Challenges of Attributing Criminal Responsibility for Artificial Intelligence in Hybrid Warfare: An International Humanitarian Law Perspective

1. Arash Arbabi: Department of Law, KI.C., Islamic Azad University, Kish, Iran
2. Maryam Afshari*: Department of Law, Dmv.C., Islamic Azad University, Damavand, Iran

***Correspondence: maryam.afshari@iau.ac.ir**

Abstract

The increasing integration of artificial intelligence into hybrid warfare has created significant challenges for the attribution of criminal responsibility under international humanitarian law and international criminal law. Autonomous and semi-autonomous systems can participate in surveillance, target identification, operational planning, and lethal decision-making, while their technical complexity may weaken the traditional connection between human intention and harmful outcomes. This study examines the legal difficulties arising from the use of artificial intelligence in hybrid warfare, particularly the accountability gap created when an autonomous system produces conduct amounting to a serious violation of international humanitarian law without an immediately identifiable human perpetrator possessing the required mental element. The study adopts a descriptive-analytical and comparative legal approach and focuses on the relationship between operational autonomy, mens rea, command responsibility, individual participation, and state responsibility. The analysis demonstrates that artificial intelligence cannot presently be regarded as an independent bearer of criminal responsibility because it lacks human-like intention, knowledge, free will, moral understanding, and the capacity to experience punishment. Accordingly, technological autonomy should not be equated with criminal or moral agency. The study argues that responsibility should instead be reconstructed through a hierarchical human-centered model. At the operational level, military commanders remain responsible for lawful deployment, supervision, and meaningful human control. At the technical level, programmers, developers, and manufacturers may incur responsibility where their intentional or legally culpable contribution satisfies the relevant legal requirements. At the sovereign level, states remain responsible for the deployment and consequences of military artificial intelligence attributable to them. The study further emphasizes the importance of meaningful human control, algorithmic auditability, traceability, weapons review, preservation of operational records, and mechanisms for compensation. It concludes that artificial intelligence should remain legally classified as an instrument of warfare rather than an autonomous criminal subject and that international law should prevent technological complexity from becoming a mechanism for avoiding human accountability.

Keywords: Artificial Intelligence; Hybrid Warfare; Criminal Responsibility; International Humanitarian Law; Autonomous Weapons Systems; Mens Rea; Command Responsibility; Meaningful Human Control; State Responsibility

Received: 17 June 2026
Revised: 15 September 2026
Accepted: 23 September 2026
Initial Publication: 26 September 2026
Final Publication: 01 November 2027

Citation: Arbabi, A. & Afshari, M. (2027). Challenges of Attributing Criminal Responsibility for Artificial Intelligence in Hybrid Warfare: An International Humanitarian Law Perspective. *Legal Studies in Digital Age*, 6(6), 1-20.

1. Introduction

The transformation of warfare in the twenty-first century has substantially altered the legal assumptions upon which the traditional regulation of armed conflict was constructed. Conventional armed conflicts were generally conceptualized around identifiable parties, geographically recognizable battlefields, relatively clear chains of military command, and human actors who selected targets, issued orders, operated weapons, and could subsequently be held personally accountable for unlawful conduct. Hybrid warfare complicates each of these assumptions because it combines conventional military force with cyber operations, information manipulation, economic coercion, proxy activity, psychological operations, autonomous technologies, and other instruments that may be deployed simultaneously or sequentially in pursuit of strategic objectives. Hoffman describes hybrid warfare as a form of conflict in which conventional and unconventional methods become increasingly integrated, making the distinction between traditional military confrontation and non-conventional forms of coercion substantially less stable (Hoffman, 2007). Such an environment does not merely produce operational complexity; it also creates significant difficulties for international humanitarian law because the identification of the responsible actor, the classification of conduct, and the reconstruction of the causal chain connecting human decision-makers to harmful outcomes can become exceptionally difficult. When artificial intelligence is incorporated into this environment, particularly through autonomous and semi-autonomous weapons, machine-learning-based target recognition, predictive systems, cyber capabilities, and algorithmic decision-support mechanisms, the traditional relationship between human will and military action becomes even more complicated.

Artificial intelligence has progressively moved from providing informational assistance to playing an increasingly influential role in surveillance, intelligence assessment, target identification, threat classification, operational planning, and weapons deployment. Scharre explains that autonomous weapons represent an important transformation in military technology because machines can increasingly perform tasks that previously required direct human judgment, including the identification and engagement of targets (Scharre, 2018). This transformation is particularly significant in hybrid warfare, where military operators may be confronted with massive volumes of information, rapidly changing operational environments, cyber deception, irregular forces, civilian infrastructure used for military purposes, and adversaries intentionally attempting to obscure the distinction between civilian and military activity. Artificial intelligence can process these conditions at a speed and scale that human operators cannot replicate, but the very capabilities that make such technologies militarily attractive also create substantial legal problems. Machine-learning systems may produce outputs through complex computational processes that are not fully explainable even by their developers, while adaptive systems may respond to environmental conditions in ways that were not individually predetermined. Chesterman observes that increasingly autonomous artificial intelligence systems challenge conventional legal models precisely because regulatory systems ordinarily assume that meaningful decisions can ultimately be traced to legally accountable human actors (Chesterman, 2021). In military environments, this difficulty acquires exceptional importance because an erroneous or unlawful output may result not merely in economic or administrative harm but in death, destruction, or conduct amounting to a war crime.

International humanitarian law is constructed around normative duties imposed upon human beings and states. Its central rules, including distinction, proportionality, precautions in attack, protection of civilians, and restrictions on means and methods of warfare, presuppose that legally relevant judgments can be made and that those who make or supervise such judgments can be held responsible for violations. Solis emphasizes that the law of armed conflict imposes duties upon combatants and commanders that require the application of legal judgment to concrete operational circumstances (Solis, 2016). A military commander deciding whether an object constitutes a military objective must evaluate factual circumstances, available intelligence, anticipated military advantage, potential civilian harm, and the feasibility of precautions. Artificial intelligence, by contrast, may classify a target through pattern recognition, probability estimation, optimization functions, or other computational mechanisms. Although such systems may technically outperform human beings in some forms of detection, this does not necessarily mean that they possess legal understanding or moral judgment. The distinction between a computational classification and a legally meaningful determination therefore lies at the center of the contemporary debate concerning artificial intelligence in warfare.

The most difficult legal problem arises when autonomous or highly automated systems contribute directly to conduct that would constitute an international crime if performed intentionally by a human actor. International criminal law ordinarily requires the convergence of a prohibited material act and a legally sufficient mental element. Cassese explains that individual

criminal responsibility in international law remains fundamentally connected to personal culpability and to the requirement that prohibited conduct be attributable to an individual whose state of mind satisfies the relevant legal standard (Cassese, 2013). Article 30 of the Rome Statute establishes intent and knowledge as the default mental elements for crimes within the jurisdiction of the International Criminal Court. This structure becomes problematic when an artificial intelligence system independently generates a targeting outcome leading to unlawful civilian casualties. The machine may have caused or executed the relevant physical act, yet it possesses neither consciousness nor a human mental state. A programmer may have written the underlying code without foreseeing the specific attack. A commander may have authorized the system's deployment without knowing precisely how it would behave in the relevant situation. A manufacturer may have developed the technology for a broader range of lawful applications. The state may nevertheless benefit from and control the military operation. In such circumstances, criminal law is confronted with an apparent division between the physical manifestation of the crime and the subjective element traditionally necessary for punishment.

The resulting problem is frequently described as an accountability or responsibility gap. The concept refers to circumstances in which harmful or unlawful conduct is produced by an autonomous system but existing doctrines of individual criminal responsibility encounter difficulties in identifying a human actor who possessed the required intention, knowledge, or level of control. Amoroso notes that autonomous weapons create distinctive challenges for international law because increasing autonomy may complicate the identification of the human decision-maker legally responsible for the consequences of force (Amoroso, 2020). The problem is not necessarily that no human beings are involved in the broader process, but that multiple actors participate in different stages of development, procurement, programming, authorization, deployment, supervision, and operation. Consequently, responsibility may become dispersed throughout a complex technological and institutional chain. The legal difficulty is therefore one of attribution: determining which acts and mental states are sufficiently connected to the eventual unlawful outcome to justify criminal responsibility.

One proposed response is to recognize artificial intelligence itself as a legal or even criminal actor. The idea of artificial legal personality has long been debated in legal theory. Solum examined whether artificial intelligences might conceivably become legal persons and demonstrated that legal personality is not conceptually identical to biological humanity, since legal systems already recognize artificial entities such as corporations for specific purposes (Solum, 1992). Nevertheless, criminal responsibility in the context of international crimes presents substantially more demanding requirements than the mere allocation of legal rights or obligations. International criminal punishment is deeply connected to blameworthiness, moral agency, intention, knowledge, deterrence, and the capacity to understand the normative meaning of conduct. A machine may process information about protected objects or prohibited targets, but computational representation is not equivalent to moral understanding. The attribution of criminal intention to an algorithm therefore risks confusing functional autonomy with moral or juridical agency.

A second response is to preserve the human-centered character of criminal responsibility and reconstruct the chain of attribution by examining those persons who design, authorize, deploy, supervise, or facilitate the use of autonomous systems. Werle and Jessberger emphasize that international criminal responsibility is individual and must be established through recognized modes of liability and appropriate mental elements rather than by generalized assumptions about collective responsibility (Werle & Jessberger, 2020). Within this framework, Article 25 of the Rome Statute becomes particularly relevant to developers, programmers, and other persons who intentionally contribute to criminal conduct, while Article 28 provides a basis for examining the responsibility of military commanders and other superiors who fail to prevent or repress crimes committed within structures under their effective authority and control. These doctrines do not provide automatic solutions to every case involving artificial intelligence, but they offer legal mechanisms through which the human actors surrounding autonomous technologies may be examined.

Hybrid warfare creates additional complications because attribution is already difficult even before artificial intelligence is introduced. Cyber operations can be conducted through proxies, compromised infrastructure, anonymized networks, private companies, and non-state groups. Schmitt's analysis of international law applicable to cyber operations demonstrates that questions of attribution, state control, and legal characterization can become extremely complex when hostile conduct occurs through digital infrastructures rather than conventional armed forces (Schmitt, 2017). The combination of cyber operations and artificial intelligence can multiply these challenges. An algorithmic system may operate across borders, rely on external

data, interact with civilian networks, or be developed by private actors before being integrated into military operations. The legal system must consequently address not merely what the machine did but who exercised legally relevant control over the process through which the machine acted.

The problem also has a philosophical dimension. Criminal responsibility has historically been connected to the ability to exercise free will and to understand the normative significance of one's conduct. Religious and natural-law traditions similarly connect responsibility to conscious choice. In Islamic jurisprudence, the distinction between direct action and causation, as well as the concept of guarantee for harm caused through instruments, provides an alternative conceptual lens for addressing machine-mediated harm. Al-Hilli's treatment of responsibility and causation in classical Islamic jurisprudence places significant emphasis on the relationship between the direct actor, the person who creates the cause, and the resulting harm (Al-Hilli, 1998). Khomeini likewise discusses legal responsibility in ways that preserve the importance of attribution, causation, and human agency when harm occurs through instrumental means (Khomeini, 1990). These principles suggest that the absence of moral agency in the instrument does not necessarily create the absence of responsibility; instead, responsibility may return to the human or institutional actor who caused, controlled, or benefited from the instrument's operation.

Natural-law thought similarly emphasizes the distinctive moral status of human choice. Francis criticizes an excessively technocratic understanding of human affairs in which technological capability displaces moral responsibility and human judgment (Francis, 2015). This concern is particularly important in the military context because lethal decisions implicate fundamental questions of human dignity, moral agency, and responsibility for life and death. The increasing delegation of targeting functions to artificial intelligence therefore raises not only technical and doctrinal legal questions but also a deeper normative issue concerning whether decisions of lethal force can legitimately be detached from human judgment.

Against this background, the present study seeks to analyze the challenges of attributing criminal responsibility for artificial intelligence-related conduct in hybrid warfare from the perspective of international humanitarian law and international criminal law. It examines the relationship between autonomous technological action and the mental element required for criminal responsibility, evaluates the possibility of treating artificial intelligence as an independent criminal actor, and analyzes the legal positions of commanders, programmers, manufacturers, and states within the responsibility chain. It further draws upon Islamic jurisprudential concepts of causation and guarantee and broader natural-law understandings of free will to clarify why technological autonomy does not necessarily entail independent moral or criminal agency. The central argument developed in this article is that the responsibility gap should not be resolved by attributing criminal personality to artificial intelligence but through a human-centered model that reconnects unlawful algorithmic outcomes to the persons and institutions responsible for designing, authorizing, supervising, and deploying military artificial intelligence.

2. Artificial Intelligence and the Transformation of Responsibility in Hybrid Warfare

Hybrid warfare is distinguished from classical warfare not simply by the use of new technologies but by the simultaneous operation of multiple strategic instruments across military and non-military domains. Hoffman explains that hybrid threats combine conventional capabilities, irregular tactics, terrorist methods, and potentially disruptive technologies within a single conflict environment (Hoffman, 2007). This multidimensional structure challenges legal systems because the categories traditionally used to regulate conflict may not correspond neatly to the practical organization of hybrid operations. Hostile actions may move continuously between conventional military strikes, cyber interference, political manipulation, economic pressure, intelligence operations, disinformation, sabotage, and activity conducted through state-supported non-state actors. Some operations may remain deliberately below the conventional threshold of armed conflict, while others may occur simultaneously with recognized hostilities. The resulting grey zone enables actors to exploit uncertainty regarding attribution, classification, and response. When artificial intelligence is integrated into these operations, the same ambiguity that characterizes hybrid warfare becomes embedded within technological systems capable of collecting information, identifying vulnerabilities, predicting behavior, prioritizing targets, and executing military functions.

Artificial intelligence possesses particular value in hybrid warfare because hybrid operations are fundamentally dependent on information. Successful hybrid strategies frequently require the rapid integration of military intelligence, open-source information, cyber data, satellite imagery, communications intelligence, social-media activity, demographic information, and real-time operational data. Machine-learning systems can identify patterns in such data at a speed that exceeds ordinary human

analytical capability. Scharre observes that advances in autonomous systems increasingly permit machines to perform functions of sensing, deciding, and acting with decreasing levels of immediate human intervention (Scharre, 2018). This technical capacity can generate substantial operational advantages. A system may detect missile launchers within imagery, identify patterns associated with hostile cyber activity, prioritize objects for surveillance, or recommend potential targets based on multiple indicators. However, these advantages depend heavily upon the quality of the data, assumptions embedded in system design, contextual validity of classifications, and the reliability of the machine-learning process.

The difference between traditional programmed systems and contemporary machine learning is especially important for responsibility analysis. In conventional rule-based software, programmers prescribe explicit instructions and the system follows predetermined logical structures. If a particular factual condition occurs, a corresponding pre-coded response is generated. The relationship between programming and output is therefore relatively transparent. By contrast, machine-learning systems may derive internal patterns from training data and use these patterns to generate predictions or classifications that were not explicitly programmed in advance. Chesterman explains that increasingly sophisticated artificial intelligence systems create regulatory difficulties because their outputs can become less directly traceable to discrete human decisions (Chesterman, 2021). Deep-learning systems may include numerous computational layers whose internal operations are difficult to interpret in ordinary causal language. The system may correctly identify thousands of military objects while misclassifying a civilian object in circumstances that neither the operator nor the original programmer specifically anticipated.

This phenomenon is frequently described as the black-box problem. The phrase does not imply that absolutely nothing can be known about the technical functioning of the system; rather, it refers to circumstances in which the precise basis for an individual output is difficult to reconstruct or explain. In criminal responsibility analysis, this creates a problem because criminal law ordinarily seeks to establish a causal and mental connection between a human actor and a prohibited consequence. If a programmer could not reasonably predict a particular machine-learning output, it may be difficult to characterize that person as intending or knowingly causing the result. If a commander relied upon the system because it had performed reliably in previous operations, the unlawful output may similarly be difficult to connect to the commander's subjective state of mind. Yet if such unpredictability were treated as automatically eliminating responsibility, the increasing deployment of complex systems could progressively weaken accountability precisely because the technology was designed to operate with greater independence.

The difficulties become particularly acute in relation to the principle of distinction. International humanitarian law requires parties to distinguish civilians and civilian objects from combatants and military objectives. Solis identifies distinction as one of the foundational principles of the law of armed conflict because the legal regulation of targeting depends upon the ability to identify those objects and persons against whom force may lawfully be directed (Solis, 2016). Human targeting decisions often require contextual interpretation rather than simple visual identification. A person carrying a weapon may be a combatant, a police officer, a civilian temporarily possessing a firearm, or an individual who has surrendered. A building may normally serve civilian functions but temporarily become a military objective. A vehicle may resemble a military platform while transporting civilians. These assessments cannot always be reduced to stable visual characteristics or statistical correlations.

Artificial intelligence may classify objects through features extracted from data rather than through a legal appreciation of their status. This creates the possibility that training-data bias, incomplete datasets, erroneous labeling, or environmental differences will produce systematic targeting errors. Scharre notes that autonomous weapons operate according to technical parameters that may fail in complex and unpredictable environments, making reliability a central concern in discussions of military autonomy (Scharre, 2018). A system trained predominantly on one geographical environment may perform differently in another. A model trained on conventional military formations may be unreliable in an urban hybrid conflict where combatants intentionally resemble civilians. An algorithm designed to associate certain behavioral patterns with hostile activity may reproduce assumptions embedded within the data rather than identify legally relevant status. In such circumstances, the problem is not that the machine consciously disregards the law but that the technical process through which it classifies the world does not necessarily correspond to the normative categories employed by humanitarian law.

The principle of proportionality creates an even more difficult problem because proportionality requires evaluative judgment. A lawful attack against a military objective may nevertheless be prohibited if the expected incidental civilian harm would be excessive in relation to the concrete and direct military advantage anticipated. Solis explains that proportionality

demands judgment rather than mechanical arithmetic because military advantage and civilian harm are not reducible to a common mathematical unit (Solis, 2016). Artificial intelligence might estimate probabilities of civilian presence, predict blast effects, calculate distances, or process intelligence concerning military importance, but the ultimate legal assessment requires a normative comparison. The fact that a system can optimize variables does not necessarily mean that it can determine what degree of civilian harm is legally excessive. If states attempt to encode such decisions through predetermined thresholds, the normative judgment has merely been displaced from the battlefield to the earlier design process.

Precautions in attack create another point at which human responsibility and machine performance intersect. Commanders and operators must take feasible precautions to verify targets, select appropriate means and methods, and minimize civilian harm. When an artificial intelligence system provides targeting recommendations, the legal significance of human review depends on whether the human decision-maker has sufficient information, time, authority, and technical understanding to challenge the algorithmic output. A formal confirmation procedure may provide only an appearance of human control if the operator has no meaningful opportunity to evaluate the recommendation. Amoroso emphasizes that debates about autonomous weapons increasingly focus on the degree of human control necessary to maintain compliance with international legal requirements (Amoroso, 2020). Meaningful human control therefore cannot be understood as the mere physical presence of a person in the decision chain. It requires substantive capacity to understand the operational context, evaluate the system's recommendation, anticipate relevant consequences, and intervene where necessary.

Hybrid warfare further intensifies these concerns because adversaries may intentionally manipulate the technological environment. Cyber operations can corrupt datasets, interfere with sensors, alter communications, compromise software, or create misleading inputs. Schmitt's treatment of cyber operations demonstrates that digital environments permit hostile actors to produce effects indirectly through information systems and network infrastructures (Schmitt, 2017). Artificial intelligence may consequently be exposed to adversarial manipulation designed to cause misclassification. An opposing force might manipulate visible characteristics, electronic signatures, communications patterns, or network behavior to deceive a system. The resulting targeting decision could appear technically justified according to the data available to the algorithm while being operationally false. These conditions complicate both compliance and responsibility because the immediate source of error may be hostile manipulation, inadequate system design, insufficient testing, defective intelligence, negligent deployment, or some combination of these factors.

The question of predictability consequently becomes central to the attribution of responsibility. Traditional criminal law frequently relies upon foreseeability when distinguishing accidents from culpable conduct. Yet artificial intelligence introduces several levels of foreseeability. A programmer may foresee that the system has a measurable error rate but not foresee the specific circumstances of a future attack. A commander may know that the system can fail under certain environmental conditions but believe that those conditions are absent. A manufacturer may understand general limitations but lack knowledge about the particular theatre of deployment. The state may nevertheless authorize the use of the system as part of its military doctrine. These distributed forms of knowledge make it difficult to identify a single actor as the sole responsible party. Werle and Jessberger emphasize that international criminal law requires careful differentiation between individual modes of participation and cannot replace personal attribution with generalized assumptions of collective guilt (Werle & Jessberger, 2020). The complexity of AI-enabled warfare therefore requires responsibility analysis to examine the distinct contributions and mental states of multiple actors.

The hybrid environment also affects the distinction between state and private activity. Artificial intelligence systems may be developed by commercial technology companies, defense contractors, university laboratories, or collaborative public-private projects. The same underlying software may have civilian and military applications. Chesterman notes that artificial intelligence regulation increasingly involves complicated relationships between private technological development and public authority (Chesterman, 2021). In hybrid warfare, private entities may provide cyber capabilities, intelligence analytics, satellite information, surveillance platforms, data-processing tools, or software incorporated into targeting systems. Responsibility becomes difficult when the technology was not designed exclusively for unlawful operations but is subsequently used in a manner producing violations. The fact that the programmer is geographically and institutionally distant from the battlefield does not eliminate the possibility of legal responsibility, but it makes proof of contribution, knowledge, and intent more complicated.

The technological transformation of warfare therefore does not simply introduce a new type of weapon. It reorganizes the structure through which decisions are made. In traditional warfare, the soldier or commander often constituted the visible point at which intelligence, judgment, and physical action converged. In AI-enabled hybrid warfare, these functions can be distributed. Developers construct technical architectures, data specialists train systems, private companies maintain software, intelligence agencies supply information, commanders establish operational parameters, operators supervise the system, and algorithms process inputs and generate outputs. The final act may consequently be the product of a network rather than a single decision. This distribution of agency explains why responsibility analysis must move beyond the simplistic question of whether the machine itself is guilty. The more important legal question is how international law can identify and reconnect the human decisions that made the machine's harmful action possible.

The legal classification of artificial intelligence as an instrument remains important in this respect. Although highly autonomous systems may behave in ways that appear independent, their operation remains embedded in human-created purposes, architectures, deployment decisions, data environments, and institutional structures. Solum's analysis of artificial legal personality demonstrates that functional sophistication alone does not automatically determine legal personality (Solum, 1992). The fact that an algorithm selects among alternatives does not necessarily mean that it possesses legal will. Similarly, the fact that a system learns from data does not mean that it acquires moral responsibility. This distinction becomes essential in warfare because otherwise the attribution of responsibility may shift from accountable human institutions to a technological object incapable of punishment in any meaningful criminal-law sense.

Hybrid warfare thus magnifies the responsibility problem in two interconnected ways. First, the strategic environment already obscures the identity, legal status, and control relationships of relevant actors. Second, artificial intelligence separates information processing and operational action from immediate human judgment. The combination creates a multilayered attribution problem in which legal responsibility can be obscured by both strategic ambiguity and technological opacity. The appropriate response cannot consist merely of demanding more accurate algorithms. Even a highly accurate system can produce unlawful consequences if deployed under inappropriate operational conditions, programmed according to defective assumptions, or used without adequate human supervision. International humanitarian law therefore requires a responsibility framework capable of addressing not only the technical error itself but also the institutional choices surrounding development and deployment.

For this reason, the central legal problem is not whether artificial intelligence can replace humans in certain operational functions, but whether human legal responsibility can remain effective when machines increasingly mediate those functions. Artificial intelligence may accelerate targeting, expand surveillance, improve predictive capacity, and reduce certain forms of human error. Nevertheless, technological efficiency cannot eliminate the requirement that the use of military force remain subject to legal judgment. Meaningful human control should consequently be understood not merely as a technical configuration but as a responsibility-preserving structure. It maintains a legally recognizable human relationship to the decision to use lethal force and prevents technological autonomy from becoming a mechanism through which the attribution of international crimes is dissolved.

3. The Mens Rea and Attribution Gap in AI-Enabled Warfare

The central conceptual obstacle to attributing criminal responsibility directly to artificial intelligence lies in the mental element of crime. International criminal law does not ordinarily impose punishment merely because a prohibited result has occurred. Personal culpability requires a legally sufficient relationship between the accused person's state of mind and the material elements of the offense. Cassese explains that international criminal law maintains the foundational principle that liability should be individualized and connected to the accused person's culpable conduct and mental state (Cassese, 2013). Article 30 of the Rome Statute reflects this principle by establishing intent and knowledge as the general mental elements applicable unless otherwise provided. The introduction of artificial intelligence into lethal decision-making creates a profound doctrinal difficulty because autonomous systems can produce outcomes resembling intentional action without possessing a human mind capable of intention or knowledge in the juridical sense.

The problem begins with the meaning of intent. Intent in criminal law is not simply the production of an outcome through a process directed toward a goal. It is ordinarily connected to the conscious orientation of a person toward conduct or

consequences. Werle and Jessberger emphasize that the mental elements of international crimes concern the subjective relationship of the individual perpetrator to the relevant material elements (Werle & Jessberger, 2020). An algorithm can be designed to optimize a particular objective and may select actions that maximize the probability of achieving that objective. Yet this type of computational goal-seeking does not necessarily constitute intention. The algorithm does not desire the result, morally approve it, consciously accept its legal consequences, or understand that it is performing a prohibited act. It processes inputs according to an architecture and generates outputs based on mathematical operations. Describing this process as intention risks translating technical functionality into a psychological category that was developed for human agents.

Knowledge presents a similar problem. Artificial intelligence systems can contain and process enormous amounts of information. They may recognize that an object has characteristics statistically associated with a hospital, identify that civilians are likely to be present in a location, or calculate that a strike will produce a particular level of collateral damage. Nevertheless, legal knowledge is not identical to data storage or pattern recognition. A system may correctly classify an object as a hospital without understanding the normative significance of the hospital's protected status. It may estimate the probability of civilian casualties without appreciating that excessive incidental harm would render an attack unlawful. Cassese's discussion of criminal culpability illustrates that knowledge in criminal law concerns the accused's awareness of relevant factual circumstances, not merely the mechanical possession of information (Cassese, 2013). The attribution of knowledge to artificial intelligence therefore requires caution because the word may be used metaphorically in computer science while carrying a far more specific meaning in criminal law.

This distinction reveals the risk of anthropomorphism. Solum's examination of artificial intelligence and legal personhood demonstrates that debates over machine legal status often depend upon assumptions concerning the characteristics that justify legal personality (Solum, 1992). In ordinary language, people frequently say that a computer "knows," "decides," "learns," or "chooses." These expressions are useful descriptions of system behavior, but they should not automatically be treated as evidence that machines possess human mental states. In criminal law, such metaphorical language can create doctrinal confusion. If a targeting system "decides" to classify an individual as hostile, the legal question is not whether the software generated an output corresponding to an attack decision but whether it possessed the type of conscious volition to which criminal blame can meaningfully attach.

The problem becomes more complicated in advanced machine-learning systems because their outputs may not be directly determined by specific human instructions. This operational autonomy creates the appearance of independent agency. Scharre explains that autonomous weapons can select and engage targets without further human intervention once activated, raising difficult questions concerning control and accountability (Scharre, 2018). Yet operational independence does not necessarily establish moral independence. The system's capacity to respond dynamically to environmental data indicates technical sophistication, but it does not demonstrate consciousness, moral understanding, or free will. Consequently, the concept of machine autonomy must be separated from the concept of legal agency. A system may be operationally autonomous while remaining legally an instrument embedded within human decision-making structures.

Theories of artificial moral agency attempt to challenge this conclusion by arguing that responsibility might attach to entities capable of autonomous behavior and significant causal influence. Such theories emphasize functional capacities rather than subjective consciousness. If a system can perceive aspects of its environment, evaluate alternatives, select actions, and adapt to experience, it might appear functionally similar to an agent. Nevertheless, international criminal law requires more than causal agency. Punishment is directed toward persons considered blameworthy. Amoroso's treatment of autonomous weapons highlights the continuing centrality of human responsibility even where weapons systems possess advanced autonomous functions (Amoroso, 2020). A machine does not experience punishment, moral condemnation, stigma, suffering, remorse, or deterrence in the human sense. Deactivating a system may prevent future harm, but this is more accurately understood as a technical safety measure than as criminal punishment.

The purposes of punishment therefore provide an additional reason for rejecting direct criminal responsibility of artificial intelligence. Retribution assumes that punishment is justified because the offender deserves condemnation. Deterrence assumes that the offender or others capable of similar choices may modify future behavior to avoid punishment. Rehabilitation presupposes that the offender can change through moral, psychological, or social intervention. None of these theories applies straightforwardly to an algorithm. A model can be retrained, modified, restricted, or removed from service, but these measures

operate upon technology rather than moral personality. Chesterman's broader analysis of AI regulation suggests that legal responses to artificial intelligence must distinguish between regulating technological systems and attributing responsibility to persons and institutions surrounding those systems (Chesterman, 2021). Conflating these two functions could weaken accountability rather than strengthen it.

The concept of free will reinforces this distinction. Although international criminal law is not formally dependent upon any particular theological doctrine, the historical and philosophical foundations of criminal responsibility have consistently associated culpability with meaningful choice. Natural-law approaches similarly regard human moral responsibility as arising from the ability to distinguish right from wrong and to choose among alternatives. Francis criticizes technological paradigms that subordinate moral judgment to technical efficiency, warning against circumstances in which human responsibility is displaced by systems treated as autonomous sources of authority (Francis, 2015). In warfare, delegating lethal decisions to algorithms without preserving genuine human judgment creates precisely this concern. If machines cannot engage in moral choice, attributing criminal guilt to them would obscure rather than resolve the normative problem created by technological delegation.

Islamic jurisprudential principles provide a comparable conceptual distinction between the responsible human actor and the instrument through which harm occurs. Al-Hilli's discussions of causation and liability distinguish between direct action and responsibility arising through the creation of causes that produce injury (Al-Hilli, 1998). This framework is significant for artificial intelligence because it does not require the physical instrument of harm to possess moral agency. Instead, legal analysis may identify the person whose conduct generated, directed, or controlled the relevant cause. Khomeini's treatment of liability similarly reflects the importance of attribution and human responsibility where damage occurs through means or instruments (Khomeini, 1990). From this perspective, the absence of criminal agency in the machine does not create a normative vacuum. It redirects legal attention toward the human actor or institution that generated the risk, deployed the system, or exercised control over its operation.

The more difficult situation arises when no individual human actor appears to possess the mental element required for the resulting crime. Suppose that a commander deploys an autonomous targeting system that has successfully passed testing. The manufacturer reasonably believes the system complies with technical specifications. The programmers did not design it to attack civilians. During an operation, however, the system encounters an unusual combination of environmental conditions and misclassifies a protected civilian object as a legitimate military target. The attack produces grave civilian casualties. The material consequences may correspond to conduct prohibited by international humanitarian law, but the prosecution of any particular individual for an intentional war crime may be difficult. This is the core of the responsibility gap.

The possibility of dolus eventualis or recklessness is sometimes invoked to address such situations. If a commander knows that a system has a substantial probability of misidentifying civilians but deploys it nonetheless, the commander's acceptance of risk may potentially become relevant to criminal responsibility. Werle and Jessberger note that the precise scope of mental elements in international criminal law requires careful attention to the statutory language and applicable jurisprudence (Werle & Jessberger, 2020). Article 30's formulation does not simply equate every form of risk-taking with intention. Consequently, national criminal concepts of recklessness or dolus eventualis cannot automatically be imported into every Rome Statute offense. Nevertheless, knowledge of serious technological limitations may be highly relevant when assessing whether a human actor possessed the mental state required under a particular mode of liability.

The problem of foreseeability is therefore central. If the harmful output was genuinely unforeseeable, criminal liability based on intention or knowledge becomes difficult. Yet legal systems must distinguish genuine unforeseeability from avoidable ignorance. A commander cannot reasonably deploy an opaque lethal system and then argue that its behavior was unknowable if adequate testing, supervision, or technical expertise would have revealed relevant risks. Scharre emphasizes that autonomous systems create operational risks precisely because complexity and unpredictability can affect reliability in real-world environments (Scharre, 2018). The decision to deploy a system whose limitations are insufficiently understood may itself become legally significant. The relevant question is therefore not only whether the specific unlawful attack was predicted but whether the responsible human actor knowingly accepted an inadequately controlled risk.

The same reasoning applies to programmers and manufacturers. A developer may not know which individual target will eventually be attacked, but knowledge that a system systematically misclassifies civilian objects could establish a more serious

relationship to subsequent harm. Conversely, a programmer working on a general-purpose component with no knowledge of its eventual military application may be too remote from the crime for criminal responsibility. Amoroso recognizes that autonomous weapons create complex responsibility chains in which the legal positions of designers, programmers, operators, commanders, and states must be differentiated (Amoroso, 2020). International criminal law therefore requires a granular analysis of contribution, knowledge, purpose, and causation rather than a generalized attribution of blame to everyone involved in technological development.

The theory of electronic legal personality presents an apparently simple but ultimately problematic solution. If artificial intelligence could be treated as an independent legal person, the material and mental elements might theoretically be attributed directly to the system. However, such a solution would alter the foundations of criminal law without resolving the practical objectives of accountability. Solum's analysis demonstrates that legal personhood is a normative construction that can be assigned for particular purposes, but this does not establish that every form of artificial intelligence should receive the same responsibilities as human persons (Solum, 1992). Corporate personality, for example, serves organizational and economic functions while operating through human decision-makers and legal representatives. An autonomous weapon lacks assets, social interests, physical liberty, and moral experience comparable to those upon which criminal sanctions typically operate.

More importantly, granting criminal personality to artificial intelligence could generate perverse incentives. Governments, commanders, or corporations might characterize unlawful outcomes as the independent actions of the machine, thereby weakening attention to human decision-making. The machine could become a convenient endpoint for legal attribution. Instead of asking why the system was designed in a particular way, why it was deployed in a certain environment, why warning indicators were ignored, or why human intervention mechanisms were absent, investigators might focus on the machine as the nominal perpetrator. This would contradict the principle that technological complexity should not become a shield against human accountability.

The responsibility gap is therefore better understood as an attribution problem than as evidence that criminal law requires a new category of machine offender. The harmful act may be physically executed by an autonomous system, but the system exists within a chain of human decisions. Someone defines the operational objective, acquires the technology, determines deployment conditions, supplies data, authorizes use, establishes supervision procedures, and determines whether humans can interrupt the system. The challenge is to identify which of these decisions satisfy the legal requirements for responsibility. This approach preserves the fundamental connection between criminal liability and human culpability while recognizing that the relevant human conduct may occur before the final attack rather than at the moment the weapon acts.

From this perspective, meaningful human control becomes closely connected to mens rea and attribution. If a human commander retains sufficient understanding and authority to review, modify, or cancel a system's proposed action, the legal relationship between human judgment and the ultimate attack remains stronger. If the system is deployed in a human-out-of-the-loop configuration, attribution becomes more difficult because immediate human decision-making is absent. However, the decision to create that configuration is itself a human act. Amoroso's discussion of human control over autonomous weapons illustrates that responsibility must therefore be examined across the entire operational lifecycle rather than only at the instant of target engagement (Amoroso, 2020). The law should ask who decided to remove human intervention, under what conditions, with what knowledge of system limitations, and whether reasonable safeguards were available.

The proper legal response to the mens rea problem is consequently not to manufacture an artificial form of intent for machines. Criminal intention should retain its human meaning. Where an autonomous system produces an unlawful result, responsibility should be reconstructed by examining human intention, knowledge, supervision, contribution, and institutional control. Where the evidence does not establish the mental element required for individual criminal punishment, other forms of responsibility may remain applicable, including state responsibility, civil liability, compensation, regulatory sanctions, and preventive obligations. The distinction between criminal guilt and broader legal responsibility is essential. Not every harmful algorithmic outcome will constitute an international crime attributable to an individual, but neither should the absence of individual criminal intent render the harm legally invisible.

The responsibility gap is thus neither entirely fictional nor legally insurmountable. It is real insofar as technological autonomy may make traditional proof of subjective culpability more difficult. Yet it becomes misleading if interpreted to mean that the machine itself must become the criminal offender. The more coherent approach is to preserve the human-centered

foundation of international criminal law while adapting doctrines of attribution and supervision to the realities of distributed technological decision-making. The next stage of analysis therefore concerns the existing legal mechanisms through which responsibility may be assigned to commanders, programmers, manufacturers, and states when artificial intelligence participates in unlawful conduct during hybrid warfare.

4. Reconstructing the Chain of Criminal Responsibility under International Humanitarian and Criminal Law

Once direct criminal responsibility of artificial intelligence is rejected, the legal analysis must return to the human and institutional actors surrounding the technology. The central question is no longer whether the algorithm itself committed a crime but which persons exercised legally relevant authority, control, knowledge, or contribution in relation to the system and its unlawful effects. International criminal law already contains several doctrines capable of addressing distributed responsibility, although their application to artificial intelligence requires careful interpretation. The most important are individual participation under Article 25 of the Rome Statute, command responsibility under Article 28, and the general mental-element requirements associated with Article 30. These mechanisms must also be distinguished from international state responsibility and from compensatory or regulatory forms of liability that do not require individual criminal guilt.

Command responsibility is particularly important because military artificial intelligence is ordinarily deployed within organized chains of authority. Article 28 addresses the responsibility of military commanders and other superiors who fail to exercise appropriate control over forces under their effective command and control. Werle and Jessberger explain that command responsibility is not simply vicarious liability for the acts of subordinates; it is connected to the superior's own failure to perform legally required duties of prevention, repression, or reporting (Werle & Jessberger, 2020). In the context of autonomous systems, a conceptual difficulty immediately appears because an artificial intelligence system is not a human subordinate. Nevertheless, the relevant responsibility need not depend upon treating the machine as a legal subordinate. The more persuasive approach is to focus on the commander's duty to supervise the military operation, including the weapons, processes, personnel, and technological systems through which force is applied.

A commander who authorizes the use of autonomous targeting technology retains responsibility for ensuring that the operation complies with international humanitarian law. Solis emphasizes that commanders remain responsible for the lawful conduct of military operations and for compliance with targeting rules (Solis, 2016). The use of artificial intelligence does not eliminate this duty. If anything, highly complex technologies may increase the importance of adequate supervision because commanders must understand the capabilities and limitations of the means they choose to employ. A commander who deploys a system known to perform unreliably in densely populated areas cannot necessarily avoid responsibility by claiming that the machine independently selected the unlawful target. The legally relevant conduct may lie in the earlier decision to employ the system under circumstances in which the risk to civilians was evident.

The "knew or should have known" standard associated with command responsibility is therefore highly relevant. Artificial intelligence may create informational asymmetries between technical specialists and commanders, but military authority cannot reasonably be exercised through complete technological ignorance. Scharre demonstrates that autonomous weapons pose risks associated with brittleness, unpredictability, interaction effects, and failures outside expected operating conditions (Scharre, 2018). A commander does not need to become a software engineer, but the duty of command requires sufficient understanding of operational limitations to make lawful deployment decisions. If the system has known difficulty distinguishing combatants from civilians under specific conditions, the commander should be informed of that limitation. If uncertainty cannot be adequately managed, the commander may have a duty to restrict or suspend deployment.

Meaningful human control provides a useful framework for evaluating these duties. The concept seeks to preserve a substantive human relationship to decisions concerning lethal force. Amoroso explains that debates on autonomous weapons frequently revolve around the level and quality of human control required to ensure legal compliance (Amoroso, 2020). Control is meaningful only when the human actor possesses adequate information, sufficient time for evaluation, authority to intervene, and a genuine capacity to alter or cancel the operation. A nominal human confirmation process does not satisfy this standard if the operator is effectively compelled to accept machine-generated recommendations without understanding them. Similarly, a human-on-the-loop system may offer little genuine oversight if the speed of operations prevents intervention.

From the standpoint of command responsibility, meaningful human control should therefore be connected to the commander's obligation to structure operations lawfully. The commander determines or influences rules of engagement, system configuration, geographical limitations, mission parameters, review procedures, intervention thresholds, and reporting mechanisms. If the operational design removes all realistic opportunities for human supervision, the commander's responsibility may arise not because the machine is a subordinate but because the commander knowingly created an inadequately controlled lethal process. The attribution of responsibility thus shifts from the moment of algorithmic output to the prior human decision concerning the architecture of control.

Article 25 of the Rome Statute becomes especially relevant to programmers, developers, manufacturers, and other persons who contribute to the commission of crimes. International criminal law recognizes various forms of participation, including ordering, soliciting, inducing, aiding, abetting, or otherwise assisting criminal conduct, subject to the required mental elements. Cassese emphasizes that secondary participation requires careful attention to both the contribution made and the participant's state of mind (Cassese, 2013). In the context of artificial intelligence, a programmer who intentionally designs a system to identify and attack protected civilian objects would present a relatively straightforward case of criminal contribution. The program itself would constitute an instrument through which the human participant facilitates unlawful conduct.

The more common and difficult scenario involves dual-use or generally lawful technology. Machine-learning systems may be developed for surveillance, object recognition, logistics, cybersecurity, navigation, or decision support without any original intention that they be used to commit international crimes. Chesterman notes that artificial intelligence frequently develops across overlapping civilian, commercial, and governmental contexts, complicating regulatory efforts (Chesterman, 2021). A programmer who writes a generic image-recognition module may be several institutional stages removed from a subsequent military strike. Criminal responsibility should not be inferred merely from technical contribution. The prosecution would need to demonstrate a legally sufficient connection between the individual's assistance and the crime, as well as the mental element required under the relevant mode of liability.

Knowledge of defects may nevertheless become legally significant. Suppose that engineers discover that a targeting model systematically misidentifies ambulances or hospitals under particular conditions but conceal the defect so that military deployment continues. Their conduct differs substantially from that of developers who had no reason to know of the problem. Similarly, if a manufacturer intentionally alters safety testing, suppresses performance failures, or provides false information concerning the reliability of an autonomous system, those facts may strengthen the basis for legal responsibility depending upon the subsequent crime and required mental element. Scharre's analysis of autonomous weapons illustrates why testing and reliability are central to responsible deployment (Scharre, 2018). Technical actors who knowingly undermine these safeguards may therefore occupy a more direct position within the causal chain.

However, criminal law must distinguish defective design from intentional criminal facilitation. A poorly designed algorithm can cause grave harm without the programmer possessing criminal intent. This distinction is important because international criminal law should not be transformed into a general product-liability system. Werle and Jessberger stress the principle of personal culpability underlying international criminal responsibility (Werle & Jessberger, 2020). Where negligence, design failure, or inadequate quality control does not satisfy the mental element of a relevant international crime, other legal mechanisms may be more appropriate. Civil liability, regulatory sanctions, professional responsibility, contractual obligations, compensation schemes, and state responsibility can address harms that fall below the threshold of individual criminal culpability.

This differentiation between criminal and compensatory responsibility is especially important for preserving doctrinal coherence. The responsibility gap often appears most severe when victims suffer serious harm but no individual possesses the intention required for criminal conviction. Yet the absence of criminal liability does not imply the absence of legal responsibility. Chesterman's work on artificial intelligence regulation illustrates the importance of allocating institutional responsibility even when traditional criminal categories do not apply (Chesterman, 2021). Military artificial intelligence could be subjected to rigorous testing obligations, documentation requirements, audit trails, compulsory risk assessments, and mechanisms ensuring that victims can obtain compensation. These mechanisms serve different functions from criminal punishment and should not be confused with it.

State responsibility therefore forms the third major layer of the responsibility structure. International criminal law focuses upon individuals, whereas public international law addresses the internationally wrongful acts of states. Hybrid warfare frequently involves private companies, proxy organizations, and technologically mediated operations whose formal relationship to the state may be intentionally obscured. Schmitt's discussion of attribution in cyber operations demonstrates that international law must distinguish between private conduct and conduct legally attributable to states on the basis of authority, instruction, direction, or control (Schmitt, 2017). This problem becomes especially significant where states outsource military artificial intelligence to private contractors. The state should not be able to evade responsibility merely because a technologically complex system was developed or operated by a commercial entity.

State responsibility also differs from individual criminal responsibility because it does not require proof that the state itself possessed a human mental state in the same sense demanded for criminal conviction. If conduct attributable to the state breaches an international obligation, the consequences of state responsibility may include cessation, assurances of non-repetition, and reparation. This makes state responsibility particularly important in situations where an autonomous system causes unlawful harm but the individual mental elements required for criminal prosecution cannot be established. The state remains responsible for the lawful organization of its military activities and for the conduct of organs or entities whose behavior is legally attributable to it.

The relationship between state responsibility and command responsibility is therefore complementary rather than mutually exclusive. A commander may be individually criminally responsible where the statutory requirements are satisfied, while the state may separately bear international responsibility for the same underlying events. Similarly, a programmer or corporate officer may incur criminal liability if they intentionally facilitate the crime, while the state remains internationally responsible for employing the system. This layered approach avoids the false assumption that responsibility must be allocated to only one actor. Hybrid operations are often produced through networks of decision-making, and legal responsibility may correspondingly arise at several levels.

The black-box problem nevertheless complicates proof across all of these categories. Criminal investigations depend upon evidence demonstrating what the system did, why it generated the relevant output, what information was available to human participants, and what they knew at particular times. If technical systems do not preserve logs, training histories, confidence levels, sensor inputs, model versions, human interventions, or operational configuration, reconstruction of events may become extremely difficult. Chesterman's analysis of algorithmic governance underscores the importance of transparency and accountability mechanisms in systems whose internal operations are difficult to explain (Chesterman, 2021). In military applications, transparency cannot necessarily mean public disclosure of classified code, but it must at minimum mean that competent investigators and reviewing authorities can reconstruct the decision process after a harmful event.

Algorithmic transparency should therefore be understood as a responsibility infrastructure. Its purpose is not merely to make technology intellectually understandable but to preserve evidence. If the system identified a hospital as a military target, investigators should be able to determine what inputs produced that classification, which model version was operating, what confidence level was generated, whether human operators received warnings, whether alternative interpretations were available, and whether commanders knew of previous similar errors. Without such information, the technological architecture itself can unintentionally create impunity.

Weapons review processes are also important. The deployment of novel military technologies should be preceded by legal assessment concerning whether their normal or expected use can comply with international humanitarian law. Solis's broader account of the law of armed conflict demonstrates that weapons regulation is inseparable from the obligation to conduct military operations through lawful means and methods (Solis, 2016). Artificial intelligence requires weapons review to extend beyond the physical characteristics of a weapon to the decision-making functions performed by software. Review should examine training data, reliability, explainability, predictable failure modes, cybersecurity vulnerabilities, human-machine interaction, and the operational conditions for which the system is approved.

The comparative jurisprudential perspective strengthens the argument for layered attribution. Islamic jurisprudence distinguishes between the direct performer of an act and the person who causes the conditions through which harm occurs. Al-Hilli's treatment of causation provides a conceptual basis for transferring responsibility from an incapable instrument to the legally relevant human cause (Al-Hilli, 1998). Khomeini similarly preserves the significance of causation and guarantee in the

allocation of responsibility for damage (Khomeini, 1990). Applied to artificial intelligence, this means that the machine's lack of culpability does not require the legal system to accept an accountability vacuum. Attention shifts instead to those who introduced the causal risk, possessed authority over the instrument, or benefited from its use.

This perspective also supports a distinction between punishment and guarantee. Criminal punishment should remain tied to personal culpability, whereas compensation can operate on broader principles of causation and responsibility for risk. A state that chooses to obtain the strategic advantages of military artificial intelligence also assumes obligations to address the harms produced by that technology. This principle is normatively attractive because it prevents victims from bearing the entire cost of technological uncertainty. It also creates incentives for states to invest in safer systems, stronger oversight, and more effective human control.

Natural-law concerns similarly reinforce the necessity of maintaining human responsibility. Francis's critique of the technocratic paradigm emphasizes the danger of allowing technological systems to dominate moral judgment rather than remain subordinate to human responsibility (Francis, 2015). In military artificial intelligence, the relevant issue is not simply whether the machine is accurate but whether human beings remain answerable for decisions affecting life and death. A system that performs efficiently but makes responsibility practically impossible to reconstruct creates a legal and ethical deficit even if its aggregate error rate is comparatively low.

A coherent responsibility framework should therefore operate hierarchically. At the operational level, commanders and operators must retain sufficient control and understanding to ensure lawful use. At the technical level, developers and manufacturers must be examined according to their contribution, knowledge, intent, and compliance with safety obligations. At the sovereign level, states remain responsible for the lawful deployment of military technology and for conduct attributable to them under international law. These levels should not be treated as rigid alternatives. The same event may produce responsibility at more than one level where different legal requirements are met.

Such a hierarchical approach also clarifies the proper place of meaningful human control. Human control is not merely an ethical aspiration or engineering preference. It functions as the connecting mechanism that prevents responsibility from dissolving between technology and institutions. Where commanders understand the system's limitations, establish appropriate parameters, retain intervention capacity, ensure documentation, and respond to warning indicators, attribution remains possible. Where those safeguards are intentionally or negligently removed, the decision to eliminate meaningful human control becomes itself legally significant.

The ultimate objective should be to prevent the technological chain from becoming an accountability shield. Artificial intelligence can distribute operational functions across many actors, but the law can respond by tracing authority, contribution, risk creation, supervision, and institutional control. This requires evidentiary mechanisms capable of reconstructing algorithmic decisions and legal standards that identify the responsibilities associated with deployment. It does not require attributing intention or guilt to the machine. The more durable approach is to preserve the distinction between artificial intelligence as a technologically sophisticated instrument and humans as the bearers of criminal and moral responsibility.

5. Comparative Foundations and the Development of a Human-Centered Responsibility Model

The challenge of artificial intelligence in hybrid warfare cannot be fully understood through a narrow reading of positive international criminal law because the controversy ultimately concerns the foundations of responsibility itself. The apparent accountability gap arises from a conflict between the architecture of contemporary technology and legal concepts developed around human agency. International criminal law presumes that meaningful conduct can be attributed to individuals who possess intention, knowledge, control, or another legally sufficient mental relationship to the offense. Artificial intelligence complicates this model because a machine may function as the immediate decision-producing mechanism while lacking the human attributes that traditionally justify punishment. Comparative examination of Islamic jurisprudence, natural-law thought, and contemporary international law therefore provides a useful means of clarifying which elements of responsibility should remain indispensable even when the means through which harm occurs become technologically complex.

Islamic jurisprudence offers an important conceptual distinction between moral duty and financial or compensatory guarantee. The attribution of criminal blame is closely associated with the characteristics of the accountable human subject, whereas responsibility for damage may also arise through causation. Al-Hilli's jurisprudential analysis of liability recognizes

the importance of distinguishing direct conduct from causative conduct and provides mechanisms for allocating responsibility where harm occurs through an intervening instrument or condition (Al-Hilli, 1998). Artificial intelligence fits naturally within this analytical structure because the machine can be understood as the immediate technical instrument while legal responsibility remains attached to the person who creates, directs, controls, or deploys the causal mechanism. This approach avoids the false choice between treating the machine as guilty and accepting that nobody is responsible.

Khomeini's treatment of responsibility similarly reflects the principle that legal consequences may arise from the relationship between human conduct and harmful outcomes even where the human actor does not physically perform the final act (Khomeini, 1990). In the military context, the relevance of this principle is substantial. A commander who activates an autonomous weapon may not personally select the individual target later attacked by the machine, but the commander has nevertheless introduced a lethal causal mechanism into the operational environment. A programmer who knowingly embeds unlawful targeting criteria may not be physically present when the attack occurs, yet the code may become a causal component of the harm. A state that deploys autonomous systems obtains strategic advantages from the technology and therefore cannot plausibly characterize resulting harms as entirely disconnected from its own conduct.

The Islamic jurisprudential distinction between the accountable agent and the instrument also reinforces the argument against independent criminal responsibility for artificial intelligence. Moral and criminal responsibility presupposes a subject capable of understanding obligation and violating it through conscious conduct. A machine may technically deviate from programmed expectations, but deviation is not equivalent to disobedience in the moral or legal sense. It does not understand that a command is lawful or unlawful, nor does it choose wrongdoing because of greed, hatred, political ideology, fear, or another human motive. Accordingly, attributing criminal guilt to the machine would transform an instrumental failure into a fictional form of moral agency.

This does not imply that machines are irrelevant to legal analysis. The technical characteristics of artificial intelligence determine the foreseeability and controllability of the risk, which in turn influence the responsibility of human actors. A highly deterministic system may create stronger causal links to its programmer than a complex adaptive model. A system deployed despite repeated evidence of civilian misclassification may create stronger grounds for responsibility than one whose failure could not reasonably have been predicted. The jurisprudential concept of causation therefore requires a technologically informed inquiry. The question is not merely who touched the machine but whose legally relevant conduct created or controlled the risk that materialized.

Natural-law traditions approach the issue from the concept of free will. Criminal blame presupposes the capacity for morally meaningful choice. Francis's criticism of technocratic reasoning is relevant because it emphasizes that technology should remain an instrument ordered toward human purposes rather than becoming a substitute for moral discernment (Francis, 2015). In the military context, the decision to take human life cannot be treated as a purely technical optimization problem. A targeting process may involve quantitative information, probability assessments, sensor data, and military calculations, but the determination that force should be used against a particular person or object involves legal and moral judgment.

Artificial intelligence can support that judgment without possessing it. This distinction provides an important basis for meaningful human control. The requirement of human control is not based on nostalgia for traditional warfare or rejection of technological innovation. It reflects the normative proposition that decisions governed by humanitarian law must remain connected to actors who can understand legal obligations and be held responsible for their violation. Amoroso's discussion of autonomous weapons demonstrates that human control has emerged as a central concept precisely because autonomy threatens to separate legal judgment from the technical execution of force (Amoroso, 2020). Meaningful control therefore serves both humanitarian and accountability functions.

International criminal law can incorporate these comparative insights without abandoning its positive legal structure. The Rome Statute already centers responsibility upon natural persons and requires recognized forms of participation. Werle and Jessberger emphasize that international criminal responsibility is based on individualized culpability and carefully differentiated modes of liability (Werle & Jessberger, 2020). The challenge presented by artificial intelligence is therefore not that international criminal law lacks any conceptual basis for responsibility but that existing doctrines must be applied to increasingly distributed decision-making structures.

A human-centered model should begin by identifying responsibility at the operational level. Commanders occupy a central position because they authorize missions, select means and methods, establish rules of engagement, and determine whether autonomous functions will be subject to human intervention. A commander should not automatically become criminally responsible for every malfunction of military technology. Such an approach would amount to strict criminal liability and conflict with fundamental principles of personal culpability. Nevertheless, command responsibility becomes relevant where the commander knew or should have known of serious risks, failed to adopt reasonable preventive measures, ignored evidence of unlawful system behavior, or deliberately removed meaningful supervision.

The second level concerns technical actors. Programmers, data scientists, system designers, manufacturers, and contractors possess forms of expertise that commanders may lack. Their responsibility should depend upon the nature of their contribution and mental relationship to the unlawful conduct. Cassese's analysis of participation in international crimes demonstrates why criminal liability cannot be based solely on causal contribution (Cassese, 2013). Modern military operations depend upon countless causal contributions, many of which are remote from criminal intent. A manufacturer of computer hardware, for example, should not incur responsibility merely because its equipment was later incorporated into a weapons system. The relevant inquiry must consider whether the technical actor intentionally facilitated criminal conduct, knowingly concealed dangerous defects, designed unlawful targeting rules, or otherwise satisfied the elements of a recognized mode of participation.

The third level concerns the state. Military artificial intelligence is ordinarily developed or deployed because a state seeks strategic advantages in speed, scale, accuracy, persistence, or reduced exposure of personnel. State responsibility should therefore accompany the benefits of technological deployment. Schmitt's analysis of cyber operations is particularly relevant because hybrid warfare often involves private technological actors operating under varying degrees of state control (Schmitt, 2017). A state cannot avoid international responsibility merely by structuring its military technology through private contracts, proxy organizations, or formally independent operators when the applicable rules of attribution connect their conduct to the state.

The state also has preventive obligations independent of responsibility for a completed violation. The procurement and deployment of artificial intelligence should include rigorous legal and technical review. Scharre's examination of autonomous weapons emphasizes the significance of testing, reliability, operational predictability, and human-machine interaction (Scharre, 2018). A state that deploys a system without understanding its failure modes creates risks that may be difficult to attribute after the fact. Preventive regulation is therefore essential. The objective should be to reduce both humanitarian harm and future evidentiary uncertainty.

A central component of this preventive model is algorithmic auditability. Military systems need not disclose sensitive source code publicly, but the relevant authorities should be able to reconstruct the system's operation. Auditability requires documentation of training data, system versions, performance limitations, mission parameters, human interventions, sensor inputs, and relevant outputs. Chesterman's treatment of AI regulation underscores the wider governance value of transparency and accountability in algorithmic systems (Chesterman, 2021). In military contexts, such documentation should be treated as a legal safeguard because without it investigators may be unable to determine whether the unlawful result originated in defective design, corrupted data, operator misuse, command negligence, hostile manipulation, or an unforeseeable technical anomaly.

Explainability also has importance, but it should not be treated as an absolute requirement in a simplistic sense. Some advanced machine-learning systems may never provide complete human-readable explanations of every internal computation. The legally relevant question is whether decision-makers can obtain enough information to assess reliability, limitations, and operational suitability. A system that cannot be meaningfully evaluated before deployment or investigated after harm presents a serious accountability problem. Accordingly, the degree of required explainability should be proportionate to the seriousness of the function performed. Lethal targeting should demand exceptionally high standards because the consequences of error are irreversible.

Meaningful human control should likewise be legally defined through functional criteria rather than symbolic human participation. A person should have sufficient situational awareness to understand the relevant operational context, adequate information concerning the system's confidence and limitations, real authority to intervene, and enough time to exercise judgment. If the process is designed so that human approval becomes automatic because machine outputs arrive too rapidly or in excessive volume, human control is not meaningful. Amoroso's work demonstrates that the legal debate surrounding

autonomous weapons increasingly recognizes the importance of preserving genuine human involvement rather than merely nominal oversight (Amoroso, 2020).

The human-centered model should also distinguish among forms of liability. Individual criminal responsibility should remain fault-based and tied to recognized mental elements. State responsibility can operate independently of individual criminal conviction. Civil and compensatory mechanisms can address victims' losses even when personal criminal intent cannot be proven. Regulatory measures can sanction inadequate compliance systems, documentation failures, or unsafe deployment before those failures rise to the level of international crimes. This differentiation avoids the temptation to force every harmful event into the category of criminal responsibility.

The concept of strict state liability can play an important role within this architecture, although it should be distinguished carefully from strict criminal liability. Where states introduce exceptionally hazardous autonomous systems into armed conflict, there is a strong normative argument that victims should not bear the consequences of technological uncertainty merely because no individual perpetrator can be identified. The state is institutionally better positioned to distribute risk, compensate harm, regulate contractors, and improve future systems. Such an approach also reflects the jurisprudential principle that those who benefit from a dangerous instrument should bear responsibility for harms generated by its use.

This model can be described as hierarchical responsibility because it recognizes different forms of responsibility according to the actor's position in the technological and operational chain. The operational level concerns those who authorize and supervise the use of force. The technical level concerns those who design, configure, maintain, or knowingly facilitate the system's unlawful functions. The sovereign level concerns the state that adopts, deploys, directs, or benefits from the military technology. The layers overlap where the legal requirements for more than one form of responsibility are satisfied.

Such a model provides several advantages. It preserves the human-centered character of criminal law and avoids fictional attribution of mens rea to machines. It acknowledges that artificial intelligence can create genuinely difficult causal and evidentiary problems without treating those problems as reasons for impunity. It distinguishes criminal punishment from compensation and institutional responsibility. It also creates incentives for states and companies to maintain human control, documentation, transparency, testing, and oversight because failures in these areas cannot simply be attributed to an autonomous machine.

The proposal also addresses the specific characteristics of hybrid warfare. Because hybrid operations frequently involve ambiguous actors, cyber systems, private contractors, and distributed technological infrastructures, responsibility cannot depend exclusively on identifying the person who physically executes the final attack. Schmitt's treatment of cyber attribution demonstrates that international law increasingly confronts situations in which state involvement must be evaluated through relationships of control and direction rather than conventional military uniforms or territorial presence (Schmitt, 2017). Artificial intelligence makes this approach even more necessary. The relevant legal inquiry should follow authority and causation through the technological chain rather than stop at the machine.

Human dignity provides the final normative justification for this model. Francis's critique of technological domination emphasizes that technical capability should not be allowed to replace moral responsibility (Francis, 2015). Warfare necessarily involves decisions with profound human consequences. Artificial intelligence may assist in making those decisions more informed, rapid, or precise, but responsibility for them must remain human. A legal order that permits states to benefit from machine autonomy while attributing resulting crimes to the machine would invert the relationship between technology and responsibility. The machine should remain the instrument; the humans and institutions that create, deploy, and govern it should remain answerable.

The proposed model therefore calls for the development of clearer international standards concerning meaningful human control, algorithmic auditability, legal review of autonomous systems, documentation of machine decision processes, responsibilities of private military technology providers, and mechanisms of state compensation. It does not require the abandonment of existing international humanitarian law. Rather, it requires that existing principles be interpreted and supplemented in a manner capable of preserving attribution in technologically distributed warfare. The central objective should be to ensure that increasing autonomy in weapons systems does not produce decreasing responsibility among human actors.

6. Conclusion

The growing integration of artificial intelligence into hybrid warfare has created one of the most difficult contemporary challenges for international humanitarian law and international criminal law. The challenge does not arise merely because artificial intelligence represents a new military technology, but because it alters the structure through which information is processed, targets are identified, decisions are generated, and lethal force is applied. Hybrid warfare already complicates attribution through the use of cyber operations, non-state actors, private contractors, proxies, information warfare, and activities conducted within the grey zone between peace and conventional armed conflict. Artificial intelligence adds another layer of complexity by allowing computational systems to perform functions that were traditionally associated with immediate human judgment. The result is an increasingly distributed decision-making environment in which the physical act, technical process, command decision, and mental element may be located in different persons or institutions.

The central finding of this study is that operational autonomy should not be equated with criminal agency. Artificial intelligence can process data, identify patterns, classify objects, generate recommendations, and in some configurations select or engage targets without immediate human intervention. Nevertheless, these capacities do not establish that the system possesses intention, knowledge, free will, moral understanding, or consciousness in the sense required by criminal law. Computational optimization is not equivalent to criminal intention, and the possession or processing of information is not equivalent to legal knowledge. Consequently, directly attributing mens rea to artificial intelligence would distort the foundations of individual criminal responsibility rather than resolve the accountability problem.

Granting independent criminal personality to artificial intelligence would also create significant practical and normative difficulties. Criminal punishment is directed toward morally accountable subjects and serves purposes such as retribution, deterrence, and rehabilitation. Artificial intelligence cannot experience blame, remorse, suffering, moral condemnation, or fear of punishment. Deactivation, deletion, retraining, or technical restriction may be necessary regulatory responses to unsafe systems, but such measures are not criminal punishment in the ordinary juridical sense. Treating the machine as the offender could also weaken accountability by encouraging states, commanders, and corporations to portray unlawful outcomes as the independent behavior of an artificial agent.

The responsibility gap should therefore be understood as a problem of attribution rather than as proof that international criminal law requires a new category of artificial offender. The harmful result produced by an autonomous system remains connected to a broader chain of human conduct. Systems are designed, trained, tested, procured, authorized, configured, deployed, monitored, and maintained by human beings and institutions. Even where the final operational output is not directly chosen by a human operator, the circumstances that permit the algorithm to exercise lethal autonomy are themselves products of human decisions. The task of international law is therefore to reconstruct those relationships in a sufficiently precise manner to determine which actors possess the authority, causal contribution, knowledge, or culpability necessary for the relevant form of responsibility.

Commanders occupy a central position within this framework. The use of artificial intelligence does not eliminate command responsibility or transform commanders into passive recipients of machine outputs. Military authority includes responsibility for selecting lawful means and methods of warfare, understanding significant operational limitations, establishing appropriate supervision mechanisms, responding to warning indicators, and ensuring that the use of force remains consistent with international humanitarian law. A commander should not be automatically criminally liable for every technical malfunction, but the deliberate or negligent removal of meaningful oversight, the deployment of systems known to be unreliable under particular conditions, and the failure to respond to recurring unlawful outcomes may become legally significant.

Meaningful human control therefore functions as more than an ethical aspiration. It is an accountability mechanism. Human control is meaningful only when decision-makers possess sufficient information to understand the relevant circumstances, adequate knowledge of system limitations, actual authority to intervene, and sufficient time to exercise independent judgment. A formal requirement that a person press a confirmation button cannot preserve responsibility if the operational process makes substantive review impossible. International standards should therefore define meaningful human control through functional criteria rather than symbolic human participation.

Programmers, developers, manufacturers, data specialists, and private technology providers constitute a second layer of responsibility. Their criminal liability should remain dependent upon recognized modes of participation and appropriate mental

elements. Technical contribution alone should not create international criminal responsibility because modern military systems depend upon numerous remote contributors who may lack knowledge of particular unlawful uses. Nevertheless, responsibility may arise where technical actors intentionally facilitate criminal conduct, knowingly design unlawful targeting functions, conceal serious defects, manipulate testing, or otherwise contribute to crimes with the required culpable state of mind. The legal analysis must distinguish intentional criminal facilitation from negligent design, ordinary technical error, and unforeseen malfunction.

Where criminal culpability cannot be established, the legal system should not treat victims' losses as irrelevant. Civil, compensatory, regulatory, and state-responsibility mechanisms can address harms that fall outside the requirements of individual criminal liability. This distinction is essential because the pursuit of accountability does not require every form of technological harm to be transformed into a criminal offense. Criminal punishment, compensation, prevention, and institutional regulation serve different purposes and should operate through different legal standards.

State responsibility forms the third and broadest level of the proposed framework. States choose to develop, acquire, and deploy military artificial intelligence because they expect strategic benefits from speed, scale, precision, persistence, and reduced exposure of personnel. They therefore cannot rely upon the technological autonomy of their systems or the private status of contractors to avoid responsibility for internationally wrongful conduct. Outsourcing technological development should not become outsourcing of legal responsibility. Where conduct is attributable to the state and breaches an international obligation, state responsibility should remain available regardless of whether an individual criminal conviction can be obtained.

The comparative analysis undertaken in this study reinforces this human-centered model. The jurisprudential distinction between the responsible person and the instrument through which harm occurs demonstrates that the absence of agency in the immediate instrument does not necessarily create an absence of responsibility. Concepts of causation and guarantee provide a coherent basis for tracing responsibility to the person or institution that creates, directs, or controls the relevant risk. Natural-law understandings of free will similarly confirm that culpability is connected to meaningful moral choice rather than mechanical decision-making. These approaches converge around a common proposition: sophisticated technology may mediate human conduct, but it should not replace the human subject as the bearer of legal and moral responsibility.

A sustainable legal response to artificial intelligence in warfare must also address evidence. Without adequate documentation, the most carefully designed doctrines of responsibility may become ineffective. Military artificial intelligence should therefore be subject to requirements of auditability, traceability, preservation of operational logs, documentation of training and system versions, recording of human interventions, and identification of known performance limitations. Such requirements should be integrated into weapons review, procurement, testing, and operational deployment. The purpose is not unrestricted public disclosure of sensitive military technology but preservation of sufficient information to allow competent legal authorities to reconstruct the decision-making process when serious harm occurs.

The same logic supports stronger legal review before deployment. Artificial intelligence should not be evaluated solely according to whether the physical platform constitutes a lawful weapon. Review should consider the decision-making functions performed by the software, the reliability of target classification, performance under degraded conditions, vulnerability to manipulation, quality of training data, explainability, human-machine interaction, and the operational environment in which deployment is authorized. Systems performing lethal functions should be subject to particularly rigorous standards because errors may be irreversible and may directly implicate protected civilians.

On this basis, the article proposes a hierarchical model of responsibility. At the operational level, commanders and relevant operators remain responsible for the lawful organization, supervision, and use of artificial intelligence in military operations. At the technical level, developers and manufacturers may bear responsibility where their intentional or otherwise legally culpable contributions satisfy the conditions of applicable law, while other forms of liability should address defective systems that do not meet the threshold of criminal culpability. At the sovereign level, the state remains internationally responsible for the lawful deployment of its military capabilities and should provide mechanisms of reparation where technological systems cause unlawful harm attributable to it.

This hierarchical structure does not eliminate every difficult case. Advanced artificial intelligence may continue to produce unforeseen outputs, and highly distributed development processes may make individual culpability difficult to prove. International criminal law should not respond to these difficulties by relaxing fundamental requirements of personal guilt. Instead, legal development should ensure that the absence of individual criminal liability does not create a complete

accountability vacuum. State responsibility, compensation, preventive regulation, technical auditing, and institutional oversight can operate alongside criminal law and preserve a broader structure of accountability.

The future regulation of autonomous military systems should therefore move toward a clear legal requirement of meaningful human control, stronger mechanisms for technical transparency and post-incident reconstruction, enhanced legal review of artificial intelligence before operational deployment, and explicit rules concerning the responsibilities of commanders, private contractors, technology companies, and states. International law should also resist legal structures that encourage humans to surrender meaningful authority over lethal force while retaining the strategic benefits of autonomous systems.

The ultimate problem is not that artificial intelligence has become too human to regulate through existing law, but that human beings may attempt to become too distant from the consequences of technologies they design and deploy. The law should prevent that distancing from becoming a pathway to impunity. Artificial intelligence may transform the speed, scale, and organization of warfare, but it should not transform responsibility into an abstraction. The legitimacy of international humanitarian law depends upon preserving a recognizable connection between the exercise of lethal force and persons or institutions capable of being held answerable for it. For that reason, the future of accountability in algorithmic warfare should remain fundamentally human-centered.

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

Al-Hilli, M. (1998). *Shara'i al-Islam fi Masa'il al-Halal wa al-Haram* (Vol. 4).
Amoroso, D. (2020). *Autonomous Weapons Systems and International Law*. Cambridge University Press.
Cassese, A. (2013). *International Criminal Law*. Oxford University Press.
Chesterman, S. (2021). *We, the Robots? Regulating Artificial Intelligence*. Cambridge University Press.
Francis, P. (2015). *Laudato si': On Care for Our Common Home*. Holy See.
Hoffman, F. G. (2007). *Conflict in the 21st Century: The Rise of Hybrid Wars*.
Khomeini, R. (1990). *Tahrir al-Wasilah*. Institute for the Compilation and Publication of Imam Khomeini's Works.
Scharre, P. (2018). *Army of None: Autonomous Weapons and the Future of War*. W. W. Norton & Company.
Schmitt, M. N. (2017). *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations*. Cambridge University Press.
Solis, G. D. (2016). *The Law of Armed Conflict: International Humanitarian Law in Theory and Practice*. Cambridge University Press.
Solum, L. B. (1992). Legal Personhood for Artificial Intelligence. *North Carolina Law Review*.
Werle, G., & Jessberger, F. (2020). *Principles of International Criminal Law*. Oxford University Press.