understanding and Researching Audit Quality. *Audit. A J. Pract. Theory* **2011**, *30*, 125–152. [[CrossRef](#)]
34. DeFond, M.; Zhang, J. A Review of Archival Auditing Research. *J. Account. Econ.* **2014**, *58*, 275–326. [[CrossRef](#)]
35. Knechel, W.R.; Krishnan, G.V.; Pevzner, M.; Shefchik, L.B.; Velury, U.K. Audit Quality: Insights from the Academic Literature. *Audit. A J. Pract. Theory* **2013**, *32*, 385–421. [[CrossRef](#)]
36. International Auditing and Assurance Standards Board. *2025 Handbook of International Quality Management, Auditing, Review, Other Assurance, and Related Services Pronouncements*; International Federation of Accountants: New York, NY, USA, 2025.
37. Rudin, C. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nat. Mach. Intell.* **2019**, *1*, 206–215. [[CrossRef](#)] [[PubMed](#)]
38. Miller, T. Explanation in Artificial Intelligence: Insights from the Social Sciences. *Artif. Intell.* **2019**, *267*, 1–38. [[CrossRef](#)]
39. Guidotti, R.; Monreale, A.; Ruggieri, S.; Turini, F.; Pedreschi, D.; Giannotti, F. A Survey of Methods for Explaining Black Box Models. *ACM Comput. Surv.* **2018**, *51*, 93. [[CrossRef](#)]
40. Nelson, M.W. A Model and Literature Review of Professional Skepticism in Auditing. *Audit. A J. Pract. Theory* **2009**, *28*, 1–34. [[CrossRef](#)]
41. Hurtt, R.K. Development of a Scale to Measure Professional Skepticism. *Audit. A J. Pract. Theory* **2010**, *29*, 149–171. [[CrossRef](#)]
42. Jobin, A.; Ienca, M.; Vayena, E. The Global Landscape of AI Ethics Guidelines. *Nat. Mach. Intell.* **2019**, *1*, 389–399. [[CrossRef](#)]
43. Singh Sewpersadh, N. Adaptive Structural Audit Processes as Shaped by Emerging Technologies. *Int. J. Account. Inf. Syst.* **2025**, *56*, 100735. [[CrossRef](#)]
44. Black, S.; Gerard, G.J. Accelerating the Future of Audit Technologies: Introducing the Special Issue and Emphasizing Future Research Directions. *Int. J. Account. Inf. Syst.* **2025**, *56*, 100740. [[CrossRef](#)]
45. Li, Y.; Goel, S. Artificial Intelligence Auditability and Auditor Readiness for Auditing Artificial Intelligence Systems. *Int. J. Account. Inf. Syst.* **2025**, *56*, 100739. [[CrossRef](#)]
46. Torroba, M.; Bertomeu Sánchez, J.R.; López, L.; Callejón Callejón, A.M. Investigating the Impacting Factors for Audit Professionals to Adopt Data Analysis and Artificial Intelligence: Empirical Evidence for Spain. *Int. J. Account. Inf. Syst.* **2025**, *56*, 100738. [[CrossRef](#)]
47. Lombardi, D.R.; Kim, M.; Sipior, J.C.; Vasarhelyi, M.A. The Increased Role of Advanced Technology and Automation in Audit: A Delphi Study. *Int. J. Account. Inf. Syst.* **2025**, *56*, 100733. [[CrossRef](#)]
48. Kokina, J.; Blanchette, S.; Davenport, T.H.; Pachamanova, D. Challenges and Opportunities for Artificial Intelligence in Auditing: Evidence from the Field. *Int. J. Account. Inf. Syst.* **2025**, *56*, 100734. [[CrossRef](#)]
49. Lee, E.J.; Tahmoush, D. Auditors' Decision-Making Aid for Going Concern Audit Opinions through Machine Learning Analysis. *Int. J. Account. Inf. Syst.* **2025**, *56*, 100732. [[CrossRef](#)]
50. Li, P. Application of Consumer Search Volume in Auditing. *Int. J. Account. Inf. Syst.* **2025**, *56*, 100736. [[CrossRef](#)]
51. Liu, Z.; Wu, H. A Framework for Dynamic Blockchain-Based Data Auditing. *Int. J. Account. Inf. Syst.* **2025**, *56*, 100737. [[CrossRef](#)]
52. Bertomeu, J.; Cheynel, E.; Floyd, E.; Pan, W. Using Machine Learning to Detect Misstatements. *Rev. Account. Stud.* **2021**, *26*, 468–519. [[CrossRef](#)]
53. Cheynel, E.; Cianciaruso, D.; Zhou, F.S. Fraud Power Laws. *J. Account. Res.* **2024**, *62*, 833–876. [[CrossRef](#)]
54. Lee, M.-C.; Zhao, Y.; Wang, A.; Liang, P.J.; Akoglu, L.; Tseng, V.S.; Faloutsos, C. AutoAudit: Mining Accounting and Time-Evolving Graphs. In Proceedings of the 2020 IEEE International Conference on Big Data (Big Data), Atlanta, GA, USA, 10–13 December 2020; pp. 950–956. [[CrossRef](#)]
55. Liang, P.J. Bookkeeping Graphs: Computational Theory and Applications. *Found. Trends Account.* **2023**, *17*, 77–172. [[CrossRef](#)]
56. European Parliament and Council of the European Union. Directive 2006/43/EC of 17 May 2006 on Statutory Audits of Annual Accounts and Consolidated Accounts, as Amended. *Off. J. Eur. Union* **2006**, *L 157*, 87–107.
57. European Parliament and Council of the European Union. Regulation (EU) No 537/2014 of 16 April 2014 on Specific Requirements Regarding Statutory Audit of Public-Interest Entities. *Off. J. Eur. Union* **2014**, *L 158*, 77–112.
58. NIST AI 100-1; Artificial Intelligence Risk Management Framework (AI RMF 1.0). National Institute of Standards and Technology, U.S. Department of Commerce: Gaithersburg, MD, USA, 2023. [[CrossRef](#)]

59. ISO/IEC 42001:2023; Information Technology—Artificial Intelligence—Management System. ISO: Geneva, Switzerland, 2023.
60. Debreceeny, R.; Lee, S.-L.; Neo, W.; Toh, J.S. Employing Generalized Audit Software in the Financial Services Sector: Challenges and Opportunities. *Manag. Audit. J.* **2005**, *20*, 605–618. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.