ACIG
APPLIED CYBERSECURITY & INTERNET GOVERNANCE

NASK

Auditing AI Governance under ISO/IEC 42001 Certification: EU AI Act, Product Liability Directive, and Multi-Dimensional Security and Liability Considerations

Natalija Parlov | Dr. Franjo Tudjman Defense and Security University, Zagreb, Croatia | ORCID: 0000-0002-8468-7465

Mihael Mišo Mudrić | Faculty of Law, University of Zagreb, Zagreb, Croatia | ORCID: 0000-0001-5364-4385

Blanka Mateša | Faculty of Law, University of Split, Split, Croatia | ORCID: 0000-0002-7591-7484

Abstract

Artificial intelligence (AI) governance increasingly relies on management systems, independent assurance, and risk and security processes, yet the evidentiary and legal significance of artificial intelligence management system (AIMS) certification remains insufficiently defined. This paper examines what International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) 42001 third-party audits assess, how ISO/IEC 42006 competence requirements affect the credibility and limits of audit evidence, and when AIMS documentation and certification outputs may be relevant in civil, criminal, professional-liability, and product-liability assessments. The study applies a conceptual, standards-interpretive, and doctrinal legal design, analysing standards, the European Union (EU) Artificial Intelligence Act, the revised

Received: 13.06.2026

Accepted: 03.08.2026

Published: 24.08.2026

Cite this article as:
N. Parlov, M.M. Mudrić, B. Mateša "Auditing AI governance under ISO/IEC 42001 certification: EU AI Act, Product Liability Directive, and multi-dimensional security and liability considerations," ACIG, vol. 5, no. 1, 2026, doi: 10.60097/ACIG/226764

Corresponding author:
Natalija Parlov, Dr. Franjo Tudjman Defense and Security University, Zagreb, Croatia; E-mail: una@apicura.hr
0000-0002-8468-7465

product liability directive and governance, cybersecurity, and auditing and liability literature. It develops a *governance-evidence-liability* framework distinguishing technical AI systems, AI-enabled products and services, organisational governance, certification, and legal accountability. The analysis finds that AIMS audits assess organisational governance, including security-risk oversight, through scope-bound, sampled, and time-bound evidence. Multidisciplinary auditor competence may strengthen the credibility of judgements but does not transform management-system certification into technical cybersecurity validation, product certification, or a binding compliance determination. AIMS documentation may constitute important contextual evidence in civil and criminal liability because cybersecurity risk assessments, vulnerability records, monitoring logs, and incident-response records may show whether foreseeable threats were identified and managed. However, its probative value depends on scope, timing, reliability, factual connection to the system or incident, and corroborating technical and operational evidence. It cannot independently establish cybersecurity resilience, technical correctness, legal compliance, defectiveness, fault, causation, or liability. The framework supports interpretation of security-related AIMS evidence without overstating the assurance provided by certification.

Keywords

artificial intelligence, EU AI Act, ISO/IEC 42001, AI auditing, multidimensional liability

1. Introduction

The adoption of artificial intelligence (AI) across public and private sectors has reshaped organisational decision-making, operational control, accountability, and risk management. AI systems are increasingly embedded in products, services, critical business processes, and public-sector functions, where their outputs may affect individuals, organisations, and society. The discussion therefore extends beyond technical performance, model accuracy, and data- or hardware-related concerns. It increasingly concerns governance, oversight, and organisations' ability to identify, treat, monitor, and manage AI-related risks systematically [1–4].

This development is reflected in comprehensive regulatory and standardisation initiatives. In the European Union (EU), the Artificial Intelligence Act (AI Act) establishes a risk-based framework and allocates obligations among actors in the AI value chain [3].

Organisations also use international standards to structure internal governance. International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) 42001 is the ISO management-system standard designed specifically for AI governance [2]. Demand for independent assurance has therefore grown. ISO/IEC 42006 certification specifies requirements for bodies auditing and certifying an artificial intelligence management system (AIMS), including competence requirements for audit teams [5]. Together with ISO/IEC 17021-1 and ISO 19011, these standards provide the framework for third-party certification audits of AI governance systems [5–7].

Despite this growing body of standards and regulation, an important conceptual problem remains: AI systems as technical artefacts, AI-enabled products and services, organisational governance mechanisms, and legal accountability frameworks are often treated as interchangeable. That conflation overstates what management-system certification can establish. An ISO/IEC 42001 certification audit assesses governance structures, organisational controls, documented processes, risk management, and assigned responsibilities; it is not product testing, technical validation, a sectoral conformity assessment of an AI system, or a legal opinion. ISO/IEC 42006 certification requires multidisciplinary competence in AI, relevant sectors, risk and impact assessment, and applicable legal obligations [5]. Such competence enables auditors to evaluate whether governance mechanisms are adequate within the audit scope; it does not convert the audit into independent verification of algorithmic performance or legal compliance.

This distinction is also relevant from a legal perspective. Product, professional, civil, and criminal liability frameworks often rely on concepts, such as due diligence, organisational control, foreseeability, and documented decision-making. As these concepts are closely connected to AIMS governance mechanisms, governance records and certification audit outputs may become relevant evidence in regulatory, judicial, or administrative proceedings. Yet the relationship between management system audits, due diligence, and legal responsibility remains only partially explored in the literature. The existing research extensively addresses AI governance, risk management, trustworthy AI, and regulatory compliance. By contrast, limited attention has been paid to the evidentiary and liability implications of AIMS audits. The distinction between technical AI systems, AI-enabled products and services, organisational governance, and legal accountability also remain insufficiently examined, creating uncertainty in both implementation and interpretation of audit outcomes.

Recent scholarship distinguishes legal, ethical, and technical approaches to AI auditing; separates governance, model, and application audits; conceptualises accountability as a socio-technical relationship among actors, forums, and consequences; and shows that ethics-based auditing produces heterogeneous forms of knowledge for different stakeholders [8–11]. These contributions support a more precise account of what an AIMS audit can evidence and why its outputs should not be treated as general proof of accountability.

This paper does not assess the technical correctness of AI systems or determine legal compliance. Instead, it examines the role of ISO/IEC 42001 and ISO/IEC 42006 within the broader context of AI governance, conformity assessment, and accountability. Through standards interpretation, governance analysis, and legal examination, it clarifies what AIMS audits assess, where their limitations lie, and how audit outputs may be considered in regulatory and liability contexts. To address this objective, the paper develops a *governance-evidence-liability* perspective that distinguishes between technical AI systems, AI-enabled products and services, organisational governance structures, and legal accountability mechanisms.

Based on this perspective, the analysis is guided by the following research questions:

> RQ1: What organisational object is assessed, and what forms of governance and security-related evidence are reviewed, in third-party AIMS certification audits?
>
> RQ2: How do the competence requirements of ISO/IEC 42006 certification affect the credibility and limits of an AIMS audit evidence?
>
> RQ3: Under what conditions AIMS documentation and certification outputs could be relevant, without being determinative, in regulatory, civil, criminal, professional-liability and product-liability assessments?

By examining these questions, the paper contributes to the ongoing discussion on AI governance by clarifying the scope, limitations, and practical significance of AIMS certification. It also helps to explain how governance-oriented conformity assessment relates to technical evidence, organisational accountability, and liability considerations in AI-enabled environments.

The topic is relevant beyond academic debate. The AI Act, the product liability directive (PLD), sector-specific regulations, and emerging harmonised standards create a growing set of obligations for actors involved throughout the AI lifecycle, including developers, providers, deployers, manufacturers, auditors, and conformity assessment bodies. As AI becomes more integrated into business processes, public services, and safety-critical environments, questions of accountability and liability become increasingly important. Yet the distinction between technical responsibility, governance responsibility, and the evidentiary role of audits remains insufficiently understood. By distinguishing governance assessment from technical validation, compliance determination, and liability attribution, the paper clarifies the role of AIMS certification audits and the limits of the conclusions that can reasonably be drawn from them.

2. Methodology and Scientific Contribution

This study adopts a conceptual, standards-interpretive, and doctrinal legal research design. This design is appropriate because the research questions concern the normative scope of AIMS certification, the meaning of audit-team competence and evidence requirements, and the possible legal relevance of governance records – not the measured performance of a population of certified organisations. Standards interpretation identifies the object, process, and limits of third-party management-system auditing. Doctrinal analysis then examines how those outputs could be characterised under the AI Act, the revised PLD, and applicable national liability principles [1–7,12–14].

Sources were selected through a targeted and iterative review conducted between February and July 2026. The review was designed to identify authoritative legal, standards-based, and peer-reviewed sources directly relevant to the research questions rather than to constitute a systematic literature review. Searches were conducted in Scopus, Web of Science, and Google Scholar, while legal and regulatory materials were identified through EUR-Lex, official European Commission sources, and other official EU institutional repositories. Standards and related implementation materials were identified through the ISO catalogue and relevant official standardisation sources.

The searches combined terms including ‘artificial intelligence auditing’, ‘AI auditing’, ‘AI management system’, ‘AIMS’, ‘ISO/IEC 42001’, ‘ISO/IEC 42006’, ‘conformity assessment’, ‘AI governance’, ‘AI accountability’, ‘algorithmic auditing’, ‘audit evidence’, ‘evidentiary

value', 'professional liability', 'product liability', 'civil liability', 'due diligence', and 'AI Act'. Search terms were used individually and in combinations appropriate to the respective database or repository.

Sources were included where they directly informed at least one of the following: the scope and object of AI or management-system auditing; certification and conformity-assessment architecture; audit-team competence; AI governance and accountability; the evidentiary relevance of governance or audit records; or regulatory, professional-, product, and civil-liability analysis. Preference was given to primary legal sources, published international standards, official institutional materials, and peer-reviewed academic literature. Sources addressing only the technical testing, benchmarking, or performance evaluation of AI models were excluded where they did not establish a substantive connection with organisational governance, conformity assessment, accountability, or liability. Additional relevant materials were identified through backward reference checking and forward citation searching of the most directly relevant publications.

The primary standards corpus comprises ISO/IEC 22989, ISO/IEC 42001, ISO/IEC 42006, ISO/IEC 17021-1, ISO 19011, ISO/IEC 42005, ISO/IEC 23894, and ISO 31000 because these instruments define the terminology, management-system requirements, audit process, competence, impact-assessment, and risk-management concepts examined in RQ1 and RQ2 [1,2,5–7,15–17]. The legal corpus comprises the AI Act, the revised PLD, official EU implementation material, and relevant national doctrine [3,13,14]. Peer-reviewed AI-auditing, governance, and accountability literature was included where it directly informed audit scope, assurance architecture, accountability, ethics-based auditing, evidentiary relevance, or risk governance [8–11,18–26].

The analysis proceeded in four stages. First, it separated the technical AI system, the AI-enabled product or service, the organisational AIMS, the certification activity, and the legal attribution of responsibility. Second, it derived the relevant forms of audit evidence and competence constraints from the standards corpus. Third, it compared the proposed *governance-evidence-liability* framework with the existing auditing and governance models. Fourth, it applied the resulting distinctions to regulatory, professional-, product, and civil-liability analysis, consistently distinguishing evidentiary relevance from proof of compliance, defect, fault, causation, or liability. This staged procedure provides a transparent chain from source interpretation to answers to RQ1–RQ3.

The legal examples are illustrative applications of established concepts, such as duty of care, defectiveness, causation, foreseeability, and disclosure; they are not presented as settled case law on ISO/IEC 42001 audits because reported judicial experience with AIMS certification is very limited, and each legal proposition is framed as a doctrinal inference whose force depends on the applicable jurisdiction, contract, audit scope, and facts.

The unit of analysis is the third-party certification audit of an organisational AIMS, not the technical correctness, performance, or legal compliance of an individual AI system. Particular attention is paid to ISO/IEC 42006 certification because its collective competence requirements may affect the credibility of audit evidence while leaving the audit's scope, sampling, and time limitations intact [5]. This boundary is necessary to answer RQ2 without treating auditor expertise as a mandate to validate every technical or legal property of an AI system. The paper therefore makes three principal contributions:

- Theoretical contribution: The paper distinguishes the technical-system, product-or-service, organisational-governance, certification, and legal-accountability layers and connects them through the *governance-evidence-liability* framework. The framework explains why evidence can move between those layers without the conclusion at one layer determining the conclusion at another.
- Methodological contribution: The paper combines standards interpretation with doctrinal legal analysis through an explicit four-stage procedure. It also introduces contextual evidentiary value – scope, timing, factual nexus, and corroboration – as a bridge between management-system audit evidence and a later regulatory or liability assessment.
- Practical implications: The analysis provides certification bodies, organisations, regulators, and legal practitioners with criteria for interpreting AIMS policies, risk and impact assessments, audit records, and certification outputs. Those materials may support an assessment of organisational governance or due diligence but cannot independently establish technical correctness, legal compliance, defectiveness, causation, fault, or liability.

2.1. Proposed Methodological Limitations

The study is conceptual and doctrinal, and does not test the framework against a sample of certified organisations, certification bodies, audit files, or judgements. The standards and EU implementation environment are recent, while liability remains partly national

and contract-specific. Cross-jurisdictional examples therefore illustrate possible reasoning rather than universal outcomes. Future research should test the framework through anonymised audit-file analysis, interviews with certification bodies and regulators, comparative national case studies, and longitudinal examination of whether particular audit findings are later used in supervisory, contractual, or judicial proceedings. Such studies should record audit scope, sampling, team competence, timing, and the factual connection between reviewed controls and the incident, so that contextual evidentiary value can be evaluated rather than assumed. The literature review was targeted rather than systematic and was limited to sources identified between February and July 2026 through databases, institutional repositories, and citation-searching procedures described above. Accordingly, the study does not claim exhaustive coverage of all legal, technical, or governance scholarship. The identified research gap should therefore be understood as a gap within the reviewed corpus and search scope, rather than as a proof that no related analysis exists in any jurisdiction or disciplinary field.

3. Governance Object of Artificial Intelligence Audits

Discussions on AI governance often blur the distinction between AI systems, AI-enabled products, organisational governance structures, and certification activities. Although closely connected, they operate at different levels and fulfil different functions. As a result, certification, audit findings, and responsibility are sometimes interpreted more broadly than intended. The conceptual foundation for understanding these distinctions can be found in ISO/IEC 22989, which defines an AI system as an engineered system capable of generating outputs, such as predictions, recommendations, decisions, or actions in pursuit of human-defined objectives [1]. This definition focuses on the technical and functional characteristics of AI systems rather than on organisational responsibilities, regulatory obligations or accountability mechanisms. An AI system therefore exists primarily as a technical artefact whose behaviour is influenced by factors such as training data, algorithms, validation methods, operational environments, and human oversight.

At a different level, AI systems may be embedded into products, services, or operational processes. Once deployed in a social, business, industrial, or public sector environment, the AI system becomes part of a broader socio-technical context in which outputs may influence decisions affecting individuals, organisations, or society. It is at this level that issues such as product safety, regulatory

obligations, operational performance, contractual responsibilities and liability considerations typically emerge. The technical existence of an AI system alone does not determine how it is used, interpreted, or relied upon in practice.

The organisational layer introduces a further distinction. ISO/IEC 42001 is a management-system standard, rather than a standard for the certification of the technical properties of an individual AI system or AI-enabled product. It does not prescribe or independently verify specific thresholds for model performance, accuracy, robustness, cybersecurity, or safety. Instead, it establishes requirements for an AIMS through which organisations govern relevant technical and organisational matters, including AI lifecycle processes, data management, risk and impact assessment, monitoring, and continual improvement [2]. The standard focuses on organisational responsibilities, risk management, impact assessment, monitoring, incident management, documented information, and continual improvement. Its purpose is to structure how organisations govern AI rather than how AI systems themselves operate.

This distinction is particularly important because an AIMS governs the development, deployment, and use of AI systems but does not directly constitute the AI system itself. The management system exists at an organisational level and functions as a framework for decision-making, oversight, and oriented towards accountability (e.g. its outputs are policies, procedures, risk registers, impact assessments, records of monitoring activities, internal audit reports, and management review outputs primarily rather than algorithmic predictions, recommendations, or automated decisions). Because of this, the effectiveness of an AIMS cannot be measured through model performance metrics alone but must instead assessed through the effectiveness of the governance mechanisms that support organisational control over AI-related activities.

The governance-oriented nature of ISO/IEC 42001 is also reflected in its relationship with risk management and organisational accountability. Rather than prescribing specific technical controls or performance thresholds, the standard requires organisations to identify AI systems within scope, understand their intended purpose, assess foreseeable risks and impacts, assign responsibilities, implement appropriate controls, and establish mechanisms for monitoring and continual improvement [2]. The risk management approach embedded within ISO/IEC 42001 is broadly consistent with the principles established by ISO 31000 [15]. The emphasis therefore lies on the existence and operation of governance

processes rather than on the direct verification of technical outcomes. This governance perspective becomes increasingly relevant as organisations operate within complex digital ecosystems characterised by cloud computing, external AI providers, software supply chains, and interconnected data environments. In many situations, organisations deploy AI systems that they neither fully develop nor control. Nevertheless, they remain responsible for determining how those systems are selected, integrated, monitored, and relied upon within their operational environment. ISO/IEC 42001 acknowledges this reality by focusing on organisational control rather than technical ownership. Similar governance challenges have already been identified in critical infrastructure environments where AI governance, supply chain dependencies, and regulatory obligations intersect within broader resilience frameworks [27].

Beyond governance and accountability, AI systems introduce security-specific risks, including data poisoning, adversarial manipulation, model extraction, prompt injection, unauthorised model access, supply chain compromise, and integrity degradation. While ISO/IEC 42001 does not prescribe technical security controls, an AIMS requires organisations to establish governance mechanisms capable of identifying, assessing, and managing such security-related risks throughout the AI lifecycle. Understanding this distinction is essential for interpreting the role of auditing. If the object of assessment is incorrectly assumed to be the AI system itself, certification may be misinterpreted as a form of technical validation or assurance regarding algorithmic performance. Conversely, if the governance nature of an AIMS is properly understood, certification can be recognised for what it is: an independent assessment of the organisational mechanisms established to govern AI-related activities. This distinction forms the foundation for understanding the scope and limitations of ISO/IEC 42001 certification audits. Before examining how such audits are conducted and what competence they require, it is necessary to recognise that the primary object of assessment is neither the AI system nor the AI-enabled product but rather the governance framework through which the organisation exercises control over both. This governance layer subsequently becomes the principal source of evidence regarding organisational oversight, risk awareness, and due diligence, and therefore occupies a central position in subsequent discussions concerning accountability and liability.

4. Auditing Artificial Intelligence Governance

The governance-oriented nature of an AIMS directly influences the character, objectives, and limitations of certification

auditing. Unlike technical assessments of AI systems, software testing activities, conformity assessments of AI-enabled products, or regulatory inspections, audits conducted under ISO/IEC 42001 focus on the organisational mechanisms established to govern AI throughout its lifecycle. Understanding this distinction is essential for correctly interpreting both the significance and the limitations of certification outcomes.

The certification of an AIMS is performed within the broader framework of management system conformity assessment established by ISO/IEC 17021-1 and supported by the auditing principles contained in ISO 19011 [6,7]. In accordance with ISO auditing terminology, certification audits performed under ISO/IEC 42001 constitute third-party audits, whereas internal audits conducted by the organisation represent first-party audits and customer or supplier assessments may be regarded as second-party audits. The present paper focuses exclusively on third-party certification audits performed by accredited certification bodies. As with other management system certifications, the purpose of the audit is to determine whether the organisation has established, implemented, maintained, and continually improved a management system conforming to the applicable standard requirements. Consequently, the primary object of assessment is the management system itself rather than the technical artefacts governed by that system.

This distinction has important implications for the nature of audit evidence. Certification audits rely on objective evidence obtained through interviews, observation of activities, review of documented information, and sampling of records [6,7]. Consistent with the principles of ISO/IEC 17021-1, certification provides confidence that a management system conforms to specified requirements and is capable of supporting the achievement of organisational policy and objectives but it does not guarantee technical outcomes, future performance or absence of risk. Audit conclusions are therefore based on the evaluation of selected evidence available at a specific point in time rather than on exhaustive examination of all organisational activities or technical components. Both ISO/IEC 17021-1 and ISO 19011 explicitly recognise that management system audits are evidence- and sample-based activities performed under conditions of finite time and finite resources [6,7]. As a result, certification conclusions should not be interpreted as guarantees of future performance, complete conformity, or absence of risk.

The introduction of ISO/IEC 42006 further expands the competence requirements applicable to certification bodies auditing AIMS [5].

Unlike traditional management system audits, AIMS certification requires multidisciplinary competence that extends beyond generic auditing skills. Certification bodies are required to ensure that audit teams collectively possess knowledge relating to AI concepts, AI lifecycle processes, risk management, impact assessment, governance structures, applicable legal obligations, relevant business sectors, and technologies associated with the audited scope [5].

Security competence becomes particularly relevant where AI systems are deployed within critical infrastructure, healthcare, transportation, defence, financial services, and other security-sensitive environments. In such contexts, audit teams may need sufficient understanding of cybersecurity, AI-specific attack vectors, resilience measures, and incident response processes to evaluate whether governance mechanisms appropriately address security risks associated with AI deployment. This requirement is particularly significant because it addresses a common misconception regarding AIMS audits. The existence of technical and regulatory competence within the audit team does not transform the certification audit into either a technical validation exercise or a legal compliance assessment. Rather, such competence enables auditors to evaluate whether the organisation has established appropriate governance mechanisms for identifying, assessing, and managing technical and regulatory issues relevant to the scope of an AIMS. Where security risks are material to the audited scope, the collective competence of the audit team should enable an informed assessment of the organisation's security-governance processes without transforming the AIMS audit into a technical cybersecurity assessment.

ISO/IEC 42006 explicitly requires knowledge of AI-related legal obligations as part of audit team competence [5]. However, neither ISO/IEC 42006 nor ISO/IEC 17021-1 requires certification bodies to issue legal opinions or determine legal compliance in a judicial or regulatory sense. This distinction is further reinforced by ISO 19011, which recognises the importance of awareness of statutory and regulatory requirements while simultaneously emphasising that management system audits should not be treated as legal compliance audits [7]. The difference between legal awareness and legal compliance assessment is therefore fundamental. Audit teams must understand the regulatory environment applicable to the organisation in order to evaluate whether relevant obligations have been identified appropriately, and considered and incorporated into the management system. Nevertheless, the purpose of the audit remains limited to assessing the effectiveness of governance

processes established to manage such obligations. Determining the definitive interpretation of legislation, resolving legal disputes, or establishing compliance in a judicial sense remains the responsibility of regulators, courts, and other competent authorities.

A similar distinction applies to technical evaluation. Certification auditors may review technical artefacts, validation records, performance reports, monitoring outputs, incident records, and other technical evidence where such information is relevant to the assessment of governance processes. However, the audit does not constitute independent verification of algorithmic correctness, model accuracy, robustness, cybersecurity performance, or safety characteristics. Technical information is primarily examined as evidence of governance activities rather than as an object of certification itself. The competence requirements introduced by ISO/IEC 42006 therefore create a distinctive audit model situated between purely organisational auditing and highly specialised technical assessment. Audit teams are expected to possess sufficient technical, sector-specific, and regulatory understanding to evaluate governance effectiveness while remaining within the boundaries of management system conformity assessment. This balance preserves the independence and objectivity of certification activities while ensuring that governance evaluations remain informed by the technical and regulatory realities of AI.

From an accountability perspective, this distinction is particularly important. The presence of multidisciplinary competence strengthens the credibility of audit findings and can increase the perceived evidentiary relevance of certification outcomes. At the same time, the governance-oriented nature of the audit limits the conclusions that may reasonably be drawn from certification. An audit may provide evidence that an organisation established and operated governance processes intended to manage AI-related risks, but it cannot independently demonstrate technical correctness, legal compliance, absence of negligence, or absence of liability. Certification audits conducted under ISO/IEC 42001 should be understood as governance assessments supported by technical, organisational, and regulatory evidence. Their primary contribution lies not in validating AI systems directly but in evaluating the organisational structures through which AI-related activities are governed. This governance layer subsequently becomes the principal source of evidence through which organisational control, oversight, and due diligence may be examined in regulatory, judicial, and liability contexts.

5. A *Governance-Evidence-Liability* Framework for Artificial Intelligence Auditing

The preceding discussion demonstrated that AI systems, AI-enabled products, organisational governance mechanisms, and legal accountability frameworks represent distinct but interconnected layers of AI management. While these layers are frequently discussed together, the targeted review conducted for this study did not identify a framework that specifically explains how ISO/IEC 42001 management-system audit outputs may move beyond the certification setting and acquire contextual evidentiary relevance in subsequent regulatory or liability assessments. In response to this gap, this paper proposes a *governance-evidence-liability* framework intended to support the interpretation of AIMS audits within broader accountability and liability contexts.

Figure 1 illustrates five interconnected layers through which AI governance, auditing, and accountability can be analysed. The model is not intended to represent a linear process. Instead, it highlights how technical AI operation, governance activities, independent auditing, and liability assessment remain distinct while continuously influencing one another.

At the foundation of the framework is the AI system itself. This technical layer includes data, models, algorithms, training and

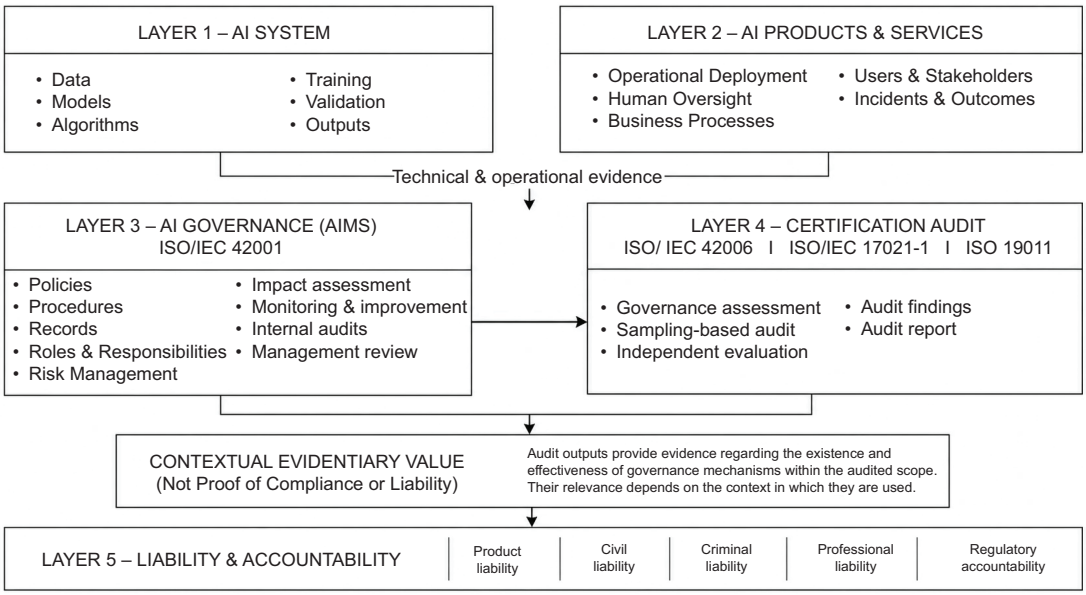

Figure 1. *Governance-evidence-liability* framework for AIMS certification and accountability assessment.
Source: Author's work.

validation activities, generated outputs, and supporting technical infrastructure. It is also the layer where issues related to performance, robustness, explainability, bias, drift, and AI-specific cybersecurity risks may originate.

The next layer comprises AI-enabled products and services. Here, AI systems interact with users, business processes, and operational environments. Product safety, operational performance, security incidents, regulatory obligations, and liability considerations emerge through the practical use of AI outputs in real-world contexts.

Above this sits the AI governance layer, corresponding to the AIMS established under ISO/IEC 42001. It includes policies, procedures, records, assigned responsibilities, risk and impact assessments, monitoring activities, internal audits, and management reviews. Governance mechanisms also support the identification and management of AI-related security risks, including vulnerabilities, misuse, supply-chain dependencies, and cybersecurity incidents throughout the lifecycle.

The fourth layer is the certification audit function performed under ISO/IEC 42006, ISO/IEC 17021-1, and ISO 19011. Unlike governance activities operated by the organisation, certification audits provide an independent conformity assessment of the management system. Audit activities evaluate governance arrangements, risk management processes, security oversight mechanisms, documented evidence, and the effectiveness of organisational controls. In this way, the audit layer acts as an intermediary between organisational governance and external accountability.

Between the audit and liability layers sits a dedicated evidentiary component labelled contextual evidentiary value. It reflects a central premise of this paper: governance documentation and audit outputs may provide evidence of organisational control, governance maturity, risk awareness, due diligence, and security oversight. Such evidence does not, however, establish legal compliance, technical correctness, or liability on its own. Its relevance depends on the specific circumstances of a case and must be considered together with technical, operational, factual, and legal evidence.

The final layer concerns liability and accountability. It includes product, civil, criminal, and professional liability as well as regulatory accountability mechanisms. Questions of responsibility are ultimately assessed through judicial, regulatory, or administrative

processes, drawing on a broader body of evidence than governance records or certification findings alone.

Figure 1 also highlights two important evidence relationships. Technical and operational evidence generated by AI systems and AI-enabled products flows towards the governance and audit layers, where it becomes part of documented information, risk assessments, monitoring activities and audit evidence. In the opposite direction, governance assessments performed through certification audits generate findings and reports that may later acquire evidentiary relevance in accountability and liability proceedings. The framework therefore illustrates that certification audits do not determine liability directly. Instead, they contribute governance-related evidence that may be considered alongside technical, operational, and legal evidence.

The framework further identifies three interconnected flows across the layers. The first is a control flow. Organisational decisions, governance arrangements, resource allocation, security measures, and risk management activities influence how AI systems are selected, deployed, monitored, and maintained. The second is an evidence flow. Technical records, validation results, monitoring logs, incident reports, security events, and operational observations generate evidence that may later appear in governance documentation, risk assessments, audit records, and management reviews. Certification audits evaluate this material through a sampling-based approach and focus on its relevance to governance effectiveness rather than technical correctness. The third is a liability flow. When adverse events occur, assessments of responsibility typically examine technical behaviour, operational decisions, governance arrangements, and organisational controls. Liability therefore extends across multiple layers and cannot be attributed solely to a single component of the AI ecosystem.

Within this model, ISO/IEC 42001 certification audits occupy an intermediary position between technical AI operation and legal accountability. Audit findings may contribute evidence relevant to organisational control, risk awareness, due diligence, and security governance. Certification outcomes alone, however, cannot establish technical correctness, legal compliance, causation, or liability.

5.1. Comparison with the Existing Conceptual Models of AI Auditing and Governance

Recent AI-auditing and governance literature can be grouped into four complementary strands.

First, scope and architecture: Mökander distinguishes legal, ethical, and technical approaches to AI auditing [8]; Mökander et al. distinguish governance, model, and application audits [9]; Raji et al. propose lifecycle-oriented internal algorithmic auditing [18]; and Mökander et al. place conformity assessment and post-market monitoring within an EU auditing ecosystem [19].

Second, accountability and ethics: Novelli et al. conceptualise accountability through relations among actors, forums, and consequences [10] while Laine et al. show that ethics-based auditing produces different forms of knowledge for stakeholders [11]; and Birhane et al. caution that audit findings do not necessarily produce accountability outcomes [21].

Third, organisational and methodological work: It addresses legal, ethical, and technological risk taxonomies, organisational AI governance, and continuous auditing [20,22,23].

Fourth, risk- and regulation-centred models: They operationalise AI Act risk categories, Article 9 risk management, and the role of civil society access in the audit ecosystem [24–26].

The *governance-evidence-liability* framework shares the premise that accountability requires structured, documented, and assessable governance, but it asks a different question. The existing models primarily classify audit objects, methods, actors, or accountability mechanisms. The present framework traces what happens when a management-system audit output leaves the assurance setting and is invoked in a regulatory or liability assessment. Its distinctive component is contextual evidentiary value: the weight of an audit output depends on audit scope, timing, connection between the audited process and the event, and corroboration by technical and factual evidence. Multidisciplinary competence may strengthen the reliability of an assessment within that boundary; it does not erase the boundary.

The contribution is therefore deliberately limited. The framework is not a new technical audit method, a scoring model for audit quality, or a liability rule. It is an analytical structure for distinguishing the governance object assessed in RQ1, the credibility and limits of competence-informed evidence in RQ2, and the conditions of legal relevance examined in RQ3. Its originality and limitations should be tested empirically against certification practice and, as decisions emerge, regulatory and judicial use of AIMS material.

6. Evidentiary Significance of AIMS Documentation and Certification Audits

The governance-oriented nature of ISO/IEC 42001 certification audits raises an important question that extends beyond conformity assessment itself: what evidentiary significance, if any, may be attributed to AIMS documentation and certification audit outputs when organisational conduct is subsequently examined by regulators, courts, or other competent authorities? Security-related governance records may include vulnerability assessments, security monitoring outputs, incident response records, adversarial testing results, cybersecurity risk assessments, and supply chain assurance documentation. Such records may become particularly relevant when regulators or courts seek to determine whether foreseeable AI-related security risks were governed appropriately.

Neither ISO/IEC 42001 nor ISO/IEC 42006 establishes legal presumptions regarding compliance, fault, negligence, or liability [2,5]. Certification confirms that a management system conforming to specified requirements has been assessed and found compliant at a particular point in time within a defined scope. Consequently, certification should not be interpreted as a proof that an AI system is technically correct, legally compliant, safe under all circumstances, or free from foreseeable failures.

Nevertheless, management system certification produces a substantial body of documented information that may become relevant in subsequent accountability assessments. Typical examples include governance policies, risk assessments, and AI impact assessments performed in accordance with emerging practices, such as ISO/IEC 42005 [16] statements of applicability, incident records, monitoring records, internal audit reports, management reviews, corrective action records, and third-party certification audit reports [2]. Such risk management activities are increasingly structured using dedicated AI risk management methodologies, including ISO/IEC 23894 [17]. Collectively, these artefacts provide insight into how the organisation structured its governance activities, allocated responsibilities, and addressed AI-related risks.

From an evidentiary perspective, such documentation may assist in reconstructing organisational decision-making processes. It may indicate which risks were identified, which controls were implemented, how responsibilities were assigned, what monitoring activities were performed, and how incidents were addressed. The documentation therefore provides information about the existence and operation of governance mechanisms rather than

direct evidence regarding the technical behaviour of AI systems themselves.

Certification audit reports introduce an additional layer of potentially relevant evidence. Unlike internal governance records, third-party audit reports represent independent evaluations conducted according to established conformity assessment procedures [5–7]. Certification audits are performed by multidisciplinary audit teams possessing collective competence in AI, governance processes, risk management, impact assessment, applicable legal obligations, and relevant business sectors [5]. As a result, certification findings may be perceived as more independent than purely internal organisational records and may therefore constitute one element among broader evidentiary material when assessing the maturity and effectiveness of governance arrangements.

However, the evidentiary value of certification outputs remains inherently limited by the nature of management system auditing. Audit findings are derived from sampling, interviews, observation, and review of selected evidence available during the audit period [6,7]. Certification conclusions therefore represent professional assessments of governance conformity, rather than exhaustive evaluations of all technical, operational, or legal aspects of AI-related activities. The absence of identified non-conformities cannot be interpreted as a proof that no deficiencies existed, nor can certification be regarded as confirmation that future incidents could not occur.

This limitation becomes particularly important when considering concepts frequently encountered in legal and regulatory proceedings. Certification audits may provide information relevant to organisational control, risk awareness, governance maturity, and due diligence. They may assist in understanding whether the organisation established structured processes for identifying and managing AI-related risks. However, they do not independently establish legal compliance, causation, fault, negligence, foreseeability, or liability. These determinations require broader factual, technical, and legal assessments extending beyond the scope of management system certification.

The distinction between governance evidence and liability evidence is therefore critical. Governance evidence demonstrates how an organisation intended to manage and oversee AI-related activities. Liability evidence seeks to establish whether specific conduct, omissions, or failures contributed to a legally relevant outcome.

While governance evidence may contribute to such assessments, it cannot replace technical investigation, factual analysis, or legal interpretation.

This distinction is particularly relevant in the context of emerging AI-related liability frameworks. As regulators, courts and other competent authorities increasingly examine the governance of AI, and management system documentation and certification audit outputs may become important sources of contextual evidence regarding organisational conduct. Their significance, however, derives not from proving legal responsibility directly but from helping to explain how responsibility, oversight, and control were structured and exercised within the organisation.

ISO/IEC 42001 certification should be understood neither as a shield against accountability nor as a determinant of liability. Rather, it represents a governance mechanism capable of generating evidence regarding organisational control and due diligence. The legal significance of such evidence ultimately depends on how it is interpreted within the broader framework of applicable regulatory requirements and liability doctrines. It is this interaction between governance evidence and legal accountability that forms the focus of the following analysis.

7. Inter-Connected Landscape of Criminal, Product, Professional and Civil Liability in AI Governance

The AI Act, as amended by Regulation (EU) 2026/1744, imposes regulatory obligations, especially for high-risk AI systems, and administrative penalties [3,28], but it does not create a general civil-liability regime for AI-related harm. Civil and professional liability, therefore, remains governed principally by the revised EU product-liability regime and applicable national contract and tort law; criminal liability remains a matter of national law. Because Directive (EU) 2024/2853 applies only to products placed on the market or put into service on or after 9 December 2026, and no reported case law concerning ISO/IEC 42001 certification audits was identified under the selected sources, the analysis below is a doctrinal and comparative projection rather than a statement of settled law [13,29–31].

As a general proposition, compliance with the AI Act may be evidentially relevant but does not determine civil or criminal liability. Voluntary standards and guidance may help articulate expected governance practice, particularly where they are incorporated into

contracts or recognised conformity-assessment arrangements, but their legal weight is jurisdiction- and context-dependent. Courts and regulators remain responsible for applying the governing legal tests, normally with technical expert evidence where necessary [32–36].

The AI Act establishes a risk-based regulatory framework and allocates duties to defined operators, including providers, deployers, importers, distributors, and product manufacturers. It does not create a general category of 'AI-system auditors'. Articles 57 and 60 address regulatory sandboxes and real-world testing; liability for resulting harm remains governed by applicable Union and national law [3,37].

The AI Act may nevertheless affect later liability assessments indirectly through its requirements on risk management, technical documentation, record keeping, transparency, and human oversight. Article 31 requires an applicant conformity-assessment body to provide information concerning liability insurance unless liability is assumed by the member state; this is not a general rule that every 'auditor' must carry insurance. Compliance records may be relevant evidence of organisational conduct, but neither compliance nor non-compliance automatically proves or disproves negligence, causation, or defectiveness.

7.1. Application of RQ1–RQ3 to the Legal Analysis

The legal analysis applies the three research questions stated in the Introduction. RQ1 identifies the organisational object and forms of evidence assessed in third-party AIMS certification audits. RQ2 examines how ISO/IEC 42006 competence requirements affect the credibility and limits of that evidence. RQ3 asks under what conditions AIMS documentation and certification outputs may be relevant, without being determinative, in regulatory, civil, criminal, professional-, and product-liability assessments.

For professional-liability analysis, RQ1 and RQ2 direct attention to the agreed audit object, scope, sampling, objective evidence, and competence of the certification body and audit team. Those factors may inform whether the applicable standard of care was met. An implementer's position depends separately on its agreed role and applicable law.

For regulatory and product-liability analysis, RQ3 requires consideration of whether a governance record is timely, within scope,

connected to the relevant AI system and incident, and supported by technical or operational evidence. Under Directive (EU) 2024/2853, such material may be relevant to disclosure, statutory presumptions or defectiveness, but does not by itself establish defectiveness, compensable damage, or causation [13]. Under the AI Act, where relevant provisions apply, it may evidence process compliance but does not determine civil or criminal liability [3]. The PLD's evidentiary mechanisms do not determine criminal fault or causation.

These regimes apply different legal tests. Professional liability concerns an applicable standard of care, while product liability concerns defectiveness, compensable damage, and causation without requiring proof of fault. The AI Act imposes regulatory duties but does not create a general civil-liability cause of action. Criminal liability depends on the applicable national offence and attribution rules. The analysis therefore remains conceptual and jurisdiction-sensitive, consistent with the methodological limits stated above.

7.2. Product Liability Regime

Under Directive (EU) 2024/2853, an economic operator may be liable without proof of fault for damage caused by a defective product. The claimant must establish damage, defectiveness, and causation, subject to the directive's disclosure and rebuttable-presumption rules. Fault is not an element of this cause of action and should not be conflated with the separate assessment of damages or with criminal responsibility [13].

Directive (EU) 2024/2853 expressly includes software in the definition of a 'product', whether supplied on a device or accessed through a communications network or cloud technology. An AI system is therefore covered insofar as it constitutes software or is integrated into another product; the directive does not create 'AI systems' as a separate product category. Member States must transpose the directive by 9 December 2026, and it applies to products placed on the market or put into service on or after that date [13].

The AI Act and the revised PLD are complementary but apply different legal tests. Under the PLD, a software developer or provider may qualify as a manufacturer or other liable economic operator in the circumstances specified by the directive. Failure to provide safety-related updates within the manufacturer's control may be relevant to defectiveness, but liability still requires the directive's conditions, including damage and causation; it does not

follow merely from a person being labelled a provider under the AI Act [13].

The revised PLD does not impose a general reversal of the burden of proof. Article 9 permits proportionate disclosure of relevant evidence, subject to protection of confidential information and trade secrets. Article 10 establishes rebuttable presumptions of defectiveness and/or causation only in specified circumstances, including non-compliance with mandatory safety requirements, failure to disclose evidence, obvious malfunction during reasonably foreseeable use, or excessive technical or scientific complexity. The claimant must still establish the elements not covered by a presumption [13].

Two mechanisms deserve particular attention. AIMS records may be within the scope of a proportionate disclosure order where they are relevant and controlled by a defendant, but this is case-specific. Certification evidence is not automatically one-sided: it cannot conclusively establish product safety, yet it may support either party depending on what it shows and may be considered among the circumstances relevant to defectiveness. Any claim of evidentiary asymmetry is therefore a hypothesis requiring empirical validation rather than a settled consequence of Articles 9 and 10 [13,38,39].

In summation, whereas the new European framework for product liability places a considerably elevated burden of accountability of the production/service side, it, at the same time, enables the ingress protection (IP) safety zone environment that the producers can control and direct. The position of auditors and implementers in relation to IP is sensitive from the perspective of required non-disclosure agreements and trade secrets, where a certain level of trust is required to enable the performance of such services. Any effect of insurance on an auditor's or implementer's exposure depends on the policy wording and applicable law; insurance held by a producer does not, by itself, reduce another actor's legal liability.

7.3. Exposure to Criminal Liability

Criminal liability in the AI context continues to be governed primarily by traditional criminal-law concepts. The AI Act establishes regulatory penalties, including administrative fines, but does not create a harmonised EU criminal-liability regime. Criminal liability remains primarily governed by domestic law. The elements to be proved depend on the specific national offence, including whether it requires intent, negligence, causation, endangerment, or a harmful result.

If a person knowingly and intentionally (dolus) uses an AI system to commit a crime, traditional intent requirements are still equally applicable as in other non-AI-related offences. For instance, using a deepfake generated by genAI tools to falsely represent someone may result in fraud or identity theft, provided that it is possible to prove that the actor intended to deceive another person(s). Another obvious example is that of a developer who deliberately introduces harmful code snippet into an AI system (malicious backdoor, crawler, and similar), and who could be liable for digital misuse if intent is proven. Croatia is a specific national example: Article 215.a of the Criminal Code, introduced by NN 136/2025, criminalises intentional or negligent endangerment of life or limb, or property of considerable scale through specified conduct involving an AI system [40]. Liability of a developer, manager, auditor, implementer, or deployer is not automatic; the conduct of a particular accused must satisfy the statutory offence and applicable attribution rules. Whistleblower material or commercial communications may be relevant evidence in an individual case but neither is a general method of proving intent, and dolus eventualis should not be presented as a civil-law analogue.

In cases of criminal negligence, different standards will apply pending domestic law variations. Depending on national law, offences, such as negligent homicide or bodily harm, may apply where a product defect resulting from inadequate care causes the legally required harm [41]. In general, main emphasis is placed on the control aspect: who is responsible for the risky action or lack of required action, and whether the action or lack of appropriate action led to a foreseeable consequence.

European Union's product-safety and consumer-protection rules may help define regulatory duties or expected conduct but they do not harmonise corporate criminal liability. Member States differ in whether and how individuals and legal persons may be held criminally, quasi-criminally, or administratively liable [41]. Should, for example, an auditor be induced to falsify AI findings, and should such facts become known, the conduct could be assessed under applicable criminal or administrative rules, but only if their constituent and attribution elements are proved.

7.4. Assigning the Blame

In the realm of AI-powered products or services, such determination is more complex, given the nature of algorithms, AI explainability, and lack of technical expertise (in certain areas)

and related technical standards (under continuous development). For example, the growing number of accidents involving motor vehicles controlled by (dynamic control) AI system opens a plethora of causation and foreseeability-related questions. The key issues under examination are centred on the role of AI navigation/advanced driving assistance systems developers and backend operators in vehicle operation. The principal question that requires practical determination is the realistic ability of vehicle users to control the behaviour of an AI system, assume control over the motor vehicle, and, ultimately, predict the behaviour of the AI system well in advance to resume control (automated transition request scenarios or voluntary control assumption) and prevent the occurrence of accident, or mitigate negative consequences. If the path to causation leads to the software (AI) development phase, a question may be raised whether there existed information or indication of foul play or inadequate quality and risk management checks and controls. Such evidence may raise questions about the conduct and responsibility of relevant actors, including auditors and implementers, within their defined roles.

To what extent it may be possible to attribute a substantive role of a developer, or potentially auditor/implementer, in the causation link leading to the damaging outcome, especially having in mind the proximate causation doctrine (necessity to establish a direct link), significantly depends on the ability of experts to understand, evaluate, and attest to such connection and accountability. The PLD provides limited disclosure and rebuttable-presumption mechanisms for civil product liability claims, but these do not resolve criminal proof. AIMS material may be relevant under RQ3, without determining criminal fault or causation. More precisely, damage suffered by a client due to an AI system whose safety-related risks were allegedly not adequately governed by the provider and where governance deficiencies were not identified during the audit could only potentially be linked to the auditor's negligence where it can be demonstrated that the auditor failed to apply a reasonable risk- and sampling-based audit approach, ignored clear objective indicators requiring further investigation, or otherwise departed from the expected professional standard of care. The mere fact that a governance deficiency was not detected during a sampling-based audit would not, in itself, constitute evidence of negligence.

However, as explained in previous text, the role of an auditor and implementer is not focused on product/service safety/quality but rather on processes and procedures employed by producers/service providers. This distinction directly addresses RQ1 and corresponds

to the distinction drawn in the recent AI auditing literature between governance audits of the organisation that develops or deploys a system, model audits of the system itself, and application audits of a specific deployed use case; of these, only the first falls within the object of an ISO/IEC 42001 certification audit [9]. Given the scope-bound and management-system nature of ISO/IEC 42001 certification audits, establishing a proximate causal link between audit findings and subsequent AI system malfunction would typically require evidence that the alleged failure relates directly to governance processes that were explicitly within the certified AIMS scope and demonstrably overlooked despite clear and objective indicators. The audit provides a time-bound, sampling-based assessment of evidence within the defined AIMS scope and does not guarantee continuing product or service quality. The applicable causation test varies by jurisdiction: a but-for inquiry may address factual causation, while legal causation or normative attribution requires a separate assessment. It would, hence, be necessary to link any actions or omissions on the side of auditors and implementers at the time of their service with the damage occurring during the scope of product/service lifecycle. In practice, it is very difficult to establish such direct links due to sheer number of other potentially relevant factors and causes.

The principal question to that end for the auditor is whether the auditor's contractual service stipulations promise a definitive (and time-durable) confirmation of validity, or whether they, as argued throughout this paper, point to a temporary confirmation of the existing evidence, or lack of evidence regarding certain indicators related to AIMS. The causation inquiry should assess whether, absent the alleged audit failure, the client would probably have acted differently and prevented or reduced the harm, subject to the applicable legal test. Given the nature of an ISO audit, as thoroughly examined through this paper, it would appear that, in practice, it would be very difficult to successfully create a valid proximate cause link.

The same question is much more complex regarding the implementers, where the implementation of risk management systems may prove to be detrimental on further behaviour of AI providers (or operators in general), as the client might generally rely on AIMS when placing the product/service on the market. In this scenario, the 'but-for' test would have to carefully analyse the role of implementer with that of developer and deployer, and assert, to what extent the role of implementer potentially differs from auditor's to the extent that this role becomes an irreplaceable component

of product/service placement (level on reliance on implementer's service as an important component of final product/service).

Coming back to the example of semi-autonomous motor vehicles, especially in high-impact cases involving fatalities (as, unfortunately, track record in the United States confirms), the investigation following such an incident may point to buggy sensor software (hardware–software error matrix). Such a defect has the potential to lead to manufacturer's strict liability in accordance with PLD, regardless of the level of due diligence on the side of motor vehicles or automated driving system producer, or fleet operator's conduct (especially if remote operation is available). From the perspective of criminal law, a CEO, CTO, fleet operator, or another person could face charges only under a specific national offence and where that person's relevant duty, conduct or omission, required fault, and causal or endangerment link are proved [42]. In such a scenario, the auditors and implementers would generally face significant challenges in being linked directly to the causation chain unless it can be demonstrated that governance deficiencies within the certified scope were clearly identifiable and materially connected to the harmful outcome.

The matter is further complicated with the above-mentioned AI explainability and the inter-connected AI-related foreseeability capacity. Bearing in mind the technological levels or readiness of many AI systems (many of them still situated in the early stages of development), and lack of mid- to long-term practice and experience in utilising such systems, it may prove to be very difficult to apply traditional standard of foreseeability for every misuse or failure. The latter is true not just for the developers and providers of AI-related services and products but also for the auditors and implementers who are faced with the challenges of evaluating and implementing novel systems and procedures for which relevant standards or performance evidence may still be incomplete. This is especially true for incidents arising outside of the reasonably contemplated utilisation scenarios or processes that (currently) are increasingly difficult to answer and resolve (such as, for example, the problem related to the hallucination of AI general models). Once a risk is documented, repetition may become more relevant to later foreseeability and fault assessments without automatically establishing liability.

For example, bias related to the inadequate level of data diversity representation present in a data training set may, if undetected and unconsidered, produce unwanted and potentially harmful consequences through AI generation. A widely reported illustration

is the image-classification incident in which photographs of individuals belonging to a particular ethnic group were assigned an animal label. Publicly available accounts attribute the outcome to the under-representation of certain groups in the training data set, and the provider's initial remedial step consisted of suppressing the affected labels rather than correcting the underlying data. Two features of the incident are analytically relevant here. First, the deficiency was located in a technical artefact, the training data set, which lies outside the object of an AIMS certification audit. Second, an AIMS audit could nonetheless be expected to establish whether the organisation maintained any documented process for assessing data representativeness and for escalating classification errors identified after release. The incident therefore illustrates how governance findings and technical defects operate at different layers of the framework proposed in this paper, and why evidence generated at one layer cannot be substituted for evidence required at another.

Following a documented incident, comparable risks may become more foreseeable, but the legal consequences remain case-specific. Auditors and implementers are increasingly under pressure to keep track of technological progress and developments in the AI realm to be able to recognise good and bad practices, and what is expected from diligent and reasonable AI systems' developers and providers. It is necessary to note that such expectations still remain limited to the auditor's competence in assessing AI-related governance processes within the defined scope of an AIMS, and do not extend to technical verification of AI model performance, algorithmic correctness, or product safety characteristics; this competence boundary directly informs RQ2. Absence of the noted qualification may lead to a failure to recognise instances where certain types of oversight related to risk management may, in turn, lead to allegations of negligence on the side of their clients, consequently raising questions of their own accountability. Such a conclusion triggers further contemplations related to the relevance of actual AI products and services produced or provided by the client, and the unusual relevance of such factors for provision of auditing and implementation services. Under the revised product-liability framework, post-market updates, upgrades, or connected services may be relevant where they remain within the manufacturer's control; this is not a general guarantee of all promised functions throughout the product lifecycle.

7.5. Exposure to Professional Liability

Professional liability generally depends on the duty undertaken by the professional, the applicable standard of care, breach,

causation, and damage, as determined by the governing contract and applicable law. Auditors and implementers may therefore face civil claims for professional negligence where those elements are established. The content of the applicable standard of care is jurisdiction- and contract-specific. In civil-law systems it may be shaped substantially by statutory and professional rules, while contractual undertakings may further define the required level of skill and care. In common-law systems, contractual formulations such as reasonable skill and care, reasonable endeavours, or best endeavours may also affect the content of the undertaking. Accordingly, the required standard should be assessed by reference to the applicable law, professional rules, the agreed scope of services, and the specific contractual undertaking.

In general, should an AI developer negligently design or implement an AI system that causes harm, they could potentially be sued for damage due to professional negligence. Equally so, should a physician use AI tools and AI-enhanced medical devices in everyday practice, the physician remains subject to the applicable professional standard of care [43]. Should a recommendation coming from the AI system be faulty, and provided that a physician 'blindly' follows such a recommendation (absence of independent medical judgement), resulting in an injury to a patient, the physician (not the AI) could potentially face malpractice claims. A related concern has already reached the Court of Justice of the European Union in the context of automated credit scoring. In SCHUFA Holding, the Court held that the automated generation of a probability value concerning a person's ability to meet future payment obligations constitutes an automated individual decision within the meaning of Article 22 of the general data protection regulation, where the lender draws strongly on that value [44]. The judgement does not address professional liability as such. It does, however, indicate a judicial willingness to look behind the formal allocation of decision-making authority and to ask who, in substance, determined the outcome, which is the same question that professional negligence analysis must answer.

The question that arises is whether, and to what extent, such a practitioner may reasonably rely on audit findings or implementation outputs indicating that governance mechanisms intended to support the safe and reliable operation of AI systems were implemented and maintained. The unresolved question, which future case law may clarify, is the extent to which an audit finding or implementation output may be relevant when liability is attributed for harm arising from an AI-enabled product or service supplied by

the client or a third party. It is also necessary to remind that the role of an AIMS implementer must also be distinguished from that of a certification body. While implementers may assist in designing governance structures, they do not assume operational control over AI systems unless explicitly contracted to do so. Liability assessment must, therefore, carefully distinguish between advisory support provided by the implementer and decision-making authority retained by the organisation.

If a physician blindly follows recommendations of an AI medical device that points to the necessity of conducting a surgery, in cases where AI system errs, the physician is expected to notice such an error to the extent that a reasonably good physician would notice the same. If a developer has tested the noted AI system prior to deployment, and found a higher level of false positive and false negative results, and despite knowing this proceeded to offer the product on the market, those facts may be relevant to knowledge, breach or fault, but do not by themselves establish deception or criminal intent. If an auditor or implementer identifies a lack of evidence supporting the existence or effectiveness of governance mechanisms intended to identify, assess, monitor or mitigate safety-related risks, and nevertheless issues positive audit findings or implementation outputs, their performance may very well fall under the scrutiny of adjudicatory examination of professional misconduct.

Many professions and professional associations impose (voluntary) ethical or competence standards (non-binding instruments that may become contractually binding or evidentially relevant when incorporated into applicable contractual terms). In direct reference to the AI systems' development, engineers and IT professionals are regularly bound by codes of practice (such as IEEE Code of Ethics, ISO standards, American National Standards Institute [ANSI], and similar) that insist on prioritising safety, supervision, risk mitigation, and reporting. In line with the previous examples, lawyers and physicians have confidentiality and diligence guidelines and ordinances that are slowly starting to incorporate new technologies. Failure to meet such standards, pending internal acts, and contractual stipulations can constitute a breach of professional duty that, in itself, and depending on the jurisdictions, constitutes either a direct breach of contract (common law) or represents a key element in establishing a breach of contract (continental law). The European ethics guidelines for trustworthy AI [45] and industry standards (although non-binding in principle) also set expectations of transparency and accountability, especially when organisations

voluntarily incorporate them into their general terms and conditions, and/or require their implementation as a prerequisite for entering a certain guild/association.

ISO/IEC 42001:2023 and National Institute of Standards and Technology's (NIST) AI risk management framework [46] are prime examples of frameworks that organisations can use to support structured AI governance without thereby demonstrating compliance with all ethical, technical, or legal requirements, as discussed in detail throughout this paper. Auditors, implementers, and certification bodies working within their respective AIMS roles could, as noted earlier, also potentially be liable if their performance falls below professional standards. Under ISO/IEC 42006:2025, certification bodies auditing AIMS are required to fulfil the relevant competence requirements (including ISO 17021-1). If an AIMS certification auditor negligently assesses the conformity of the AIMS within its defined scope (e.g. a failure to identify a clear absence of required governance processes or documented risk controls may be relevant to an alleged breach of professional duty), liability could potentially arise under general professional negligence principles.

In general, the concept of due diligence permeates AI governance. When the relevant AI Act provisions become applicable under the amended phased timetable, high-risk operators will be subjected to specified mandatory obligations. Compliance may then be evidence-relevant to an assessment of due diligence [3,28,47,48]. A caveat is warranted here. Regulatory compliance and trustworthiness are not interchangeable notions: the AI Act operationalises trustworthiness largely as the acceptability of residual risk, so evidence of compliance speaks to the quality of risk management rather than to the absence of harm [49,50]. Compliance evidence should accordingly be assessed for what it demonstrates about the process followed, not as a proxy for a safe outcome. Courts will likely assess whether all reasonable precautions were taken and safety mechanisms were introduced (reasonable measures, in line with the risk tolerance principle where mitigation measures must be proportional to potential probability and scope of harm). Existence of a thorough audit trail, adherence to relevant technical, legal, and ethical standards, and established response mechanisms may be relevant to, but does not determine, liability. Conversely, absence of documentation or ignoring best practices (e.g. lack of impact assessment) may point to less diligent, potentially negligent behaviour.

The diligent audit trail must demonstrate the systemic approach as promulgated by ISO/IEC 42001 [2]. Potentially multi-stage audit(s)

must examine the readiness of policies and documentation, and evaluate actual implementation and control effectiveness [5–7]. Successful findings may provide evidence relevant to the assessment of organisational due diligence. Where the relevant high-risk provisions apply, providers must keep up-to-date risk assessments (AI Act Article 9). Where relevant to the audit objectives and within scope, auditors may examine records of hazard analyses and mitigation actions. Confirming through evidence that risks were systematically identified and addressed may support the provider's argument that reasonable governance measures were implemented. Furthermore, the developers are encouraged to maintain an 'audit trail' of system design and maintenance, including documentation on training data sources, model scope, limitations, and update procedures and history. In potential litigation, these records could help establish the facts concerning what was known and done at what stage of product development and deployment. An analysed and qualified risk that results in controlled and mitigated damage (known risk, or system limitation) may result in a more favourable adjudication outcome. Similarly, update and AI model re-training logs, especially those connected to the incident response procedures, might prove particularly relevant when adjudicators determine whether an implementer has adequately addressed the emerging threats (e.g. analysis of security vulnerabilities and implementation of appropriate mitigation measures). Hence, it may be surmised that comprehensive documentation reviewed by auditors or prepared by implementers may serve as evidence relevant to the assessment of whether implementers and operators exercised an appropriate standard of due care, without itself eliminating or reducing liability exposure. This evidence addresses RQ1, while the following competence analysis addresses RQ2. The competence requirements of ISO/IEC 42006, read together with the audit methodologies described in the assurance literature [19,20,23], supply reference points against which a court could assess whether a reasonably competent auditor would have pursued a given objective indicator further. What they do not supply is a duty to verify technical performance, and the absence of such a duty remains the principal limit on the exposure of auditors.

7.6. Civil Liability for Damage

The following discussion uses Croatian civil law as an illustrative example of broader European liability principles applicable to AI-related damage.

The concept of damage occupies a central place in civil law, particularly in tort law, precisely because the existence of damage

represents a *conditio sine qua non* for the obligation to provide compensation. In a broader sense, damage refers to any reduction of someone's property or the violation of personal rights and interests, and it is defined in Article 1046 of the Croatian Civil Obligations Act [14] as: a reduction of someone's property (actual damage), prevention of its increase (lost profit), and infringement of personal rights (non-pecuniary damage) [51–58].

Identifying the type of damage is not only of theoretical significance but also directly affects judicial practice, since the method of compensation depends on the type of damage [19]. As Klarić and Vedriš emphasise, a systematic classification of damage enables consistent application of the principle of monetary compensation and also ensures that arbitrariness in court assessments is avoided [54].

The primary and most practically significant classification is that of material (pecuniary and property) and non-material (non-pecuniary and non-property) damage. Material damage encompasses any reduction or prevention of an increase in the victim's property, while non-material damage refers to violations of personal rights [58]. Pecuniary damage, according to the Croatian Obligations Act (*Zakon o obveznim odnosima* or ZOO), manifests in the form of actual damage (*damnum emergens*) and lost profit (*lucrum cessans*). Non-material damage is not expressed in the property (monetary) sphere but through the violation of an individual's moral, physical, and spiritual values. The Croatian Obligations Act, in Article 1046, defines it as a violation of personal rights under Article 19 of the Act. For such a relationship to arise at all, the following cumulative conditions must be met: subjects of the obligation – the injured party and the tortfeasor, harmful act, occurrence of damage, causal link (causal nexus) between the act and the damage, illegality (unlawfulness).

In the case of subjective or fault-based liability, in addition to the general conditions, the fault of the tortfeasor (intent or negligence) is also required [59]. The basic rule is that fault is presumed (Article 1045, paragraph 1 of the Croatian Obligations Act), so it is up to the injured party to prove, in support of their claim, the harmful act, the damage, the causal link, and the elements of unlawfulness. The tortfeasor, on the other hand, can exonerate themselves by proving that the damage occurred without their fault, or that one of the basic conditions is missing. On the other hand, in the case of strict (objective) liability, fault is not required [56]. That is, liability arises solely on the basis of the fact that the general conditions are met (harmful act, damage, causal link, and unlawfulness). Strict liability

must rest on a specific statutory basis and remains subject to the elements and defences of that regime.

The development of autonomous vehicles and AI-based systems brings significant advantages in terms of safety, efficiency, and mobility. However, with the increasing functionality of autonomous systems, new legal challenges also emerge, particularly in the area of civil liability for damage caused by such vehicles. The traditional legal framework of liability, based on fault or the principle of strict product liability, faces challenges due to the complexity of algorithms, the unpredictable behaviour of systems, and their interaction with users and the environment [55,60].

The following autonomous vehicle illustrations distinguish product liability from possible operator and professional liability. They are analytical examples, not predictions of how a court would allocate responsibility in a particular case.

This paper focuses on the legal analysis of cases in which autonomous vehicles cause damage to third parties or property, and examines the liability of manufacturers, operators, implementers, and AIMS certification auditors. Special attention is given to the principle of strict liability under PLD and the role of professional duty of care (due diligence) in the governance implementation and AIMS certification. The aim is to demonstrate how civil liability can be applied in practice, taking into account the technical, operational, and legal aspects of using autonomous vehicles [61].

Civil liability for damage caused by an AI-operated vehicle is based on the principles of strict product liability and liability for professional negligence. Under the revised EU PLD, a economic operator may be held liable without proof of fault only where a defective product caused compensable damage and the other statutory conditions are met. Software is included within the definition of a product, and an AI system is covered insofar as it constitutes software or is integrated into another product. Accordingly, where an autonomous vehicle allegedly causes damage, a claim may be brought against the manufacturer or, where applicable under the directive, another liable economic operator, provided that defectiveness, compensable damage, and causation are established. An erroneous AI output does not by itself establish defectiveness, causation, or liability.

For example, imagine a situation in which an autonomous vehicle of brand X, equipped with an AI-based vehicle control system and an

advanced driver assistance system, fails to stop in front of a pedestrian due to a sensor algorithm error. The pedestrian is injured, and the vehicle suffers property damage; under the revised PLD, damage to the allegedly defective product itself is excluded from compensable property damage. From a civil liability perspective, the incident raises the question whether the malfunction constitutes a product defect and whether it caused damage attributable to a liable economic operator. If it is established that the manufacturer failed to carry out an adequate risk assessment, omitted appropriate safety testing, or failed to implement necessary software updates, such omissions may constitute relevant evidence in assessing defectiveness. However, strict product liability arises only where the statutory conditions, including defectiveness, compensable damage, and causation, are satisfied [62–64].

An AIMS auditor or implementer is not a manufacturer merely by performing that role. Potential professional liability would require a duty arising under the applicable contract or law, a breach of that duty and the fulfilment of the other legal requirements necessary to establish liability for damage. Audit trails, testing records, and maintenance documentation may help to establish what the relevant actors knew and did; such records do not automatically excuse a manufacturer or operator. This distinction is relevant to RQ3 because AIMS evidence may be relevant without being determinative.

On the other hand, AIMS auditors and implementers, whose work concerned governance processes within their agreed scope, are not directly liable for the resulting damage unless it is shown that they ignored obvious deficiencies in the risk management system and thereby breached their professional duty of care (due diligence). Such records may be relevant to either party; they do not automatically mitigate liability or demonstrate that all reasonable measures were taken.

This example illustrates how product liability may arise from a technical defect when the statutory conditions are met, whereas professional liability of implementers and auditors requires an applicable duty, breach, causation, and damage. Such an approach allows for a clear distinction between product liability and liability for oversight and management, which is essential for the proper allocation of legal responsibility in the event of damage.

Another example involving an autonomous taxi and civil liability is analysed below. An autonomous taxi of brand Y is driving through

city streets without a human driver. During the ride, the pedestrian recognition system incorrectly interprets a group of people at a crosswalk, causing the vehicle to continue moving and strike a pedestrian. The pedestrian is seriously injured, and the taxi sustains property damage.

From a civil liability perspective, a claim against the vehicle manufacturer or software developer depends on whether each is a liable economic operator and whether defectiveness, damage, and causation are established. Since the AI system is considered high-risk, the manufacturer had a legal obligation to assess and mitigate risks, monitor safety tests, and update the software. Such omissions may be relevant evidence, but high-risk classification or non-compliance does not by itself establish defectiveness, causation, or PLD liability [57].

On the other hand, the vehicle operator (if remote control is available) may also be liable if proper supervision or timely intervention was not ensured, whereas an AIMS implementer's or auditor's exposure depends on the role-specific duty, breach, causation, and damage established under the applicable law. For instance, an unreasonable failure to pursue clear objective evidence of an absent risk-governance process within the defined AIMS scope may be relevant to a professional negligence claim; technical assessment of the safety model itself is not ordinarily the object of AIMS certification.

This example demonstrates how different actors in the responsibility chain may bear civil liability: manufacturers for technical defects, operators for operational failures, and implementers/auditors only in the context of proven negligence in process management. In practice, documentation and record-keeping may help to distinguish roles, but they do not themselves determine liability or reduce litigation risk.

In the context of AI applications, the most difficult aspect is establishing the causal link. The problem of causation is not limited to identifying a cause [60] but also to selecting the correct one among the multitude of events that may be related to the resulting damage. Damage is almost never the result of a single cause; it is the outcome of a complex interplay of various circumstances. For this reason, it is important to identify the cause that is legally relevant, that is, the one that is commonly and specifically associated with the occurrence of the particular damage. How to decide which direction to take among a multitude of potential possibilities? Croatian legal doctrine has, in this regard, adopted the adequacy theory of

causation. According to the adequacy theory, among the various events that could be considered as causes, only the event that is specific to the occurrence of the particular damage is regarded as the cause. Specificity is determined based on life experience; if a particular event regularly leads to a concrete harmful consequence, it is considered a legally relevant cause.

The adequacy theory certainly does not provide an absolute and infallible measure. However, its requirement of specificity (typicality) offers a practical criterion for selecting the cause in complex real-life situations. Legally relevant causation may concern acts, omissions, product defects, dangerous things, or dangerous activities, depending on the legal basis. The burden of proving the causal link, since it is not presumed, falls on the injured party. Article 1063 of the Croatian Obligations Act provides a specific presumption for damage arising in connection with a dangerous thing or activity, unless the contrary is proved; its application to a particular AI-enabled vehicle or activity is case-specific [14].

Finally, the court determines the legally relevant causal link and applicable basis of liability. Strict defective-product liability applies only where the statutory conditions and defences under national law implementing the PLD are satisfied [14].

In response to RQ3**,** the analysis demonstrates that, in the context of AI and civil liability, the establishment of causation is one of the most challenging legal issues. It is essential to determine the legally relevant role of each participant, the degree of that participant's control over the relevant event, and their contribution to the occurrence of the damage. AIMS documentation may be relevant, without being determinative, in distinguishing these roles and reconstructing the decision-making process.

8. Evidentiary Relevance of AIMS Documentation and Audits in Regulatory and Judicial Contexts

The AIMS material should be evaluated through four cumulative factors: scope, timing, factual nexus, and corroboration. Policies, role descriptions, risk registers, and impact assessments may show how an organisation intended to allocate responsibility and address foreseeable risks. Their probative value increases only where the documented process covered the system and event in question, existed at the relevant time, and is consistent with the records of actual implementation.

Third-party audit reports add independent, but still limited, evidence. Their weight depends on the sampled evidence, audit scope, timing, identified non-conformities, and competence of the team. A **r**esponsible, **a**ccountable, **c**onsulted, and **i**nformed (RACI) matrix or certificate shows intended governance, not necessarily actual conduct. A regulator or court should therefore examine discrepancies between documented and observed practice and consider the material alongside technical investigation, witness evidence, and the governing legal test.

9. Discussion

The analysis directly answers the three research questions stated in the Introduction. RQ1 identifies the organisational AIMS and the scope-bound governance and security-related evidence assessed in third-party certification audits. RQ2 shows that multidisciplinary competence may strengthen the credibility of audit evidence without overcoming sampling, timing, and scope limitations. RQ3 proposes four contextual factors (scope, timing, factual nexus, and corroboration) for assessing the weight of AIMS material in regulatory, civil, criminal, professional-liability, and product-liability contexts, without treating that material as determinative.

These findings preserve the distinction between AI governance, conformity assessment, technical validation, and liability attribution. They also explain the framework's theoretical contribution: evidence can move across the governance, assurance, and legal layers, but a conclusion at one layer does not automatically determine another. A certificate is therefore an evidence of management-system conformity within a defined scope and at a particular time, not proof of technical correctness, cybersecurity resilience, product safety, legal compliance, or absence of liability.

Methodologically, the *governance-evidence-liability* framework makes that movement of evidence traceable through standards interpretation and doctrinal analysis. Practically, it directs certification bodies to define scope and evidence limits clearly, organisations to preserve records of actual implementation rather than certificates alone, and regulators or courts to test AIMS material against the specific system, event, and applicable legal test. ISO/IEC 42006 multidisciplinary competence may support more informed assessment, but it strengthens professional judgement within management-system auditing and does not itself expand the audit mandate into technical product assessment, product certification, or legal adjudication.

10. Conclusions

This paper examined ISO/IEC 42001 and ISO/IEC 42006 through the *governance-evidence-liability* framework. It concludes that a third-party AIMS audit assesses organisational governance through scope-bound, sampled, and time-bound governance and security-related evidence (RQ1); that multidisciplinary competence can strengthen the credibility of that evidence without eliminating scope, sampling, and timing limitations (RQ2); and that the evidentiary weight of AIMS records depends on scope, timing, factual nexus, and corroboration, while their legal effect remains governed by the applicable regime (RQ3).

ISO/IEC 42001 certification can support organisational oversight, documented risk management, and informed decision-making. It cannot by itself establish technical performance, cybersecurity resilience, product safety, legal conformity, causation, or liability. That restrained interpretation is particularly important in critical business processes, public services, and safety- or security-sensitive sectors. Empirical research should now test how certification bodies apply ISO/IEC 42006 competence requirements and how regulators, contracting parties, and courts actually use AIMS evidence.

Acknowledgements

To ensure clarity and consistency in English, AI-based translation tools were used for language translation and style polishing. The content, structure, and scientific contributions remain the sole work of the authors.

Conflict of Interest

The authors declared no conflict of interest.

References

[1] International Organization for Standardization/International Electrotechnical Commission (ISO/IEC), *Information Technology – Artificial Intelligence – Concepts and Terminology (ISO/IEC 22989:2022)*. Geneva: ISO, 2022.

[2] International Organization for Standardization/International Electrotechnical Commission (ISO/IEC), *Artificial Intelligence – Management System (ISO/IEC 42001:2023)*. Geneva: ISO, 2023.

[3] *AI Act. Regulation 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending certain Union legislative acts (Artificial Intelligence Act).* [Online]. Available: http://data.europa.eu/eli/reg/2024/1689/oj. [Accessed: Feb 18, 2026].

[4] European Commission. (2025). Commission guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act), C(2025) 924 final. [Online]. Available: https://ec.europa.eu/newsroom/dae/redirection/document/112455. [Accessed: Feb 18, 2026].

[5] International Organization for Standardization/International Electrotechnical Commission (ISO/IEC), *Conformity Assessment – Requirements for Bodies Auditing and Certifying Artificial Intelligence Management Systems (ISO/IEC 42006:2025)*. Geneva: ISO, 2025.

[6] International Organization for Standardization/International Electrotechnical Commission (ISO/IEC), *Conformity Assessment – Requirements for Bodies Providing Audit and Certification of Management Systems – Part 1 (ISO/IEC 17021-1:2015)*. Geneva: ISO, 2015.

[7] International Organization for Standardization (ISO), *Guidelines for Auditing Management Systems (ISO 19011:2026)*. Geneva: ISO, 2026.

[8] J. Mökander, "Auditing of AI: Legal, ethical and technical approaches," *Digital Society*, vol. 2, Art. no. 49, 2023, doi: 10.1007/s44206-023-00074-y.

[9] J. Mökander, J. Schuett, H.R. Kirk, L. Floridi, "Auditing large language models: A three-layered approach," *AI and Ethics*, vol. 4, pp. 1085–1115, 2024, doi: 10.1007/s43681-023-00289-2.

[10] C. Novelli, M. Taddeo, L. Floridi, "Accountability in artificial intelligence: What it is and how it works," *AI & Society*, vol. 39, pp. 1871–1882, 2024, doi: 10.1007/s00146-023-01635-y.

[11] J. Laine, M. Minkkinen, M. Mäntymäki, "Ethics-based AI auditing: A systematic literature review on conceptualizations of ethical principles and knowledge contributions to stakeholders," *Information & Management*, vol. 61, no. 5, Art. no. 103969, 2024, doi: 10.1016/j.im.2024.103969.

[12] European Commission. (2025). *Supporting the implementation of the AI Act with clear guidelines*. [Online]. Available: https://digital-strategy.ec.europa.eu/en/news/supporting-implementation-ai-act-clear-guidelines. [Accessed: Feb 18, 2026].

[13] *Directive (EU) 2024/2853 of 23 October 2024 on liability for defective products*. [Online]. Available: https://eur-lex.europa.eu/eli/dir/2024/2853/oj. Corrigendum of 7 May 2026. http://data.europa.eu/eli/dir/2024/2853/corrigendum/2026-05-07/oj. [Accessed: Feb 18, 2026].

[14] Republic of Croatia. *Croatian Civil Obligations Act, Official Gazette Nos. 35/05, 41/08, 125/11, 78/15, 29/18, 126/21, 114/22, 156/22,155/23, 69/26*. Zagreb: *Supreme Court of the Republic of Croatia.*

[15] International Organization for Standardization (ISO), *Risk Management – Guidelines (ISO 31000:2018)*. Geneva: ISO, 2018.

[16] International Organization for Standardization/International Electrotechnical Commission (ISO/IEC), *Artificial Intelligence – Impact Assessment of AI Systems (ISO/IEC 42005:2025)*. Geneva: ISO, 2025.

[17] International Organization for Standardization/International Electrotechnical Commission (ISO/IEC), *Information Technology – Artificial Intelligence – Guidance on Risk Management (ISO/IEC 23894:2023)*. Geneva: ISO, 2023.

[18] I.D. Raji, A. Smart, R.N. White, M. Mitchell, T. Gebru, et al. "Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing," in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. 2020, pp. 33–44, doi: 10.1145/3351095.3372873., New York, NY, USA: Association for Computing Machinery, 2020

[19] J. Mökander, M. Axente, F. Casolari, L. Floridi, "Conformity assessments and post-market monitoring: A guide to the role of auditing in the proposed European AI regulation," *Minds and Machines*, vol. 32, no. 2, pp. 241–268, 2022, doi: 10.1007/s11023-021-09577-4.

[20] A. Koshiyama, E. Kazim, P. Treleaven, P. Rai, L. Szpruch, et al. "Towards algorithm auditing: Managing legal, ethical and technological risks of AI, ML and associated algorithms," *Royal Society Open Science*, vol. 11, no. 5, Art. no. 230859, 2024, doi: 10.1098/rsos.230859.

[21] A. Birhane, R. Steed, V. Ojewale, B. Vecchione, I.D. Raji, "AI auditing: The broken bus on the road to AI accountability," in *2024 IEEE Conference on Secure and Trustworthy Machine Learning (SaTML)*. 2024, pp. 612–643, doi: 10.1109/SaTML59370.2024.00037.

[22] M. Mäntymäki, M. Minkkinen, T. Birkstedt, M. Viljanen, "Defining organizational AI governance," *AI and Ethics*, vol. 2, pp. 603–609, 2022, doi: 10.1007/s43681-022-00143-x.

[23] M. Minkkinen, J. Laine, M. Mäntymäki, "Continuous auditing of artificial intelligence: A conceptualization and assessment of tools and frameworks," *Digital Society*, vol. 1, no. 3, Art. no. 21, 2022, doi: 10.1007/s44206-022-00022-2.

[24] C. Novelli, F. Casolari, A. Rotolo, M. Taddeo, L. Floridi, "Taking AI risks seriously: A new assessment model for the AI Act," *AI & Society*, vol. 39, pp. 2493–2497, 2024, doi: 10.1007/s00146-023-01723-z.

[25] J. Schuett, "Risk management in the Artificial Intelligence Act," *European Journal of Risk Regulation*, vol. 15, no. 2, pp. 367–385, 2024, doi: 10.1017/err.2023.1.

[26] D. Hartmann, J.R. Laranjeira de Pereira, C. Streitbörger, B. Berendt, "Addressing the regulatory gap: Moving towards an EU AI audit ecosystem beyond the AI Act by including civil society," *AI and Ethics*, vol. 5, pp. 3617–3638, 2025, doi: 10.1007/s43681-024-00595-3.

[27] N. Parlov, G. Akrap, J. Esterhajer, "Supply chain security and AI risk governance model for critical infrastructure under NIS2, CER, and CRA," Applied Cybersecurity & Internet Governance, vol. 4, no. 1, pp. 147-184, 2025, doi: 10.60097/ACIG/211823.

[28] *Regulation (EU) 2026/1744 of the European Parliament and of the Council of 8 July 2026 amending regulations (EU) 2024/1689, (EU) 2018/1139 and (EU) 2023/1230 as regards the simplification of the implementation of harmonised rules on artificial intelligence (Digital Omnibus on AI).* [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32026R1744. [Accessed: Jul 28, 2026].

[29] European Commission. (2022). *Liability rules for artificial intelligence*. [Online]. Available: https://commission.europa.eu/topics/business-and-industry/doing-business-eu/contract-rules/digital-contracts/liability-rules-artificial-intelligence_en. [Accessed: Jun 05, 2026].

[30] P. Hacker, "The European AI liability directives – Critique of a half-hearted approach and lessons for the future," *Computer Law & Security Review*, vol. 51, p. 105871, 2023, doi: 10.1016/j.clsr.2023.105871.

[31] S. Wachter, "Limitations and loopholes in the EU AI Act and AI liability directives: What this means for the European Union, the United States, and beyond," *Yale Journal of Law & Technology*, vol. 26, no. 3, pp. 671–718, 2024. [Online]. Available: https://law.yale.edu/sites/default/files/area/center/isp/documents/wachter_26yalejltech671.pdf. [Accessed: Jun 05, 2026].

[32] European Commission. (2025). *Standardisation request to the European standardisation organisations in support of Regulation (EU) 2024/1689 on artificial intelligence, C(2025) 3871*. [Online]. Available: https://ec.europa.eu/transparency/documents-register/detail?ref=C(2025)3871&lang=en. [Accessed: Jun 15, 2026].

[33] European Commission. (2025). *AI Act standardisation and harmonised standards.* [Online]. Available: https://digital-strategy.ec.europa.eu/en/policies/ai-act-standardisation. [Accessed: Jun 15, 2026].

[34] European Commission. (2022). "The 'Blue Guide' on the implementation of EU product rules (2022 edition)," *Official Journal of the European Union*, vol. C247/01. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:C:2022:247:TOC. [Accessed: Mar 10, 2026].

[35] CEN-CENELEC Joint Technical Committee 21. (JTC 21). *Overview and Architecture of Standards in Support of the EU Artificial Intelligence Act*. Brussels: CEN-CENELEC.

[36] CEN-CENELEC. *Artificial intelligence standardisation – CEN-CENELEC JTC 21*. [Online]. Available: https://www.cencenelec.eu/areas-of-work/cen-cenelec-topics/artificial-intelligence/. [Accessed: Jun 15, 2026].

[37] European Commission. (Jul. 27, 2026). *Navigating the AI Act. Shaping Europe's digital future*. [Online]. Available: https://digital-strategy.ec.europa.eu/en/faqs/navigating-ai-act. [Accessed: Jul 28, 2026].

[38] M. Almada, N. Petit, "The EU AI Act: Between the rock of product safety and the hard place of fundamental rights," *Common Market Law Review*, vol. 62, no. 1, pp. 85–120, 2025. [Online]. Available: https://kluwerlawonline.com/journalarticle/Common%2BMarket%2BLaw%2BReview/62.1/COLA2025004. [Accessed: Jun 12, 2026].

[39] C. Novelli, F. Casolari, P. Hacker, G. Spedicato, L. Floridi, "Generative AI in EU law: Liability, privacy, intellectual property, and cybersecurity," *Computer Law & Security Review*, vol. 55, p. 106066, 2024, doi: 10.1016/j.clsr.2024.106066.

[40] Republic of Croatia, Act on Amendments to the Criminal Code (Zakon o izmjenama i dopunama Kaznenog zakona), Narodne novine, no. 136/2025, Art. 11 (inserting Art. 215.a), 2025. [Online]. Available: https://narodne-novine.nn.hr/clanci/sluzbeni/2025_11_136_2018.html. [Accessed: Jun 12, 2026].

[41] S. Kämpfer, A. Kreis, P.N. Kroner, T. Lauth. (Feb. 20, 2025). *Criminal product liability in Germany: Impact of the new EU product liability directive and product safety regulation. Freshfields risk & compliance*. [Online]. Available: https://riskandcompliance.freshfields.com/post/102k13a/criminal-product-liability-in-germany-impact-of-the-new-eu-product-liability-dir. [Accessed: Jun 12, 2026].

[42] M. Mudrić, "Advanced and autonomous driving systems in motor vehicles: Paving the responsibility road from drivers to algorithm systems,"

in Transporte, Competencia y Nuevas Tecnologías, M. V. Petit Lavall and A. Puetz, Dirs., S. Boboc, Coord. Madrid, Spain: Marcial Pons, Ediciones Jurídicas y Sociales, 2022, pp. 173–199. [Online]. Available: https://dialnet.unirioja.es/servlet/articulo?codigo=9226858. [Accessed: Jun 12, 2026].

[43] W.N. Price II, S. Gerke, I.G. Cohen, "Potential liability for physicians using artificial intelligence," *Journal of the American Medical Association*, vol. 322, no. 18, pp. 1765–1766, 2019, doi: 10.1001/jama.2019.15064.

[44] *Judgment of 7 December 2023, SCHUFA Holding (Scoring), Case C-634/21, ECLI:EU:C:2023:957*. [Online]. Available: https://curia.europa.eu/juris/liste.jsf?num=C-634/21. [Accessed: Mar 15, 2026].

[45] European Commission. (2019). *High-level expert group on artificial intelligence. Ethics guidelines for trustworthy AI*. [Online]. Available: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai. [Accessed: Mar 15, 2026].

[46] National Institute of Standards and Technology. *Artificial Intelligence Risk Management Framework (AI RMF 1.0) NIST AI 100-1*. Washington DC: US Department of Commerce, 2023, doi: 10.6028/NIST.AI.100-1.

[47] European Commission. (2020). *Report on the safety and liability implications of artificial intelligence, the Internet of Things and robotics, COM(2020) 64 final*. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0064. [Accessed: Mar 15, 2026].

[48] European Parliament. (2020). *Resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence (2020/2014(INL))*. [Online]. Available: https://www.europarl.europa.eu/doceo/document/TA-9-2020-0276_EN.html. [Accessed: Mar 15, 2026].

[49] J. Laux, S. Wachter, B. Mittelstadt. "Trustworthy artificial intelligence and the European Union AI Act: On the conflation of trustworthiness and acceptability of risk," *Regulation & Governance*, vol. 18, no. 1, pp. 3–32, 2024, doi: 10.1111/rego.12512.

[50] M. Mudrić, "Semi-automated and automated vehicles, insurance, and private international law," in *Insurance in Private International Law*, M. Wałachowska, M. Fras, P. Marano, Eds. Cham: Springer, 2024, pp. 485–516, doi: 10.1007/978-3-031-74511-9_19.

[51] European Commission. *AI Act service desk and AI Act single information platform*. [Online]. Available: https://ai-act-service-desk.ec.europa.eu/. [Accessed: Jun 15, 2026].

[52] H. Kačer, A. Perkušić. In *Commentary on the Civil Obligations Act*, V. Gorenc et al., Eds. Zagreb: RRiF plus, 2005, p. 1605.

[53] H. Kačer, A. Radolović, Z. Slakoper, *Law on Obligations with Commentary*. Zagreb: Poslovni Zbornik, 2006.

[54] P. Klarić, M. Vedriš, *Civil Law*. Zagreb: Narodne Novine, 2025.

[55] M. Ebers, "Liability for artificial intelligence and EU consumer law," *Journal of Intellectual Property, Information Technology and E-Commerce Law*, vol. 12, no. 2, pp. 204–220, 2021.

[56] Z. Slakoper, S. Nikšić, *New Regulation of Liability for Material Defects in Croatian Obligations Law.* Rijeka: Faculty of Law, University of Rijeka, vol. 43, no. 3, 2022, pp. 531–558.

[57] I. Crnić, J. Matić, *Tort Law – Collection of Court Decisions.* Zagreb: Zgombić & Partneri, 2008.

[58] B. Mateša, *Artificial Intelligence and Some Legal Issues, Current Issues of Croatian Legislation and Legal Practice, Yearbook 32*. Zagreb: Organizator, 2025.

[59] I. Grbin, "Liability assumptions and fault-based liability," Naša Zakonitost, no. 9–10, 1987, p. 978. (Zagreb: Savez društava pravnika Hrvatske, Udruženje za upravne znanosti i praksu Hrvatske, Republički zavod za javnu upravu SR Hrvatske.)

[60] J. De Bruyne and W. Ooms, "Tort liability and artificial intelligence: Some challenges and (regulatory) responses," in The Cambridge Handbook of the Law, Ethics and Policy of Artificial Intelligence, N. A. Smuha, Ed. Cambridge, UK: Cambridge University Press, 2025, pp. 158–173, doi: 10.1017/9781009367783.011.

[61] M. Proso, "Reform of the rules on liability for defective products," in *Proceedings of the 10th International Conference on Current Issues of Civil Procedure Law – National and Comparative Theoretical and Practical Achievements*, D. Šago, V. Rijavec, and A. Galič, Eds. Split, Croatia: Faculty of Law, University of Split, 2024, pp. 209–231.

[62] N. Parlov, B. Mateša, and A. Mladinić, "Structuring AI risk management framework: EU AI Act FRIA, GDPR DPIA and ISO 42001/23894," in 2025 14th Mediterranean Conference on Embedded Computing (MECO), R. Stojanović, L. Jóźwiak, and B. Lutovac, Eds. Piscataway, NJ, USA: IEEE, 2025, pp. 1–8, doi: 10.1109/MECO66322.2025.11049196.

[63] P. Hense, T. Mustać, *AI Act Compact: Compliance, Management & Use Cases in Corporate Practice.* Frankfurt am Main: Fachmedien Recht und Wirtschaft, dfv Medien Gruppe, 2025.

[64] I. Crnić, *Obligations Act: Extensive Case Law, Notes and Comments, Detailed Alphabetical Index of Terms*. Zagreb: Organizator, 2016.