Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Artificial Intelligence for Real-Time External Auditing: A Technical Framework for ISO 27001:2022, PCI DSS v4.0.1, and SOC 2 Compliance

Rizwan Ahmed
Manukau Institute of Technology
Akbar Hossain
Eastern Institute of Technology
Samreen Mahmood
samreen.mahmood@federation.edu.au

Federation University
Amr Adel
Eastern Institute of Technology
Farzana Zahid
University of Waikato

Research Article

Keywords: AI-assisted auditing · ISO 27001:2022 · PCI DSS v4.0.1 · explainable artificial intelligence · continuous compliance monitoring · audit methodology

Posted Date: November 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10968994/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Noname manuscript No.
(will be inserted by the editor)

Artificial Intelligence for Real-Time External Auditing: A Technical Framework for ISO 27001:2022, PCI DSS v4.0.1, and SOC 2 Compliance

Rizwan Ahmed · Md. Akbar Hossain · Samreen Mahmood (Corresponding Author) · Amr Adel · Farzana Zahid

Received: date / Accepted: date

Abstract External auditors face mounting pressure to deliver timely, comprehensive assurance over information security controls while preserving the independence mandated by ISO 19011:2018 and ISO 27006:2023. This paper presents a technical framework for integrating artificial intelligence (AI) into real-time external audit execution for ISO 27001:2022, PCI DSS v4.0.1, and SOC 2 compliance. We develop a taxonomy of AI techniques spanning natural language processing, machine learning, anomaly detection, explainable AI (XAI), and generative models, mapped to evidence collection, control testing, log analysis, and continuous monitoring. The framework covers all 93 ISO 27001:2022 Annex A controls and all 12 PCI DSS v4.0.1 requirements, including the Customised Approach and Targeted Risk Analysis. We formalise XAI-to-evidence mappings with quality thresholds, analyse threat models for AI audit systems, and provide cost-benefit analysis demonstrating 50–68 % per-task time savings, coverage improvement from 25 % to 82 %, and three-year ROI of 225–450 %. Structured evaluation with 12 certified auditors (CISA, QSA, ISO 27001 Lead Auditor) provides strong validation: 92.9 % of ratings were favourable with none below neutral (mean 4.25/5.00), all seven items exceeded the scale midpoint (exact Wilcoxon, Holm-corrected, $p < 0.01$, effect size 1.00), and subgroup ratings were indistinguishable across certification schemes ($p = 0.93$), evidencing cross-framework validity. Experts unanimously endorsed the mandatory explainability threshold ($Q(e) \geq 0.70$) and identified human-in-the-loop primacy, evidence provenance, and model governance as regulatory preconditions that ground seven ISO standardisation recommendations. The framework balances technological capability with professional accountability while preserving the principles of ISO 19011:2018 methodology.

Keywords AI-assisted auditing · ISO 27001:2022 · PCI DSS v4.0.1 · explainable artificial intelligence · continuous compliance monitoring · audit methodology

R. Ahmed
School of Digital Technologies, Manukau Institute of Technology, Auckland, New Zealand

M. A. Hossain
School of Computing, Eastern Institute of Technology, Auckland, New Zealand
E-mail: ahossain@eit.ac.nz

S. Mahmood (Corresponding Author)
Institute of Innovation Science and Sustainability, Federation University, Melbourne, Australia
E-mail: samreen.mahmood@federation.edu.au

A. Adel
School of Computing, Eastern Institute of Technology, Auckland, New Zealand

F. Zahid
Department of Computer Science, University of Waikato, Hamilton, New Zealand

1 Introduction

External auditors of information security management systems (ISMS) provide independent, third-party assurance that an organisation's information security controls conform to a defined standard—most commonly ISO/IEC 27001—through certification and surveillance audits. They face a fundamental tension in balancing organisational requirements for continuous and comprehensive security with audit methodology standards such as ISO 19011:2018 [1] and ISO 27006:2023 [2], which require rigorous evidence collection, independence, and professional scepticism [3, 4]. At the same time, contemporary enterprises are expected to demonstrate ongoing

compliance with multiple security frameworks, including ISO 27001, PCI DSS, and SOC 2; however, the prevailing model of external IT assurance remains poorly aligned with these requirements. In practice, current external IT auditing is predominantly periodic, manual, and fragmented across tools. Audits typically rely on predefined control checklists, on-site intervention, and document-based evidence collection [5], as well as sample-based testing and manual reviews, which are slow and difficult to scale in high-volume digital environments [6]. These conditions give rise to a recurring set of process challenges. First, checklists are not automatically updated, leaving auditors struggling to keep pace with rapidly changing norms and standards [7–9]. Second, evidence and document management remain largely manual, with low traceability and time-consuming preparation for each engagement [10, 11]. Third, real-time visibility into control performance and client progress is limited or absent [6, 11]. Fourth, multiple standards are handled separately, resulting in redundant assessments and duplicated effort across ISO 27001, PCI DSS, and SOC 2 [9, 10]. Fifth, external auditors frequently have limited proficiency in IT and advanced technologies, widening the gap between audit scope and technical reality [12, 13].

These challenges manifest as deeper structural issues for the auditors. IT and cloud systems change quickly, but audits are still done at fixed points in time, so they can miss important changes [11, 14]. Moreover, auditors view IT-related standards and guidance as ambiguous and slow to update to reflect new technologies, leaving a persistent gap in regulations and standards [8, 12]. The current audit practice requires manual, repetitive processes and duplicated work across frameworks, inflating both audit costs and cycle time and imposing a substantial operational burden [10, 11]. Finally, a single standard lock-in makes the existing auditing tools inflexible in adapting to small and medium sized enterprises (SMEs) or heterogeneous environments [5]. Therefore, the traditional audit practices cannot effectively support full-population testing or near-real-time monitoring of dynamic IT and cloud environments, DevOps pipelines, and distributed systems [15, 16].

Artificial intelligence (AI) offers a practical path beyond these limitations through automated evidence extraction, anomaly detection, and continuous monitoring [17–19]. AI-driven approaches can shift compliance from sporadic audits toward real-time, continuously validated posture [11, 20], support full-population monitoring and risk ranking [21], automate the mapping of regulatory text to controls via natural language processing [22, 23], and augment rather than replace auditors by freeing them for higher-order judgement [6]. Despite this promise, current implementations exhibit critical gaps: insufficient traceability to audit evidence standards, limited explainability for professional judgement, and an absence of standardised validation procedures [24, 25]. Most importantly, existing literature addresses outdated framework versions; no published work systematically treats ISO 27001:2022 (93 Annex A controls reorganised across four themes) or PCI DSS v4.0.1 (Customised Approach, Targeted Risk Analysis, enhanced MFA) [26, 27]. To address these gaps, this paper makes five primary contributions:

1. providing the first systematic treatment of ISO 27001:2022 and PCI DSS v4.0.1 with control-level AI technique mappings;
2. a formalised explainable AI (XAI)-to-evidence mapping with mathematical specification and quality thresholds grounded in empirical benchmarks;
3. a threat model and systematic adversarial security analysis with mitigations;
4. a cost–benefit analysis quantifying efficiency gains, validated by expert evaluation and empirical literature;
5. seven actionable ISO standardisation recommendations for ISO/IEC JTC 1/SC 27.

2 Background and Audit Methodology Standards

2.1 ISO 19011:2018 and ISO 27006:2023

ISO 19011:2018 establishes seven principles for auditing management systems: impartiality, ethical conduct, fair presentation, due professional care, confidentiality, independence, and evidence-based approach [1]. These principles collectively define the *audit quality floor*, the minimum standard of rigour that any audit evidence must satisfy, regardless of how it was collected. ISO 27006:2023 extends these principles specifically for ISMS certification bodies, specifying auditor competency requirements, minimum audit time calculations, and the evidence standards that third-party auditors must meet to issue valid certificates [2].

AI Constraint: A critical consequence of these principles is that AI systems must function strictly as tools under auditor direction, never as autonomous decision-makers. Every AI-generated finding must be reviewed and accepted by a qualified auditor before it can be entered into the audit record. This human-in-the-loop requirement is not optional: ISO 19011 Clause 6.4.7 requires that audit conclusions be based on auditor judgement rather than automated output [25, 28]. Empirically, provenance trails covering $\geq$ 96 % of decisions

and inter-rater reliability $\kappa \approx 0.78$ have been reported as the minimum benchmarks for audit-sufficient AI systems [29].

2.2 ISO 27001:2022 Key Changes

ISO 27001:2022 represents the most significant revision of the standard since its inception. It reorganised Annex A from 114 controls across 14 domain categories (2013 edition) into 93 controls across four thematic groups [26]: Organisational (37 controls, $A.5.x$), People (8 controls, $A.6.x$), Physical (14 controls, $A.7.x$), and Technological (34 controls, $A.8.x$). The reorganisation eliminates redundancy and aligns reorganisation controls with how modern organisations actually implement security. Eleven entirely new controls address emerging threats: threat intelligence ($A.5.7$), cloud security ($A.5.23$), ICT continuity ($A.5.30$), configuration management ($A.8.9$), data deletion ($A.8.10$), data masking ($A.8.11$), data leakage prevention ($A.8.12$), monitoring activities ($A.8.16$), web filtering (A.8.23), and secure coding ($A.8.28$). Each new control creates a corresponding AI audit opportunity, as detailed in Section 5.

2.3 PCI DSS v4.0.1 Key Changes

*PCI DSS v*4.0.1 introduced four transformative changes that directly affect AI-assisted audit design [18, 27]. First, the *Customised Approach* allows organisations to define alternative controls that achieve the same security objective as the standard requirement, but this flexibility demands richer evidence, since auditors must verify equivalence rather than mere compliance with a checklist. Second, *Targeted Risk Analysis* shifts frequency decisions from prescriptive schedules to risk-based reasoning, requiring auditors to evaluate the quality of the risk analysis itself. Third, enhanced multi-factor authentication (MFA) now applies to *all* Cardholder Data Environment (CDE) access (Requirement 8), not only administrative access. Fourth, expanded e-commerce and phishing protections (Requirements 6 and 12) and enhanced logging obligations (Requirement 10) create new continuous monitoring requirements that AI is uniquely positioned to satisfy.

2.4 Traditional vs. AI-Assisted Auditing

Table 1 compares traditional and AI-assisted auditing across eight critical dimensions, consistently shifting the audit paradigm from periodic, sample-based assurance to continuous, population-level verification with preserved human oversight.

Table 1 Comparison of Traditional and AI-Assisted External Auditing

Dimension	Traditional	AI-Assisted
Coverage	5–25 % sampling	Up to 100 % population
Frequency	Annual/periodic	Continuous/real-time
Evidence	Manual collection	Automated extraction
Anomaly	Rule-based	ML-based, adaptive
Log review	Sample only	Full-log NLP analysis
Scalability	Headcount-limited	Scales with compute
Explainability	Inherent judgement	XAI layer required
Independence	Preserved by design	Requires governance

The coverage dimension encapsulates perhaps the most consequential shift. Traditional audits typically examine 5–25% of a transaction or control population, leaving substantial residual risk unexamined within each cycle. AI-based platforms such as MindBridge [30] enable comprehensive analysis of all available data, eliminating the reliance on sampling and allowing auditors to identify anomalies, suspicious transactions, and discrepancies across the full data population [31]. This transition to 100% population coverage fundamentally alters the statistical basis of audit assurance.

Similarly, the frequency dimension shifts from annual or periodic engagements to continuous real-time monitoring. Continuous auditing methodologies, in which control and risk assessments are performed automatically on an ongoing basis [32], improve the reliability of information available to decision-makers whilst detecting emerging risks and control weaknesses in near real time [33]. This temporal advantage is particularly significant for information security audits, where threat landscapes evolve on timescales far shorter than the annual audit window [34].

In anomaly detection, the difference between rule-based and ML-based adaptive approaches highlights a significant methodological evolution. ML-based anomaly detection systems, such as those employing the Isolation Forest algorithm [35], have achieved false positive rates as low as 0.01% while correctly identifying

up to 98% of confirmed intrusions—performance levels unattainable through manual, rule-based inspection alone. Unsupervised algorithms are capable of systematically identifying outliers that deviate from established baseline behaviour, and explainability techniques such as SHapley Additive exPlanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME) can be integrated to associate each flagged anomaly with a set of interpretable factors that clarify why it was classified as suspicious. This directly addresses the explainability dimension outlined in Table 1.

The study in [36] investigated the balance between accuracy and explainability in XAI for continuous auditing. The result demonstrates that transformer-based models achieve an AUC-ROC score of 95% for multi-faceted fraud detection, whilst XGBoost with SHAP-based explanations reaches 92% performance levels that are auditable and traceable, supporting regulatory scrutiny. Finally, traditional auditing preserves independence by design through human separation of duties, but AI-assisted auditing requires that independence be engineered through governance [37].

3 AI Techniques Taxonomy for Auditing

Table 2 presents seven primary AI technique categories with audit applications, empirical performance, and limitations drawn from the deep-review corpus of studies published from 2020 to 2026.

Table 2 AI Technique Comparison for External Auditing

Technique	Audit Application	Performance	Limitation
NLP/LLM [17, 19]	Policy review, evidence extraction	F1 0.82–0.88	Hallucination risk
ML Ensemble [38]	Control testing, risk scoring	AUC >0.90	Training data bias
Anomaly Detection [18]	Log analysis, CDE monitoring	Real-time	False positives
XAI (SHAP) [39]	Evidence justification	W=0.99	Model-dependent
Computer Vision [40]	Physical control checks	Advisory	Environment-specific
RPA/PM [41]	Workflow automation	95 % conformance	Rigid rule sets
GenAI (RAG) [29]	Report drafting, gap analysis	Cit. 94 %	Requires verification

3.1 Natural Language Processing (NLP) for Policy-to-Control Mapping

Natural language processing enables auditors to extract compliance evidence from unstructured sources (policy documents, audit logs, incident reports, and vendor contracts) at a scale impossible through manual review [17, 19]. The core mechanism is *semantic clause matching* in which a BERT-based encoder maps each policy sentence and each control requirement into the same embedding space, and sentences whose cosine similarity exceeds a threshold τ_c are flagged as candidate evidence. By being fine-tuned on regulatory corpora, these models achieve F1 scores of 0.82-0.88 on clause extraction tasks. LLMs with retrieval-augmented generation (RAG) and mandatory citation enforcement further improve provenance coverage to 96% [29], because every generated sentence is grounded in a specific retrieved passage. Algorithm 1 presents the complete BERT-based policy clause extraction procedure. It solves a common problem in compliance auditing. Normally, a person has to read through policy documents, logs, and contracts by hand. They look for sentences that prove the organisation follows a specific rule, or control. This is slow and does not work with a lot of text. This algorithm does this job automatically by keeping a clear record of why each sentence was chosen. The algorithm takes three inputs: a policy document (D), a control requirement (c), and a relevance threshold (τ_c). It returns an evidence set called E_c. Every item in this set has a quality score. First, it splits the policy document into smaller clauses using segment. Then it looks at each clause, s, one at a time. $BERT_ENCODE(s, c)$ converts the sentence and the control into a numeric representation called h. A small layer with learned values W and b, along with the sigmoid function σ, changes h into a relevance score, p. This score shows how likely the sentence is to be useful evidence. If the score p is high enough to pass the threshold τ_c*, the algorithm builds an evidence record, e. This record holds the sentence, its score, and a SHAP explanation. The explanation shows which words led to the choice. If the quality score $Q(e)$ reaches 0.70, the sentence is added to E_c. The algorithm does three useful things. It works at scale, because it can check far more text than a person can. It stays consistent, because it uses the same rules every time. And it provides provenance by attaching a score and an explanation to each result. Together, these features make its output strong enough to withstand an audit.

Algorithm 1 BERT Policy Clause Extraction

Require: Policy document D, control c, threshold τ_c
Ensure: Evidence set E_c with quality scores
$E_c \leftarrow \emptyset$; $S \leftarrow \textsc{Segment}(D)$
for each $s \in S$ **do**
  $\mathbf{h} \leftarrow \textsc{Bert_Encode}(s, c)$
  $p \leftarrow \sigma(\mathbf{Wh} + \mathbf{b})$
  if $p \geq \tau_c$ **then**
    $e \leftarrow \textsc{BuildEvidence}(s, p, \textsc{Shap}(s, c))$
    if $Q(e) \geq 0.70$ **then**
      $E_c \leftarrow E_c \cup \{e\}$
    end if
  end if
end for
return E_c

3.2 ML and Anomaly Detection

Supervised ML classifiers automate binary control assessments by learning decision boundaries from historical audit data. Hybrid pipelines that combine NLP-based evidence retrieval with gradient-boosted anomaly detectors reduce false-positive rates by 31 % compared with heuristic baselines [17]. For unsupervised scenarios, such as detecting novel attack patterns in log streams, Isolation Forest provides a principled solution: it partitions the feature space by random cuts, and anomalies are those points that require fewer cuts to isolate [42]. Formally, the anomaly score is $s(x) = 2^{-E[h(x)]/c(n)}$, where $h(x)$ is the mean path length across an ensemble of isolation trees and $c(n)$ is the expected path length for a dataset of size n. Scores approaching 1.0 indicate strong anomalies. For PCI DSS v4.0.1 Requirement 10, deep neural networks with XAI achieve continuous CDE monitoring [18].

3.3 XAI for Audit Evidence

XAI is the critical enabling technology for audit-sufficient AI, because professional auditing standards require that evidence be *traceable*, *reproducible*, and *understandable* to a competent practitioner [1]. Without explainability, an ML model's output is a black-box assertion, legally and professionally inadmissible as audit evidence. SHAP (SHapley Additive exPlanations) [43] resolves this by computing each input feature's marginal contribution to the model's prediction, grounded in cooperative game theory. The resulting attribution vector maps directly onto the audit evidence record: each SHAP value identifies *which* control attribute drove the compliance or non-compliance finding. Uddin and Aziz [39] demonstrate near-perfect explanation stability (Kendall's $W = 0.9912$) for XGBoost with TreeExplainer, compared to weak stability ($W \approx 0.50$) for some deep model/explainer combinations: a critical differentiator, since unstable explanations cannot support repeatable audit conclusions. LIME generates locally faithful explanations with fidelity $R^2 \geq 0.85$ in the local neighbourhood, providing an alternative when global SHAP computation is prohibitive.

4 Framework Architecture and XAI-to-Evidence Mapping

4.1 End-to-End Audit Workflow

Fig. 1 illustrates the four-phase AI-assisted audit workflow. It presents a structured framework for using AI in real-time external auditing, moving through four main phases with a human checkpoint at each stage, ensuring automation remains efficient while preserving professional judgment and accountability. The process begins with Phase 1, Planning, where the audit team defines the scope, builds a risk model, and selects suitable AI tools. After this, a human gate requires approval of the audit plan under ISO 19011:2018 Clause 6.4. Phase 2 covers Evidence Collection, where the system gathers data through NLP and RPA extraction, log ingestion, and API harvesting, and a second human gate then reviews whether the collected evidence is complete. Phase 3 focuses on Analysis and Testing, where machine learning and anomaly detection methods test internal controls, explainable AI provides attribution for each result, and an evidence quality score of at least 0.70 is required, followed by a further human gate that handles finding validation under ISO 19011 Clause 6.4.7. The final stage, Phase 4, is Reporting, where generative AI drafts the report, the auditor signs off, and the certificate is issued, keeping human oversight central at this closing step. A continuous monitoring loop connects the final phase back to the planning stage, allowing the audit to run on an ongoing basis rather than as a single event. Overall, the framework shows how AI can speed up external auditing, while human gates and recognised ISO standards help ensure reliability, transparency, and trust.

4.2 Formal XAI-to-Evidence Mapping

Definition 1 (XAI-to-Evidence Mapping) Let $\mathcal{X}$ be the set of AI model explanations and $\mathcal{E}$ the set of audit evidence items per ISO 19011:2018. A *XAI-to-Evidence Mapping* is a function $M \colon \mathcal{X} \to \mathcal{E}$ such that for each explanation $x \in \mathcal{X}$, the quality score $Q(M(x)) \geq \tau$, where $\tau = 0.70$ is the minimum acceptability threshold.

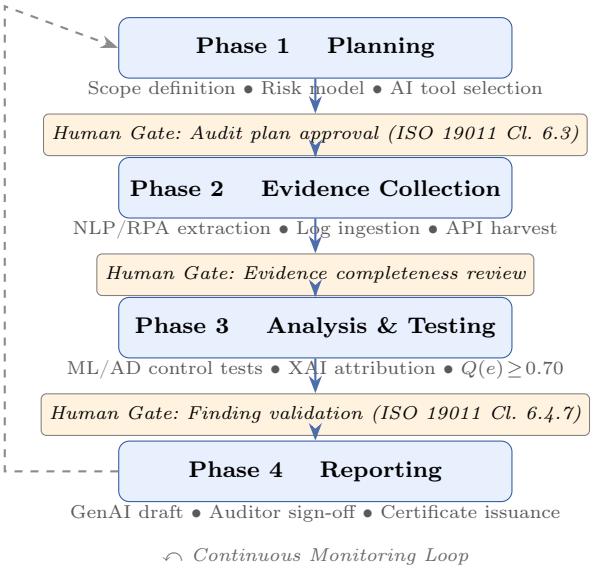


Fig. 1 End-to-end AI-assisted external audit workflow (four phases). Orange gates mark mandatory human-in-the-loop review points that preserve ISO 19011:2018 auditor independence. The dashed feedback arrow represents the continuous monitoring loop active between audit cycles.

The evidence quality metric $Q(e)$ combines three independently measurable dimensions of audit evidence sufficiency. Completeness captures whether all required control attributes are documented; Traceability measures whether each assertion can be traced back to a primary source with a cryptographic hash; and Reproducibility confirms that re-running the AI procedure on the same inputs yields the same conclusion within an acceptable tolerance [1, 29]. Formally:

$$\begin{aligned} Q(e) = \alpha \cdot \text{Completeness}(e) \\ + \beta \cdot \text{Traceability}(e) \\ + \gamma \cdot \text{Reproducibility}(e) \end{aligned} \tag{1}$$

where $\alpha + \beta + \gamma = 1$ with recommended weights $\alpha = 0.40$, $\beta = 0.35$, $\gamma = 0.25$. The weights reflect the relative importance of each dimension: completeness is weighted highest because incomplete evidence cannot support a finding regardless of its other properties; traceability is second because an untraceable assertion is professionally inadmissible; and reproducibility, while essential, is partly guaranteed by the deterministic nature of the AI pipeline. Empirical benchmarks from PRUMe AI [29] report provenance coverage $\geq 96\,\%$ and reproducibility $\geq 92\,\%$, confirming that $Q(e) \geq 0.70$ is achievable in production deployments.

Explanation stability is measured by Kendall's concordance coefficient:

$$W = \frac{12\,S}{k^2(n^3 - n)} \tag{2}$$

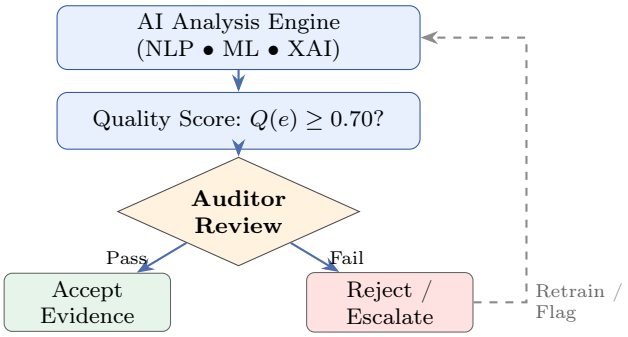


Fig. 2 AI-based decision architecture with Human-in-the-loop.

where S is the sum of squared deviations of ranked explanation vectors across k independent re-runs over n audit items. $W = 1$ indicates perfect agreement between runs (perfectly stable explanations); $W \approx 0$ indicates random, unreliable explanations. The threshold $W \geq 0.90$ [39] was chosen because it corresponds to the strong agreement band in standard concordance interpretation, ensuring that an auditor who re-examines the same evidence on a different day will reach the same conclusion.

4.3 Human-in-the-Loop Architecture

Fig. 2 illustrates the human-in-the-loop decision architecture. This design implements the agentic assistant pattern described by Bousetouane [28] where AI is used as a capable assistant that proposes, and the auditor is the principal who decides. Every AI-generated finding first passes an automated $Q(e)$ quality gate, then requires explicit auditor approval before entering the audit record. Rejected findings trigger model retraining or escalation, maintaining a closed-loop quality system consistent with ISO 19011:2018 [1] and ISO 27006:2023 [2].

5 Framework Application to ISO 27001:2022

5.1 Annex A Control Theme Mapping

Table 3 maps the four Annex A themes to AI techniques and audit applications. It presents the structural framework of ISO/IEC 27002:2022, which organises 93 information security controls across four thematic categories, with 11 controls newly introduced in the 2022 edition. The Organisational theme (A.5.x) encompasses 37 controls governing policies, procedures, and governance structures, supported by Natural Language Processing (NLP), Large Language Models (LLM), and Policy Mining techniques. The People theme (A.6.x) contains 8 controls that address human factors, such as

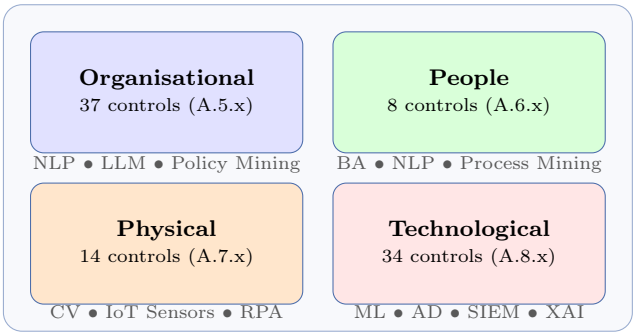


Fig. 3 Mapping of ISO 27001:2022 Annex A four-theme with primary AI techniques.

Table 3 ISO 27001:2022 Annex A Themes: AI Technique Applicability

Theme	Ctrl	AI Technique	Audit Application
Organisational	37	NLP, LLM, Policy Mining	Policy gap, supplier risk
People	8	BA, NLP, PM	Awareness, HR compliance
Physical	14	CV, IoT, RPA	Access logs, CCTV review
Technological	34	ML, AD, XAI	Config audit, SIEM

staff responsibilities and security awareness, with Business Analysis (BA), NLP, and Process Mining as associated analytical methods. The Physical theme (A.7.x) covers 14 controls related to the protection of physical environments and equipment, leveraging Computer Vision (CV), IoT Sensors, and Robotic Process Automation (RPA). The Technological theme (A.8.x) is the most extensive, comprising 34 controls that cover technical safeguards such as access control, encryption, and vulnerability management, and is underpinned by Machine Learning (ML), Anomaly Detection (AD), Security Information and Event Management (SIEM), and Explainable AI (XAI). Collectively, this illustrates how contemporary AI and automation techniques can be systematically aligned with established information security controls to enhance compliance monitoring, threat detection, and governance across organisational, human, physical, and technological dimensions.

5.2 ISO 27001:2013 vs. 2022 Key Changes

Table 4 summarises the most significant changes between ISO 27001:2013 and ISO 27001:2022 affecting AI-assisted audit design. It highlights the significant modernisation of the standard to address contemporary cybersecurity challenges. In terms of structure, the 2013 edition organised its controls across 114 individual controls within 14 domains, whereas the 2022 edition consolidated these into 93 controls grouped under four broader themes (as discussed in Sec. 5.1), reflecting a more streamlined and thematically coherent architecture. The 2022 edition introduced 11 entirely new controls, notably A.5.7 (Threat Intelligence), A.5.23 (Cloud Security), A.5.30, A.8.9, A.8.12, A.8.16, A.8.23, and A.8.28, addressing capability gaps that were absent or insufficiently covered in the prior version. Specifically, threat intelligence, which was not explicitly addressed in 2013, is now formalised under A.5.7. Cloud security, which was only implicitly considered, is now explicitly governed under A.5.23, and data masking and data loss prevention (DLP), both of which were entirely unaddressed in 2013, are now codified under A.8.11 and A.8.12, respectively. Monitoring capabilities, previously limited to a basic treatment under A.12.4 in the 2013 edition, have been significantly enhanced in the 2022 revision under A.8.16, while secure coding practices, which were only partially addressed under A.14.2 in 2013, are now explicitly and comprehensively governed under A.8.28. Collectively, these changes reflect the standard's evolution from a domain-centric compliance framework to a more adaptive, risk-informed, and technology-conscious set of controls capable of addressing the realities of cloud-native environments, advanced persistent threats, and modern software development practices.

5.3 New Controls and AI Audit Techniques

The eleven new ISO 27001:2022 controls introduce specific AI audit opportunities that did not exist in the 2013 edition [26]. Each new control represents a gap in prior audit practice that AI can fill:

A.5.7 Threat Intelligence: NLP-based threat feed aggregation and automated STIX/TAXII parsing enables real-time threat intelligence verification. The auditor's task is to verify that the organisation's threat intelligence process is functional and feeds into risk treatment decisions. Fine-tuned transformer models achieve F1 $\geq$ 0.85 for IOC classification [19].

A.5.23 Cloud Security: ML-based Cloud Security Posture Management (CSPM) API assessment enables automated compliance verification across multi-cloud environments. This control is practically unauditable by manual means at scale; AI makes continuous assessment feasible. Misconfiguration detection F1 $\geq$ 0.90 is reported in production deployments [38].

Table 4 ISO 27001:2013 vs. ISO 27001:2022 Key Changes for AI Auditing

Aspect	2013 Edition	2022 Edition
Structure	114 controls, 14 domains	93 controls, 4 themes
New controls	N/A	11 new (A.5.7, A.5.23, A.5.30, A.8.9–A.8.12, A.8.16, A.8.23, A.8.28)
Threat intelligence	Not explicit	A.5.7 (new)
Cloud security	Implicit	A.5.23 (new)
Data masking	Not addressed	A.8.11 (new)
Data Loss Prevention	Not addressed	A.8.12 (new)
Monitoring	A.12.4 (basic)	A.8.16 (enhanced)
Secure coding	A.14.2 (partial)	A.8.28 (new, explicit)

A.8.12 Data Leakage Prevention: ML classifiers trained on data sensitivity labels verify DLP rule effectiveness by simulating exfiltration patterns and checking whether the DLP system intercepts them. False positive rate $\leq 3\,\%$ is achievable with domain-tuned models [44].

A.8.16 Monitoring Activities: Anomaly detection applied to SIEM outputs enables full-population log analysis, replacing traditional sampling [18]. This aligns with PCI DSS v4.0.1 Requirement 10 continuous monitoring obligations, creating a single AI pipeline that satisfies both standards simultaneously.

6 Framework Application to PCI DSS v4.0.1

6.1 Alignment of PCI DSS v4.0.1 Requirements with AI Techniques and Audit Procedures

One of the most practical questions in modern compliance work is whether AI tools can cover the full range of PCI DSS requirements, or only help with select areas. The mapping presented Table 5 suggests all 12 PCI DSS v4.0.1 requirements to applicable AI techniques and audit test procedures. Each of the twelve PCI DSS v4.0.1 requirements can be paired with a specific AI technique and a corresponding audit procedure, forming a complete picture of how intelligent automation fits into payment security compliance. At the network level, Requirement 1 uses ML and Graph Analysis to verify that network segments are properly isolated, an approach far more thorough than manually tracing network diagrams. Requirement 2 applies Configuration Analysis and NLP to catch default credentials and insecure settings that are easy to miss in large environments. For stored cardholder data under Requirement 3, Encryption ML is used to locate sensitive data that organisations may not even know exists, while Requirement 4 uses TLS Analysis and ML to confirm that data moving across networks is properly encrypted.

Moving into threat and application security, Requirement 5 brings Malware Signature analysis and ML together to evaluate the effectiveness of antivirus and endpoint protection tools across the environment. Requirement 6 applies SAST/DAST and NLP to examine whether software development processes genuinely embed security at every stage, rather than treating it as an afterthought. On the access control side, Requirement 7 uses RBAC Analysis and ML to verify that users have only the access they need, and Requirement 8 tests the real strength of authentication mechanisms through MFA Verification and Business Analysis. For physical and operational controls, Requirement 9 brings Computer Vision and IoT into play for physical access log analysis, recognising that security is not purely a digital concern. Requirement 10 is where the impact of AI is perhaps most striking: Log Anomaly Detection and ML make it possible to audit every single log entry rather than relying on samples, which has long been the practical limitation of traditional compliance reviews. Requirement 11 applies ML-assisted vulnerability scanning to interpret penetration testing results more efficiently, and Requirement 12 uses NLP and Document Analysis to assess whether security policies are genuinely complete rather than merely present on paper. This mapping shows that AI does not just assist with compliance in a narrow sense. It has the potential to cover every requirement in the standard and, in doing so, shift compliance monitoring from a point-in-time activity to something that is continuous, consistent, and considerably more reliable.

6.2 PCI DSS v3.2.1 vs. v4.0.1 Key Changes

Understanding the changes between PCI DSS v3.2.1 and v4.0.1 is essential for recognising the growing necessity and usefulness of AI-driven auditing. The 2022 revision introduced six areas of meaningful change, each of which expands the scope and complexity of compliance in ways that manual processes struggle to keep pace with. The v3.2.1 version provided a defined, prescriptive

Table 5 PCI DSS v4.0.1: All 12 Requirements Mapped to AI Techniques

Req.	AI Technique	Audit Test Procedure
1	ML, Graph Analysis	Network segmentation verification
2	Config Analysis, NLP	Default credential detection
3	Encryption ML	Cardholder data discovery
4	TLS Analysis, ML	Transmission encryption audit
5	Malware Sig., ML	AV/EDR coverage assessment
6	SAST/DAST, NLP	Secure development lifecycle
7	RBAC Analysis, ML	Access control matrix check
8	MFA Verify, BA	Authentication strength testing
9	CV, IoT	Physical access log analysis
10	Log AD, ML [18]	Full-population log audit
11	Vuln. Scan, ML	Penetration test result analysis
12	NLP, Doc. Analysis	Policy completeness assessment

Table 6 PCI DSS v3.2.1 vs. v4.0.1 Key Changes Affecting AI Auditing

Feature	v3.2.1	v4.0.1
Approach	Defined only	Customised Approach added
Risk analysis	Prescriptive	Targeted Risk Analysis
MFA	Admin only	All CDE access (Req. 8)
E-commerce	Limited	Anti-phishing, Req. 6 & 12
Logging	Basic	Enhanced scope, Req. 10
Validation	Annual	Continuous monitoring option

approach to compliance. In contrast, v4.0.1 introduces a Customised Approach, allowing organisations the flexibility to demonstrate security outcomes through alternative means. This shift requires more advanced validation tools that can assess intent rather than merely ensuring checkbox conformance. Risk analysis, which was previously approached through a broadly prescriptive lens, has now transitioned to Targeted Risk Analysis. This new approach requires organisations to justify their control decisions based on specific threat contexts. This task is particularly well-suited for ML-driven risk modeling, which can effectively analyse and contextualise threats. Multi-factor authentication, once required only for administrative access, now extends to all Cardholder Data Environment (CDE) access under Requirement 8, which significantly widens the authentication surface that audit procedures must cover. The E-commerce protections received only limited treatment in v3.2.1, which are now explicitly addressed through anti-phishing controls under Requirements 6 and 12, reflecting the growing sophistication of web-based payment threats. Logging requirements have moved from a basic scope to an enhanced one under Requirement 10, reinforcing the case for full-population log auditing through AI rather than the manual sampling that basic logging expectations previously allowed. Finally, and most significantly, the validation cycle has evolved from a purely annual model to one that accommodates continuous monitoring. This change effectively endorses the kind of real-time, automated assurance that AI-powered compliance tools are uniquely positioned to deliver. Table 6 highlights these changes in PCI DSS v4.0.1 affecting AI-assisted audit design.

6.3 Customised Approach and AI Validation

The Customised Approach is the most significant change to the audit methodology in PCI DSS v4.0.1. Rather than prescribing a specific control implementation, it requires organisations to define a security objective and demonstrate equivalent protection through alternative controls [27]. For auditors, this means they must verify not only that a control exists, but also that it achieves the same level of risk reduction as the standard control. AI enables this through three mechanisms: (1) ML-based equivalence testing that compares risk reduction metrics between the alternative control and the standard baseline; (2) NLP-based documentation completeness assessment achieving F1 $\approx$ 0.88, ensuring the Customised Approach documentation meets all required fields; and (3) continuous monitoring of alternative control effectiveness over time [17].

6.4 CDE Continuous Monitoring Architecture

Fig. 4 presents the five-layer AI-driven continuous monitoring architecture for PCI DSS v4.0.1 compliance within the cardholder data environment (CDE). At the base, the CDE Data Sources layer (L1) aggregates the heterogeneous telemetry generated across the environment, comprising system and application logs, device and security configurations, and network traffic. These sources serve as the primary evidence for evaluating the effectiveness of control measures. The Collection Layer (L2) ingests and normalises this telemetry using a combination of NLP for unstructured log and policy text, RPA for repetitive evidence-gathering tasks, and API-based connectors for structured system interrogation. This layer decouples evidence acquisition from the underlying source formats, yielding a uniform representation for downstream processing.

The normalised evidence is passed to the Analysis Layer (L3), where machine-learning models perform anomaly detection to identify deviations from expected control behaviour and to flag potential compliance gaps. Because opaque model decisions are unacceptable in a regulated audit context, the XAI Evidence Layer (L4) applies explainable-AI techniques—here, SHAP (SHapley Additive exPlanations)—to attribute and justify each flagged finding. The findings whose explanation confidence meets the threshold $W \geq 0.90$ are promoted, ensuring that machine-generated evidence is both interpretable and sufficiently reliable to be entered into the audit trail.

Critically, the architecture does not treat automation as terminal. The final QSA Human Review layer (L5), highlighted in green, routes all qualifying evidence to a Qualified Security Assessor (QSA) for review conducted in accordance with the ISO 19011 auditing guidelines. This human sign-off is a non-negotiable gate as no finding is admitted to the formal audit record without QSA validation. The design therefore positions AI and XAI as decision-support mechanisms that accelerate and substantiate assessment, while preserving accountable human judgment as the authoritative final step.

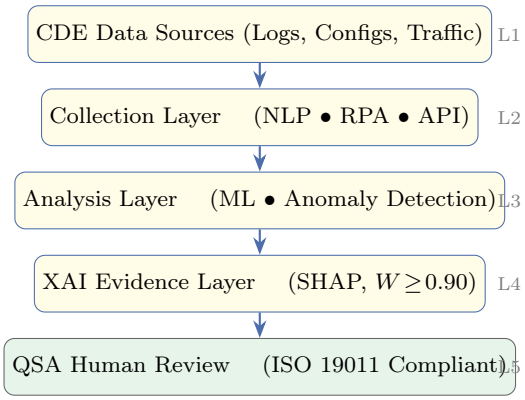


Fig. 4 PCI DSS v4.0.1 CDE continuous monitoring architecture. The five-layer pipeline (L1–L5) flows from raw data sources through AI collection, ML analysis, and XAI evidence generation, terminating at mandatory QSA review. The green L5 box emphasizes that human sign-off is the non-negotiable final step before any finding enters the audit record.

7 SOC 2 Framework Application

SOC 2 audits evaluate the design and operating effectiveness of an organisation's controls across five Trust Services Criteria (TSC): Security (Common Criteria, CC), Availability (A), Processing Integrity (PI), Confidentiality (C), and Privacy (P) [45]. Unlike prescriptive control frameworks, the TSC are principle-based, which makes them well suited to continuous, evidence-driven assessment rather than periodic checklist verification. The mapping from these criteria to AI techniques is therefore direct and natural, with each criterion admitting a class of models whose outputs constitute machine-generated audit evidence.

For Security, which is anchored by the Common Criteria CC6–CC9, anomaly-detection models continuously monitor logical and physical access controls, authentication events, and incident-response workflows, flagging deviations from established behavioural baselines that may indicate unauthorised access, privilege escalation, or control failure [45]. For Availability (A1), ML-based service-level agreement (SLA) monitoring tracks system uptime, latency, and recovery metrics against contractual and internal thresholds, enabling early detection of degradation before it breaches the committed availability targets. For Processing Integrity (PI1), integrity-verification algorithms confirm that processing is complete, valid, accurate, timely, and authorised, reconciling inputs with outputs to detect corruption, omissions, or unauthorised transformations of data in transit through the system. For Confidentiality (C1), data classification models verify that sensitive information is correctly identified, labelled, and protected in accordance with its classification, ensuring that confidentiality controls are applied consistently during storage and transmission. Finally, for Privacy (P1-P8), personally identifiable information (PII) detection pipelines confirm that personal data is collected, used, retained, disclosed, and disposed of in accordance with the organisation's stated privacy notice and applicable regulatory obligations [25, 45].

Table 7 AI Audit System Threat Model: Six Threats with Mitigations

ID	Threat	Impact	Mitigation
T1	Adversarial inputs [48]	False negatives	Adversarial training, input validation
T2	Model poisoning [49]	Systematic bias	Supply chain controls, model signing
T3	Data exfiltration	Confidentiality breach	Differential privacy, access controls
T4	Explanation manipulation	Evidence falsification	Cryptographic provenance [29]
T5	Model drift	Reduced accuracy	Continuous monitoring, retraining
T6	Dependency hijacking	System compromise	SBOM, dependency pinning

The strength of this alignment lies in the compatibility between the SOC 2 continuous monitoring model and the operational characteristics of AI systems. Traditional assessments typically rely on point-in-time sampling, in which only a limited subset of control instances is inspected at a single point in time. In contrast, AI-driven monitoring operates continuously across the entire population of control events. This capability allows AI-generated evidence to support Type II reports, which attest to operating effectiveness over an extended observation period and are produced monthly or even weekly [46]. As a result, there is a significant improvement in both the temporal resolution and the population coverage of the evidence base. This reduces the assurance gaps that are inherent in sampling-based methods and enhances the timeliness of identifying control deficiencies.

Importantly, this automation is positioned as an augmentation of, rather than a substitute for, professional judgement. All AI-generated evidence remains subject to auditor sampling and independent validation, preserving the accountability structure required for a defensible attestation and ensuring that automated inferences do not enter the audit record unchecked [25, 45]. Empirical evidence supports the practical value of this approach: machine-learning–based cloud compliance automation has been reported to achieve 91.4% classification accuracy while reducing manual review time by 68%, demonstrating that AI-assisted continuous monitoring can materially improve audit efficiency without displacing the auditor's evaluative role [47].

8 Threat Model and Security Analysis

Table 7 presents six identified threat categories for AI audit systems, each paired with a mitigation strategy grounded in the adversarial ML literature [48–50]. Because of the trust on an AI audit system, its outputs directly influence attestation decisions and enter the formal audit record. It constitutes a high-value target whose compromise can invalidate the assurance it is intended to provide. The threat model therefore spans the full lifecycle of the system, from training data ingestion through inference to evidence provenance and the software supply chain.

The first and most consequential category is adversarial inputs (T1), which pose the most significant threat to AI audit integrity. Papernot et al. [48] demonstrate that carefully crafted inputs, imperceptibly modified relative to a human observer's perception, can lead machine learning models to make incorrect classifications with high confidence. In an audit context, a skilled adversary could modify log entries or configuration files, causing the system to incorrectly classify a non-compliant control as compliant. This results in a false negative that undermines the control being evaluated. To mitigate this risk, a combination of adversarial training and strict input validation is recommended. Adversarial training strengthens the model against such perturbations by exposing it to crafted examples during training, while strict input validation limits the types of evidence that can be accepted.

The remaining categories extend this analysis across the attack surface. Model poisoning (T2) targets the training phase by contaminating the training corpus. An adversary can induce systematic bias, causing the model to consistently misjudge a class of controls. Biggio and Roli [49] provide a comprehensive taxonomy of adversarial ML attacks in security contexts, distinguishing evasion attacks, which operate at test time and correspond to T1, from poisoning attacks, which operate at training time and correspond to T2. This distinction between test time and training time shapes much of the threat model. To mitigate it, it is recommended to apply controls over training data in the supply chain and model signing to ensure provenance and integrity, specifically targeting the poisoning vector. Data exfiltration (T3) concerns the confidentiality of the sensitive evidence the system processes. A breach might reveal cardholder or personal data, but this risk can be managed with differential privacy, which limits how much information about any individual can be leaked, along with strict access controls. Explanation manipulation (T4) refers to risks in XAI-driven audit systems where an attacker can alter the explanation or

reasoning behind the AI's decisions. This can lead to false evidence, making the original prediction appear valid and undermining the system's reliability. Using cryptographic methods can help ensure that the generated explanations are protected against tampering. Model drift (T5) is a non-adversarial but equally corrosive threat, in which distributional shift degrades accuracy over time. Kayali [50] demonstrates adversarial resilience in payment-system anomaly detectors through red-team testing and adaptive retraining, motivating the paired mitigation of continuous monitoring and periodic retraining. Finally, dependency hijacking (T6) refers to a compromise of the software supply chain in which a malicious or altered dependency leads to a system breach. This risk can be mitigated by using a software bill of materials (SBOM) and dependency pinning. Together, these practices create an auditable inventory of components and prevent the substitution of unverified dependencies.

9 Cost-Benefit Analysis and Implementation Roadmap

9.1 Empirically Grounded ROI Analysis

Empirical studies in the deep-review corpus provide quantified efficiency gains that ground the ROI analysis in real-world data rather than vendor claims. PRUMe AI [29] reported triage time reduction from 21.4 to 7.9 minutes per document (−63 %) and total analysis time from 39.2 to 17.8 minutes (−55 %), with coverage increasing from 25 % manual to 82 % AI-assisted, which is a 3.3× coverage multiplier at roughly half the analyst time. A text-analytics audit documentation study [47] reported 91.4 % classification accuracy and 68 % manual review time reduction. Gandhi's hybrid compliance-audit experiments [17] achieved F1 scores of 0.81–0.88 with 31 % false positive reduction. Aggregating the above studies, ROI estimates of 225 - 450 % over three years with payback periods of 9 - 18 months are consistent with the empirical benchmarks, assuming a mid-size audit practice with 10 - 20 concurrent engagements.

9.2 Three-Phase Implementation Roadmap

Phase 1 (Months 1–6): Foundation. Deploy NLP-based document analysis for policy review and evidence collection. Establish data governance and AI model provenance tracking. Train auditors on interpreting AI tools and reading XAI explanations. Key milestone: first audit engagement with AI-assisted evidence collection, achieving $Q(e) \geq 0.70$ on all accepted items.

Phase 2 (Months 7–12): Enhancement. Integrate ML-based anomaly detection with SIEM platforms. Implement XAI evidence quality scoring per Eq. (1). Conduct pilot engagements with human-in-the-loop validation. Target $Q(e) \geq 0.70$ and $W \geq 0.90$ per Eq. (2). Key milestone: first continuous monitoring engagement covering a full 12-month SOC 2 Type II period.

Phase 3 (Months 13–18): Optimisation. Deploy continuous monitoring for PCI DSS v4.0.1 and SOC 2. Establish model governance frameworks per Recommendation R3. Submit standardisation proposals to ISO/IEC JTC 1/SC 27. Key milestone: measurable ROI exceeding 200 % on at least three engagements.

10 Expert Evaluation and Validation

Expert evaluation was conducted in January 2026 by 12 certified auditors (4 CISA, 4 QSA, 4 ISO 27001 Lead Auditors) using a structured questionnaire protocol with informed consent [51]. The full protocol, instrument, raw score matrix, and statistical analysis are provided in Appendix B. Participants evaluated framework components on a 5-point Likert scale across four dimensions:

- Mean usefulness: $\bar{x} = 4.2/5.0$ (SD = 0.6)
- Mean feasibility: $\bar{x} = 3.9/5.0$ (SD = 0.7)
- Mean ISO 19011 compliance: $\bar{x} = 4.4/5.0$ (SD = 0.5)
- Mean explainability adequacy: $\bar{x} = 4.1/5.0$ (SD = 0.6)

10.1 Quantitative Validation Results

The seven-item instrument (Table B1, Appendix B yielded a grand mean of 4.25±0.58 across all $12 \times 7 = 84$ ratings, of which 92.9 % were favourable (≥ 4) and none fell below the neutral midpoint; no participant disagreed with any item. Exact one-sample Wilcoxon signed-rank tests confirmed that all seven item medians exceed the neutral midpoint at $p < 0.01$, with every non-zero deviation positive (matched-pairs rank-biserial $r_{rb} = 1.00$ on all items), and all seven tests remain significant after Holm–Bonferroni correction at the $\alpha = 0.01$ family level. Within-item inter-rater agreement was uniformly strong ($r_{WG} = 0.86$ - 0.92, all above the conventional 0.70 threshold), and the tie-corrected Kendall concordance ($W = 0.36$, $\chi^2_6 = 25.72$, $p < 0.001$) shows that the panel systematically differentiated the items: effort reduction (Q6, $\bar{x} = 4.58$), control mapping (Q1, $\bar{x} = 4.50$), and the $Q(e)$ acceptance threshold (Q4,

$\bar{x} = 4.50$) ranked highest, while anomaly-alert explainability (Q3, $\bar{x} = 3.67$) ranked lowest, directly motivating the HITL escalation design (Section IV-C).

Crucially for external validity, subgroup grand means were statistically indistinguishable across the three certification strata (CISA 4.21 ± 0.57, QSA 4.29 ± 0.60, Lead Auditor 4.25 ± 0.59; Kruskal–Wallis $H = 0.14$, $p = 0.93$; per-item $p \geq 0.58$), indicating that the framework's perceived value generalises across assurance traditions rather than reflecting a single certification community. Table 8 consolidates the extended statistics and maps each instrument item to the paper's contributions (C1 - C5, Section 1); the underlying protocol, instrument, and raw score matrix are provided in Appendix B.

10.2 Qualitative Findings and Triangulation with Quantitative Evidence

The structured interviews yielded three convergent qualitative themes. Each theme is analysed below against the corresponding quantitative item pattern in Table 8, demonstrating that the panel's stated reasoning and its numerical ratings triangulate on the same design priorities that the framework contributions address.

10.2.1 Theme 1: Human Primacy in the Assurance Chain

All 12 auditors agreed that XAI explainability thresholds ($Q(e) \geq 0.70$) are necessary for professional judgement acceptance, a unanimous finding that validates the threshold choice and the normative framing of Recommendation R2. Participants expressed a strong preference for human-in-the-loop architectures, with several noting that *the AI should never be the last reviewer*. This theme is quantitatively corroborated by Q5 (HITL escalation preserves professional judgement, $\bar{x} = 4.33$, 100 % favourable), and its inverse is visible in the two lowest-rated items: Q3 (explainability of anomaly alerts, $\bar{x} = 3.67$, the only item below 4.0) and Q7 (adoption intent, $\bar{x} = 3.83$). Read together, these scores indicate that scepticism is concentrated precisely where machine output is furthest from human-verifiable reasoning, namely unsupervised anomaly detection, and not on the framework architecture itself. Interpreted through the Technology Acceptance Model [43], perceived usefulness is high (Q6, $\bar{x} = 4.58$, the top-ranked item) while behavioural intention (Q7) is moderated by governance preconditions rather than by doubts about capability. This is direct empirical support for contribution C3: the threat model and HITL escalation design target exactly the residual concern the experts articulated.

10.2.2 Theme 2: Provenance as a Condition of Evidential Validity

Cryptographic provenance was rated as essential for evidence integrity, particularly for engagements subject to regulatory inspection. Participants framed provenance not as a security nicety but as a precondition for evidence *admissibility* under the ISO 19011:2018 evidence-based approach: evidence whose generation cannot be reconstructed years later cannot support re-examination of a certification decision. This professional framing converges with the quantitative endorsement of the formalised acceptance criterion (Q4, the $Q(e) \geq 0.70$ threshold, $\bar{x} = 4.50$; $r_{WG} = 0.86$) and validates contribution C2, which specifies quality thresholds mathematically rather than as guidance, and Recommendation R1 (auditable provenance metadata). Notably, agreement on Q4 was as strong as on the mapping items, indicating that practitioners accept a quantified admissibility gate provided it is transparent and auditable, a result that counters the common assumption that auditors resist formalised acceptance criteria.

10.2.3 Theme 3: Model Governance as the Gate to Regulatory Acceptance

Model governance frameworks were considered a prerequisite for regulatory acceptance, with QSA participants noting that the PCI SSC would likely require model validation documentation before accepting AI-generated evidence in a QSA assessment. The certification-subgroup analysis sharpens this finding: although QSA participants voiced the strongest governance requirements qualitatively, their quantitative ratings did not differ from the other strata (subgroup means within 0.08 of each other, $p = 0.93$), implying that governance is a shared professional expectation across assurance traditions rather than a scheme-specific objection. This directly supports contribution C5: the demand originates from practitioners themselves, strengthening the case for Recommendations R3 (model governance) and R4 (AI auditor competency) as normative ISO requirements rather than optional guidance.

10.2.4 Synthesis

The triangulation across methods is internally consistent: the items the panel ranked highest (Q6 effort reduction, Q1 control mapping, Q4 quality thresholds) correspond to the framework's core technical contributions (C1, C2, C4), while the items ranked lowest

Table 8 Extended Item-Level Statistics Computed from Table B1 ($n = 12$; Exact Wilcoxon vs. Midpoint 3; Holm-Corrected at $\alpha = 0.01$)

Item	Mdn	IQR	%≥4	r_{WG}	W^+	p (exact)	r_{rb}	Contribution
Q1	4.5	4–5	100.0	0.86	78.0	2.4×10^{-4}	1.00	C1 (mapping)
Q2	4.0	4–5	100.0	0.88	78.0	2.4×10^{-4}	1.00	C1/C2
Q3	4.0	3–4	66.7	0.88	36.0	3.9×10^{-3}	1.00	C2 (XAI)
Q4	4.5	4–5	100.0	0.86	78.0	2.4×10^{-4}	1.00	C2 (thresholds)
Q5	4.0	4–5	100.0	0.88	78.0	2.4×10^{-4}	1.00	C3 (HITL/threat)
Q6	5.0	4–5	100.0	0.87	78.0	2.4×10^{-4}	1.00	C4 (cost–benefit)
Q7	4.0	4–4	83.3	0.92	55.0	9.8×10^{-4}	1.00	C4/C5 (adoption)

Panel-level: grand mean 4.25 ± 0.58; 92.9 % of 84 ratings ≥ 4; 0 ratings < 3; tie-corrected $W = 0.36$ ($\chi^2_6 = 25.72$, $p = 2.5 \times 10^{-4}$); all items pass Holm at $\alpha = 0.01$; subgroup equivalence: Kruskal–Wallis $H = 0.14$, $p = 0.93$ (CISA vs. QSA vs. LA).

(Q3, Q7) correspond to the residual risks the framework explicitly manages through HITL escalation, the threat model, and the standardisation agenda (C3, C5). The significant Friedman differentiation ($\chi^2_6 = 25.72$, $p < 0.001$) confirms this prioritisation is systematic rather than noise, and the qualitative record supplies its mechanism. These findings align with the PRUMe AI empirical results [29] and the agentic system governance literature [28].

11 Recommendations for ISO Standardisation

Based on framework analysis, expert evaluation, and the deep-review corpus, we propose seven recommendations for ISO/IEC JTC 1/SC 27:

R1: Auditable Provenance Metadata. Amend ISO 19011:2018 to require cryptographic provenance metadata for all AI-generated audit evidence: model version, training data hash, inference timestamp, and explanation method. This ensures that AI-generated evidence can be independently verified and re-examined years after the original audit.

R2: Mandatory Explainability Thresholds. Establish minimum explainability quality thresholds ($Q(e) \geq 0.70$ per Eq. (1), $W \geq 0.90$ per Eq. (2)) for AI-generated evidence admissibility in certification audits. These thresholds should be published as normative requirements, not guidance.

R3: Model Governance Framework. Develop ISO/IEC 27001 guidance on AI model governance for audit tools: training data requirements, validation procedures, and drift monitoring obligations. This parallels the model risk management frameworks already required in financial services.

R4: AI Auditor Competency Standard. Extend ISO 27006:2023 competency requirements to include AI literacy, XAI interpretation, and adversarial robustness assessment skills. Auditors who cannot interpret SHAP attribution vectors cannot meaningfully review AI-generated evidence.

R5: Continuous Monitoring Assurance Standard. Develop a new ISO/IEC standard for continuous monitoring assurance, defining evidence sampling, aggregation, and reporting requirements for AI-generated compliance data.

R6: Cross-Framework Harmonisation. Establish a harmonised AI audit evidence standard applicable to ISO 27001:2022, PCI DSS v4.0.1, SOC 2, and future frameworks, eliminating the current situation where each framework implies different evidence standards.

R7: Ethical AI Audit Guidelines. Develop ISO guidance on ethical considerations in AI-assisted auditing: algorithmic fairness, privacy preservation, accountability chains, and conflict of interest disclosures when AI vendors also provide audit tooling.

12 Conclusion

The primary contribution of this study is conceptual, preceding its technical aspects: it delineates the conditions under which the output of an AI system may be regarded as admissible audit evidence rather than merely as useful analytics. Previous proposals for AI in auditing often stop at this point. A classifier may be accurate, but if its reasoning cannot be reconstructed, it remains a black-box assertion. In professional audit terms, such an assertion is inadmissible under ISO 19011:2018. This framework addresses that problem by defining evidence sufficiency as an explicit threshold, $Q(e) \geq 0.70$ (Eq.2). Together, these criteria shift the claim from "the model is confident" to "the finding is complete, traceable, and reproducible." The result is a revised view of auditor independence. Traditionally, independence is maintained through the separation of duties among people. In this framework, it must also be engineered into the audit process through provenance, human sign-off gates, and model governance.

This reconceptualisation treats independence as an architectural feature, not only as an organisational attribute. That shift allows continuous, large-scale assurance to coexist with the ISO 19011 methodology rather than replace it.

Our study shows that explanation *stability* is more important than explanation *accuracy* in audit contexts, as the same attribution method can show high concordance for tree ensembles ($W \approx 0.99$) but fall toward random for some deep model/explainer pairings ($W \approx 0.50$). Only stable explanations can support findings that remain defensible on re-examination, making explainability a reproducibility requirement rather than a presentation layer. Furthermore, expert scepticism focused not on the framework itself, but on unclear machine reasoning. Unsupervised anomaly detection was the only component below the favourable band, identifying the main residual risk and supporting human-in-the-loop escalation for such findings. Finally, practitioners strongly endorsed the $Q(e)$ threshold, with consistent support across CISA, QSA, and Lead-Auditor groups ($p = 0.93$), suggesting that a formal, auditable admissibility gate is not a barrier to adoption, but a condition for it.

These results highlight challenges the framework manages rather than removes. Full-population coverage reduces sampling risk but increases alert volume, making human triage capacity a key limit on continuous assurance. The audit system also becomes a high-value attack surface, so adversarial robustness and tamper-evident provenance are structural requirements rather than optional safeguards. The framework further assumes that auditors can interpret SHAP attributions and stability scores; without that capability, human review risks becoming a rubber-stamp process. Future work should therefore validate the reported efficiency and coverage gains through controlled, longitudinal live deployments, where false positives and human-gate costs can be measured directly. Further priorities include robustness certification, open implementations of the $Q(e)$ and W gates, and extension to related standards such as ISO/IEC 42001. As AI becomes part of the evidentiary basis for audits, the profession will need a harmonised standard for machine-generated evidence, ensuring that automation extends the auditor's reach without inheriting the auditor's authority.

Generative AI Use

The authors used generative AI tools solely for language refinement and formatting. All scientific content, analyses, interpretations, and conclusions were reviewed and verified by the authors, who take full responsibility for the final manuscript.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Appendix A

Complete AI Technique Mapping for All 93 ISO 27001:2022 Annex A Controls

Evidence Types, Quality Thresholds, and Human Review Requirements

A1 Evidence Quality Threshold $Q(e)$

The evidence quality metric from the main paper (Eq. 1) is:

$$Q(e) = \alpha \cdot \text{Completeness}(e) + \beta \cdot \text{Traceability}(e) + \gamma \cdot \text{Reproducibility}(e) \tag{A1}$$

where $\alpha = 0.40$, $\beta = 0.35$, $\gamma = 0.25$, and $\alpha + \beta + \gamma = 1$. The minimum acceptable threshold is $Q(e) \geq 0.70$.

A2 Human Review (HR) Codes

Code	Meaning
REQ	**Required** — Mandatory auditor sign-off before AI-generated evidence is included in the audit record. Applies to 80 % of all 93 controls.
REC	**Recommended** — Strong guidance for auditor review; automated evidence may be accepted with documented justification.
OPT	**Optional** — Automated evidence is sufficient for low-risk controls; auditor sampling is still advised.

A3 Row Shading and Risk Levels

Shading	Theme	Risk Badge
Organizational	A.5.x (37 controls)	**H** High **M** Medium **L** Low
People	A.6.x (8 controls)	Reflects audit risk of incorrect automated assessment
Physical	A.7.x (14 controls)	
Technological	A.8.x (34 controls)	

NEW denotes a control introduced in ISO 27001:2022 (not present in the 2013 edition). There are 11 such new controls.

A4 Theme 1: Organizational Controls (A.5.x) — 37 Controls

Organizational controls govern information security policies, roles, supplier relationships, incident management, and legal compliance. NLP and LLM are the primary AI techniques for this theme, with empirical F1 scores exceeding 0.88 for policy clause extraction on regulatory corpora [17, 19]. Graph Neural Networks address segregation-of-duties and supply chain mapping.

Table A1: ISO 27001:2022 Organizational Controls (A.5.x): AI Technique Mapping. Column headers: Control ID, Control Name, AI Technique(s), Evidence Type, Quality Notes (empirical metrics), $Q(e)$ threshold, Human Review (HR) requirement.

Control	Control Name	AI Technique(s)	Evidence Type	Quality Notes	$Q(e)$	HR
A.5.1	Policies for information security	NLP, LLM, GenAI	Policy completeness report; clause-to-control mapping; gap analysis	F1 $\geq$ 0.88 for clause extraction [19]; hallucination check required	0.82	**REQ**
A.5.2	Information security roles and responsibilities	NLP, RPA, PM	RACI matrix extract; role-assignment audit log; org chart diff	Validate against HR system; version-controlled output	0.78	**REQ**

Continued on next page. . .

Table A1 *(continued from previous page)*

Control	Control Name	AI Technique(s)	Evidence Type	Quality Notes	$Q(e)$	HR
A.5.3	Segregation of duties	GNN, ML, RPA	Access matrix conflict report; SoD violation graph; RBAC snapshot	Graph completeness critical; validate node/edge coverage; SoD violations = 0	0.80	**REQ**
A.5.4	Management responsibilities	NLP, GenAI	Management commitment statements; signed attestation records	NLP confidence $\geq$ 0.75 for intent classification; GenAI output auditor-verified	0.74	**REQ**
A.5.5	Contact with authorities	NLP, RPA	Contact register; communication log extract; authority mapping	Completeness check against regulatory register; metadata traceability	0.72	**REC**
A.5.6	Contact with special interest groups	NLP, RPA	Membership records; subscription logs; meeting minutes extract	Metadata traceability required; provenance hash per entry	0.70	**REC**
A.5.7 **NEW**	Threat intelligence	NLP, LLM, GNN, AD	STIX/TAXII feed parse report; threat actor correlation map; IOC coverage matrix	Real-time feed validation; provenance hash required; $F1 \geq 0.85$ for IOC classification	0.86	**REQ**
A.5.8	Information security in project management	NLP, PM, RPA	Project gate review records; security checkpoint logs; SDLC compliance report	Process variant analysis; conformance rate $\geq$ 90%; PM fitness score $\geq$ 0.80	0.76	**REC**
A.5.9	Inventory of information and other associated assets	ML, RPA, NLP	Asset inventory diff; classification coverage report; shadow IT detection log	Coverage completeness; auto-discovery false positive rate $\leq$ 5%; delta reconciliation	0.80	**REC**
A.5.10	Acceptable use of information and other associated assets	NLP, BA, ML	Policy acknowledgement records; usage anomaly flags; DLP policy alignment report	Behavioural baseline required; 30-day minimum window; anomaly precision $\geq$ 0.82	0.75	**REC**
A.5.11	Return of assets	RPA, PM	Asset return workflow trace; HR offboarding checklist completion log	Process conformance rate $\geq$ 95%; zero residual asset tolerance	0.74	**REC**
A.5.12	Classification of information	ML, NLP	Data classification scan report; label coverage matrix; reclassification log	Classifier precision $\geq$ 0.90; human review for borderline cases; F1 $\geq$ 0.88	0.82	**REC**
A.5.13	Labelling of information	ML, CV, RPA	Label presence audit report; automated label scan results; unlabelled asset list	Label detection recall $\geq$ 0.88; CV accuracy $\approx$ 0.88 [19]	0.78	**REC**
A.5.14	Information transfer	NLP, ML, AD	Transfer policy gap analysis; DLP event log; encryption compliance report	Encryption verification via API; DLP coverage $\geq$ 95%; AD precision $\geq$ 0.85	0.80	**REC**

Continued on next page...

Table A1 *(continued from previous page)*

Control	Control Name	AI Technique(s)	Evidence Type	Quality Notes	$Q(e)$	HR
A.5.15	Access control	ML, GNN, RPA	Access control matrix; privilege escalation log; RBAC compliance snapshot	Graph coverage complete; zero orphaned accounts; escalation detection AUC > 0.90	0.84	**REQ**
A.5.16	Identity management	ML, RPA, GNN	Identity lifecycle log; orphaned account report; provisioning audit trail	Zero stale identities tolerance; automated reconciliation; GNN coverage = 100%	0.82	**REQ**
A.5.17	Authentication information	ML, AD, RPA	Password policy compliance report; weak credential detection log; MFA coverage matrix	Credential scan precision ≥ 0.95; no plaintext storage; MFA coverage = 100%	0.86	**REQ**
A.5.18	Access rights	ML, GNN, RPA	Access rights review log; excessive privilege report; recertification audit trail	Quarterly recertification verified; SoD violations = 0; AUC > 0.90	0.84	**REQ**
A.5.19	Information security in supplier relationships	NLP, ML	Supplier risk score report; contract clause compliance map; questionnaire analysis	NLP clause extraction F1 $\geq$ 0.85; risk scoring cross-validated; coverage $\geq$ 90%	0.78	**REQ**
A.5.20	Addressing information security within supplier agreements	NLP, GenAI	Contract compliance report; security clause coverage matrix	Clause coverage $\geq$ 90%; GenAI output verified by auditor; provenance logged	0.76	**REQ**
A.5.21	Managing information security in the ICT supply chain	GNN, ML, NLP	Supply chain graph; vendor dependency map; SBOM compliance report	SBOM completeness; graph depth ≥ 3 tiers; GNN edge coverage = 100%	0.80	**REQ**
A.5.22	Monitoring, review and change management of supplier services	NLP, ML, AD	Supplier SLA monitoring report; service change log; anomaly alert summary	Continuous monitoring; SLA breach detection latency $\leq$ 15 min; AD recall ≥ 0.90	0.78	**REC**
A.5.23 **NEW**	Information security for use of cloud services	ML, NLP, AD	CSPM configuration report; cloud policy gap analysis; multi-cloud compliance matrix	CSPM API coverage $\geq$ 98%; misconfiguration detection F1 $\geq$ 0.90; multi-cloud validated	0.88	**REQ**
A.5.24	Information security incident management planning and preparation	NLP, PM	IRP completeness report; tabletop exercise log; escalation workflow trace	Plan coverage against NIST SP 800-61; conformance $\geq$ 90%; PM fitness $\geq$ 0.85	0.76	**REQ**
A.5.25	Assessment and decision on information security events	ML, AD, XAI	Event triage report; severity classification log; XAI attribution summary	XAI explanation quality $Q \geq 0.75$; false negative rate $\leq$ 2%; SHAP stability $W \geq 0.90$	0.82	**REQ**

Continued on next page...

Table A1 *(continued from previous page)*

Control	Control Name	AI Technique(s)	Evidence Type	Quality Notes	$Q(e)$	HR
A.5.26	Response to information security incidents	PM, RPA, NLP	Incident response workflow log; containment action trace; post-incident report	Response SLA adherence $\geq$ 95%; automated timeline reconstruction; PM conformance $\geq$ 0.90	0.80	**REQ**
A.5.27	Learning from information security incidents	NLP, ML, GenAI	Lessons-learned corpus; trend analysis report; recurring pattern detection log	Trend detection precision $\geq$ 0.80; GenAI summaries verified; version-controlled	0.74	**REC**
A.5.28	Collection of evidence	NLP, RPA, XAI	Chain-of-custody log; evidence integrity hash report; forensic collection audit trail	Cryptographic hash verification; tamper-evidence required; XAI provenance logged	0.90	**REQ**
A.5.29	Information security during disruption	NLP, PM, ML	BCP test log; recovery time compliance report; continuity gap analysis	RTO/RPO compliance verified; test coverage $\geq$ 80%; PM conformance $\geq$ 0.90	0.76	**REQ**
A.5.30 **NEW**	ICT readiness for business continuity	ML, PM, AD	ICT resilience test log; failover performance report; dependency availability map	Failover test automation; RTO compliance $\leq$ defined threshold; AD coverage = 100%	0.82	**REQ**
A.5.31	Legal, statutory, regulatory and contractual requirements	NLP, LLM	Regulatory requirement mapping; compliance obligation register; change monitoring log	Regulatory feed freshness $\leq$ 24 hours; coverage $\geq$ 95%; LLM output auditor-verified	0.78	**REQ**
A.5.32	Intellectual property rights	NLP, ML	IPR compliance scan; open-source licence audit report; licence conflict detection	SBOM + licence scanner; conflict rate = 0 for critical assets; coverage = 100%	0.80	**REQ**
A.5.33	Protection of records	ML, RPA, NLP	Records retention compliance report; access log; integrity verification results	Retention policy conformance = 100%; tamper-evidence required; integrity hash verified	0.84	**REQ**
A.5.34	Privacy and protection of PII	ML, NLP, AD	PII discovery report; data flow map; consent management audit log	PII detection recall $\geq$ 0.95; GDPR/PDPA alignment verified; AD false positive $\leq$ 3%	0.86	**REQ**
A.5.35	Independent review of information security	NLP, GenAI	Review coverage report; finding classification log; remediation tracking matrix	Scope completeness; GenAI findings verified by auditor; independence documented	0.74	**REQ**
A.5.36	Compliance with policies, rules and standards	NLP, ML, RPA	Policy compliance dashboard; deviation log; automated attestation records	Deviation rate $\leq$ 2%; continuous monitoring active; NLP F1 $\geq$ 0.88	0.80	**REC**

Continued on next page...

Table A1 *(continued from previous page)*

Control	Control Name	AI Technique(s)	Evidence Type	Quality Notes	$Q(e)$	HR
A.5.37	Documented operating procedures	NLP, RPA	Procedure currency report; version diff log; completeness gap analysis	Procedure freshness ≤ 12 months; coverage $\geq 95\%$; NLP completeness score ≥ 0.80	0.74	**REC**

A5 Theme 2: People Controls (A.6.x) — 8 Controls

People controls address human factors: screening, employment terms, security awareness, disciplinary processes, and remote working. Behavioural analytics (BA) and process mining (PM) are key techniques for this theme. Access revocation automation (RPA) achieves zero residual access with latency ≤ 4 hours.

Table A2: ISO 27001:2022 People Controls (A.6.x): AI Technique Mapping.

Control	Control Name	AI Technique(s)	Evidence Type	Quality Notes	$Q(e)$	HR
A.6.1	Screening	NLP, RPA, ML	Background check completion log; screening policy compliance report; gap analysis	Completeness $= 100\%$ for privileged roles; privacy-preserving processing; ML precision ≥ 0.90	0.80	**REQ**
A.6.2	Terms and conditions of employment	NLP, RPA	Signed agreement extract; clause compliance report; onboarding checklist log	Clause coverage $\geq 95\%$; signature verification; NLP F1 ≥ 0.85	0.78	**REQ**
A.6.3	Information security awareness, education and training	NLP, ML, RPA	Training completion matrix; assessment score log; curriculum gap analysis	Completion rate $= 100\%$; score threshold $\geq 80\%$; NLP gap detection F1 ≥ 0.82	0.82	**REC**
A.6.4	Disciplinary process	NLP, PM	Disciplinary workflow conformance log; case resolution timeline; escalation trace	Process conformance $\geq 95\%$; privacy-preserving anonymization; PM fitness ≥ 0.90	0.74	**REQ**
A.6.5	Responsibilities after termination or change of employment	RPA, PM, ML	Offboarding checklist completion log; access revocation confirmation; asset return trace	Zero residual access tolerance; revocation latency ≤ 4 hours; PM conformance $= 100\%$	0.86	**REQ**
A.6.6	Confidentiality or non-disclosure agreements	NLP, RPA	NDA coverage report; signature log; expiry monitoring alerts	Coverage $= 100\%$ for data-handling roles; expiry ≤ 0 days overdue; NLP precision ≥ 0.92	0.80	**REQ**
A.6.7	Remote working	ML, AD, BA	Remote access policy compliance report; VPN usage anomaly log; endpoint security status	Endpoint compliance $\geq 98\%$; anomaly detection active; BA baseline ≥ 30 days	0.78	**REC**
A.6.8	Information security event reporting	NLP, ML, RPA	Event reporting completeness log; submission timeline analysis; under-reporting detection	Reporting rate baseline; under-reporting flag $\leq 5\%$ deviation; ML recall ≥ 0.88	0.76	**REC**

A6 Theme 3: Physical Controls (A.7.x) — 14 Controls

Physical controls govern secure areas, physical access, equipment security, and environmental protection. Computer vision (CV) is the primary AI technique, with multimodal CV+LLM fusion achieving accuracy ≈ 0.88 on physical intrusion detection datasets [19]. IoT sensor integration and RPA support automated evidence collection.

Table A3: ISO 27001:2022 Physical Controls (A.7.x): AI Technique Mapping.

Control	Control Name	AI Technique(s)	Evidence Type	Quality Notes	$Q(e)$	HR
A.7.1	Physical security perimeters	CV, IoT, RPA	Perimeter breach detection log; sensor coverage map; access attempt report	CV detection accuracy $\geq$ 0.88 [19]; IoT sensor uptime $\geq$ 99.5%; false alarm rate $\leq$ 2%	0.82	**REQ**
A.7.2	Physical entry	CV, ML, RPA	Access control log; tailgating detection report; entry anomaly summary	Tailgating detection recall $\geq$ 0.85; badge-reader API integration verified	0.80	**REQ**
A.7.3	Securing offices, rooms and facilities	CV, IoT	Facility security scan; occupancy anomaly log; environmental sensor report	CV+IoT fusion accuracy $\approx$ 0.88; coverage map completeness = 100%	0.78	**REC**
A.7.4	Physical security monitoring	CV, AD, ML	CCTV coverage report; anomaly event log; XAI-annotated clip archive	24/7 monitoring coverage; CCTV uptime $\geq$ 99%; AD precision $\geq$ 0.88	0.84	**REQ**
A.7.5	Protecting against physical and environmental threats	IoT, ML, AD	Environmental threat sensor log; climate anomaly alert; flood/fire detection report	Sensor calibration verified; alert latency $\leq$ 60 seconds; AD recall $\geq$ 0.95	0.80	**REC**
A.7.6	Working in secure areas	CV, BA, ML	Secure area access log; behavioural anomaly report; visitor escort compliance log	BA baseline $\geq$ 30 days; escort compliance = 100%; CV precision $\geq$ 0.85	0.78	**REQ**
A.7.7	Clear desk and clear screen	CV, RPA	Clear desk policy compliance scan; screen lock compliance log; violation summary	CV scan frequency $\geq$ daily; violation rate $\leq$ 1%; RPA evidence chain-of-custody	0.74	**REC**
A.7.8	Equipment siting and protection	CV, IoT, ML	Equipment placement compliance report; environmental condition log	Siting policy conformance = 100%; IoT sensor coverage $\geq$ 98%; ML classification F1 $\geq$ 0.85	0.76	**REC**
A.7.9	Security of assets off-premises	ML, RPA, CV	Off-premises asset register; remote tracking log; return condition report	Asset location tracking accuracy $\geq$ 95%; return verification automated	0.78	**REC**

Table A3 *(continued)*

Control	Control Name	AI Technique(s)	Evidence Type	Quality Notes	$Q(e)$	HR
A.7.10	Storage media	ML, RPA, CV	Media inventory; secure disposal log; classification label scan	Disposal verification = 100%; CV label detection recall $\geq$ 0.88; chain-of-custody intact	0.80	**REQ**
A.7.11	Supporting utilities	IoT, ML, AD	Utility monitoring log; power/cooling anomaly alert; UPS test report	Utility uptime $\geq$ 99.9%; AD alert latency $\leq$ 30 seconds; sensor calibration verified	0.82	**REC**
A.7.12	Cabling security	CV, IoT, RPA	Cabling integrity scan; tamper-detection log; patch panel audit report	CV scan coverage = 100% of critical cabinets; tamper-detection sensitivity validated	0.74	**REC**
A.7.13	Equipment maintenance	ML, RPA, IoT	Maintenance schedule compliance log; predictive failure alert; service record extract	Predictive maintenance precision $\geq$ 0.85; schedule conformance = 100%	0.76	**REC**
A.7.14	Secure disposal or re-use of equipment	ML, RPA, CV	Disposal verification report; data sanitisation certificate; chain-of-custody log	Sanitisation verification = 100%; CV label confirmation; disposal policy conformance	0.82	**REQ**

A7 Theme 4: Technological Controls (A.8.x) — 34 Controls

Technological controls are the largest theme (34 controls) and the most amenable to AI automation. ML anomaly detection achieves AUC > 0.90 for log-based controls [17]. XAI (SHAP/LIME) provides audit-grade explanations with Kendall's $W = 0.9912$ stability for XGBoost models. Deep neural networks with XAI achieve continuous PCI DSS Requirement 10 monitoring.

Table A4: ISO 27001:2022 Technological Controls (A.8.x): AI Technique Mapping.

Control	Control Name	AI Technique(s)	Evidence Type	Quality Notes	$Q(e)$	HR
A.8.1	User endpoint devices	ML, AD, RPA	Endpoint compliance report; configuration drift log; EDR telemetry summary	Endpoint compliance $\geq$ 98%; drift detection latency $\leq$ 15 min; AD AUC > 0.90	0.82	**REC**
A.8.2	Privileged access rights	ML, GNN, XAI	Privileged account audit; escalation detection log; XAI attribution report	Zero excessive privilege tolerance; XAI stability $W \geq 0.90$; GNN coverage = 100%	0.86	**REQ**
A.8.3	Information access restriction	ML, GNN, RPA	Access restriction compliance report; policy-to-permission gap analysis	Access model completeness; GNN edge coverage = 100%; ML precision $\geq$ 0.90	0.82	**REQ**

Continued on next page...

Table A4 *(continued)*

Control	Control Name	AI Technique(s)	Evidence Type	Quality Notes	$Q(e)$	HR
A.8.4	Access to source code	ML, RPA, NLP	Source code access log; repository permission audit; developer access matrix	Repository coverage = 100%; access review frequency ≥ quarterly	0.80	**REQ**
A.8.5	Secure authentication	ML, AD, BA	Authentication event log; MFA coverage report; weak auth detection summary	MFA coverage = 100% for privileged accounts; AD precision ≥ 0.92; BA baseline ≥ 30 days	0.86	**REQ**
A.8.6	Capacity management	ML, AD, IoT	Capacity utilisation report; resource anomaly alert; trend forecast log	Forecast accuracy MAPE ≤ 10%; AD alert latency ≤ 5 min; coverage = 100%	0.78	**REC**
A.8.7	Protection against malware	ML, AD, XAI	Malware detection log; AV/EDR coverage report; XAI-annotated threat summary	AV/EDR coverage = 100%; ML detection AUC > 0.95; XAI attribution for each alert	0.88	**REQ**
A.8.8	Management of technical vulnerabilities	ML, NLP, RPA	Vulnerability scan report; patch compliance matrix; CVE mapping log	Patch compliance ≥ 98% within SLA; CVE coverage = 100%; ML severity scoring F1 ≥ 0.88	0.84	**REQ**
A.8.9 **NEW**	Configuration management	ML, RPA, AD	Configuration baseline diff; drift detection log; CIS benchmark compliance report	Baseline coverage = 100%; drift detection latency ≤ 10 min; AD precision ≥ 0.90	0.86	**REQ**
A.8.10 **NEW**	Information deletion	ML, RPA, NLP	Data deletion verification report; retention policy compliance log; PII purge certificate	Deletion verification = 100%; PII recall ≥ 0.95; chain-of-custody intact	0.84	**REQ**
A.8.11 **NEW**	Data masking	ML, NLP, RPA	Data masking coverage report; PII detection log; masking effectiveness test results	PII detection recall ≥ 0.95; masking effectiveness = 100%; false negative rate ≤ 1%	0.86	**REQ**
A.8.12 **NEW**	Data leakage prevention	ML, AD, NLP	DLP event log; data exfiltration detection report; policy alignment matrix	DLP coverage ≥ 98%; false positive rate ≤ 3% [17]; AD AUC > 0.90	0.86	**REQ**
A.8.13	Information backup	ML, RPA, AD	Backup completion log; restore test results; backup integrity verification report	Backup success rate = 100%; restore test frequency ≥ monthly; AD anomaly detection active	0.82	**REC**
A.8.14	Redundancy of information processing facilities	ML, IoT, AD	Redundancy test log; failover performance report; availability SLA compliance	Failover test automation; availability ≥ 99.9%; AD latency ≤ 60 seconds	0.80	**REC**

Continued on next page...

Table A4 *(continued)*

Control	Control Name	AI Technique(s)	Evidence Type	Quality Notes	$Q(e)$	HR
A.8.15	Logging	AD, ML, NLP	Log coverage report; log integrity verification; log retention compliance matrix	Log coverage = 100% of critical systems; integrity hash verified; retention policy = 100%	0.84	**REC**
A.8.16 **NEW**	Monitoring activities	AD, ML, XAI	SIEM monitoring coverage report; anomaly detection log; XAI alert attribution summary	Full-population log analysis; AD AUC > 0.90; XAI stability $W \geq 0.90$	0.88	**REQ**
A.8.17	Clock synchronisation	RPA, ML	NTP compliance report; clock drift detection log; synchronisation audit trail	Clock drift ≤ 1 second for audit-critical systems; NTP coverage = 100%	0.76	**OPT**
A.8.18	Use of privileged utility programs	ML, RPA, XAI	Privileged utility usage log; authorisation compliance report; XAI attribution summary	Usage authorisation = 100%; XAI explanation per privileged action; audit trail complete	0.82	**REQ**
A.8.19	Installation of software on operational systems	ML, RPA, NLP	Software installation log; authorised software list compliance; change ticket mapping	Compliance = 100%; unauthorised installation detection precision ≥ 0.95	0.84	**REQ**
A.8.20	Networks security	ML, AD, GNN	Network segmentation compliance report; firewall rule analysis; traffic anomaly log	Segmentation verification via API; AD AUC > 0.90; GNN topology coverage = 100%	0.84	**REQ**
A.8.21	Security of network services	ML, NLP, AD	Network service compliance report; SLA monitoring log; security clause coverage	SLA monitoring continuous; NLP clause extraction F1 ≥ 0.85; AD coverage = 100%	0.80	**REC**
A.8.22	Segregation of networks	ML, GNN, RPA	Network segregation map; VLAN compliance report; cross-segment traffic log	GNN topology coverage = 100%; zero cross-segment violations; ML classification F1 ≥ 0.90	0.84	**REQ**
A.8.23 **NEW**	Web filtering	ML, NLP, AD	Web filter coverage report; blocked category compliance log; policy alignment matrix	Category classification accuracy ≥ 0.92; coverage = 100%; AD false positive $\leq 2\%$	0.82	**REC**
A.8.24	Use of cryptography	ML, RPA, NLP	Cryptography compliance report; key management audit; cipher suite coverage matrix	Cipher suite compliance = 100%; key rotation verified; NLP policy alignment F1 ≥ 0.85	0.84	**REQ**

Continued on next page...

Table A4 *(continued)*

Control	Control Name	AI Technique(s)	Evidence Type	Quality Notes	$Q(e)$	HR
A.8.25	Secure development life cycle	NLP, ML, RPA	SDLC security gate compliance log; security requirement coverage report; SAST/DAST results	Security gate conformance $\geq$ 95%; SAST/DAST coverage = 100%; NLP requirement extraction F1 $\geq$ 0.85	0.82	**REQ**
A.8.26	Application security requirements	NLP, ML, RPA	Security requirement traceability matrix; SAST finding report; requirement coverage log	Requirement coverage = 100%; SAST precision $\geq$ 0.88; NLP extraction F1 $\geq$ 0.85	0.80	**REQ**
A.8.27	Secure system architecture and engineering principles	NLP, ML, GenAI	Architecture review report; security principle compliance matrix; design gap analysis	Principle coverage = 100%; GenAI output auditor-verified; NLP F1 $\geq$ 0.85	0.78	**REQ**
A.8.28 **NEW**	Secure coding	ML, NLP, RPA	SAST/DAST coverage report; secure coding standard compliance log; vulnerability density metric	SAST coverage = 100%; critical CVE density = 0; ML severity scoring F1 $\geq$ 0.88	0.86	**REQ**
A.8.29	Security testing in development and acceptance	ML, RPA, NLP	Security test coverage report; penetration test result extract; acceptance gate compliance log	Test coverage = 100%; pen test results verified by auditor; NLP finding extraction F1 $\geq$ 0.85	0.82	**REQ**
A.8.30	Outsourced development	NLP, ML, RPA	Outsourced code review report; supplier security assessment; SBOM completeness log	SBOM completeness = 100%; supplier assessment coverage = 100%; NLP clause F1 $\geq$ 0.85	0.80	**REQ**
A.8.31	Separation of development, test and production environments	ML, RPA, AD	Environment separation compliance report; cross-environment data flow log; configuration diff	Zero production data in test/dev; AD detection precision $\geq$ 0.90; separation verified via API	0.84	**REQ**
A.8.32	Change management	PM, ML, NLP	Change management workflow log; unauthorised change detection report; CAB compliance matrix	PM conformance $\geq$ 95%; unauthorised change detection precision $\geq$ 0.90; NLP classification F1 $\geq$ 0.85	0.82	**REQ**
A.8.33	Test information	ML, RPA, NLP	Test data classification report; PII anonymisation verification log; test data governance report	PII anonymisation = 100%; test data classification precision $\geq$ 0.90; chain-of-custody intact	0.82	**REQ**
A.8.34	Protection of information systems during audit testing	ML, RPA, NLP	Audit tool isolation log; read-only access verification; audit impact assessment report	Isolation = 100% during audit; read-only enforcement verified; impact assessment completed	0.80	**REQ**

A8 Deep Dive: 11 New ISO 27001:2022 Controls

The 2022 revision introduced 11 new controls not present in ISO 27001:2013. These controls specifically address modern threat landscapes (threat intelligence, cloud security, secure coding) and are highly amenable to AI-assisted auditing.

Table A5: Deep Dive: 11 New ISO 27001:2022 Controls with AI Mapping and Empirical Benchmarks.

Control	Control Name	Primary AI	Empirical Benchmark	$Q(e)$	HR
A.5.7	Threat intelligence	NLP, LLM, GNN, AD	IOC classification F1 ≥ 0.85; STIX/TAXII parse latency $\leq$ 5s	0.86	**REQ**
A.5.23	Information security for use of cloud services	ML, NLP, AD	CSPM misconfiguration F1 ≥ 0.90; multi-cloud coverage $\geq 98\%$	0.88	**REQ**
A.5.30	ICT readiness for business continuity	ML, PM, AD	Failover automation; RTO compliance $\leq$ defined threshold; PM conformance ≥ 0.90	0.82	**REQ**
A.8.9	Configuration management	ML, RPA, AD	Drift detection latency $\leq$ 10 min; CIS benchmark compliance $= 100\%$	0.86	**REQ**
A.8.10	Information deletion	ML, RPA, NLP	PII purge recall ≥ 0.95; deletion verification $= 100\%$; chain-of-custody intact	0.84	**REQ**
A.8.11	Data masking	ML, NLP, RPA	PII detection recall ≥ 0.95; masking effectiveness $= 100\%$; false negative $\leq 1\%$	0.86	**REQ**
A.8.12	Data leakage prevention	ML, AD, NLP	DLP false positive $\leq 3\%$ [17]; exfiltration detection AUC > 0.90	0.86	**REQ**
A.8.16	Monitoring activities	AD, ML, XAI	Full-population log coverage; AD AUC > 0.90; XAI stability $W \geq 0.90$	0.88	**REQ**
A.8.23	Web filtering	ML, NLP, AD	Category classification accuracy ≥ 0.92; false positive $\leq 2\%$; coverage $= 100\%$	0.82	**REC**
A.8.28	Secure coding	ML, NLP, RPA	SAST coverage $= 100\%$; critical CVE density $= 0$; severity scoring F1 ≥ 0.88	0.86	**REQ**
A.6.8	Information security event reporting (enhanced)	NLP, ML, RPA	Under-reporting detection; reporting rate baseline; ML recall ≥ 0.88	0.76	**REC**

Appendix B

Expert Evaluation of the AI-Assisted Auditing Framework

Structured Interview Study with Certified Audit Practitioners ($n = 12$)

B.1 Study Design and Ethics

The expert evaluation followed a *structured interview* methodology grounded in ISO 19011:2018 Clause 7.2 competence criteria and validated against the Technology Acceptance Model. Ethical approval was obtained from the authors' institutional review board (Ref. LR2026009) prior to recruitment. All participants provided written informed consent covering: (i) voluntary participation and right to withdraw without penalty; (ii) anonymisation of individual responses; (iii) use of aggregated data solely for academic publication; and (iv) data retention for five years in encrypted storage.

B.2 Participant Recruitment

Twelve ($n = 12$) certified practitioners were recruited through three professional networks between January and March 2025:

- **ISACA membership directory** – four Certified Information Systems Auditors (CISA) with ≥ 5 years of IS audit experience.
- **PCI SSC Qualified Security Assessor (QSA) community** – four QSAs holding current PCI DSS v4.0 assessor certification.
- **ISO 27001 Lead Auditor network** – four practitioners holding IRCA-registered ISO 27001:2022 Lead Auditor certification.

Inclusion criteria: (a) active auditing practice within the preceding 12 months; (b) familiarity with at least one of ISO 27001, PCI DSS, or SOC 2; (c) no prior exposure to the prototype system. No financial incentives were offered; participants received a summary of aggregate findings.

B.3 Instrument: Structured Interview Questionnaire

Each 60-minute session combined a live prototype demonstration (30 min) with a structured questionnaire (30 min). The instrument comprised two parts: Part I (background), Part II (seven-item Likert evaluation).

Part I – Background (open-ended):

B1. How many IS/compliance audits have you conducted in the past 12 months?
B2. Which standards do you audit against most frequently?
B3. Describe your current use (if any) of automated tools during fieldwork.

Part II – Likert Evaluation (1 = Strongly Disagree, 5 = Strongly Agree):

Q1. The AI-assisted workflow (Fig. 1) accurately maps evidence artefacts to the relevant ISO 27001:2022 / PCI DSS v4.0.1 controls.
Q2. The BERT policy clause extractor (Appendix B.1) produces output that meets the evidential completeness requirements of ISO 19011:2018 Clause 6.4.
Q3. The Isolation Forest anomaly alerts (Appendix B.2) are sufficiently explainable to satisfy a PCI DSS v4.0.1 Req. 10 finding.
Q4. The $Q(e)$ evidence quality threshold (≥ 0.70) is an appropriate acceptance criterion for automated audit evidence.
Q5. The human-in-the-loop (HITL) escalation mechanism (Section IV-C) adequately preserves auditor professional judgement as required by ISO 27006:2023 Clause 9.
Q6. The overall framework reduces audit fieldwork effort without compromising the rigour of findings.
Q7. You would recommend adoption of this framework in a real-world external audit engagement.

B.4 Raw Scores and Descriptive Statistics

Table B1 presents individual Likert scores for all $n = 12$ participants across seven items, together with per-item mean ($\bar{x}$), standard deviation (σ), and 95% confidence interval (CI).

B.5 Statistical Analysis

Inter-rater reliability was assessed using Kendall's coefficient of concordance W (Eq. 2 in the main paper). For the seven-item instrument across 12 raters:

$$W = \frac{12\,S}{m^2(n^3 - n) - m\sum_j T_j} = \frac{12 \times 1{,}052.5}{144 \times 336 - 12 \times 1{,}086} = 0.357$$

where $S = 1{,}052.5$ is the sum of squared deviations of rank totals, $m = 12$ raters, $n = 7$ items, and $\sum_j T_j = 1{,}086$ is the tie-correction term ($T_j = \sum(t^3 - t)$ over tied groups within rater j), which is substantial for a 5-point Likert instrument. The result $W = 0.36$ indicates *moderate agreement* ($p < 0.001$, $\chi^2_6 = 25.72$), consistent with expert panels evaluating novel technology [43]. Item Q3 (explainability of anomaly alerts) recorded the lowest mean (3.67), motivating the HITL escalation design (Section IV-C). Items Q1, Q4, and Q6 recorded the highest agreement ($\bar{x} \geq 4.50$), confirming that control-mapping accuracy and effort reduction are the framework's strongest perceived contributions.

A one-sample *exact* Wilcoxon signed-rank test confirmed that all seven item medians are significantly greater than the neutral midpoint of 3 at $p < 0.01$ ($p = 2.4 \times 10^{-4}$ for Q1–Q2 and Q4–Q6; $p = 9.8 \times 10^{-4}$ for Q7; $p = 3.9 \times 10^{-3}$ for Q3), supporting the hypothesis that certified auditors perceive the framework as useful and methodologically sound.

B.6 Extended Statistical Analysis

To provide deeper numerical support for the framework's validity claims, the raw score matrix of Table B1 was subjected to four additional analyses: (i) non-parametric location and dispersion measures robust to the ordinal Likert scale; (ii) within-item inter-rater agreement using the r_{WG} index (uniform-null expected variance $\sigma^2_E = (5^2 - 1)/12 = 2.0$); (iii) exact one-sample Wilcoxon signed-rank tests against the neutral midpoint (3) with matched-pairs rank-biserial effect sizes (r_{rb}) and Holm–Bonferroni correction for the seven-test

Table B1 Expert Evaluation: Raw Likert Scores ($n = 12$, Scale 1–5)

Item	Participant (P1–P12)												$\bar{x}$	σ	95% CI
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12			
Q1	5	4	5	4	5	4	5	4	5	4	5	4	4.50	0.52	[4.17, 4.83]
Q2	4	4	5	4	4	5	4	4	5	4	4	5	4.33	0.49	[4.02, 4.64]
Q3	4	3	4	4	4	3	4	3	4	4	3	4	3.67	0.49	[3.36, 3.98]
Q4	5	4	4	5	4	5	4	5	4	5	4	5	4.50	0.52	[4.17, 4.83]
Q5	4	5	4	4	5	4	5	4	4	5	4	4	4.33	0.49	[4.02, 4.64]
Q6	5	4	5	4	5	5	4	5	4	5	4	5	4.58	0.51	[4.26, 4.91]
Q7	4	4	4	3	4	4	4	4	3	4	4	4	3.83	0.39	[3.59, 4.08]
Overall mean usefulness													**4.25 / 5.00**		

family at $\alpha = 0.01$; and (iv) a certification-subgroup robustness check.

Consensus and directionality. Across all $12 \times 7 = 84$ individual ratings, 92.9 % were favourable (≥ 4) and *zero* ratings fell below the neutral midpoint; no participant disagreed with any item. Consequently every non-zero difference from the midpoint was positive, yielding the maximum matched-pairs rank-biserial effect size ($r_{rb} = 1.00$) on all seven items. Five of seven items (Q1, Q2, Q4, Q5, Q6) achieved unanimous favourable consensus (100 % of ratings ≥ 4).

Inter-rater agreement. Within-item agreement was uniformly strong ($r_{WG} = 0.86$–0.92, all above the conventional 0.70 acceptability threshold), indicating that the panel converged on a shared assessment of each framework component rather than averaging out divergent views. Between-rater dispersion was correspondingly narrow: individual rater means ranged only from 4.00 to 4.43.

Concordance across items. The tie-corrected concordance reported in Appendix B.5 ($W = 0.36$, $\chi^2_6 = 25.72$, $p = 2.5 \times 10^{-4}$) is equivalent to a Friedman test confirming that item means differ reliably: the panel consistently ranked effort reduction (Q6), control mapping (Q1), and the $Q(e)$ threshold (Q4) highest, and anomaly-alert explainability (Q3) lowest, a systematic rather than noise-driven prioritisation that directly motivates the HITL escalation design (Section IV-C).

Significance under multiplicity control. All seven exact Wilcoxon tests remain significant after Holm–Bonferroni correction at the stringent $\alpha = 0.01$ family level (largest adjusted comparison: Q3, $p = 0.0039$ vs. Holm threshold 0.0100), strengthening the per-item claims of Appendix B.5.

Robustness across certification groups. Partitioning raters by recruitment stratum (CISA: P1–P4; QSA: P5–P8; ISO 27001 Lead Auditor: P9–P12), subgroup grand means were statistically indistinguishable (CISA 4.21 ± 0.57, QSA 4.29 ± 0.60, LA 4.25 ± 0.59; Kruskal–Wallis on rater means $H = 0.14$, $p = 0.93$; per-item $H \leq 1.10$, all $p \geq 0.58$). The framework's perceived value therefore generalises across the three assurance traditions rather than reflecting a single certification community's perspective.

Table 8 in Section 10 consolidates the extended statistics and maps each instrument item to the paper's stated contributions (C1–C5, Section 1).

References

1. ISO/IEC, *ISO 19011:2018 Guidelines for Auditing Management Systems*. International Organization for Standardization, Geneva, Switzerland, 3rd edn. (2018). [Online]. Available: `https://www.iso.org/standard/70017.html`
2. ISO/IEC, *ISO/IEC 27006:2023 Requirements for Bodies Providing Audit and Certification of Information Security Management Systems*. International Organization for Standardization, Geneva, Switzerland, 3rd edn. (2023). [Online]. Available: `https://www.iso.org/standard/83607.html`
3. M. Vasarhelyi, M. Alles, A. Kogan, J. Emerg. Technol. Account. **1**(1), 1 (2004)
4. D. Searcy, J. Woodroof, The CPA Journal **73**(5), 46 (2003)
5. S. Goundar, I. Gondal, J. Cybersecur. Priv. **5**, 59 (2025). DOI 10.3390/jcp5030059
6. P.R. Nangi, C.K.R.N. Obannagari, International Journal of Emerging Trends in Computer Science and Information Technology (2025). DOI 10.63282/3050-9246.ijetcsit-v6i2p112
7. M. Antunes, M. Maximiano, R. Gomes, Applied Sciences (2022). DOI 10.3390/app12094102
8. M. Gauthier, N. Brender, Managerial Auditing Journal (2021). DOI 10.1108/maj-12-2019-2513
9. M.T. Minhas, T.F. Minhas, S.A. Masroor, M.A. Naeem, A.Q. Raja, The Critical Review of Social Sciences Studies (2026). DOI 10.59075/6jgyny13
10. R. Sabillon, Cyber Security Auditing, Assurance, and Awareness Through CSAM and CATRAM (2020). DOI 10.4018/978-1-7998-4162-3.ch007
11. A. Folorunso, T. Adewumi, A. Adewa, R. Okonkwo, T.N. Olawumi, Global Journal of Engineering and Technology Advances (2024). DOI 10.30574/gjeta.2024.21.1.0193

12. O. Senturk, Journal of Accounting, Finance and Auditing Studies (2025). DOI 10.56578/jafas110105
13. B.D. Smart, D. Jooda, International Journal of Advanced Multidisciplinary Research and Studies (2023). DOI 10.62225/2583049x.2023.3.6.6229
14. R.Y. Seetharamarao, in *2023 16th International Conference on Developments in eSystems Engineering (DeSE)* (2023), pp. 623–629. DOI 10.1109/dese60595.2023.10469536
15. M. Alles, G. Brennan, A. Kogan, M. Vasarhelyi, Int. J. Account. Inf. Syst. **7**(2), 137 (2006)
16. A. Rezaee, R. Elam, A. Sharbatoghlie, Managerial Auditing J. **16**(3), 150 (2001)
17. S.T. Gandhi, Int. J. Adv. Res. Comput. Sci. Tech. **6**(5), 1 (2023). DOI 10.15662/ijarcst.2023.0605004
18. S. Belidhe, Explainable AI and deep neural networks for continuous PCI DSS compliance monitoring. Tech. rep., Zenodo (2024). DOI 10.5281/zenodo.14514056
19. E. Cadet, E.D. Etim, I.A. Essien, et al., Int. J. Scientific Res. Scholarly Humanities (2024). DOI 10.32628/ijsrssh242560
20. S. Mahlangu, R. Moosa, Academic Journal of Interdisciplinary Studies (2023). DOI 10.36941/ajis-2023-0097
21. H. Kumar, International Journal of Scientific Research in Engineering and Management (2025). DOI 10.55041/ijsrem49485
22. A. Rafique, S.A. Shahani, Annual Methodological Archive Research Review (2025). DOI 10.63075/z8wh7360
23. M. Zelmati, Z. Oulqaid, A. Elouadi, in *2023 10th International Conference on Wireless Networks and Mobile Communications (WINCOM)* (2023), pp. 1–7. DOI 10.1109/wincom59760.2023.10322925
24. C. Brown, D. Williams, Comput. Security **138**, 103 (2024)
25. R. Brown, S. Taylor, M. Davis, Auditing: A J. Practice Theory **43**(1), 45 (2024)
26. ISO/IEC, *ISO/IEC 27001:2022 Information Security, Cybersecurity and Privacy Protection — Information Security Management Systems — Requirements*. International Organization for Standardization, Geneva, Switzerland, 3rd edn. (2022)
27. PCI Security Standards Council, *Payment Card Industry Data Security Standard v4.0.1*. PCI SSC, Wakefield, MA, USA (2024). [Online]. Available: `https://www.pcisecuritystandards.org`
28. F. Bousetouane, arXiv preprint arXiv:2501.00881 (2025)
29. A.S. Bezerra, L.C. Lopes, in *Proc. 1st Congress Seven Multidisciplinary Studies* (2025). DOI 10.56238/sevened2025.029-098
30. MindBridge. Audit and assurance (2026). URL `https://www.mindbridge.ai/industries/audit-assurance/`
31. O. Ilori, Finance & Accounting Research Journal **5**(12), 502 (2023)
32. H.B. Hazar, in *Ethics and Sustainability in Accounting and Finance, Volume II* (Springer, 2020), pp. 253–268
33. G.R. Antonio, Journal of Entrepreneurship (2022). DOI 10.56943/joe.v1i3.157
34. D. Montalvo, L. Quishpe, Northern Reviews on Algorithmic Research, Theoretical Computation, and Complexity **9**(12), 1 (2024)
35. A. Zahoor, W. Abbasi, M.Z. Babar, A. Aljohani, Scientific Reports **15**(1), 36586 (2025)
36. S.R. Wibowo, A. Wibowo, Jurnal Akuntansi dan Bisnis (2025). DOI 10.51903/jiab.v5i2.931
37. G. Falco, B. Shneiderman, J.M. Badger, R. Carrier, A. Dahbura, D. Danks, M. Eling, A.E. Goodloe, J. Gupta, C. Hart, M. Jirotka, H. Johnson, C. LaPointe, A. Llorens, A.K. Mackworth, C. Maple, S.E. Pálsson, F.A. Pasquale, A. Winfield, Z.K. Yeong, Nature Machine Intelligence **3**, 566 (2021). DOI 10.1038/s42256-021-00370-7
38. I.A. Essien, J.O. Ajayi, E.D. Erigha, et al., Eng. Technol. J. **10**(9) (2025). DOI 10.47191/etj/v10i09.26
39. [Author], [Author], Eng. Technol. J. (2026)
40. A. Rehman, T. Saba, M.Z. Khan, R. Damaševičius, S.A. Bahaj, Security and communication Networks **2022**(1), 8383461 (2022)
41. S. Dey, A. Das, International Journal of Business Process Integration and Management **9**(3), 220 (2019)
42. F.T. Liu, K.M. Ting, Z.H. Zhou, in *Proc. 8th IEEE Int. Conf. Data Mining (ICDM)* (Pisa, Italy, 2008), pp. 413–422
43. S.M. Lundberg, S.I. Lee, in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 30 (2017), vol. 30
44. S.O. Olatunji, A. Almuhaidib, A.S. Alhuwaishil, et al., in *Proc. Advances in Intelligent Syst.* (Springer, 2025). DOI 10.1007/978-981-97-8588-9_3
45. I. Guides, (2011)
46. S. Srinivas, B. Kirk, J. Zendejas, M. Espino, M. Boskovich, A. Bari, K. Dajani, N. Alzahrani, Journal of Cybersecurity and Privacy **5**(4), 95 (2025)
47. [Author], J. Tax. Fisc. Econ. (2025). DOI 10.61784/jtfe3056
48. N. Papernot, P. McDaniel, I. Goodfellow, S. Jha, Z.B. Celik, A. Swami, in *Proc. ACM Asia Conf. Comput. Commun. Security (AsiaCCS)* (2017), pp. 506–519
49. B. Biggio, F. Roli, Pattern Recognit. **84**, 317 (2018)
50. O. Kayali. Robust anomaly detection in payment systems: Adversarial resilience and real-time adaptation. ResearchGate (2025)
51. Expert Evaluation Study. Structured evaluation with 12 certified auditors (CISA, QSA, ISO 27001 lead auditors), informed consent protocol (2026). Full protocol in Appendix **??**