Received: 13 March 2026; **Revised:** 18 October 2026; **Accepted:** 12 November 2026; **Published:** 20 November 2026
Archs Sci. (2026) Volume 76, Issue 5 Pages 151-157, Paper ID 2026515.
https://doi.org/10.68304/as/76515

Attributing Criminal Liability for Harm Caused by Semi-Autonomous Artificial Intelligence: Subjectlessness and International Legal Benchmarks

Nataliia Kozak[1,*], Vitalij Grabovets[2], Oleh Karas[3], Olha Hanych[1] and Yuliia Fidria[4]
[1]Department of Law, Lutsk National Technical University, Lutsk, Ukraine
[2]Department of Transport Logistics, Lutsk National Technical University, Lutsk, Ukraine
[3]Doctoral Programme in Economics and International Economic Relations, Lutsk National Technical University, Lutsk, Ukraine
[4]Department of Criminal Justice and Law Enforcement, Lesya Ukrainka Volyn National University, Lutsk, Ukraine

Correspondence: n.riabyh@lntu.edu.ua

Abstract This article re-examines the attribution of criminal liability for harm caused by semi-autonomous artificial intelligence (AI) in light of international legal principles governing human rights protection and the regulation of AI-related activities. It defines subjectlessness as a criminal-law situation in which a socially harmful outcome is directly associated with an algorithmically generated decision, yet cannot immediately be attributed to a single identifiable person without compromising the principles of culpability, personal liability, and the prohibition of liability without fault. Functional autonomy, data processing, adaptation, and independent selection of means do not establish criminal-law personhood. The study examines the limits of conventional legal concepts of the offender, fault, and causation, while distinguishing individual criminal liability from states' obligations to protect human rights, ensure accountability, and provide effective remedies. A developer, manufacturer, supplier, deployer, owner, operator, or user cannot be held criminally liable solely because of a formal connection with a system. The article proposes a multilevel functional attribution model that identifies the lifecycle stage at which an impermissible risk arose, the person competent to address that risk, their knowledge and control, a culpable breach of duty, and its causal significance. The proposed directions for regulatory development include differentiated lifecycle duties, risk-based requirements, legally specified meaningful human control, narrowly defined offences addressing particularly dangerous breaches involving high-risk systems, and requirements for algorithmic traceability. The analysis supports a functional, anthropocentric approach that neither personifies algorithms nor permits human actors to invoke algorithmic autonomy to conceal culpable human decisions. Its consistency with international human rights safeguards requires effective protection for affected persons while preserving the substantive and procedural conditions of individual criminal liability.

Index Terms artificial intelligence; semi-autonomous systems; criminal liability; subjectlessness; attribution of liability; international law; international human rights standards; meaningful human control; algorithmic risk

I. Introduction

Artificial intelligence is increasingly used in transport, medicine, industry, defence, and critical infrastructure management. These applications are changing the mechanisms through which socially harmful outcomes may arise and the ways in which those outcomes must be investigated. Semi-autonomous systems operate within human-defined goals but may process information, adapt to external conditions, and select the means of performing a task without obtaining separate human approval for every operation. Harm may therefore follow a decision that no particular human formulated or endorsed in advance. This complicates the identification of the offender, the assessment of fault, and the establishment of causation, while leaving the requirements of personal liability and legal certainty intact. Cross-border development and supply chains introduce further difficulties in determining applicable duties and obtaining technical evidence.

An international legal perspective requires effective protection for affected persons alongside robust safeguards for individuals who may face criminal proceedings. The inability to identify a culpable developer or operator immediately does not remove the need to regulate dangerous activities, investigate the circumstances of an incident, and provide accessible and effective remedies. Conversely, accountability cannot justify presuming criminal fault from a formal connection with a technology. The attribution problem must therefore be examined through human decisions across the system lifecycle and assessed against international human rights benchmarks.

The study aims to justify a functional, anthropocentric model for attributing criminal liability for harm caused by semi-autonomous AI and to identify the international legal benchmarks relevant to its development. It examines subjectlessness and the limits of personal attribution, differentiates lifecycle duties, distinguishes individual liability from states' protective obligations, and develops proposals concerning risk management, meaningful human control, and algorithmic traceability. Its central contribution is a structured sequence of attribution inquiries linking competence at different lifecycle stages to culpable conduct and the materialisation of an impermissible risk.

Matthias' responsibility-gap account explains how learning automata can undermine the control and predictability on which conventional attribution depends [1]. Hallevy identifies three models of criminal liability: perpetration via another, natural-probable consequences, and direct liability of the AI entity [2]. Gless, Silverman, and Weigend examine intentional instrumental use and the duties of care relevant to robot-related negligence [3]. Giannini and Kwik compare regulatory responses to automated vehicles and emphasise differentiated legal roles, liability, and immunity [4]. Together, these approaches clarify why neither technological involvement nor physical proximity to a system, considered alone, is sufficient to establish criminal liability.

Santoni de Sio and Mecacci distinguish gaps in culpability, moral accountability, public accountability, and active responsibility [5]. Santoni de Sio and van den Hoven develop meaningful human control through tracking and tracing conditions [6]. Asaro relates human judgement to decisions involving lethal force and the requirements of humanitarian law and due process [7]. In Ukrainian scholarship, Kamensky examines the multiplicity of actors potentially involved in AI-related offending [8], while Radutnyi explores electronic personhood and criminal-law measures applicable to AI [9]. Further doctrinal analysis is needed to connect these accounts with effective remedies, without equating broader institutional accountability with personal criminal culpability.

II. Materials and Methods

The study adopts a doctrinal research design. It combines formal legal analysis, systems analysis, functional analysis, and comparative legal reasoning to distinguish the identity of an offender, fault, factual causation, legally imposed duties, and effective control. Legal modelling is used to construct an attribution sequence based on the lifecycle stage, the nature of the risk, and the competence of the relevant person. The illustrative situations are hypothetical analytical examples, not findings derived from an empirical survey or a systematic dataset of decided cases. The research materials comprise the scholarly works reviewed above and the legal instruments identified below; the study uses neither a primary empirical dataset nor an experimental intervention.

The legal analysis considers the European Convention on Human Rights [10], the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law [11], and Regulation (EU) 2024/1689 [12]. Their different legal character, respective addressees, substantive scope, and conditions of applicability are distinguished throughout the analysis. Treaty obligations and EU regulatory requirements are examined as benchmarks for prevention, accountability, and remedies; they are not treated as independently established criminal offences or as substitutes for proving an individual's fault under applicable national law.

The analysis proceeds from the concept of subjectlessness to a differentiated lifecycle model, then to international safeguards and legislative implications. Each attribution inquiry asks whether an identifiable person was subject to a relevant legal duty, possessed the necessary information and practical capacity to act, and committed a culpable breach with causal significance for the harmful outcome. This structure separates technical reconstruction from legal attribution and distinguishes existing rules from proposals de lege ferenda.

III. Analytical Results

A. Subjectlessness and the Limits of Personal Attribution

For the purposes of this study, semi-autonomy denotes a system's capacity to select particular means of performing a task within goals and constraints established by human actors. It is a working doctrinal concept rather than a separate statutory category. An algorithm may generate an operational decision without approval of every action, but functional autonomy does not establish criminal responsibility, culpability, or responsiveness to criminal condemnation. Consequently, the immediate technical decision-making mechanism and the legally recognised bearer of criminal liability may be different. Assessment must encompass both the incident and earlier decisions concerning training, testing, deployment, and control.

Subjectlessness does not imply an absence of people involved in creating and using AI. Developers determine architecture; manufacturers integrate models into products; suppliers place them on the market; deployers select the operating environment; and operators provide supervision. Nevertheless, none of these participants may have formulated the particular operational decision that directly produced the harm. Kamensky connects the identification problem with the multiplicity of potential participants [8]. The legal task is to identify whose conduct and fault cover the relevant elements of the outcome, rather than select any person associated with the technology.

The distribution of functions across an AI system's lifecycle is not always accompanied by adequate information exchange between the actors concerned. A manufacturer may lack access to a model's internal logic, an operator may be unaware of training-data limitations, and a user may reasonably rely on advertised characteristics. Assessment must address both actual knowledge and any duty to communicate relevant information. A breach of that duty may have independent causal significance. An inability to reconstruct the entire process

does not authorise investigators to select the person who is easiest to identify or who last interacted with the system as a presumptively guilty party.

Matthias explains the responsibility gap by reference to changes in machine behaviour that manufacturers or operators may be unable to predict [1]. For criminal law, this account cautions against attributing every harmful outcome to a person merely because that person created or used the system. However, the unforeseeability of one particular operation does not exclude an earlier culpable decision to permit a sufficiently foreseeable and legally impermissible risk. A possible gap must be assessed against a particular person and duty; it is not a universal legal consequence of AI autonomy.

Recognising an algorithm as a legal person would not, by itself, resolve the criminal-law requirement to establish culpability. Radutnyi considers electronic personhood and special criminal-law measures [9], but such a legal fiction would require separate justification of responsibility, fault, and punishment. Hallevy's three models are doctrinal constructions [2]; their introduction into national law cannot bypass statutory requirements concerning the offender and the relevant mental element. For the semi-autonomous systems examined here, an identifiable, legally significant, and culpable human contribution remains central to the analysis.

Recognising a prohibited option computationally or calculating a probability of harm does not, without more, establish a legally relevant mental attitude towards conduct in the criminal-law sense. Nor does the technical capacity to change code or stop a system establish the possibility of punishment serving retributive, corrective, and preventive functions. Disabling a dangerous system or limiting its use should be assessed as risk management rather than evidence of the algorithm's criminal fault. Criminal condemnation must therefore remain connected to a person in respect of whom the substantive statutory conditions of liability have actually been established.

Where a person sets an unlawful objective, programs a system to cause harm, or deliberately uses its functions with the required intent, AI operates as an instrument. Technological complexity does not, in itself, change the legally responsible human actor or displace the established elements of the offence. If algorithmic behaviour departs from the initial plan, the inquiry must establish the relevant fault and whether prevention of the risk fell within the person's duties. Gless, Silverman, and Weigend examine these questions through reasonable care and control [3]. General awareness that a technical error is possible is not automatically equivalent to the legally required foreseeability of the particular harmful outcome or class of outcomes.

A finding of negligence requires an individualised assessment of what the person ought to have foreseen, what they could reasonably have foreseen, and what preventive conduct was available. Known limitations, test results, prior reports of failures, the operating environment, and available intervention mechanisms are relevant. Autonomy does not excuse an actor who knew of a danger and was competent to prevent it. Conversely, an operator should not be blamed for unforeseeable behaviour they objectively could not control. The required care must correspond to accessible information, professional preparation, and feasible conduct; a formal role cannot replace those conditions.

Factual causation, standing alone, is also insufficient to establish criminal liability. Harm may result from architecture, training data, updates, hardware integration, settings, user behaviour, and external conditions. Identifying each factor as a condition of system operation does not identify a culpable person. Attribution requires a relevant human breach that created or increased an impermissible risk materialising in the outcome. An independent algorithmic decision does not necessarily break the chain of causation, although its atypicality, unforeseeability, and lack of controllability may affect whether the harmful outcome can be legally attributed to a particular person.

A software defect may be technically connected to a crash, but developer liability requires inquiry into mandatory professional requirements, detectability, and whether prevention of that risk fell within the developer's competence. An operator's failure to intervene is relevant where a duty, sufficient information, and a practical opportunity to stop the operation existed. A defective stop mechanism or objectively inadequate response time instead requires examination of earlier design and organisational decisions. Mere participation in the wider causal process must therefore be distinguished from culpably creating or increasing the impermissible risk.

Subjectlessness is therefore a situation in which an algorithmically mediated harmful outcome cannot immediately be attributed to a single identifiable agent consistently with culpability, personal liability, and the prohibition of liability without fault. It denotes a problem of initial attribution, not a final finding that criminal liability is impossible. Resolving it requires a careful investigation of each participant's duties, capacities, and potentially culpable contribution, while distinguishing the technical explanation of the incident from the independent legal justification for criminal condemnation.

B. A Functional Lifecycle Model of Criminal Liability

The proposed functional model connects possible criminal liability with the lifecycle segment over which the relevant person exercised legally significant competence and control. Developers, manufacturers, suppliers, deployers, owners, operators, and users have different powers and information. Their duties cannot be reduced to an undifferentiated obligation to answer "for AI". For each actor, the inquiry must establish the source of the duty, the decision or omission, knowledge, practical control, and causal significance of the breach. Organisational participation may warrant investigation, but it does not independently or automatically establish individual criminal liability.

For developers, manufacturers, and integrators, the inquiry concerns impermissible technological risk, including concealed critical defects, ignored test findings, or breaches of mandatory safety requirements. An algorithmic error alone

cannot establish the requisite personal fault or prove a culpable breach of an applicable legal duty. The defect must be assessed against the person's competence, available detection methods, and the type of danger that materialised. Supplier liability may concern culpable failures to disclose known limitations or satisfy applicable duties governing system transfer. None of these roles entails an absolute criminal-law guarantee of error-free operation.

Contemporaneous test records and documentation of known limitations support assessment of conduct at the time of the relevant decision, rather than exclusively through hindsight after harm has occurred. An adverse outcome must not retrospectively transform an unknown deviation into an obvious danger. Equally, ignoring recorded critical failures cannot be excused by the complexity of machine learning. Knowledge, available verification methods, and professional duties must be considered together. For manufacturers and suppliers, conformity with advertised characteristics and accurate information are relevant because other participants may rely on them in deciding how to use the product safely.

Deployers determine the sphere of operation, supervision, staff preparation, and admission to high-risk environments. The inquiry concerns the reasonableness of using a model for tasks whose demands could turn its limitations into impermissible danger. An owner or manager may be liable for breaching a concrete safety duty where information and preventive means were available. Organisational shortcomings do not automatically establish the fault of every employee. Individual criminal liability and measures applicable to legal persons must be established separately under the governing national law.

Deployment decisions encompass permitted system tasks, user-access permissions, and the practical limits of available human supervision. A system tested for advisory use may create a different risk when its output directly controls a physical action. Changes in environment or purpose therefore require assessment by the competent actor. Applicable duties must make it possible to determine who was responsible for checking the compatibility of the new use with known capabilities. Professional or managerial status without a concrete breach and the required fault is insufficient for criminal proceedings.

Operator assessment must include training, comprehensible warnings, response time, technical means, and authority to interrupt a dangerous operation. Merely nominal supervision is inadequate where a system's operating speed or opacity prevents effective human intervention. Giannini and Kwik demonstrate the importance of differentiating legal roles, liability, and immunity in automated-vehicle regulation [4]. If the operator lacked practical capacity, earlier organisational decisions should be investigated, with separate proof of the decision-makers' duties and fault. Attributing every consequence to the last human participant in the control loop would undermine the requirement of individualised personal liability.

User liability depends on the purpose and manner of use. An unlawful task or deliberate use of dangerous functionality may constitute intentional instrumental use. Where the purpose is lawful, disregarding known restrictions or warnings, or continuing operation after danger becomes apparent, requires assessment under negligence rules. If several decisions contribute, each person's conduct is examined separately. Distributed attribution must not be confused with collective fault: the absence of an offence by one participant cannot be offset by increasing another participant's presumed blame. Interaction alone does not establish complicity.

The proposed model identifies the stage at which an impermissible risk arose; the person competent to address it; their knowledge, control, and capacity to prevent the outcome; and a concrete culpable breach with its causal significance. It concerns human conduct rather than abstract algorithmic autonomy. Cross-border cases additionally require identification of the applicable duties, determination of jurisdiction, and lawful access to relevant technical evidence held in another state. The location of a company or server does not itself determine fault. Functional competence and individualised proof together guard against unjustified criminal condemnation while preventing culpable human decisions from being concealed behind technological complexity.

This sequence must not be used to apportion one outcome mechanically among all participants. A developer may create a defect, a manager admit an untested system to a dangerous setting, an operator ignore a clear warning, and a user alter the task. Each contribution has legal significance only in relation to the applicable offence. Involvement at another lifecycle stage does not exclude liability where a culpable breach materialised in harm, but neither establishes it without that connection. The approach accommodates multiple human decisions while preserving individualised fault and evidential requirements.

C. International Legal Benchmarks for Accountability and Remedies

International legal assessment must protect affected persons and individuals facing prosecution. The European Convention on Human Rights protects the right to a fair trial, the presumption of innocence, the principle of no punishment without law, and the right to an effective remedy for Convention violations (Articles 6, 7, and 13) [10]. In the present doctrinal analysis, these guarantees support accessible and effective examination of relevant incidents without replacing proof of personal fault with general assertions about technological danger. Effective protection is not confined to criminal conviction; civil, administrative, and regulatory mechanisms may also contribute to restoring rights.

States' obligations under international human rights law must be distinguished from the substantive grounds for holding a particular developer or operator criminally liable. Inadequate regulation or supervision does not prove a private person's fault, and supplier conduct is not attributable to the state solely because the state has a supervisory obligation. Nevertheless, absence of an individual criminal offence does not eliminate the need to organise dangerous activities appropriately, preserve incident information, and provide remedies

within applicable international obligations. Subjectlessness in criminal attribution must not become a denial of protection for a person harmed by AI.

The Council of Europe Framework Convention, opened for signature on 5 September 2024 (CETS No. 225), addresses activities across the AI lifecycle. It provides for effective remedies, documentation, challenges to decisions, and procedural safeguards (Articles 14–15), together with differentiated risk assessment and mitigation (Article 16) [11]. It establishes a treaty framework for the obligations of its parties, subject to the applicable rules on ratification and entry into force; it does not create a universal criminal prohibition. Its use as a legal benchmark requires identification of the relevant duty holders and consequences, with careful attention to the treaty's binding status and applicability in the jurisdiction concerned. For the proposed model, prevention must be combined with information enabling an incident to be examined.

Relevant information must be sufficient for a competent body to investigate system use and for an affected person to challenge its outcome. Protection should not depend exclusively on a supplier's voluntary explanation after an incident. Disclosure must nevertheless reflect the subject matter, personal-data protection, and others' legitimate interests. Opacity calls for appropriate verification procedures; it does not justify treating the participant responsible for the system as guilty without establishing the offence.

Regulation (EU) 2024/1689 is an EU legislative act whose legal nature, territorial scope, and implementation timetable differ from those of an international treaty. Its human-oversight requirements for high-risk systems are proportionate to risk, autonomy, and use context, and enable understanding of limitations, interpretation of outputs, disregard of recommendations, and intervention (Article 14) [12]. They help specify adequate control but do not automatically prove criminal fault. Its territorial, material, and temporal applicability, including any relevant transitional or amended provisions, must be established in each case. For Ukraine, these obligations also provide a relevant comparative benchmark without displacing the substantive grounds of criminal liability prescribed by Ukrainian law.

Uses of force require separate determination of applicable humanitarian and criminal-law rules. Asaro distinguishes the technical capacity to automate a decision from the permissibility of delegating decisions over life and death, emphasising human judgement and due process [7]. This is a doctrinal argument rather than proof of a general treaty ban on autonomous systems. It supports assessment of delegation decisions before harm occurs. Civilian AI regulation cannot be transferred mechanically to armed-conflict settings, which require a separate assessment of the governing legal framework.

Santoni de Sio and van den Hoven connect meaningful human control with tracking and tracing: responsiveness to relevant human reasons and circumstances, and the ability to trace outcomes to human participants in design and use [6]. These conditions can assist in identifying the competent actor and an earlier risk-creating decision. Traceability, however, is not equivalent to culpability and cannot replace an independent assessment of fault. A demonstrated connection still requires proof of duty, breach, feasible proper conduct, and the necessary mental elements. International legal and doctrinal benchmarks can improve the conditions for attribution without replacing its constitutive legal elements or evidential requirements.

IV. Discussion

Modernisation of criminal-law regulation should retain a functional, anthropocentric approach to human decisions, acts, and omissions while recognising the technological mediation of those actions through AI systems. Priority should be given to differentiated duties concerning testing, disclosure of limitations, admission to operation, supervision, and incident response. Duties must be tied to specific powers and feasible means of compliance. Technical standards may help define the relevant standard of due care, but they cannot independently create criminal prohibitions in the absence of a lawful statutory basis. Certification does not confer unconditional immunity, while a regulatory breach does not automatically constitute a criminal offence.

Risk-based regulation requires preventive obligations proportionate to potential harm, autonomy, and the operating environment. Transport, medicine, and critical infrastructure warrant stricter safeguards than auxiliary information tasks. Classification as high-risk supports intensified regulation; it is not proof of fault. Prosecution still requires the statutory conduct, mental elements, and, for result offences, causation. Permissible technological risk must be distinguished from culpably creating or increasing impermissible danger by considering foreseeability and available preventive measures.

Meaningful human control requires sufficient information, adequate response time, and actual authority to assess a particular risk and intervene effectively. The four gaps identified by Santoni de Sio and Mecacci show why accountability cannot be reduced to the presence of an operator [5]. The governance framework must also allow a decision against deployment where supervision is demonstrably inadequate. Failure to provide information or response mechanisms requires examination of organisational decisions rather than automatic attribution to the operator. Obligations should reflect actual competence and capacity.

Effective human control involves the critical evaluation of system recommendations rather than the merely formal confirmation of algorithmically generated outputs. A person who cannot understand a warning, verify it in time, or stop the system cannot prevent harm merely by being present. Interface design, staff preparation, and response organisation must therefore be distinguished from ongoing operational duties. Control must fit the risk and operating conditions. This allows assessment of whether proper performance of a particular duty could have prevented danger or whether supervision remained nominal.

Proposals *de lege ferenda* should target particularly dangerous and clearly defined breaches, including intentional

concealment of critical defects, disabling mandatory control mechanisms, unauthorised changes of purpose, or continued operation after confirmation of unacceptable risk. Criminalisation requires prior assessment of existing offences and the boundary between criminal, administrative, and civil responses. Special duty holders, prohibited conduct, fault, and proportionate sanctions must be defined clearly. An abstract obligation to "ensure AI safety" should not replace an offence defined with sufficient legal precision to satisfy the principle of legality. These are legislative proposals rather than claims that such prohibitions already exist.

Endangerment offences require especially precise limits. Absence of actual grave harm cannot be compensated for by asserting that a technology is dangerous in general. The breach, required fault, and risk characteristics must distinguish criminal conduct from other non-compliance. Otherwise, a special prohibition may become a universal sanction for technological failure. Legal certainty matters both to duty holders predicting the consequences of conduct and to competent authorities assessing incidents consistently.

Effective algorithmic traceability requires appropriate records of model versions, relevant data, operating settings, decisions, warnings, and recorded human interventions. Documentation, logging, and independent audit support assessment of fault and causation. They should protect affected persons' access to the facts and the accused's ability to challenge the technical case. Digital evidence must be assessed for integrity while personal information is safeguarded. Missing logs alone do not establish guilt for causing harm; breach of a documentation duty requires separate evaluation. Traceability supports the production and testing of evidence rather than creating any presumption of criminal liability.

Technical records must remain connected to the conditions of the particular operation. Documentation for another version or logs lacking settings may not adequately reconstruct an incident. Recorded data should be compared with updates, warnings, and interventions. Independent technical expertise can assist in establishing the relevant technical circumstances, while determination of criminal fault remains a matter for the legally competent authority. Incomplete explainability does not permit substituting an assumption of developer or operator liability for necessary evidence. Traceability thus has a protective as well as supervisory function.

The proposed model is doctrinal in nature and has not been empirically validated against a systematic dataset of criminal cases or incidents involving AI systems. Its application depends on the elements of the relevant national offence, the allocation of legal duties, and the evidence available for a particular system and incident. The international instruments examined provide different forms of binding regulation, conditional treaty obligations, or comparative guidance; they do not establish a single, uniform global regime of criminal liability. Further work should test how lifecycle duties and traceability operate in practice while maintaining the distinction between technical involvement, regulatory non-compliance, and culpable criminal conduct.

V. Conclusion

Subjectlessness reflects an initial attribution gap between an algorithmically generated harmful outcome and an immediately identifiable person who could lawfully bear individual criminal liability. It signifies neither an absence of participants nor a final impossibility of liability. Functional autonomy does not establish criminal-law personhood, and a human connection with a system does not replace proof of fault. Conventional concepts of criminal law remain indispensable, but they must be applied with careful attention to distributed competence, relevant knowledge, and actual practical control.

The multilevel functional model identifies the risk-creating lifecycle stage, the competent person, knowledge and control, a culpable breach, and its causal significance. Each participant is assessed independently; distributed functions do not create collective fault. International benchmarks add requirements of accountability, procedural safeguards, and effective remedies. States' obligations, organisations' regulatory duties, and individual criminal liability remain legally distinct.

Regulatory development should combine differentiated duties, proportionate risk prevention, meaningful human control, narrowly defined prohibitions, and algorithmic traceability. A functional, anthropocentric model neither personifies algorithms nor permits claims of algorithmic autonomy to conceal culpable human decisions. Equally, it rejects criminal condemnation for unforeseeable behaviour beyond a person's control. Effective protection must remain available where criminal conviction is unwarranted, through interaction with other legal mechanisms.

Author Contributions

All authors contributed equally to the conceptualisation and design of the doctrinal study, examination of the cited legal instruments and scholarly literature, legal analysis, drafting, and critical revision of the manuscript. All authors have reviewed and approved the final version and accept responsibility for its content.

Conflict of Interest

The authors declare that they have no known financial or non-financial conflicts of interest in relation to the research, analysis, or conclusions presented in this manuscript.

Funding Information

The authors report that they received no specific financial support from public, commercial, or not-for-profit funding organisations for the research, authorship, or publication of this article.

Data Availability

No new empirical datasets were generated or analysed for this doctrinal legal study.

References

[1] Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology, 6*(3), 175–183. https://doi.org/10.1007/s10676-004-3422-1

[2] Hallevy, G. (2010). The criminal liability of artificial intelligence entities: From science fiction to legal social control. *Akron Intellectual Property Journal, 4*(2), 171–201. https://ideaexchange.uakron.edu/akronintellectualproperty/vol4/iss2/1/

[3] Gless, S., Silverman, E., & Weigend, T. (2016). If robots cause harm, who is to blame? Self-driving cars and criminal liability. *New Criminal Law Review, 19*(3), 412–436. https://doi.org/10.1525/nclr.2016.19.3.412

[4] Giannini, A., & Kwik, J. (2023). Negligence failures and negligence fixes: A comparative analysis of criminal regulation of AI and autonomous vehicles. *Criminal Law Forum, 34*, 43–85. https://doi.org/10.1007/s10609-023-09451-1

[5] Santoni de Sio, F., & Mecacci, G. (2021). Four responsibility gaps with artificial intelligence: Why they matter and how to address them. *Philosophy & Technology, 34*, 1057–1084. https://doi.org/10.1007/s13347-021-00450-x

[6] Santoni de Sio, F., & van den Hoven, J. (2018). Meaningful human control over autonomous systems: A philosophical account. *Frontiers in Robotics and AI, 5*, Article 15. https://doi.org/10.3389/frobt.2018.00015

[7] Asaro, P. (2012). On banning autonomous weapon systems: Human rights, automation, and the dehumanization of lethal decision-making. *International Review of the Red Cross, 94*(886), 687–709. https://doi.org/10.1017/S1816383112000768

[8] Kamensky, D. V. (2026). "Alhorytm-zlochynets" abo pro perspektyvy kryminalizatsii zlovzhyvan z vykorystanniam shtuchnoho intelektu [The "criminal algorithm": On the prospects of criminalising abuses involving artificial intelligence]. *Naukovyi Visnyk Uzhhorodskoho Natsionalnoho Universytetu. Seriia: Pravo, 94*(3), 114–121. https://doi.org/10.24144/2307-3322.2026.94.3.14

[9] Radutnyi, O. E. (2017). Kryminalna vidpovidalnist shtuchnoho intelektu [Criminal liability of artificial intelligence]. *Informatsiia i Pravo, 2*(21), 124–132. https://doi.org/10.37750/2616-6798.2017.2(21).273064

[10] Council of Europe. (1950). *Convention for the Protection of Human Rights and Fundamental Freedoms* (as amended). https://www.echr.coe.int/en/european-convention-on-human-rights

[11] Council of Europe. (2024). *Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law* (CETS No. 225). https://www.coe.int/en/web/conventions/full-list?module=treaty-detail&treatynum=225

[12] European Parliament & Council of the European Union. (2024). Regulation (EU) 2024/1689 of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union, L 2024/1689*. https://eur-lex.europa.eu/eli/reg/2024/1689/oj