https://research.adra.ac.id /index.php/rjl/
P - ISSN: 2988-4454
E - ISSN: 2988-4462

The Algorithmic Accountability Paradox: Rethinking Corporate Criminal Liability for Autonomous Decision-Making Systems

Ren Suzuki[1], Daiki Nishida[2], Luis Santos[3]
[1]Nagoya University, Japan
[2]Chuo University, Japan
[3]University of the Philippines Diliman, Philippines

ABSTRACT

Background. Autonomous decision-making systems increasingly shape corporate operations, challenging criminal liability doctrines traditionally grounded in identifiable human conduct, knowledge, and culpability.

Purpose. This study examines how corporate criminal responsibility can remain legally defensible when harmful algorithmic outcomes emerge without direct human authorization.

Method. Employing qualitative doctrinal research, comparative legal analysis, and normative inquiry, the study examines United States corporate attribution principles, United Kingdom corporate offence frameworks, and European Union artificial intelligence regulation. Documentary materials are interpreted through the dimensions of attribution, culpability, causation, organizational control, and evidentiary safeguards.

Results. The analysis identifies an accountability paradox: corporations may retain substantial authority over system selection, deployment, monitoring, and intervention while lacking direct knowledge of particular algorithmic decisions. Comparative findings distinguish criminal attribution from preventive regulatory compliance and show why technological opacity neither automatically defeats liability nor establishes corporate fault.

Conclusion. The framework offers a principled basis for future legislative clarification and judicial assessment while preserving legality, proportionality, and procedural fairness in increasingly automated corporate environments. Its conclusions remain provisional pending empirical and judicial validation.

KEYWORDS

Algorithmic Accountability, Artificial Intelligence, Legal Responsibility

Citation: Suzuki, R., Nishida, D & Santos, L. (2026). The Algorithmic Accountability Paradox: Rethinking Corporate Criminal Liability for Autonomous Decision-Making Systems. *Rechtsnormen Journal of Law*, *4*(4), 334–350.
https://doi.org/10.70177/rjl.v4i4.4579

Correspondence:

Ren Suzuki,
rensuzuki@gmail.com

Received: April 16, 2026

Accepted: July 12 2026

Published: October 27, 2026

INTRODUCTION

The increasing integration of artificial intelligence into corporate governance has fundamentally transformed how organizations exercise authority, allocate resources, and manage operational risks. Autonomous decision-making systems now influence financial transactions, regulatory compliance, employment practices, and strategic management. Their growing independence from direct human intervention challenges conventional assumptions about organizational responsibility. Corporate criminal law, historically grounded in human conduct and institutional attribution, consequently faces unprecedented

questions concerning accountability when algorithmic decisions produce legally significant harm (Chaudhary, 2024; Vasudevan, 2025).

Autonomous systems introduce distinctive complexities into corporate criminal responsibility because their decisions may emerge from adaptive processes that neither developers nor executives fully anticipate. Machine learning architectures can generate outcomes through interactions among training data, optimization mechanisms, and operational environments. Such technological characteristics complicate the identification of culpable conduct, particularly where harmful outcomes arise without explicit managerial instructions. The resulting accountability dilemma exposes tensions between technological autonomy and legal doctrines requiring demonstrable connections between organizational behavior and criminal wrongdoing (Bircan, 2025; Kaur, 2023).

Corporate reliance on algorithmic governance has intensified scholarly debates concerning whether established liability frameworks remain adequate for increasingly automated enterprises. Traditional doctrines commonly attribute criminal responsibility through employees, corporate policies, or organizational failures. Algorithmic decision-making complicates these pathways by distributing operational influence across technical infrastructures and institutional actors. Understanding this transformation requires reconsidering how corporate fault, organizational control, and legally relevant knowledge should be conceptualized within technologically mediated decision-making environments.

The central legal problem concerns the difficulty of attributing criminal responsibility when autonomous systems generate harmful decisions without identifiable human authorization or awareness. Conventional corporate liability models frequently depend on individual misconduct, managerial knowledge, or failures of organizational supervision. Algorithmic processes challenge these requirements because consequential decisions may arise from computational operations that remain partially inaccessible to corporate personnel. This disconnect creates uncertainty regarding the evidentiary foundations and normative legitimacy of corporate criminal prosecution (Cheong, 2024; Economou-Zavlanos, 2024).

Algorithmic opacity further complicates assessments of corporate culpability by limiting the ability of investigators and courts to reconstruct decision-making pathways. Complex models may produce harmful outcomes through interactions that resist straightforward causal explanation. Corporate defendants may consequently invoke technological complexity to contest knowledge, foreseeability, or control. Conversely, attributing criminal liability solely because an autonomous system caused harm risks transforming fault-based accountability into punishment based on outcomes rather than legally established culpability.

Existing legal frameworks also struggle to distinguish legitimate technological delegation from organizational attempts to evade responsibility. Companies may deploy autonomous systems while retaining substantial influence over design choices, implementation conditions, monitoring practices, and corrective interventions. The absence of consistent standards for evaluating these responsibilities creates uncertainty for regulators, prosecutors, and corporate decision-makers. A coherent legal approach must therefore address both accountability gaps and the danger of disproportionate criminalization (Groh, 2022; Khan, 2025).

This study aims to critically examine the doctrinal limitations of corporate criminal liability when organizational decisions are mediated by autonomous artificial intelligence systems. Its primary objective is to determine how established principles of attribution, culpability, causation, and organizational responsibility operate under conditions of algorithmic autonomy. The analysis seeks to identify circumstances in which corporate criminal accountability remains legally defensible despite the absence of direct human participation in particular decisions.

The research further investigates whether algorithmic opacity and distributed decision-making require modifications to conventional standards of corporate fault. Particular attention is directed toward institutional knowledge, foreseeable technological risks, supervisory obligations, and failures of preventive governance. Examining these dimensions enables the development of criteria distinguishing culpable organizational conduct from technologically generated harm that lacks sufficient grounds for criminal attribution.

The study ultimately seeks to formulate a normative framework for assessing corporate criminal responsibility in autonomous decision-making environments. This framework emphasizes organizational control, risk governance, and demonstrable failures of legally relevant duties rather than technological autonomy alone. Its intended contribution is to reconcile effective accountability with fundamental criminal law principles, including legality, proportionality, and culpability, while preserving incentives for responsible technological innovation (Alzubaidi, 2023; Zhu, 2025).

Existing scholarship on artificial intelligence liability has extensively examined technological risks, algorithmic transparency, and the allocation of responsibility among developers, operators, and users. Corporate criminal liability, however, presents distinct doctrinal questions that cannot be resolved through general discussions of technological accountability. Significant analytical uncertainty remains concerning how autonomous computational conduct can be connected to organizational fault without abandoning established requirements governing criminal attribution.

Contemporary debates frequently approach algorithmic responsibility through regulatory compliance, civil liability, or proposals concerning artificial intelligence personhood. These perspectives provide valuable insights but do not fully explain how criminal culpability should be established within corporations exercising delegated algorithmic authority. The relationship between organizational governance failures and autonomous system outcomes consequently requires closer examination, particularly regarding evidentiary thresholds, institutional foreseeability, and the limits of criminal punishment (Chauhan, 2024; Khan, 2022).

Comparative discussions of corporate criminal responsibility also reveal insufficient integration between traditional attribution doctrines and the operational characteristics of adaptive algorithms. Legal analysis often treats technological opacity as an evidentiary obstacle without adequately considering its implications for organizational duties and culpability assessments. This study addresses that conceptual gap by examining algorithmic autonomy alongside corporate governance structures, rather than treating autonomous technology as an independent source of criminal responsibility.

The proposed research advances an organizationally grounded interpretation of algorithmic accountability that distinguishes autonomous computational behavior from criminally culpable corporate conduct. Its conceptual contribution lies in connecting technological delegation with legally meaningful institutional responsibilities. This approach challenges the assumption that greater algorithmic autonomy necessarily diminishes corporate accountability, while rejecting automatic liability whenever autonomous systems produce harmful consequences.

The study introduces an integrated analytical framework linking corporate control, foreseeable risks, governance deficiencies, and evidentiary attribution. Rather than assigning independent criminal agency to artificial intelligence, the framework evaluates whether organizational decisions surrounding system deployment and oversight satisfy established culpability requirements. Its originality rests in clarifying the conditions under which algorithmic autonomy becomes relevant to corporate fault without independently constituting that fault (Mökander, 2022; Rajendra, 2026).

The significance of this investigation extends to criminal law theory, corporate governance, and emerging regulatory responses to artificial intelligence. A principled accountability framework can assist lawmakers and courts in distinguishing punishable organizational failures from unavoidable technological uncertainty. The research offers a foundation for proportionate corporate criminal liability that protects fundamental legal safeguards while addressing the institutional risks associated with autonomous decision-making (Arab, 2025; Ganaie, 2026).

RESEARCH METHODOLOGY

This study adopts a qualitative doctrinal legal research design supported by comparative legal analysis and normative conceptual inquiry. The inquiry examines whether established principles of corporate criminal liability adequately address harmful decisions generated through autonomous artificial intelligence systems. Doctrinal analysis evaluates the legal requirements of attribution, culpability, causation, and organizational fault, while comparative analysis identifies how selected jurisdictions distribute responsibility within corporate structures. Normative analysis assesses the consistency of potential liability models with legality, proportionality, and the principle that criminal punishment requires an identifiable basis of fault. The research does not measure algorithmic performance or investigate corporate offending through fieldwork (Bruijn, 2022; C. Wang, 2026).

The research population comprises accessible legal materials addressing corporate criminal accountability, autonomous decision-making, and organizational governance of artificial intelligence. Relevant materials include legislation, judicial decisions, prosecutorial guidance, regulatory instruments, and peer-reviewed legal scholarship. The unit of analysis is the legal proposition or institutional duty expressed in each document, not individual persons or corporations. No human participants are recruited, and the study makes no statistical generalization concerning businesses, courts, or the prevalence of algorithmic misconduct (Alrawahna, 2025; Rachovitsa, 2022).

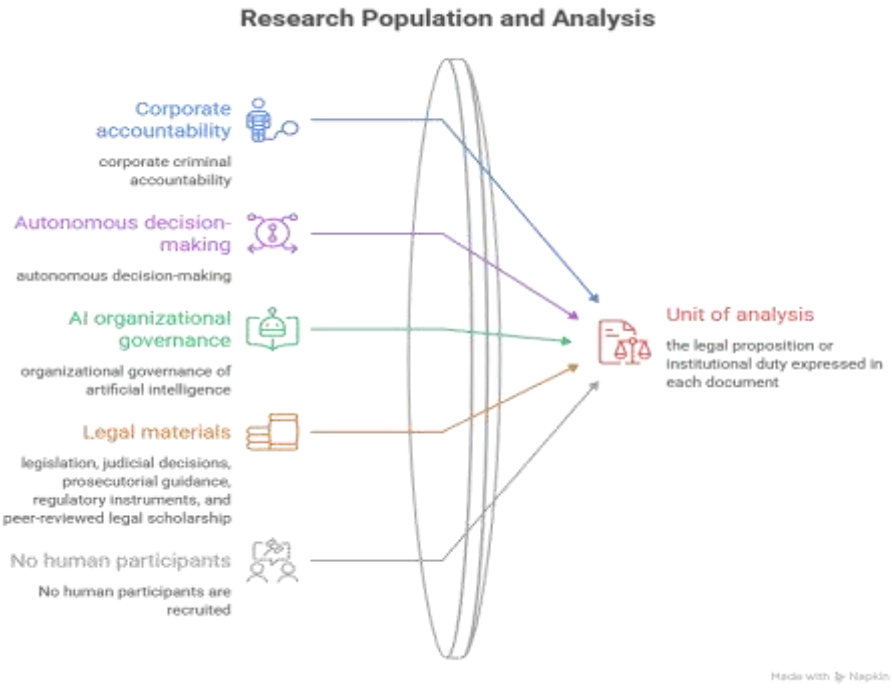


Figure 1. Research Population and Analytical Framework for Qualitative Document-Based Legal Research

Purposive sampling selects sources according to doctrinal relevance, authority, comparability, and their capacity to illuminate accountability mechanisms. Core materials include the United Kingdom Bribery Act 2010, United States Department of Justice guidance on corporate compliance, and Regulation (EU) 2024/1689. Judicial authorities and scholarly analyses supplement these instruments where they clarify corporate attribution, preventive duties, and evidentiary expectations. Documents are excluded when they discuss artificial intelligence solely as a technical innovation without addressing responsibility, governance, or legally consequential decision-making. The sample remains documentary and criterion-based rather than numerically representative.

A structured document analysis matrix serves as the primary research instrument. The matrix records each source's jurisdiction, issuing authority, legal status, publication date, relevant provisions, liability mechanism, attribution standard, fault requirement, and treatment of organizational oversight. Additional fields capture references to algorithmic opacity, risk foreseeability, human intervention, auditability, and evidentiary limitations. A source appraisal protocol distinguishes binding law from nonbinding guidance and academic interpretation. This classification safeguards the analysis against presenting aspirational compliance recommendations as enforceable criminal duties (Green, 2022; Guo, 2025).

A comparative coding guide operationalizes five analytical categories: attribution, culpability, causation, organizational control, and safeguards against excessive criminalization. Definitions are established before coding and refined through documented review of ambiguous passages. An interpretive audit trail records source selection, coding decisions, competing readings, and reasons for doctrinal conclusions. Instrument consistency is assessed by revisiting coded passages against their legal context rather than relying on numerical reliability coefficients unsuitable for this interpretive research design.

The research proceeds through systematic identification, screening, extraction, comparison, and normative synthesis of documentary sources. Materials are retrieved from official legislative and judicial repositories, governmental publications, and academic databases using combinations of corporate criminal liability, autonomous systems, algorithmic accountability, organizational fault, and artificial intelligence governance. Each document is screened against the selection criteria, checked for legal currency, and classified by jurisdiction and normative authority. Relevant passages are entered into the analysis matrix with traceable references to their original provisions or arguments (X. Li, 2025; Rad, 2025).

Doctrinal interpretation first reconstructs the elements necessary for corporate criminal responsibility within each selected legal setting. Comparative analysis then examines how those elements operate where decision-making is delegated to adaptive systems and human knowledge is incomplete. Divergent doctrines are preserved rather than collapsed into a single universal rule. Normative synthesis finally evaluates whether corporate deployment choices, foreseeable risks, monitoring failures, and available corrective capacity establish a sufficiently defensible basis for fault. Conclusions are tested against counterexamples involving unforeseeable outcomes, reasonable safeguards, and limited organizational control. The final framework is presented as a reasoned legal proposal, not as an established statement of law (Cruz-Aguilar, 2026; Porlezza, 2024).

RESULT AND DISCUSSION

The comparative examination of corporate criminal liability identifies substantial differences in how the United States, United Kingdom, and European Union address organizational responsibility arising from technologically mediated decisions. United States federal doctrine recognizes relatively broad mechanisms for attributing employee conduct to corporations, whereas United Kingdom law has historically emphasized identification of responsible individuals occupying sufficiently authoritative organizational positions. European Union artificial intelligence regulation approaches accountability primarily through preventive governance, risk management, documentation, and human oversight. These approaches establish different relationships between corporate control, legally relevant knowledge, and organizational responsibility. Their comparison provides a foundation for identifying doctrinal limitations emerging when autonomous systems undertake decisions previously exercised by corporate personnel.

The documentary evidence reveals three distinguishable dimensions of accountability: attribution of legally relevant conduct, identification of organizational fault, and verification of preventive compliance obligations. United States judicial reasoning concerning collective corporate knowledge addresses information distributed among employees, while United Kingdom statutory developments expand certain attribution mechanisms through senior management responsibility. European Union requirements establish technical and procedural obligations applicable to specified artificial intelligence systems. Table 1 summarizes these normative characteristics without treating distinct legal instruments as equivalent criminal liability provisions. The comparison is qualitative because the selected documents do not constitute a statistically representative population of corporate prosecutions or autonomous-system failures.

Table 1. Comparative Legal Frameworks for Corporate Accountability in Autonomous Decision-Making

Analytical dimension	United States	United Kingdom	European Union
Primary reference	Federal corporate criminal liability doctrine and DOJ compliance guidance	Bribery Act 2010 and Economic Crime and Corporate Transparency Act 2023	Regulation (EU) 2024/1689
Accountability approach	Employee attribution and organizational compliance assessment	Identification, senior manager attribution, and specified failure-to-prevent offences	Risk-based regulatory obligations
Organizational knowledge	Individual and, in defined contexts, collective employee knowledge	Attribution through legally relevant persons and statutory routes	Documentation and governance duties
Algorithmic governance	AI risk assessment within prosecutorial compliance evaluation	Preventive procedures relevant to specified offences	Risk management, logging, and human oversight
Principal limitation	Establishing the required offence-specific elements	Connecting autonomous conduct to statutory attribution or	Regulatory noncompliance does not automatically

		prevention requirements	establish criminal culpability
Analytical implication	Examine organizational knowledge and legally attributable conduct	Examine managerial authority and applicable preventive duties	Examine legally applicable governance obligations and supporting evidence

The comparative findings indicate that organizational accountability cannot be determined simply by identifying which technological system produced a harmful outcome. Criminal attribution ordinarily requires examination of an applicable offence, the relevant mental element, and a legally recognized connection between organizational conduct and prohibited consequences. Autonomous systems complicate this analysis because their outputs may reflect interactions among training data, operational parameters, and changing deployment environments. Corporate responsibility therefore depends on examining decisions concerning system selection, implementation, supervision, and continued operation. These factors can illuminate legally relevant human conduct without assuming that algorithmic independence necessarily constitutes an independent basis for criminal culpability.

Differences among the examined frameworks also demonstrate that regulatory obligations and criminal attribution perform distinct functions. Preventive requirements may identify reasonable organizational precautions, establish documentation duties, or reveal failures in institutional governance. Their evidentiary significance nevertheless depends on the elements of the offence under consideration and the applicable jurisdictional rules. A failure to maintain appropriate algorithmic safeguards could support an inference of organizational negligence where the law recognizes that standard, but it cannot independently establish intention or knowledge where a criminal offence requires those elements. The findings consequently distinguish evidence of defective compliance from legally sufficient proof of corporate wrongdoing.

The examination of algorithmic governance instruments identifies recurring mechanisms through which corporations retain influence over autonomous decision-making. These mechanisms include system procurement, training-data governance, operational configuration, access restrictions, logging procedures, performance monitoring, and intervention protocols. Regulation (EU) 2024/1689 addresses several related responsibilities through requirements concerning risk management, record keeping, and human oversight for covered high-risk systems. Such measures demonstrate that algorithmic autonomy does not necessarily eliminate institutional influence over technological deployment. Organizational decisions remain relevant even where particular algorithmic outputs cannot be fully reconstructed through conventional explanations of human intention or managerial instruction.

The documentary comparison identifies a further distinction between technical opacity and evidentiary insufficiency. Limited interpretability may prevent investigators from reconstructing every computational operation, yet other evidence can establish relevant organizational decisions surrounding system deployment. Internal assessments, warnings, validation reports, incident histories, and governance records may reveal whether responsible personnel recognized foreseeable risks or neglected applicable obligations. Technological opacity therefore changes the evidentiary pathway rather than automatically excluding criminal attribution. The significance of particular records remains contingent upon authenticity, admissibility, relevance, and the precise legal

elements requiring proof. This distinction prevents computational complexity from being confused with either institutional innocence or presumed culpability.

The inferential analysis suggests that the most defensible foundation for corporate criminal accountability lies in legally attributable organizational conduct rather than algorithmic outcomes considered in isolation. Where decision-makers knowingly deploy systems presenting substantial foreseeable risks, disregard documented warnings, or deliberately weaken legally required safeguards, their conduct may contribute evidence relevant to fault. Such inferences remain conditional upon the applicable offence, attribution doctrine, and evidentiary standard. Algorithmic autonomy cannot independently justify replacing established mental elements with general assertions of institutional responsibility. The proposed interpretation therefore preserves the distinction between organizational risk creation and conduct satisfying the requirements of criminal liability.

The comparative evidence also supports a qualified relationship between corporate control and potential accountability. Greater capacity to configure, monitor, suspend, or correct an autonomous system may strengthen arguments concerning organizational responsibility where relevant legal duties exist. Limited technological control, effective preventive measures, and genuinely unforeseeable outcomes may weaken those arguments. These conclusions constitute normative legal inferences rather than statistical associations established through regression, hypothesis testing, or probabilistic estimation. Their validity depends on doctrinal coherence, accurate interpretation of governing rules, and the availability of evidence connecting organizational conduct with the charged offence.

The relationship between algorithmic autonomy and corporate accountability becomes clearer when technological characteristics are examined alongside institutional governance arrangements. System opacity influences the accessibility of evidence, while organizational control determines which operational decisions remain available to responsible personnel. Risk foreseeability concerns what relevant actors could reasonably anticipate under the circumstances, whereas supervisory arrangements indicate whether identified dangers were addressed through appropriate interventions. These dimensions interact without establishing a uniform hierarchy of criminal responsibility. High technological autonomy may coexist with substantial organizational control over deployment conditions, making the absence of direct human intervention insufficient to resolve the attribution question.

The reviewed regulatory materials also reveal a functional relationship between documentation, oversight, and the capacity to investigate corporate conduct. Records concerning system testing, operational modifications, incident responses, and internal escalation procedures may assist investigators in reconstructing institutional decisions. Their evidentiary value increases when they establish when relevant information became available and which individuals possessed authority to respond. The DOJ compliance guidance specifically addresses organizational management of emerging technologies, including artificial intelligence, within corporate compliance evaluation. Such guidance offers an investigative framework but does not independently create additional criminal offences or eliminate existing requirements of proof.

The first documentary case concerns United States v. Bank of New England, 821 F.2d 844 (1st Cir. 1987), which examined corporate responsibility involving failures to comply with currency transaction reporting requirements. The appellate court upheld the conviction and accepted an approach to corporate knowledge recognizing information distributed among employees. The decision illustrates how organizational knowledge may be legally evaluated without requiring every relevant fact to reside within a single individual. The case did not concern artificial intelligence,

however, and cannot establish a direct precedent for criminal liability arising from autonomous computational decisions. Its relevance concerns the narrower question of dispersed organizational knowledge.

The second documentary case study examines United Kingdom legislative developments concerning corporate failures to prevent specified economic offences. Section 7 of the Bribery Act 2010 establishes corporate liability where qualifying bribery by an associated person occurs, subject to an adequate procedures defence. The Economic Crime and Corporate Transparency Act 2023 subsequently introduced a separate failure-to-prevent-fraud offence applicable to qualifying large organizations, effective from September 1, 2025. Both statutory models demonstrate the legal significance of preventive governance, although neither automatically criminalizes harmful outputs produced independently by artificial intelligence. Their applicability requires satisfaction of the particular statutory elements, including the relevant underlying human conduct.

The United States case illustrates a potential analytical response to fragmented organizational information. Corporate structures can distribute knowledge across personnel responsible for compliance, management, and operational activities. Autonomous systems may intensify this fragmentation by separating developers, system operators, auditors, and senior managers. The collective knowledge reasoning provides a useful comparison but cannot be mechanically extended to computational information as though an algorithm were an employee possessing legally attributable intent. The analysis indicates that organizational knowledge must remain connected to recognized legal attribution principles and the specific mental elements of the offence. Technical data alone does not establish criminally relevant corporate awareness.

The United Kingdom statutory examples illuminate a different accountability mechanism based on failures to implement adequate or reasonable preventive measures in defined circumstances. Their significance lies in demonstrating legislative alternatives to reliance exclusively on traditional identification doctrines. Such approaches may inform proposals concerning autonomous corporate systems, particularly where institutional precautions are more readily examinable than individual algorithmic decisions. Their extension to artificial intelligence would nevertheless require careful legislative specification of prohibited conduct, associated actors, organizational duties, and available defences. The cases collectively demonstrate that organizational accountability can be strengthened without attributing independent criminal personality to autonomous technology.

The combined doctrinal findings identify an algorithmic accountability paradox: autonomous decision-making may reduce direct human involvement while leaving corporate decisions about deployment, oversight, and intervention legally relevant. Traditional attribution mechanisms can become difficult to apply where operational knowledge is fragmented, yet broader liability approaches risk weakening fundamental requirements of culpability. The proposed resolution is a conditional assessment of corporate responsibility grounded in applicable criminal offences, identifiable institutional conduct, legally recognized duties, and sufficient evidence of fault. Technological autonomy remains an important contextual factor rather than an independent justification for criminal punishment.

The comparative analysis supports an organizationally grounded model combining attribution analysis, risk foreseeability, governance assessment, and evidentiary safeguards. Its principal contribution is conceptual: responsibility should be examined through the legal significance of corporate choices surrounding autonomous technologies, rather than presumed from technological sophistication or harmful consequences. The framework does not establish that existing legal systems uniformly recognize algorithmic organizational fault. Its findings instead identify the

conditions requiring further doctrinal clarification and potential legislative development. Corporate criminal accountability can remain meaningful in automated environments only when institutional responsibility is evaluated consistently with legality, proportionality, and established standards of proof (Lima, 2022; Mills, 2024).

The findings reveal a fundamental tension between increasing algorithmic autonomy and the traditional foundations of corporate criminal liability. Autonomous decision-making systems complicate the identification of legally responsible actors because harmful outcomes may emerge without direct human authorization. Comparative examination of United States, United Kingdom, and European Union frameworks indicates that existing legal mechanisms address organizational responsibility differently, particularly regarding attribution, culpability, institutional knowledge, and preventive governance obligations.

The comparative analysis demonstrates that technological autonomy does not necessarily eliminate corporate responsibility. Organizations retain substantial influence over system procurement, deployment, operational configuration, monitoring, and corrective intervention. These institutional activities remain relevant to criminal attribution when they establish conduct satisfying applicable legal requirements. The findings consequently challenge approaches that equate the absence of immediate human intervention with the disappearance of organizational accountability.

The documentary case analysis identifies contrasting mechanisms for recognizing organizational wrongdoing. United States v. Bank of New England illustrates the treatment of knowledge distributed among corporate employees, while United Kingdom failure-to-prevent legislation demonstrates the significance of specified preventive obligations. Neither framework automatically establishes criminal liability for autonomous algorithmic conduct. Their comparative relevance lies in identifying possible doctrinal foundations for evaluating legally attributable organizational failures (Bobkov, 2025; Larsson, 2024).

The overall findings support a conditional accountability framework integrating organizational control, risk foreseeability, governance deficiencies, and evidentiary attribution. Corporate criminal responsibility becomes defensible when these considerations establish the elements of a recognized offence rather than merely demonstrating technological harm. This interpretation preserves legality and culpability while acknowledging that autonomous systems require more sophisticated approaches to institutional responsibility and the assessment of corporate conduct.

The findings partially support Matthias (2004), who identified a responsibility gap arising when learning machines produce behavior that operators cannot reliably anticipate. Algorithmic unpredictability similarly complicates conventional attribution within corporate criminal law. The present analysis nevertheless qualifies this argument by distinguishing unforeseeable computational outcomes from organizational decisions concerning deployment and supervision. Unpredictability at the operational level does not necessarily exclude responsibility for independently culpable institutional conduct.

The findings also resonate with Citron (2008), whose account of technological due process demonstrates how automated decisions can weaken transparency and procedural safeguards. The present study extends this concern into corporate criminal accountability, where evidentiary uncertainty affects both prosecution and defence. Algorithmic documentation may improve investigative scrutiny, but procedural fairness requires that technological opacity never substitute for proof of legally relevant fault.

The regulatory comparison complements contemporary approaches emphasizing organizational risk management and human oversight. Regulation (EU) 2024/1689 establishes

requirements concerning risk governance, documentation, and human supervision for covered high-risk systems. These measures support the importance of preventive institutional control identified in this study. Their regulatory character, however, distinguishes them from criminal attribution doctrines, preventing direct assumptions that noncompliance necessarily establishes criminal culpability.

The research differs from accounts that propose resolving accountability deficits primarily through the independent legal personhood of artificial intelligence. Its analysis locates potentially culpable conduct within existing corporate governance relationships instead of treating autonomous systems as criminal agents. This distinction contributes a narrower doctrinal approach: technological independence matters because it changes available evidence and institutional control, not because it necessarily creates a new criminal subject (Goktas, 2025; J. Wang, 2025).

The findings signify a broader transformation in the relationship between organizational authority and technological decision-making. Corporate power increasingly operates through technical architectures rather than exclusively through direct managerial instructions. This transformation challenges legal assumptions that meaningful control must coincide with immediate human participation. Organizational responsibility may instead arise from decisions establishing the conditions under which automated processes operate, provided those decisions satisfy legally recognized requirements of culpability.

The algorithmic accountability paradox also indicates that technological complexity can redistribute visibility without eliminating institutional influence. Corporate executives may lack detailed knowledge of individual computational outputs while retaining authority over strategic implementation and oversight resources. This separation creates opportunities for legitimate operational delegation alongside potential accountability avoidance. Criminal law must distinguish these circumstances through evidence rather than presume either institutional innocence or culpability.

The findings further suggest that algorithmic accountability involves an important epistemological challenge. Investigators may establish that harmful outcomes occurred without reconstructing the precise computational pathway producing them. This limitation does not automatically defeat responsibility, since documentary evidence may demonstrate independently relevant human decisions. Criminal adjudication nevertheless requires careful differentiation between technical uncertainty, evidentiary insufficiency, and legally established organizational fault (Y. Li, 2025; Voinea, 2025).

The comparative evidence ultimately reflects a shift toward evaluating corporate governance as an important component of technological accountability. Preventive measures, incident reporting, and documented oversight can reveal how organizations respond to identifiable risks. Their significance remains conditional rather than absolute. Effective governance may provide evidence of reasonable precautions, whereas deficient governance becomes criminally relevant only when connected to applicable offences and required mental elements.

The theoretical implications concern the need to reconsider how corporate criminal law conceptualizes legally relevant organizational agency. Traditional attribution doctrines remain important but may be insufficiently sensitive to fragmented knowledge and algorithmically mediated operations. A governance-oriented approach offers additional analytical dimensions without abandoning established legal principles. Corporate fault should consequently be examined through identifiable institutional decisions, legally applicable duties, and offence-specific standards of responsibility.

The practical implications are particularly relevant to corporations deploying autonomous technologies in sensitive operational environments. Organizations should establish clear decision-making authority, documented risk assessments, monitoring arrangements, and procedures for responding to detected irregularities. The United States Department of Justice's corporate compliance guidance already identifies artificial intelligence governance as relevant to compliance evaluation. Such practices may reduce organizational exposure while strengthening the availability of reliable evidence (Bajracharya, 2023; Busuioc, 2023).

The regulatory implications concern the importance of separating preventive obligations from criminal punishment. Legislators may establish specific organizational duties addressing high-risk autonomous systems, but criminalization requires clearly defined prohibited conduct and appropriately structured fault requirements. Vague references to algorithmic irresponsibility could undermine legal certainty. Carefully designed statutory provisions should instead identify responsible organizational actors, relevant failures, applicable defences, and the connection between misconduct and protected legal interests.

The judicial implications involve developing methods for evaluating technical evidence without allowing computational complexity to determine guilt. Courts may require specialized expertise to assess algorithmic records, system limitations, and institutional control arrangements. Independent technical interpretation can assist factual reconstruction, although experts should not determine ultimate legal culpability. Judicial reasoning must preserve the prosecution's burden of proof and distinguish regulatory deficiencies from the constituent elements of criminal offences (Bajracharya, 2023; Kingsman, 2022).

The findings emerge partly because corporate criminal liability doctrines developed around human conduct and recognizable organizational relationships. Autonomous computational systems complicate these assumptions by separating operational outputs from individual instructions. Decisions generated through adaptive models may not correspond to any particular employee's intentions. Established attribution mechanisms therefore face difficulties when legally relevant behavior cannot be reconstructed through conventional evidence of human authorization.

The second explanation concerns the distributed structure of contemporary corporate governance. Artificial intelligence systems frequently involve developers, external providers, procurement personnel, compliance departments, and operational managers. Each actor may possess only partial information concerning technological capabilities and associated risks. This fragmentation creates challenges resembling conventional problems of dispersed corporate knowledge, although computational processes introduce additional distinctions that prohibit straightforward doctrinal equivalence.

The third explanation relates to differences between criminal law and preventive regulation. Regulatory frameworks generally seek to manage risks before harm occurs, whereas criminal liability requires satisfaction of particular statutory or doctrinal elements. European Union artificial intelligence regulation emphasizes prospective governance and oversight, while United Kingdom failure-to-prevent offences attach criminal consequences to specific statutory circumstances. Their different objectives explain why regulatory requirements cannot automatically function as criminal attribution rules.

The final explanation concerns the normative importance of protecting culpability as a limiting principle. Corporate punishment based exclusively on algorithmic harm would risk imposing responsibility without sufficient examination of attributable conduct or fault. Conversely, complete immunity whenever autonomous systems intervene would create opportunities to evade

established obligations. The findings therefore favor conditional responsibility because it better reconciles technological realities with the foundational constraints of criminal justice.

The findings justify developing a structured legal assessment that begins with identification of the applicable criminal offence and its constituent elements. Subsequent examination should establish the relevant human or organizational conduct, applicable attribution rules, foreseeable risks, and available governance evidence. This sequential approach would prevent investigators from treating technological harm as sufficient proof while preserving avenues for prosecuting genuinely culpable organizational behavior.

Future comparative research should examine judicial decisions involving algorithmically mediated financial misconduct, automated compliance failures, and corporate deployment of high-risk technologies. Particular attention should be directed toward whether courts distinguish system malfunction from deliberate organizational misconduct. Longitudinal analysis of prosecutorial practices could establish whether emerging technological governance requirements influence corporate criminal enforcement and whether existing attribution doctrines remain workable across different institutional contexts (Hickok, 2024).

Corporate governance research should investigate the effectiveness of algorithmic impact assessments, independent audits, incident documentation, and meaningful human intervention. Empirical studies involving compliance professionals, technology developers, prosecutors, and legal practitioners could identify operational obstacles to implementing these safeguards. Such evidence would strengthen the proposed framework by testing whether theoretical expectations concerning control, foreseeability, and supervision correspond to actual corporate decision-making practices.

The proposed accountability framework ultimately requires evaluation against concrete offences, comparative judicial reasoning, and established procedural protections before legislative adoption. Future scholarship should examine its suitability across different corporate structures and technological applications without assuming universal doctrinal compatibility. The central contribution remains the recognition that autonomous technologies transform the conditions of accountability without necessarily extinguishing organizational responsibility. Effective legal development must preserve criminal justice principles while responding to increasingly complex corporate decision-making environments.

CONCLUSION

The central finding of this study is that algorithmic autonomy does not necessarily diminish corporate criminal accountability, although it can disrupt methods of attributing conduct, knowledge, and fault. Comparative doctrinal analysis of United States and United Kingdom liability approaches, alongside European Union artificial intelligence regulation, reveals a mismatch between technologically distributed decision making and legal doctrines organized around identifiable human actors. The distinctive result is the identification of an accountability paradox: corporations may retain authority over system selection, deployment, monitoring, and intervention even when they cannot explain or predict algorithmic outputs. Criminal responsibility should therefore not arise from harmful outcomes, nor should technical opacity function as a categorical defence. The legally decisive inquiry concerns whether attributable organizational conduct satisfies the elements of a defined offence, supported by sufficient evidence of the required fault.

The study contributes a governance grounded conceptual framework for examining corporate criminal liability in autonomous decision making environments. Its innovation is analytical rather than empirical: it connects attribution doctrines with four considerations, namely organizational

control, foreseeable risk, governance failures, and evidentiary safeguards. This framework differentiates regulatory duties from criminal law requirements and clarifies why a deficient compliance program, standing alone, cannot establish intention, knowledge, or other offence specific elements. The comparative method demonstrates the value of treating United States attribution principles, United Kingdom statutory prevention models, and European Union risk governance obligations as distinct legal responses. For lawmakers, the framework offers criteria for designing precise obligations without imposing punishment solely because an autonomous system caused harm. For courts and corporate practitioners, it provides a sequence for examining institutional decisions, technical evidence, and legitimate delegation.

The study is limited by its doctrinal and documentary design, its selective jurisdictional coverage, and the absence of empirical observations involving prosecutions involving autonomous artificial intelligence decisions. The examined judicial and statutory examples illuminate analogous legal problems but do not establish how courts would resolve every algorithmic scenario. Differences in offence definitions, evidentiary rules, regulatory maturity, and corporate structures restrict generalization across jurisdictions. Future research should investigate emerging case law, compare prosecutorial decisions, and examine how organizations document model risks, allocate oversight responsibilities, and respond to system failures. Interviews with legal professionals, compliance officers, technical auditors, and regulators could test the practical feasibility of the proposed framework. Subsequent scholarship should assess whether defined legislative reforms improve accountability while preserving legality, proportionality, procedural fairness, and the requirement that criminal liability rest on demonstrable culpability.

DECLARATION OF AI AND AI ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work the author(s) used Google Gemini to assist in improving grammar, language quality, and overall readability of the text. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

AUTHORS' CONTRIBUTION

Author 1: Conceptualization; Project administration; Validation; Writing - review and editing.

Author 2: Conceptualization; Data curation; In-vestigation.

Author 3: Data curation; Investigation.

DECLARATION OF COMPETING INTEREST

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

REFERENCES

Alrawahna, A. S. (2025). The Impact of Artificial Intelligence on Public Sector Decision-Making: Benefits, Challenges, and Policy Implications. *International Review of Management and Marketing*, *15*(5), 125–138. https://doi.org/10.32479/irmm.19419

Alzubaidi, L. (2023). Towards Risk-Free Trustworthy Artificial Intelligence: Significance and Requirements. *International Journal of Intelligent Systems*, *2023*(Query date: 2026-10-13 14:23:32). https://doi.org/10.1155/2023/4459198

Arab, R. A. E. (2025). The Role of AI in Nursing Education and Practice: Umbrella Review. *Journal of Medical Internet Research*, *27*(Query date: 2026-10-13 14:23:32). https://doi.org/10.2196/69881

Bajracharya, A. (2023). Recent Advances in Algorithmic Biases and Fairness in Financial Services: A Survey. *Lecture Notes in Networks and Systems*, *559*(Query date: 2026-10-13 14:23:32), 809–822. https://doi.org/10.1007/978-3-031-18461-1_53

Bircan, T. (2025). Unmasking inequalities of the code: Disentangling the nexus of AI and inequality. *Technological Forecasting and Social Change*, *211*(Query date: 2026-10-13 14:23:32). https://doi.org/10.1016/j.techfore.2024.123925

Bobkov, A. (2025). Telepsychiatry and Artificial Intelligence: A Structured Review of Emerging Approaches to Accessible Psychiatric Care. *Healthcare Switzerland*, *13*(11). https://doi.org/10.3390/healthcare13111348

Bruijn, H. de. (2022). The perils and pitfalls of explainable AI: Strategies for explaining algorithmic decision-making. *Government Information Quarterly*, *39*(2). https://doi.org/10.1016/j.giq.2021.101666

Busuioc, M. (2023). Reclaiming transparency: Contesting the logics of secrecy within the AI Act. *European Law Open*, *2*(1), 79–105. https://doi.org/10.1017/elo.2022.47

Chaudhary, G. (2024). UNVEILING THE BLACK BOX: BRINGING ALGORITHMIC TRANSPARENCY TO AI. *Masaryk University Journal of Law and Technology*, *18*(1), 93–122. https://doi.org/10.5817/MUJLT2024-1-4

Chauhan, M. (2024). Towards a Greener Tomorrow: Exploring the Potential of AI, Blockchain, and IoT in Sustainable Development. *Nature Environment and Pollution Technology*, *23*(2), 1105–1113. https://doi.org/10.46488/NEPT.2024.v23i02.044

Cheong, B. C. (2024). Transparency and accountability in AI systems: Safeguarding wellbeing in the age of algorithmic decision-making. *Frontiers in Human Dynamics*, *6*(Query date: 2026-10-13 14:23:32). https://doi.org/10.3389/fhumd.2024.1421273

Cruz-Aguilar, M. A. (2026). The epistemic revolution of AI: reconfiguring the foundations of scientific knowledge. *AI and Society*, *41*(3), 2041–2057. https://doi.org/10.1007/s00146-025-02658-3

Economou-Zavlanos, N. J. (2024). Translating ethical and quality principles for the effective, safe and fair development, deployment and use of artificial intelligence technologies in healthcare. *Journal of the American Medical Informatics Association*, *31*(3), 705–713. https://doi.org/10.1093/jamia/ocad221

Ganaie, N. A. (2026). The role of artificial intelligence in radicalisation, recruitment and terrorist propaganda: Deconstructing violent extremism and reimagining counterterrorism in contemporary digital ecosystems. *Frontiers in Political Science*, *7*(Query date: 2026-10-13 14:23:32). https://doi.org/10.3389/fpos.2025.1718396

Goktas, P. (2025). Shaping the Future of Healthcare: Ethical Clinical Challenges and Pathways to Trustworthy AI. *Journal of Clinical Medicine*, *14*(5). https://doi.org/10.3390/jcm14051605

Green, B. (2022). The flaws of policies requiring human oversight of government algorithms. *Computer Law and Security Review*, *45*(Query date: 2026-10-13 14:23:32). https://doi.org/10.1016/j.clsr.2022.105681

Groh, M. (2022). Towards Transparency in Dermatology Image Datasets with Skin Tone Annotations by Experts, Crowds, and an Algorithm. *Proceedings of the ACM on Human Computer Interaction*, *6*(Query date: 2026-10-13 14:23:32). https://doi.org/10.1145/3555634

Guo, J. (2025). The ethical legitimacy of autonomous Weapons systems: Reconfiguring war accountability in the age of artificial Intelligence. *Ethics and Global Politics*, *18*(3), 27–39. https://doi.org/10.1080/16544951.2025.2540131

Hickok, M. (2024). Public procurement of artificial intelligence systems: New risks and future proofing. *AI and Society*, *39*(3), 1213–1227. https://doi.org/10.1007/s00146-022-01572-2

Kaur, D. (2023). Trustworthy Artificial Intelligence: A Review. *ACM Computing Surveys*, *55*(2). https://doi.org/10.1145/3491209

Khan, M. M. (2022). Toward Accountable and Explainable Artificial Intelligence Part One: Theory and Examples. *IEEE Access*, *10*(Query date: 2026-10-13 14:23:32), 99686–99701. https://doi.org/10.1109/ACCESS.2022.3207812

Khan, M. M. (2025). Towards secure and trusted AI in healthcare: A systematic review of emerging innovations and ethical challenges. *International Journal of Medical Informatics*, *195*(Query date: 2026-10-13 14:23:32). https://doi.org/10.1016/j.ijmedinf.2024.105780

Kingsman, N. (2022). Public sector AI transparency standard: UK Government seeks to lead by example. *Discover Artificial Intelligence*, *2*(1). https://doi.org/10.1007/s44163-022-00018-4

Larsson, S. (2024). The Artificial Recruiter: Risks of Discrimination in Employers' Use of AI and Automated Decision-Making. *Social Inclusion*, *12*(Query date: 2026-10-13 14:23:32). https://doi.org/10.17645/si.7471

Li, X. (2025). The ethical challenges in the integration of artificial intelligence and large language models in medical education: A scoping review. *Plos One*, *20*(10). https://doi.org/10.1371/journal.pone.0333411

Li, Y. (2025). Reducing misdiagnosis in AI-driven medical diagnostics: A multidimensional framework for technical, ethical, and policy solutions. *Frontiers in Medicine*, *12*(Query date: 2026-10-13 14:23:32). https://doi.org/10.3389/fmed.2025.1594450

Lima, G. (2022). The Conflict Between Explainable and Accountable Decision-Making Algorithms. *ACM International Conference Proceeding Series*, (Query date: 2026-10-13 14:23:32), 2103–2113. https://doi.org/10.1145/3531146.3534628

Mills, S. (2024). The autonomous choice architect. *AI and Society*, *39*(2), 583–595. https://doi.org/10.1007/s00146-022-01486-z

Mökander, J. (2022). The US Algorithmic Accountability Act of 2022 vs. The EU Artificial Intelligence Act: What can they learn from each other? *Minds and Machines*, *32*(4), 751–758. https://doi.org/10.1007/s11023-022-09612-y

Porlezza, C. (2024). The datafication of digital journalism: A history of everlasting challenges between ethical issues and regulation. *Journalism*, *25*(5), 1167–1185. https://doi.org/10.1177/14648849231190232

Rachovitsa, A. (2022). The Human Rights Implications of the Use of AI in the Digital Welfare State: Lessons Learned from the Dutch SyRI Case. *Human Rights Law Review*, *22*(2). https://doi.org/10.1093/hrlr/ngac010

Rad, A. A. (2025). The ethical considerations of integrating artificial intelligence into surgery: A review. *Interdisciplinary Cardiovascular and Thoracic Surgery*, *40*(3). https://doi.org/10.1093/icvts/ivae192

Rajendra, J. B. (2026). The role of explainability and human intervention in AI decisions: Jurisdictional and regulatory aspects. *Information and Communications Technology Law*, *35*(2), 152–183. https://doi.org/10.1080/13600834.2025.2537514

Vasudevan, A. (2025). Voices of Concern: Contemplations on Socioeconomic, Ethical, and Developmental Impacts of AI. *Applied Mathematics and Information Sciences*, *19*(1), 221–232. https://doi.org/10.18576/amis/190119

Voinea, D. V. (2025). Reconceptualizing Gatekeeping in the Age of Artificial Intelligence: A Theoretical Exploration of Artificial Intelligence-Driven News Curation and Automated Journalism. *Journalism and Media*, *6*(2). https://doi.org/10.3390/journalmedia6020068

Wang, C. (2026). The rise of algorithmic governance and the dual revolution: Applications, challenges, and governance of artificial intelligence in public administration. *Technology in Society*, *86*(Query date: 2026-10-13 14:23:32). https://doi.org/10.1016/j.techsoc.2026.103264

Wang, J. (2025). Teaching with Artificial Intelligence in Architecture: Embedding Technical Skills and Ethical Reflection in a Core Design Studio. *Buildings*, *15*(17). https://doi.org/10.3390/buildings15173069

Zhu, H. (2025). Towards responsible artificial intelligence in education: A systematic review on identifying and mitigating ethical risks. *Humanities and Social Sciences Communications*, *12*(1). https://doi.org/10.1057/s41599-025-05252-6

Copyright Holder :
© Ren Suzuki et al. (2026).

First Publication Right :
© Rechtsnormen Journal of Law

This article is under:

