# An Analysis of the Foundations of Criminal Responsibility Arising from the Operation of Artificial Intelligence in Light of Recent Developments in International Law

**1. Aree Abbas Qadir : Ph.D. Student, Department of Law, SR.C., Islamic Azad University, Tehran, Iran**
**2. Anwar Ahmadi *: Department of Law, Sa.C., Islamic Azad University, Sanandaj, Iran**
**3. Tayebeh Bijani Mirza: Department of Law, Sa.C., Islamic Azad University, Sanandaj, Iran**
**4. Ebad Rouhi : Assistant Professor of Public International, Department of Law, Cihan University of Sulaymaniyah, Kurdistan, Iraq**

*Correspondence: ahmadiyas@ymail.com

**Abstract**

Artificial intelligence increasingly performs functions traditionally dependent on human judgment, including military targeting, autonomous transportation, predictive policing, surveillance, medical decision-making, and cyber operations. This development has complicated the attribution of criminal responsibility when harmful consequences arise from systems capable of learning, adaptation, and partially autonomous action. This article examines the legal foundations of criminal responsibility arising from the operation of artificial intelligence in light of recent developments in international law. Using a doctrinal and analytical method, it evaluates technical autonomy, legal agency, causation, foreseeability, mens rea, omission, complicity, corporate fault, and command and superior responsibility. The analysis argues that contemporary AI systems lack the consciousness, moral understanding, legal intention, and capacity to experience punishment required for independent criminal responsibility. Accordingly, AI should generally be treated as an instrument, operational intermediary, or risk-generating technology rather than as an autonomous criminal subject. Responsibility should instead be attributed to the natural and legal persons involved throughout the AI lifecycle, including programmers, developers, manufacturers, deployers, operators, corporate managers, military commanders, and public authorities. The appropriate allocation of responsibility depends on each actor's authority, control, causal contribution, knowledge, ability to foresee harm, and capacity to prevent or terminate unlawful conduct. The article further demonstrates that international criminal law, international humanitarian law, and international human rights law already provide important foundations for addressing AI-related harm, although technological opacity, distributed decision-making, and fragmented organizational knowledge create substantial evidentiary and enforcement difficulties. Particular emphasis is placed on autonomous weapons systems, meaningful human control, due diligence, transparency, traceability, and organizational accountability. The study concludes that increasing technological autonomy must not lead to decreasing human accountability.

**Keywords:** Artificial Intelligence; Criminal Responsibility; International Criminal Law; Autonomous Weapons Systems; Meaningful Human Control; Corporate Criminal Liability.

Received: 02 March 2026
Revised: 07 July 2026
Accepted: 15 July 2026
Published: 30 July 2026



**Citation**: Abbas Qadir, A., Ahmadi, A., Bijani Mirza, T., & Rouhi, E. (2026). An Analysis of the Foundations of Criminal Responsibility Arising from the Operation of Artificial Intelligence in Light of Recent Developments in International Law. *Legal Studies in Digital Age,* 5(4), 1-14.

## 1. Introduction

Artificial intelligence has moved beyond its traditional role as a computational instrument and increasingly performs functions that were previously dependent on human judgment, perception, and decision-making. Contemporary AI systems can identify patterns, classify persons, generate recommendations, predict behavior, operate vehicles, detect military targets, control weapons platforms, and adapt their outputs in response to changing environments. These capabilities have created significant benefits in areas such as medicine, transportation, public administration, security, communications, and economic management. At the same time, they have introduced new forms of risk that challenge the conceptual architecture of criminal law. The possibility that an artificial system may produce harmful or unlawful consequences without an immediately identifiable human decision raises fundamental questions about agency, causation, culpability, and accountability. Matthias describes this problem as a "responsibility gap" that emerges when learning systems behave in ways that their designers and operators cannot fully predict or directly control (Matthias, 2004). The legal difficulty is therefore not limited to determining whether harm has occurred; it also concerns identifying the person or institution to whom the harmful conduct can fairly and lawfully be attributed.

Traditional criminal responsibility is primarily anthropocentric. It presupposes a human actor capable of understanding the nature of conduct, forming an intention, appreciating legal prohibitions, and controlling physical behavior. International criminal law similarly rests on the principle that individuals may be held responsible only when the material and mental elements of a crime are established in relation to their own conduct. Werle and Jessberger explain that international criminal responsibility is based on personal culpability and cannot be imposed merely because a person is connected institutionally or politically to a harmful event (Werle & Jessberger, 2020). The growth of AI complicates this framework because the conduct leading to harm may be divided among programmers, data providers, manufacturers, corporate managers, military commanders, operators, and public authorities. The final output may result from thousands of technical and organizational decisions rather than from a single identifiable act. In such circumstances, conventional doctrines of perpetration, complicity, omission, and superior responsibility must be reconsidered in light of the distributed structure of technological decision-making.

The problem becomes more acute when an AI system possesses a high degree of operational autonomy. Autonomy in this context does not necessarily mean consciousness, moral understanding, or free will. It refers instead to the capacity of a system to select actions or produce outputs without immediate human instruction. A self-learning system may modify its internal parameters after deployment and may reach results that were not specifically anticipated by its original designers. Gless, Silverman, and Weigend demonstrate that even in the context of self-driving vehicles, criminal liability becomes difficult when the harmful act cannot be attributed exclusively to the manufacturer, programmer, user, or machine (Gless et al., 2016). This difficulty is amplified in international contexts, especially where AI is incorporated into military operations, cyber activities, border management, surveillance systems, or automated security practices. The greater the distance between the human decision and the ultimate harmful output, the more challenging it becomes to establish causation and mens rea.

AI-related harms may assume many forms. An autonomous vehicle may cause death because of a defective decision model. A predictive policing system may contribute to discriminatory detention. An automated border system may incorrectly classify an asylum seeker as a security threat. A military targeting algorithm may misidentify a civilian object as a lawful military objective. A generative system may be used to produce false evidence, facilitate fraud, or coordinate cyberattacks. King and colleagues argue that artificial intelligence can be involved in both traditional criminal conduct and entirely new forms of technologically enabled wrongdoing (King et al., 2020). These risks are not confined to private relations. They may affect rights protected under international human rights law and may also contribute to conduct amounting to war crimes, crimes against humanity, or other serious violations of international law. The legal system must therefore address not only malicious uses of AI but also harmful consequences arising from negligent design, reckless deployment, insufficient oversight, or institutional failure.

The military use of artificial intelligence illustrates the seriousness of the problem. Autonomous weapons systems may be capable of identifying, selecting, and engaging targets with limited human intervention. Such technologies challenge the capacity of existing legal rules to ensure compliance with distinction, proportionality, and precautions in attack. Schmitt and Thurnher argue that autonomous weapon systems must remain subject to the law of armed conflict even where machines

perform functions traditionally undertaken by human combatants (Schmitt & Thurnher, 2013). However, the formal applicability of legal rules does not resolve the question of who is criminally responsible when an autonomous system violates those rules. The programmer may have designed the system years before its deployment, the commander may have relied on technical assurances, the operator may have had no realistic opportunity to intervene, and the manufacturer may have concealed known limitations. The system itself may have selected the particular target, but it does not necessarily possess the legal or moral qualities required for punishment.

Some scholars have proposed recognizing a form of legal personality for advanced artificial intelligence. According to this approach, highly autonomous systems might be treated as legal subjects capable of bearing certain obligations or liabilities. Chesterman, however, emphasizes the distinction between functional legal personality and genuine moral agency, warning that legal recognition of AI entities may obscure rather than solve questions of accountability (Chesterman, 2020). Criminal law is particularly resistant to artificial personality because punishment traditionally presupposes blameworthiness, deterrability, and the capacity to experience legal condemnation. A machine cannot suffer imprisonment, understand moral reproach, or reform its conduct in the human sense. Assigning criminal responsibility to AI could therefore become a symbolic exercise that allows human and corporate actors to distance themselves from the consequences of systems they created, deployed, or controlled.

Hallevy presents several theoretical models for addressing machine-related criminality, including situations in which AI acts as an innocent agent, as a foreseeable consequence of human conduct, or as a potentially independent entity (Hallevy, 2013). These models are useful because they demonstrate that the legal response need not be uniform. In some cases, AI is merely a sophisticated instrument used intentionally by a human offender. In other cases, the harmful outcome results from negligent development or reckless deployment. In still other cases, no actor possesses full knowledge of the final result, but several actors collectively create or maintain an unacceptable risk. The central challenge is therefore to develop a differentiated framework capable of identifying the relevant form of responsibility without abandoning the principle of personal culpability.

The development of international criminal law provides important doctrinal resources for this task. International crimes are frequently committed through complex organizational structures involving political leaders, commanders, corporations, intermediaries, and subordinate actors. Ambos explains that doctrines such as indirect perpetration, co-perpetration, aiding and abetting, and superior responsibility have evolved to address crimes that cannot be reduced to the physical act of a single offender (Ambos, 2013). These doctrines may be relevant where AI operates as part of a broader criminal enterprise or institutional process. A military commander who knowingly deploys an unreliable autonomous weapon, a corporation that deliberately supplies technology for unlawful targeting, or a programmer who intentionally creates a system to facilitate persecution may be responsible even if none of them directly performs the final harmful act.

Nevertheless, the application of existing doctrines must be approached cautiously. Criminal law cannot impose liability merely because an individual participated in the development or use of a dangerous technology. Innovation necessarily involves uncertainty, and not every technical failure is criminal. The distinction between ordinary error, civil negligence, regulatory noncompliance, recklessness, and intentional wrongdoing must be preserved. Overly broad criminalization could undermine technological development and violate the principle that criminal offences must be defined with sufficient clarity. At the same time, excessive reliance on technological complexity may create impunity. Where corporations or governments design opaque systems, suppress information about known defects, or deploy high-risk applications without adequate safeguards, claims of unpredictability should not automatically eliminate liability.

The objective of this article is to analyze the legal foundations on which criminal responsibility may be attributed for harmful conduct arising from the operation of artificial intelligence. The study examines the relationship between technical autonomy and legal agency, the material and mental elements of AI-related offences, the attribution of responsibility among actors involved in the AI lifecycle, and the relevance of international criminal law, international humanitarian law, and international human rights law. The central argument is that current international law does not provide a persuasive basis for treating AI as an independent criminal subject. Responsibility should instead be distributed among natural and legal persons according to their authority, control, contribution, knowledge, and capacity to prevent harm.

The article adopts a doctrinal and analytical method. It examines established principles of criminal responsibility and evaluates their application to technologically mediated conduct. Particular attention is given to autonomous weapons because

they demonstrate the interaction between AI, international humanitarian law, human rights, and individual criminal responsibility. The analysis also considers civilian applications where relevant, including automated vehicles and algorithmic decision-making systems. The proposed framework seeks to prevent an accountability gap while maintaining the requirements of causation, culpability, legality, and fair attribution. The ultimate purpose is not to create an entirely separate criminal law for artificial intelligence, but to determine how existing principles should be interpreted and developed in response to new forms of distributed technological agency.

## 2. Conceptual and Legal Foundations of AI-Related Criminal Responsibility

A coherent analysis of AI-related criminal responsibility must begin with a legally useful understanding of artificial intelligence. AI is not a single technology and should not be treated as a uniform category. The term includes rule-based systems, machine-learning applications, deep-learning models, autonomous vehicles, decision-support tools, predictive algorithms, and weapons systems capable of selecting and engaging targets. Some systems merely provide information to human users, while others generate recommendations that are routinely followed without meaningful review. More advanced systems can adapt to new data and modify their performance after deployment. King and colleagues emphasize that the criminal risks associated with AI depend heavily on the system's functional capabilities, accessibility, autonomy, and potential for misuse (King et al., 2020). A legal framework must therefore distinguish among systems according to their purpose, level of autonomy, operating environment, capacity for self-modification, and potential impact on protected interests.

The concept of autonomy is especially important but often misunderstood. Technical autonomy refers to a system's capacity to perform tasks or select actions without continuous human direction. It does not establish that the system possesses consciousness, intention, moral judgment, or legal understanding. Chesterman notes that legal personality is a construct that may be granted for practical purposes, but the existence of functional independence does not necessarily justify recognition of full legal subjectivity (Chesterman, 2020). A corporation, for example, is granted legal personality because doing so facilitates the organization of rights, duties, and assets. Criminal responsibility, however, involves more than functional convenience. It is connected to moral blame, deterrence, condemnation, and the capacity to understand the meaning of punishment. Current AI systems lack these characteristics.

The distinction between agency and autonomy has direct implications for the actus reus of AI-related offences. Criminal law traditionally requires a voluntary act or legally relevant omission. When an AI system generates an unlawful output, the immediate physical or digital act may be performed by the machine, but legal analysis must determine whether that act can be attributed to a human or institution. Hallevy's model of AI as an innocent agent treats the system as an instrument through which a human perpetrator commits an offence (Hallevy, 2013). This model is straightforward where a person intentionally programs or deploys AI to achieve a criminal purpose. An individual who uses an automated system to conduct fraud, launch cyberattacks, identify victims, or manipulate financial markets remains responsible as the principal offender. The machine does not interrupt the causal or legal relationship between the human decision and the prohibited result.

More difficult cases arise where the AI system behaves in an unanticipated manner. A learning algorithm may produce an outcome that was not specifically intended by the developer or operator. Matthias argues that adaptive systems can create situations in which neither the programmer nor the user exercises sufficient control to satisfy conventional assumptions of responsibility (Matthias, 2004). This does not mean that responsibility necessarily disappears. The relevant inquiry may shift from whether a person intended the exact output to whether that person created, increased, or knowingly tolerated a substantial risk. Criminal law frequently imposes responsibility for reckless or grossly negligent conduct even where the precise harmful consequence was not desired.

Causation is central to this inquiry. In technologically complex systems, harm may result from the interaction of software design, training data, environmental conditions, user instructions, maintenance decisions, cybersecurity vulnerabilities, and institutional policies. No single factor may be sufficient on its own. Gless, Silverman, and Weigend show that autonomous vehicle accidents can involve overlapping contributions from engineers, manufacturers, owners, and users (Gless et al., 2016). The same structural problem appears in military and governmental AI systems. A targeting error may originate in biased training data, inadequate sensor information, an unreliable classification model, insufficient testing, or inappropriate

operational parameters. The causal chain may also include a commander's deployment decision and an operator's failure to intervene.

Traditional "but for" causation may be inadequate where several actors contribute to a harmful outcome. International criminal law has developed broader concepts of contribution because collective crimes often result from coordinated structures rather than isolated acts. Ambos explains that criminal responsibility may arise where an individual makes an essential or substantial contribution to a common criminal process, depending on the applicable mode of liability (Ambos, 2013). In AI-related cases, the law may need to examine whether an actor materially contributed to the creation or continuation of an unlawful risk. This approach does not eliminate causation; it recognizes that technological harms may have multiple legally significant causes.

The mental element presents an equally serious challenge. Intent remains the clearest basis for responsibility. A person who designs an AI system to facilitate unlawful killing, persecution, fraud, or cybercrime possesses the required criminal purpose. Knowledge may also be sufficient where an actor is aware that the system will be used to commit a crime. More controversial cases involve recklessness, wilful blindness, or negligence. A manufacturer may discover that a system repeatedly misidentifies protected persons but continue distribution because correction would be costly. A military commander may authorize deployment despite reports that the system cannot reliably distinguish civilians from combatants. A public authority may use an algorithm known to produce discriminatory outcomes. In each case, the relevant mental state concerns awareness of risk rather than desire for the final result.

Foreseeability must therefore be assessed with care. The fact that an exact output was unpredictable does not necessarily mean that the category of harm was unforeseeable. Crootof argues that autonomous weapons create legal and policy challenges because their specific actions may be difficult to predict even when the general risks of their deployment are known (Crootof, 2015). A system may select an unexpected target, but the risk of target misclassification may have been apparent before deployment. Criminal responsibility should not depend on whether the accused anticipated the precise sequence of events. It should instead consider whether the actor knew or should have known that the system created a substantial and unjustifiable danger.

The degree of human control is another foundational factor. Meaningful human control is not satisfied merely because a human formally approves a machine-generated recommendation. Amoroso and Tamburrini explain that control must include adequate understanding, sufficient information, practical intervention capacity, and real authority over the system's operation (Amoroso & Tamburrini, 2020). A person who is technically "in the loop" but lacks time, expertise, or authority to challenge the AI output may not exercise genuine control. Conversely, a senior official who defines the parameters of deployment and decides whether the system will be used may possess significant control even if not involved in individual decisions.

The distinction between human-in-the-loop, human-on-the-loop, and human-out-of-the-loop systems is legally relevant but not determinative. A human-in-the-loop system requires approval before action, while a human-on-the-loop system permits supervision and intervention during operation. A human-out-of-the-loop system operates without immediate human participation. Schmitt and Thurnher observe that autonomous weapon systems may operate within the law of armed conflict if they are designed and deployed in circumstances where compliance can reasonably be ensured (Schmitt & Thurnher, 2013). However, the removal of immediate human involvement may increase the duties of those who design, authorize, and supervise the system. Responsibility may shift upward rather than disappear.

The possibility of AI legal personality has been proposed as a response to these challenges, but it is conceptually and practically problematic. Criminal punishment ordinarily serves retribution, deterrence, rehabilitation, incapacitation, and public condemnation. An AI system cannot experience suffering, shame, moral criticism, or rehabilitation. It can be deactivated or modified, but those measures are regulatory or technical rather than punitive in the conventional sense. Sparrow argues that delegating lethal decisions to machines creates a fundamental moral problem because no genuinely responsible agent may be identifiable at the moment of killing (Sparrow, 2007). Treating the machine as the offender would not resolve this moral deficit. It could instead institutionalize it.

There is also a risk that AI personality would become a liability shield. Corporations could assign responsibility to autonomous systems while retaining economic benefits from their operation. Governments could attribute unlawful targeting to algorithmic error. Developers could claim that self-learning systems departed from their original design. Chesterman

cautions that legal personality should not be used to conceal the human and institutional relationships underlying technological systems (Chesterman, 2020). In the criminal context, the central question should remain which human or organization possessed authority, knowledge, and preventive capacity.

The principle of legality also limits the expansion of criminal responsibility. Individuals cannot be punished under vague or retroactively constructed standards. AI-related offences must be defined with sufficient clarity, and existing doctrines should not be stretched beyond their legitimate boundaries. Werle and Jessberger emphasize that international criminal law must preserve strict interpretation because of the severity of criminal punishment (Werle & Jessberger, 2020). This requirement does not prevent adaptation, but it requires careful reasoning. A developer should not be criminally liable merely because a system caused harm. The prosecution must establish a recognized duty, a sufficiently significant contribution, a relevant mental state, and a legally attributable result.

The most defensible conceptual model is therefore a human-centered and institution-centered approach. AI should generally be treated as an instrument, operational intermediary, or risk-generating technology rather than as an independent criminal subject. The degree of autonomy remains relevant because it affects foreseeability, control, and causation. However, autonomy should not be allowed to dissolve accountability. Where a system becomes less predictable, the actors responsible for deployment may have stronger duties of testing, monitoring, limitation, and intervention. The legal inquiry should focus on who created the risk, who benefited from it, who possessed the power to control it, and who failed to act despite knowledge of danger.

This framework preserves the principle of culpability while acknowledging the distinctive features of AI. It recognizes that responsibility may be direct, indirect, collective, corporate, or omission-based. It also permits different forms of liability for different actors. A programmer who intentionally inserts an unlawful function is not situated in the same way as an operator who reasonably relies on an apparently safe system. A commander who deploys an unreliable autonomous weapon is not equivalent to a technician who performs routine maintenance without knowledge of the broader operation. The law must therefore move beyond a binary choice between blaming the machine and blaming every human associated with it. A differentiated analysis is required in which responsibility follows control, knowledge, contribution, and legal duty.

## 3. Attribution of Criminal Responsibility among Actors in the AI Lifecycle

The attribution of criminal responsibility for AI-related harm requires examination of the entire technological lifecycle. AI systems are not created or deployed by a single actor. They emerge from networks of programmers, data scientists, manufacturers, investors, corporate directors, regulators, users, operators, and public officials. Each actor performs a different function and possesses a different degree of knowledge and control. Criminal responsibility should therefore be allocated according to the specific contribution of each actor rather than imposed collectively without distinction. The central task is to identify the relevant conduct, the applicable legal duty, the causal relationship to the harm, and the actor's mental state.

The simplest case is the intentional use of AI as an instrument of crime. A person may employ an AI system to automate fraud, conduct cyberattacks, generate deceptive content, identify vulnerable victims, or coordinate unlawful surveillance. King and colleagues explain that AI can expand the scale, speed, and sophistication of criminal activity while lowering the technical barriers to participation (King et al., 2020). In such situations, traditional principles of direct perpetration remain applicable. The offender forms the criminal intent and uses the system as a means of execution. The fact that the system performs complex operations does not transform it into the principal perpetrator.

AI may also be used as an instrument in the commission of international crimes. A political or military authority may employ algorithmic systems to identify members of a targeted ethnic group, prioritize persons for detention, generate lists of suspected opponents, or select objects for attack. If the system is integrated into a broader plan of persecution or unlawful violence, the persons who design and implement that plan may be responsible as perpetrators or co-perpetrators. Werle and Jessberger note that international crimes often involve organized structures in which senior actors do not physically perform the criminal acts (Werle & Jessberger, 2020). The use of AI therefore does not create an entirely new form of liability; it may become part of an existing organizational apparatus of crime.

Programmers and developers occupy a more complex position. Software development is generally remote from the final harmful event, and programmers may have limited knowledge of how a system will later be used. Criminal liability should not

arise merely because code contributes to an undesirable result. Responsibility becomes more plausible where a developer intentionally designs a system for unlawful purposes, knowingly includes prohibited functions, conceals serious defects, or provides substantial assistance with awareness of intended criminal use. Hallevy's analysis suggests that a programmer may be responsible where the AI system functions as an innocent agent acting on the programmer's instructions (Hallevy, 2013). In such cases, the machine's operational autonomy does not interrupt the programmer's responsibility.

Reckless development may also justify liability in exceptional circumstances. A developer may know that a system has a high probability of causing serious harm but proceed without adequate safeguards. The threshold should remain high because software development necessarily involves uncertainty and error. Criminal negligence should be distinguished from ordinary technical imperfection. The relevant factors may include the seriousness of the foreseeable harm, the availability of safer alternatives, the extent of testing, the transparency of known limitations, and whether warnings were deliberately ignored.

Manufacturers and technology companies may bear responsibility when organizational decisions create or tolerate unlawful risks. AI development is frequently a corporate process involving multiple departments and decision-making levels. Individual responsibility may be difficult to establish because knowledge is fragmented. One team may identify a safety defect, another may approve commercial release, and senior management may prioritize market entry over further testing. Gless, Silverman, and Weigend demonstrate that autonomous systems challenge traditional criminal doctrines because responsibility may be distributed across corporate structures (Gless et al., 2016). An organizational-fault model may therefore be more appropriate than a narrow search for a single guilty employee.

Corporate criminal responsibility can be based on several theories. Under an identification model, the conduct and mental state of senior managers are attributed to the corporation. Under vicarious liability, the corporation may be responsible for offences committed by employees within the scope of their employment. An organizational-fault model examines whether corporate policies, culture, incentives, or management failures created the conditions for harm. The latter approach is particularly relevant to AI because unsafe outcomes may result from systemic practices rather than from one decision. A corporation that suppresses safety reports, discourages internal criticism, or deploys high-risk systems without adequate review may exhibit institutional culpability even where no single person possesses complete knowledge.

Operators and users may also incur responsibility, especially when they exercise direct control over deployment. An operator may intentionally enter unlawful parameters, ignore system warnings, use the technology outside its authorized environment, or rely on outputs known to be unreliable. In military settings, an operator may approve a target without conducting the required legal assessment. In civilian contexts, an official may rely on an automated recommendation despite clear evidence of bias or error. Responsibility depends on the extent to which the operator had access to relevant information and a realistic opportunity to intervene.

Automation bias complicates this analysis. Human users often place excessive trust in machine-generated recommendations, particularly where systems are presented as objective or scientifically superior. Formal human approval may therefore conceal practical dependence on the algorithm. Amoroso and Tamburrini argue that meaningful human control requires more than nominal participation (Amoroso & Tamburrini, 2020). A user must understand the system's limitations, possess adequate time and information, and have authority to reject its output. Where these conditions are absent, responsibility may lie more heavily with the institution that designed the decision-making process.

Deployers and owners occupy an important intermediate position. They determine whether a particular system is appropriate for a given context. A hospital, military unit, police agency, or corporation may acquire a system and use it in circumstances for which it was not designed. Even if the system is technically sound, inappropriate deployment may create criminally relevant risks. A targeting algorithm trained for conventional battlefield conditions may be unreliable in a densely populated urban environment. A facial-recognition system may be used despite documented disparities affecting particular groups. A predictive model may be applied to decisions involving detention or lethal force without adequate validation.

The legal duty of deployers should therefore include assessment of context. They must evaluate whether the system is suitable for the intended purpose, whether its limitations are understood, and whether adequate safeguards exist. Crootof emphasizes that autonomous weapons raise accountability concerns because the law must examine not only design but also the circumstances in which the technology is used (Crootof, 2016). A system that is lawful in one environment may be unlawful in another. Responsibility may arise when decision-makers ignore this contextual dimension.

Data providers and trainers may also influence harmful outcomes. Machine-learning systems depend on large datasets that determine how they classify information and generate predictions. Biased, incomplete, manipulated, or unlawfully obtained data can produce discriminatory or dangerous results. Criminal responsibility should not attach to every data problem, but deliberate data poisoning or knowing manipulation may constitute substantial assistance to an offence. A person who intentionally trains a system to misidentify protected persons or to facilitate unlawful targeting may be responsible as an accomplice or co-perpetrator.

The causal role of data is often difficult to prove because model behavior may result from complex statistical relationships. The law should therefore require evidence that the data-related conduct materially contributed to the harmful output. Mere participation in data collection is insufficient. Knowledge is also essential. A data provider who reasonably believes that information will be used for lawful research is not equivalent to one who knows that the dataset will support persecution or indiscriminate attack.

Aiding and abetting provides a potentially important basis for liability. A programmer, supplier, financier, or platform operator may provide practical assistance to a principal offence without directly controlling the final act. Ambos explains that accessorial responsibility generally requires a sufficiently significant contribution accompanied by the relevant knowledge or purpose, depending on the applicable legal framework (Ambos, 2013). In AI-related cases, the prosecution would need to establish that the accused's contribution facilitated the crime and that the accused possessed the required awareness of the criminal context.

Commercial suppliers present a particularly difficult issue. Dual-use AI systems may have both lawful and unlawful applications. A company may sell surveillance technology to a government that later uses it for repression. Criminal responsibility should not be imposed solely because misuse was conceivable. The supplier's knowledge must be assessed through evidence such as contractual terms, prior incidents, internal communications, public reports, customer demands, and technical customization. Responsibility becomes more plausible where the supplier modifies the system specifically for unlawful purposes or continues assistance after receiving credible evidence of abuse.

Collective forms of responsibility may apply where AI is integrated into a common criminal plan. Co-perpetration may arise when several actors jointly control essential aspects of the operation. Indirect perpetration may be relevant where a senior actor uses an organization and its technological systems as instruments. Common-purpose liability may apply when participants intentionally contribute to a group acting with a shared criminal objective. Werle and Jessberger discuss how international criminal law attributes responsibility to persons who exercise control through organizational structures rather than through immediate physical conduct (Werle & Jessberger, 2020). AI can become part of such a structure.

Command and superior responsibility are especially relevant in military contexts. A commander may authorize the use of autonomous weapons, define operational parameters, receive reports concerning system performance, and possess authority to suspend deployment. Responsibility may arise where the commander knew or should have known that subordinates were using the system to commit crimes and failed to prevent or punish them. The application of superior responsibility becomes more complex when the immediate harmful act is generated by an autonomous system rather than by a subordinate soldier.

The key issue is whether the commander retained effective control over the human and technological process. Schmitt and Thurnher maintain that commanders remain legally responsible for decisions to deploy autonomous systems and must evaluate whether their use complies with the law of armed conflict (Schmitt & Thurnher, 2013). Delegating target selection to software does not eliminate command responsibility. If anything, the use of an unpredictable system may increase the commander's duty to monitor, restrict, or terminate operations.

Omission liability is also central. Responsibility may arise where an actor has a legal duty to prevent harm but fails to act. Such duties can derive from professional roles, contractual obligations, statutory regulation, command authority, corporate office, or prior creation of risk. A manufacturer that learns of a critical defect may have a duty to issue warnings or suspend the system. A commander may have a duty to halt deployment after repeated unlawful outcomes. A regulator may have a duty to investigate serious violations. The failure to act becomes criminally relevant where the actor possesses the capacity to intervene and the required mental state.

The creation of a dangerous technological system may itself generate continuing duties. A company that releases a self-learning model cannot necessarily treat its responsibility as complete at the moment of sale. Post-deployment monitoring may

be required because the system's behavior can change over time. A state that deploys AI for security purposes may need to preserve logs, investigate incidents, and provide remedies. Crootof's accountability analysis supports the view that responsibility must extend across the operational life of autonomous systems (Crootof, 2016).

Strict criminal liability is generally inappropriate for serious AI-related offences. Although strict liability may encourage safety and simplify enforcement, it conflicts with the principle of personal culpability. Criminal punishment should ordinarily require proof of intent, knowledge, recklessness, or at least gross negligence. Administrative penalties, civil liability, mandatory insurance, and regulatory sanctions may operate under broader standards. Serious criminal sanctions should remain connected to blameworthy conduct.

An effective attribution model should therefore proceed through several analytical stages without reducing the inquiry to one formal test. The system's function and level of autonomy must first be identified. The relevant actors and their roles must then be mapped. Each actor's legal duty, contribution, authority, knowledge, and preventive capacity should be assessed. The analysis should determine whether the actor created or increased an unlawful risk, whether the harm fell within that risk, and whether the required mental state existed. The final step is to select the appropriate mode of responsibility, whether direct perpetration, indirect perpetration, complicity, omission, corporate fault, or superior responsibility.

This differentiated model avoids two opposing errors. It prevents the machine from becoming a convenient scapegoat, but it also avoids imposing liability on every person associated with the technology. Criminal responsibility must remain individualized. The complexity of AI does not justify collective punishment, yet it should not permit strategic fragmentation of responsibility. Actors who intentionally divide knowledge and control across an organization should not benefit from the resulting opacity. The law must be capable of reconstructing the entire decision-making process and identifying where culpable choices were made.

## 4. AI-Related Criminal Responsibility in Light of Recent Developments in International Law

The relevance of artificial intelligence to international law is most visible in the fields of international criminal law, international humanitarian law, and international human rights law. AI systems can contribute to serious violations by expanding the capacity of states, armed groups, and corporations to identify persons, process information, coordinate operations, and apply force. The technology may be used to support lawful objectives, but it may also facilitate systematic persecution, indiscriminate attack, arbitrary detention, unlawful surveillance, or extrajudicial killing. The central legal issue is whether existing international rules can adequately govern conduct mediated by autonomous or semi-autonomous systems.

International criminal law is based on individual responsibility for genocide, crimes against humanity, war crimes, and other offences within the jurisdiction of international tribunals. These crimes often involve large organizations and complex chains of command. AI may become embedded within these structures. A system may identify members of a protected group, prioritize targets, analyze communications, recommend detention, or automate elements of military planning. Werle and Jessberger explain that international criminal law focuses on the personal contribution of the accused to the collective commission of crimes (Werle & Jessberger, 2020). The involvement of AI does not alter the prohibited nature of the underlying conduct, but it may complicate proof of contribution and intent.

AI systems are unlikely to become independent defendants under current international criminal law. International tribunals exercise jurisdiction over natural persons, and the normative foundations of international punishment remain closely connected to human agency. Chesterman's analysis of legal personality reinforces the conclusion that artificial systems do not possess the moral and institutional characteristics necessary for criminal subjectivity (Chesterman, 2020). The more plausible approach is to examine how AI functions within the acts of human perpetrators, accomplices, commanders, and organizations.

Autonomous weapons systems have generated the most developed international debate. These systems may identify, select, and engage targets with varying degrees of human involvement. Their legality depends not solely on their technical classification but on the manner in which they are designed and used. Schmitt and Thurnher argue that autonomous weapons are not inherently unlawful, but their deployment must comply with the principles and rules of the law of armed conflict (Schmitt & Thurnher, 2013). This position emphasizes contextual legality rather than categorical prohibition.

The principle of distinction requires parties to distinguish civilians and civilian objects from combatants and military objectives. An autonomous system must therefore possess sufficient technical capacity to classify targets accurately in the

operational environment. The principle of proportionality requires assessment of whether expected civilian harm would be excessive in relation to anticipated military advantage. This judgment is inherently contextual and often depends on qualitative considerations. Wagner argues that delegating such evaluations to machines may dehumanize international humanitarian law by removing moral and interpretive judgment from decisions involving life and death (Wagner, 2014).

Precautions in attack require feasible measures to verify targets, select appropriate means and methods, and minimize civilian harm. These duties apply before and during an operation. A commander who relies on an AI system must therefore understand its capabilities and limitations. Technical reliability cannot be assumed simply because a system performs well in controlled testing. Environmental complexity, adversarial manipulation, incomplete data, and unexpected human behavior may undermine performance.

Meaningful human control has emerged as a central concept in the debate. Amoroso and Tamburrini argue that meaningful control requires informed human judgment, adequate situational awareness, and a genuine capacity to supervise or intervene (Amoroso & Tamburrini, 2020). The concept seeks to preserve human responsibility over decisions that have serious legal and moral consequences. However, meaningful control remains difficult to define. A human may formally approve an attack but rely almost entirely on machine-generated information. An operator may have authority to intervene but insufficient time to do so. A commander may understand the general system but not the technical basis of a particular output.

A legally effective concept of meaningful human control should include several elements. The human decision-maker must understand the purpose and limitations of the system. The operational environment must fall within tested and authorized conditions. The decision-maker must receive sufficient information to evaluate the recommendation. There must be adequate time and authority to intervene. Responsibility must be clearly allocated before deployment. Records must be preserved so that decisions can later be reconstructed.

Asaro argues that autonomous lethal decision-making raises human rights concerns because the decision to kill should remain subject to human moral and legal judgment (Asaro, 2012). His position reflects a broader concern that machines cannot appreciate human dignity or interpret ambiguous behavior in the same way as accountable human agents. The use of AI in lethal force may therefore threaten both substantive rights and procedural accountability.

Heyns similarly emphasizes the importance of human control in decisions concerning the use of lethal force (Heyns, 2013). The right to life requires that force be necessary, proportionate, and subject to effective accountability. An autonomous system may execute force rapidly, but speed can reduce the opportunity for reflection and correction. Where a system makes an erroneous decision, victims and investigators may be unable to identify who was responsible.

Sparrow's analysis of killer robots focuses on the moral impossibility of assigning responsibility when no human fully controls the lethal act (Sparrow, 2007). This concern has direct legal implications. International law requires accountability for unlawful killing, but autonomous systems may fragment decision-making across time and institutions. The programmer develops the system, the manufacturer integrates it, the commander authorizes it, and the operator activates it. The final target selection may be generated by the machine. Without clear rules, each actor may claim that another possessed the decisive role.

The law should respond by treating responsibility as layered rather than singular. The operator may be responsible for immediate misuse. The commander may be responsible for unlawful deployment or failure to supervise. The manufacturer may be responsible for concealment of known defects. The programmer may be responsible for intentional design features. The state may bear international responsibility for the conduct of its organs. These forms of responsibility can coexist.

International human rights law broadens the analysis beyond armed conflict. AI systems may affect the right to life, privacy, equality, freedom of expression, liberty, due process, and access to an effective remedy. Automated decision-making may be used in policing, migration, welfare administration, and criminal justice. Harm may arise even where no physical violence occurs. Discriminatory classification, arbitrary surveillance, or automated denial of legal rights can produce serious violations.

The use of AI by public authorities triggers obligations of legality, necessity, proportionality, and non-discrimination. A state cannot avoid responsibility by claiming that an algorithm made the decision. The system is part of the state's administrative machinery. If a public agency adopts an AI output, the resulting action remains attributable to the state. Individual criminal responsibility may also arise where officials intentionally use AI to facilitate persecution, arbitrary detention, or other serious abuses.

Private companies may likewise contribute to human rights violations. Technology firms develop surveillance tools, facial-recognition systems, predictive analytics, and military applications. Their conduct may support state repression or armed conflict. Criminal liability depends on the company's knowledge and contribution. Crootof's discussion of accountability for autonomous weapons illustrates the need to examine the roles of private manufacturers and suppliers alongside those of military users (Crootof, 2016).

Due diligence provides an important bridge between international legal obligations and criminal responsibility. States and corporations should assess risks before deploying high-impact AI systems. They should test performance, evaluate bias, establish oversight, preserve records, and monitor post-deployment behavior. Failure to perform these duties may constitute negligence or recklessness when serious harm is foreseeable. The precise criminal threshold will depend on the applicable law, but due diligence standards help define what reasonable conduct requires.

Transparency is essential to accountability. AI systems may be technically opaque, protected by trade secrecy, or inaccessible to victims and investigators. Without records, it may be impossible to determine why a system produced a harmful output. The law should therefore require preservation of system logs, training documentation, operational parameters, warnings, updates, and human interventions. These materials are the technological equivalent of evidence concerning orders, communications, and decision-making.

Explainability must be understood functionally. It may not be possible to translate every internal computational process into ordinary language. However, criminal investigations do not always require complete technical interpretation. They require sufficient information to identify relevant inputs, system limitations, known risks, human decisions, and causal contributions. A company should not be permitted to avoid responsibility by designing systems that cannot be audited.

The destruction or concealment of AI records should be treated seriously. An organization that intentionally deletes logs or suppresses safety reports may obstruct investigation. Such conduct may also support an inference that the organization knew of the risk. Corporate opacity should not become a structural defence. The burden of technological complexity should not fall entirely on victims.

Recent developments in international legal discourse increasingly emphasize accountability, human control, risk assessment, and preventive governance. Although binding rules remain fragmented, the direction of development is clear. Autonomous systems are not regarded as legally independent from those who design and deploy them. Human actors remain responsible for ensuring compliance with international law. Crootof notes that the emergence of autonomous weapons requires adaptation of accountability mechanisms rather than abandonment of existing law (Crootof, 2015).

The debate over prohibition and regulation reflects competing views. Some scholars argue that autonomous weapons should be banned because machines cannot make morally and legally adequate lethal decisions. Asaro supports restrictions grounded in human rights and the preservation of human judgment (Asaro, 2012). Others argue that autonomous systems may in some circumstances reduce civilian harm by improving accuracy and removing emotional factors. Schmitt and Thurnher emphasize that legality depends on capability and context rather than autonomy alone (Schmitt & Thurnher, 2013).

A criminal-responsibility framework need not resolve the entire policy debate. Whether systems are prohibited or regulated, responsibility must remain identifiable. If certain uses are categorically banned, intentional deployment may itself become criminally relevant. If use is permitted under conditions, responsibility may arise from failure to satisfy those conditions. In both models, legal duties must be specified in advance.

The accountability-gap argument should be divided into normative and evidentiary dimensions. A normative gap exists where no legal rule clearly applies to a new form of conduct. An evidentiary gap exists where the law applies but proof is difficult. Matthias's original formulation emphasizes the difficulty of assigning responsibility when learning systems behave unpredictably (Matthias, 2004). In many AI cases, however, the more serious problem may be evidentiary. Existing doctrines of intent, negligence, complicity, and superior responsibility may be conceptually available, but technical opacity prevents their effective application.

This distinction is important because the solutions differ. A normative gap may require new offences or modes of liability. An evidentiary gap requires documentation, auditability, investigative expertise, access to corporate information, and international cooperation. The law should avoid creating unnecessary new crimes where existing principles are sufficient. At the same time, existing principles must be supported by procedures capable of uncovering technological facts.

A coherent international framework should classify AI systems according to risk and function. Systems used in lethal force, detention, policing, border control, and critical infrastructure should be subject to enhanced duties. The relevant actors should be identified before deployment. Legal responsibility should be allocated across the lifecycle. Human oversight should be genuine rather than symbolic. Records should be preserved. Independent investigation should follow serious incidents.

Criminal responsibility should then be assessed through the combined elements of control, contribution, knowledge, foreseeability, and omission. A person who intentionally uses AI to commit a crime should be treated as a direct perpetrator. A person who knowingly facilitates criminal use may be an accomplice. A commander who deploys an unreliable system despite clear warnings may be responsible for reckless conduct or superior failure. A corporation that systematically suppresses known risks may bear organizational responsibility.

This approach also recognizes the relationship between individual and state responsibility. A state may be internationally responsible for unlawful AI use by its organs, while individual officials may simultaneously incur criminal liability. Corporate responsibility may coexist with the liability of managers and technical personnel. These forms of responsibility are complementary rather than mutually exclusive.

International cooperation will be necessary because AI development and deployment are frequently transnational. A system may be designed in one country, trained with data from another, sold by a multinational corporation, and used in a third state. Evidence may be distributed across several jurisdictions. Effective accountability will require cross-border access to digital records, harmonized preservation duties, mutual legal assistance, and technical expertise.

The central lesson of recent international legal development is that autonomy does not justify legal abandonment. The more complex and powerful an AI system becomes, the more carefully responsibility must be designed. Law should not wait until harm occurs and then search retrospectively for an offender. Duties of testing, supervision, documentation, and intervention should be allocated in advance. Criminal responsibility should operate as the final element of a broader preventive framework.

## 5. Conclusion

Artificial intelligence challenges criminal law because it separates harmful outcomes from the immediate and visible conduct of a single human actor. Autonomous and learning systems can generate decisions that are difficult to predict, explain, or control. Their operation may involve many participants distributed across technical, corporate, governmental, and military structures. As a result, traditional concepts of agency, causation, intent, and control must be applied with greater precision.

The analysis demonstrates that technical autonomy should not be equated with legal or moral agency. Contemporary AI systems do not possess consciousness, moral understanding, legal intention, or the capacity to experience punishment. They may perform actions that resemble decision-making, but this functional capacity does not justify treating them as independent criminal subjects. Criminal liability imposed directly on AI would have limited deterrent or retributive value and could create a dangerous mechanism through which human and institutional actors evade accountability.

The more persuasive approach is to treat AI as an instrument, intermediary, or risk-generating system through which human and organizational conduct is expressed. This does not mean that AI is legally irrelevant. The degree of autonomy affects the analysis of foreseeability, causation, control, and preventive duty. A highly autonomous system may weaken direct human involvement in a particular output, but it may simultaneously increase the responsibility of those who decide to design, authorize, or deploy it.

Criminal responsibility should therefore be distributed across the AI lifecycle. Programmers may be liable where they intentionally create unlawful functions or knowingly facilitate criminal use. Manufacturers and corporations may be responsible where organizational policies tolerate serious risks, suppress known defects, or prioritize deployment over safety. Operators may be liable when they misuse systems, ignore warnings, or approve unlawful outputs. Deployers may incur responsibility when they use a system in an inappropriate environment or without adequate oversight. Commanders and superiors may be responsible where they authorize dangerous deployment, fail to monitor operations, or disregard credible evidence of unlawful conduct.

This differentiated structure preserves the principle of personal culpability. It avoids the unjust result of blaming every person associated with an AI system, while preventing technological complexity from becoming a defence. Liability must

depend on an actor's actual role, legal duty, contribution, authority, knowledge, and capacity to prevent harm. The prosecution should be required to establish a sufficiently close relationship between the accused and the prohibited result.

Causation should be understood in a manner capable of addressing distributed technological systems. Harm may arise from several interacting factors, including programming choices, training data, system design, deployment conditions, human instructions, institutional policies, and failures of supervision. The existence of multiple causes should not eliminate responsibility. The law should determine whether each actor materially contributed to the creation or continuation of an unlawful risk and whether the resulting harm fell within the scope of that risk.

The mental element remains indispensable. Intentional use of AI for criminal purposes presents little conceptual difficulty. More challenging cases involve recklessness, wilful blindness, and gross negligence. An exact output may be unpredictable even where the general category of harm is foreseeable. The law should therefore distinguish between unforeseeable accidents and conscious disregard of substantial risk. Actors should not escape liability merely because they could not predict the precise sequence by which a known danger materialized.

Meaningful human control should become a central principle of AI governance, particularly where systems affect life, liberty, physical integrity, or fundamental rights. Human control must be substantive rather than formal. A person should possess adequate knowledge, time, authority, and practical capacity to evaluate and interrupt the system. Placing a nominal human supervisor in a technologically or institutionally powerless position does not satisfy this requirement.

International criminal law already contains doctrines capable of addressing many AI-related harms. Direct and indirect perpetration, co-perpetration, complicity, omission, corporate responsibility, and superior responsibility can be applied to conduct mediated through technology. The principal challenge is not always the absence of law. It is often the difficulty of obtaining evidence, reconstructing decision-making, and identifying how knowledge and control were distributed.

Transparency and traceability are therefore essential. High-risk AI systems should be designed in a manner that permits later review. Developers, manufacturers, and users should preserve records concerning training data, system limitations, testing results, operational parameters, updates, warnings, and human interventions. Without such evidence, criminal accountability may become practically impossible. Deliberate opacity should not be permitted to function as a shield.

The legal framework should also distinguish between different levels of risk. Systems used in routine administrative tasks do not require the same safeguards as systems used in lethal force, policing, detention, border control, or critical infrastructure. High-risk applications should be subject to stronger duties of testing, authorization, supervision, audit, and incident reporting. The severity of potential harm should determine the intensity of preventive obligations.

Corporate responsibility requires particular attention because AI development is often institutional rather than individual. Knowledge may be divided among departments, and harmful outcomes may result from organizational culture rather than from a single decision. Legal systems should be capable of attributing responsibility where corporate structures encourage risk-taking, conceal defects, or prevent effective oversight. Individual managers should not be able to escape liability by fragmenting information across an organization.

The international dimension of AI also demands cooperation. Technological development, data processing, manufacturing, deployment, and harm may occur in different jurisdictions. Effective investigation will require cross-border preservation of digital evidence, mutual legal assistance, technical expertise, and compatible standards. Without cooperation, multinational actors may exploit jurisdictional fragmentation.

The article concludes that the emergence of artificial intelligence does not require abandonment of the human-centered foundations of criminal law. It requires refinement of those foundations. Responsibility should follow authority, control, contribution, knowledge, benefit, and preventive capacity. The law should identify in advance who is responsible for testing, supervising, documenting, and interrupting high-risk systems.

The central normative proposition is that increasing technological autonomy must not result in decreasing human accountability. The delegation of decision-making to machines is itself a human and institutional choice. Those who create, authorize, control, profit from, and fail to supervise AI systems must remain subject to legal scrutiny. Artificial intelligence should never become a substitute offender used to absorb blame that properly belongs to human actors or organizations.

A coherent future framework should therefore combine personal culpability with lifecycle responsibility, meaningful human control, organizational accountability, and mandatory traceability. Such a framework would preserve the legitimacy of criminal punishment while responding to the distinctive characteristics of artificial intelligence. It would also ensure that technological

innovation develops within a legal order that protects human dignity, fundamental rights, and the foundational principle that serious harm must not occur without accountability.

## Ethical Considerations

All procedures performed in this study were under the ethical standards.

## Acknowledgments

Authors thank all who helped us through this study.

## Conflict of Interest

The authors report no conflict of interest.

## Funding/Financial Support

According to the authors, this article has no financial support.

## References

Ambos, K. (2013). *Treatise on International Criminal Law, Volume I: Foundations and General Part*. Oxford University Press. https://academic.oup.com/oxford-law-pro/book/57884

Amoroso, D., & Tamburrini, G. (2020). Autonomous Weapons Systems and Meaningful Human Control: Ethical and Legal Issues. *Current Robotics Reports*, *1*, 187-194. https://doi.org/10.1007/s43154-020-00024-3

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

Chesterman, S. (2020). Artificial Intelligence and the Limits of Legal Personality. *International & Comparative Law Quarterly*, *69*(4), 819-844. https://doi.org/10.1017/S0020589320000366

Crootof, R. (2015). The Killer Robots Are Here: Legal and Policy Implications. *Cardozo Law Review*, *36*(5), 1837-1915. https://www.cardozolawreview.com/the-killer-robots-are-here-legal-and-policy-implications/

Crootof, R. (2016). War Torts: Accountability for Autonomous Weapons. *University of Pennsylvania Law Review*, *164*(6), 1347-1402. https://repository.law.upenn.edu/Documents/Detail/war-torts-accountability-for-autonomous-weapons/155644

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

Hallevy, G. (2013). *When Robots Kill: Artificial Intelligence Under Criminal Law*. Northeastern University Press. https://books.google.com/books?id=5FvO0aeKp_AC

Heyns, C. (2013). *Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions*. https://documents.un.org/doc/undoc/gen/g13/127/76/pdf/g1312776.pdf

King, T. C., Aggarwal, N., Taddeo, M., & Floridi, L. (2020). Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions. *Science and Engineering Ethics*, *26*(1), 89-120. https://doi.org/10.1007/s11948-018-00081-0

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

Schmitt, M. N., & Thurnher, J. S. (2013). "Out of the Loop": Autonomous Weapon Systems and the Law of Armed Conflict. *Harvard National Security Journal*, *4*(2), 231-281. https://harvardnsj.org/2013/05/22/out-of-the-loop-autonomous-weapon-systems-and-the-law-of-armed-conflict/

Sparrow, R. (2007). Killer Robots. *Journal of Applied Philosophy*, *24*(1), 62-77. https://doi.org/10.1111/j.1468-5930.2007.00346.x

Wagner, M. (2014). The Dehumanization of International Humanitarian Law: Legal, Ethical, and Political Implications of Autonomous Weapon Systems. *Vanderbilt Journal of Transnational Law*, *47*(5), 1371-1424. https://scholarship.law.vanderbilt.edu/vjtl/vol47/iss5/6/

Werle, G., & Jessberger, F. (2020). *Principles of International Criminal Law* (4th ed.). Oxford University Press. https://academic.oup.com/oxford-law-pro/book/57106