



TOWARDS AN INTERNATIONAL CRIMINAL- LAW FRAMEWORK FOR ARTIFICIAL INTELLIGENCE IN DEFENCE AND SECURITY: A PROPOSAL FOR A NEW CONVENTION

Mirsulton Abbosov

Independent Researcher,

Tashkent State University of Law, Uzbekistan

E-mail: abbosovmirsulton@gmail.com

ORCID: 0000-0002-1173-6254

Abstract

The rapid development of artificial intelligence (AI) has created new challenges for international security, international humanitarian law, and criminal law. The use of increasingly autonomous AI systems in military and defense activities creates a serious legal problem: when an AI-enabled system causes unlawful death, injury, or other serious consequences, who should bear criminal responsibility? Existing international instruments provide important rules concerning the conduct of hostilities and the protection of human rights, but they do not establish a sufficiently specific criminal-law framework for highly autonomous AI-enabled weapons and systems. This article examines this legal gap and proposes a new international Convention on the Regulation of Artificial Intelligence Technologies and Their Use in Defense and Security. The proposed Convention was developed within a doctoral research project and consists of 17 articles. Its central principles include a prohibition of fully autonomous weapon systems, mandatory and effective human control over permitted AI-enabled weapons, international monitoring, algorithmic audit, explain ability, emergency shutdown mechanisms, and direct linkage between *actus reus*, *mens rea*, human conduct and individual responsibility. The article argues that the main criminal-law innovation of the proposed Convention is not the recognition of artificial intelligence as an independent criminal subject, but the establishment of a clear human-centered responsibility model for crimes committed with the



involvement of autonomous technologies. The article also discusses the proposed International Artificial Intelligence Agency as an institutional mechanism for monitoring compliance, conducting inspections and investigating serious incidents. The study concludes that a specialized international instrument could complement existing AI governance and international humanitarian law by addressing the criminal-law dimension of autonomous military technologies.

Keywords: Artificial intelligence, criminal law, international criminal law, autonomous weapons, human control, criminal responsibility, *actus reus*, *mens rea*, artificial intelligence agency, international convention.

INTRODUCTION

Artificial intelligence is rapidly changing the way modern societies, governments and military organizations operate [12,16]. AI systems are increasingly capable of analyzing large volumes of information, identifying objects, predicting events, selecting targets and making decisions with limited human intervention. These developments create significant opportunities, but they also create new risks for international peace and security. The legal problem becomes particularly serious when AI is used in defense and military operations. A conventional weapon is normally operated or controlled by a human being who can be identified and whose conduct can be legally assessed. An autonomous or highly automated AI-enabled system may, however, perform several stages of the decision-making process independently. This raises a fundamental question of criminal law: if an AI-enabled military system makes an unlawful decision that causes serious harm, who is criminally responsible for that result? This question is not merely technological. It concerns the foundations of criminal responsibility. Traditional criminal law is built around the conduct of a human offender. The existence of a crime normally requires an objective element (*actus reus*) and a subjective element (*mens rea*), together with a legally recognized human subject of the offence. AI systems challenge this structure because they may perform actions that were not specifically determined by a human operator at the moment when the harmful consequence



occurred. The existing academic debate has produced several approaches. Some scholars consider AI primarily an instrument through which a human commits a crime. Others have discussed whether highly autonomous AI could eventually acquire some form of legal personality or legal subjectivity [4,5,6,8,9,11]. A further approach considers AI a “quasi-subject” because its autonomous characteristics may influence the commission of an offence without making the system a legally responsible person [13,17]. Our research proposes underlying this article adopts the latter analytical perspective but does not transfer criminal responsibility to the AI system itself. Instead, it develops a model of human-centered distributed responsibility in which the conduct of developers, manufacturers, commanders, operators and other responsible persons is assessed according to their contribution, knowledge, intention, negligence and degree of control. Our research proposes describes this model through the concepts of “quasi-subject”, “hybrid crimes” and “distributed responsibility along the chain”. The problem is particularly acute in the military sphere. Fully autonomous weapons may identify, select and attack targets without meaningful human intervention. The proposed Convention therefore starts from the principle that autonomy cannot be allowed to become a mechanism for eliminating human criminal responsibility. The central argument of this article is that technological autonomy should not produce legal impunity.

The article therefore proposes a new international criminal-law-oriented framework for AI in defence and security. The framework is embodied in a 17-article draft Convention developed as part of the author’s doctoral research. The draft seeks to establish common international standards concerning prohibited autonomous weapon systems, permissible AI-enabled weapons, human control, international inspection, algorithmic auditing, incident investigation and institutional cooperation.

The proposed Convention should not be understood as replacing existing international humanitarian law. Rather, it is intended to address a specific regulatory gap: how to prevent the use of AI systems from creating situations in which the causal connection between human conduct and criminal responsibility becomes legally unclear.



MATERIALS AND METHODS

This study uses a doctrinal and comparative legal research methodology. The main research material consists of the proposed 17-article Convention developed within our research proposes, together with international legal instruments and academic approaches concerning artificial intelligence, autonomous weapons and criminal responsibility [17].

Our research itself examines AI from the perspective of criminal law, including its legal status, objective and subjective elements of AI-related offences, criminalization, qualification of offences and improvement of criminal legislation. The research also relies on comparative analysis of foreign legislation and international instruments. Our research proposes specifically analyses approaches from the United States, European Union, China, South Korea, the Russian Federation and Kazakhstan, as well as international instruments such as the Budapest and Lanzarote Conventions.

The study applies the following methods: Doctrinal legal analysis – to determine the relationship between AI autonomy and traditional concepts of criminal responsibility. Comparative legal analysis – to compare existing approaches to AI regulation and autonomous weapons. Legal modelling – to develop a proposed international Convention and its institutional mechanisms. Systemic analysis – to examine the relationship between international humanitarian law, international security and criminal responsibility. Formal-legal analysis – to assess the proposed definitions, prohibitions, obligations and institutional rules. Functional analysis – to determine whether existing legal mechanisms are capable of addressing criminal conduct involving highly autonomous AI systems.

The research also used dialectical, analytical, synthetic, deductive, inductive, sociological survey, historical-legal, political-legal, content-analysis and statistical methods.

RESULTS

The research demonstrates that AI creates a specific form of what may be called a responsibility gap. The traditional criminal-law model generally assumes that the person who acts, decides or controls the relevant conduct can be identified.



***Modern American Journal of Business,
Economics, and Entrepreneurship***

ISSN (E): 3067-7203

Volume 2, Issue 5, May, 2026

Website: usajournals.org

***This work is Licensed under CC BY 4.0 a Creative Commons
Attribution 4.0 International License.***

AI-based systems complicate this assumption. A system may be designed by one person, developed by another organization, deployed by a military institution and operated by another individual. At the same time, the system may adapt to its environment and produce a decision that was not expressly programmed by any particular human being. This creates a chain of potentially responsible actors. Our research therefore proposes the concept of technological participation, under which the legal assessment should not focus only on the immediate operator but should examine the entire chain of human involvement. Our research proposes also develops the concepts of “hybrid crimes” and “distributed responsibility along the chain”.

The proposed model is based on an important principle: AI autonomy may affect the mechanism through which an offence is committed, but it should not eliminate the requirement of human criminal responsibility. This principle is especially important in military operations.

“Convention on the Regulation of Artificial Intelligence Technologies and Their Use in Defense and Security”. Its purpose is to harmonies the use of AI technologies in defense, security and military activities with international humanitarian law and to ensure human control over autonomous weapon systems. The draft consists of 17 articles and is divided into provisions dealing with general rules, prohibitions and limitations, international control mechanisms, institutional governance, inspections, dispute settlement, amendments, withdrawal, entry into force and depositary functions. Its criminal-law significance can be explained through five principal innovations.

First: prohibition of fully autonomous weapons. Article 2 establishes an absolute prohibition on fully autonomous weapon systems (LAWS). It also prohibits self-learning systems capable of changing the algorithms for distinguishing civilians from combatants without human intervention, as well as AI systems given authority to control weapons of mass destruction. This is a major legal principle because the prohibition is directed not merely at a harmful result but at the creation and deployment of a technological capability capable of removing human control from lethal decision-making. The approach is therefore preventive as well as punitive.



Second: mandatory human control. The proposed Convention permits the use of AI-enabled weapons only where continuous, effective and direct human control is maintained and an emergency stop mechanism is available. This requirement has direct criminal-law significance. Human control provides a potential legal connection between the technological action and the responsible human actor. If an operator has meaningful control, receives relevant information, understands the system's operation and has the ability to intervene, the assessment of *mens rea* and negligence becomes legally more realistic. Conversely, a system designed to operate without meaningful human control creates a substantially different risk profile.

Third: algorithmic audit and explain ability. The proposed International Artificial Intelligence Agency would have authority to analyse source code, training datasets and algorithms of military and dual-use AI systems. It could also require disclosure of decision-making chains and audit trails. This is important for criminal law because modern AI systems may operate as a "black box". If the decision-making process cannot be reconstructed, establishing causation and *mens rea* may become extremely difficult. Therefore, the proposed Convention treats auditability as a component of legal accountability. The underlying logic is straightforward: a system capable of exercising potentially lethal autonomy must generate sufficient technical and documentary information to allow subsequent legal assessment.

Fourth: international investigation and monitoring The Convention proposes an independent international agency capable of inspecting military facilities, research centers and technological laboratories, including inspections in suspicious cases. It would also monitor the illegal proliferation of prohibited military AI technologies and investigate incidents involving AI weapons. The draft further requires the Agency to establish international standards concerning cybersecurity, emergency shutdown mechanisms and effective human control. This mechanism is designed to reduce the possibility that criminally dangerous AI technologies remain outside effective international supervision.

Fifth: preservation of individual criminal responsibility the most important criminal-law feature of the proposed Convention is its human-centered approach to responsibility. The Convention does not propose treating AI itself as a



criminal offender. Instead, its preamble expressly connects *actus reus* and *mens rea* with human conduct and individual responsibility. This is a significant doctrinal choice. It avoids two extremes. The first extreme is to treat AI as merely an ordinary object or tool and ignore its capacity for autonomous decision-making. The second extreme is to recognize AI as an independent criminal subject and thereby risk transferring responsibility away from the humans who designed, deployed, controlled or intentionally misused the system. The proposed model occupies an intermediate position: AI can have significant technological autonomy, but criminal responsibility remains attached to legally responsible human actors.

DISCUSSION

This is the principal question that must be addressed in any serious academic discussion. One may argue that a new international Convention is unnecessary because existing international humanitarian law already regulates the conduct of hostilities. The principles of distinction, proportionality and precautions in attack provide important legal limitations on the use of weapons, including technologically advanced weapons. This argument has substantial force. However, it does not completely resolve the criminal-law problem created by autonomous AI. Existing rules primarily regulate what states and parties to armed conflict may or may not do. The proposed Convention addresses an additional question: How should responsibility be preserved when an AI system participates in the chain leading to an unlawful act? This distinction is important. The Council of Europe Framework Convention on AI [1], for example, establishes principles concerning human dignity, individual autonomy, transparency, oversight, accountability, responsibility and safe innovation. It also requires risk and impact assessment and permits states to consider bans or moratoria for certain AI uses. However, its primary subject matter is human rights, democracy and the rule of law [2,3]. It is not a specialized international criminal-law treaty concerning autonomous military weapons. The proposed Convention therefore seeks to occupy a different normative space. It is not intended to replace general AI governance. It is intended to establish a



specialized international security and criminal-responsibility framework for AI in defense and military activities.

It is not sufficient to place a human somewhere within the technological chain and describe the system as “human-controlled”. A human operator who cannot understand the system’s decision, cannot predict its behavior and cannot stop it in time does not exercise meaningful legal control. For this reason, the proposed Convention requires control to be:

continuous; effective; direct; and supported by an emergency stop mechanism. This approach should be developed further in international criminal law. A useful distinction may therefore be made between: formal human control, a human is technically present;

and substantive human control, a human possesses the actual capacity to understand, supervise, intervene and stop the AI system. Only the second form should have significant legal value in determining responsibility.

AI autonomy and *mens rea*. The most difficult issue is *mens rea*. Criminal intent is traditionally associated with a person’s mental attitude towards conduct and consequences. An AI system may select a target or perform an action without possessing consciousness or legally recognized intent. Therefore, the criminal law should not ask whether the AI “intended” to cause the result. The relevant question should be: What did the responsible human actor know, foresee, accept, or fail to prevent? For example, if a commander knowingly deploys an AI system whose autonomous behavior creates a foreseeable and unacceptable risk to civilians, the system’s autonomy should not automatically exclude criminal responsibility. Similarly, a developer who deliberately removes safety restrictions for the purpose of enabling unlawful military use may potentially occupy a different position from an operator who uses a certified system in accordance with established rules.

This supports Our research’s concept of distributed responsibility along the technological chain. The criminal-law analysis should therefore consider: the person’s role; the degree of control; knowledge of the system’s capabilities; foreseeability of the harmful result; intention or acceptance of the risk; compliance with safety requirements; failure to monitor or intervene; modification or reprogramming of the system; and the causal relationship



***Modern American Journal of Business,
Economics, and Entrepreneurship***

ISSN (E): 3067-7203

Volume 2, Issue 5, May, 2026

Website: usajournals.org

***This work is Licensed under CC BY 4.0 a Creative Commons
Attribution 4.0 International License.***

between human conduct and the final harmful result. This approach provides a more realistic model than assigning responsibility exclusively to the final operator.

The research proposes that AI should not be regarded simply as a conventional tool or as a fully independent criminal subject. Instead, its autonomy can justify describing it as a quasi-subject in certain criminal-law analyses. This concept is theoretically useful, but it should not be confused with legal personality. A “quasi-subject” does not necessarily possess rights and duties equivalent to those of a human being or corporation.

Rather, the concept describes a technologically autonomous element within the structure of an offence. This distinction is important because criminal law cannot punish an algorithm in the same manner as a natural person. Imprisonment, moral blame, rehabilitation and other conventional criminal sanctions presuppose a particular type of legal and moral subject. The appropriate legal response is therefore not to punish AI as an independent offender, but to regulate the humans and organizations responsible for creating, deploying, controlling, modifying and intentionally misusing AI systems.

4.5. The proposed International Artificial Intelligence Agency

The Convention’s institutional innovation is the establishment of an International Artificial Intelligence Agency. According to the draft, the Agency would perform international monitoring and coordination functions and would help reconcile differences between national regulatory approaches. Its proposed powers include: inspections of military facilities; analysis of source code and training datasets; algorithmic audits; monitoring of prohibited military AI technologies; development of international cybersecurity and human-control standards; assessment of new defense AI systems; international compliance certification; investigation of AI-related military incidents; and reporting to international institutions. The proposed institutional structure consists of a General Assembly, Coordinating Council, Executive Committee and Secretariat. This institutional design attempts to combine political representation with independent technical expertise. The Executive Committee, for example, would consist of independent specialists in AI, algorithmic auditing, cybersecurity and military law and would have authority to initiate inspections and audits.



***Modern American Journal of Business,
Economics, and Entrepreneurship***

ISSN (E): 3067-7203

Volume 2, Issue 5, May, 2026

Website: usajournals.org

***This work is Licensed under CC BY 4.0 a Creative Commons
Attribution 4.0 International License.***

The proposed Convention should be viewed within the broader development of international AI law. The Council of Europe Framework Convention is currently the most significant example of a binding international treaty specifically addressing AI. It was adopted in 2024 and opened for signature on 5 September 2024. It establishes principles including human dignity, autonomy, equality, privacy, transparency, accountability, reliability and safe innovation. The proposed Convention differs in its sectoral and criminal-law orientation. Its central object is not AI in general but AI in defense, security and military operations. Its central legal concern is not only whether AI respects human rights but also whether human criminal responsibility can remain identifiable when AI systems acquire operational autonomy. This difference allows the proposed Convention to complement rather than duplicate existing instruments. The proposed Convention can be considered a criminal-law innovation in at least four respects. The first innovation is the explicit connection between technological conduct and individual responsibility. AI autonomy is recognized as a technological fact, but not as a basis for excluding human responsibility. It links *actus reus* and *mens rea* to the AI decision chain. The second innovation is methodological. The analysis of a crime involving AI should not stop at the final technological action. It should reconstruct the entire chain: design → development → programming → deployment → command → operation → AI decision → harmful consequence. The criminal-law inquiry must then determine which human actor contributed to this chain and with what mental state. It transforms algorithmic transparency into a criminal-law safeguard the requirement of audit trails and explain ability has traditionally been discussed as a technological or regulatory matter. The proposed Convention gives it a criminal-law dimension. Without adequate records, it may be impossible to establish causation, foreseeability, knowledge or negligence. Therefore: algorithmic auditability becomes a precondition for effective legal accountability. The Convention does not rely exclusively on punishment after a crime occurs. It establishes: prohibited systems; human-control requirements; technical safeguards; international certification; monitoring; inspections; investigation; a responsibility mechanism. This reflects a preventive model of criminal-law policy.



***Modern American Journal of Business,
Economics, and Entrepreneurship***

ISSN (E): 3067-7203

Volume 2, Issue 5, May, 2026

Website: usajournals.org

***This work is Licensed under CC BY 4.0 a Creative Commons
Attribution 4.0 International License.***

The proposed Convention may have several practical implications. First, it could establish a common minimum standard for states developing or deploying military AI. Second, it could reduce the possibility of a “responsibility gap” in cases involving autonomous weapons. Third, it could create common requirements for audit trails and explain ability. Fourth, it could facilitate international investigation of incidents involving AI-enabled weapons. Fifth, it could establish a common understanding that autonomy does not remove human responsibility. The draft also provides mechanisms for inspections and monitoring. Inspections may be initiated following a reasoned request by a member state, information concerning non-compliance or an initiative of the Executive Committee. In emergency situations, an inspection decision may be taken within 24 hours. The proposed Convention also contains mechanisms for peaceful dispute settlement, amendments, withdrawal, entry into force and a UN depositary. Its final article provides for authentic texts in six languages: English, Arabic, Spanish, Chinese, Russian and French.

The proposed Convention also raises several questions requiring further international discussion. First, states may be reluctant to permit international inspections of military AI systems because of national security and classified information concerns.

Second, the requirement of source-code and algorithmic access may conflict with intellectual property and trade-secret protection. Third, different states may have substantially different technological capabilities. Fourth, the concept of “effective human control” requires further technical and legal elaboration. Fifth, the relationship between the proposed Convention and existing international humanitarian law requires detailed treaty-level analysis. Sixth, the institutional powers of the proposed International Artificial Intelligence Agency would need to be carefully reconciled with the authority of the United Nations Security Council and existing international organizations. These limitations do not invalidate the proposal. Rather, they demonstrate that AI-related criminal responsibility cannot be solved solely through national legislation.

Artificial intelligence is changing the factual structure within which criminal conduct may occur. The traditional assumption that a human being directly controls the relevant act is increasingly challenged by autonomous and semi-



autonomous systems. The central legal problem is therefore not whether AI should be “punished”. The more fundamental question is how criminal law can preserve human responsibility when technological systems perform increasingly autonomous functions. The proposed Convention on the Regulation of Artificial Intelligence Technologies and Their Use in Defense and Security provides one possible answer. The Convention developed within the author’s doctoral research consists of 17 articles and introduces a specialized international framework based on five interconnected principles: prohibition of fully autonomous weapons, effective human control, algorithmic auditability, international monitoring and preservation of individual criminal responsibility. Its principal doctrinal contribution is the recognition that AI autonomy should be treated as a factor affecting the structure of criminal conduct, but not as a basis for transferring responsibility from humans to machines. The proposed model therefore rejects the idea that technological autonomy creates legal immunity. Instead, it requires criminal-law analysis to follow the technological chain and identify the human actors who designed, developed, programmed, deployed, controlled, modified or intentionally misused the AI system. In this context, *actus reus* and *mens rea* remain fundamentally human concepts. AI may participate in the causal mechanism of an offence, but the legal assessment of guilt must remain connected to a human subject capable of bearing criminal responsibility. The proposed International Artificial Intelligence Agency would complement this model by providing international monitoring, algorithmic auditing, incident investigation and compliance mechanisms. The broader international legal environment demonstrates that international AI regulation is already moving toward binding legal instruments. The Council of Europe Framework Convention represents an important development in human rights, democracy and rule-of-law protection in the AI era. However, the criminal-law dimension of autonomous military AI remains a distinct field requiring further development.

Accordingly, the proposed Convention should be viewed not as a replacement for existing international humanitarian or human-rights law, but as a specialized criminal-law and international-security framework designed to prevent technological autonomy from becoming a source of criminal responsibility gaps.



The fundamental principle can be stated simply: The more autonomous the technology becomes, the more precise the law must become in identifying the human responsibility behind its use.

REFERENCES

1. Council of Europe. Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, CETS No. 225, Vilnius, 5 September 2024.
2. Council of Europe. The Framework Convention on Artificial Intelligence. Council of Europe Artificial Intelligence Portal.
3. Council of Europe. Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law. Treaty Series No. 225.
4. Hallevy, G. When Robots Kill: Artificial Intelligence Under Criminal Law. Northeastern University Press, 2013.
5. Hallevy, G. The Basic Models of Criminal Liability of AI Systems and Outer Circles. 2019.
6. Simmler, M., & Markwalder, N. "Guilty Robots? Rethinking the Nature of Culpability and Legal Personhood in an Age of Artificial Intelligence." Criminal Law Forum, Vol. 30, 2019.
7. Duan, Z. "Artificial Intelligence and the Law: Cybercrime and Criminal Liability." British Journal of Criminology, Vol. 62, 2022.
8. Kingston, J. "Artificial Intelligence and Legal Liability." In Innovative Techniques and Applications of Artificial Intelligence. Springer, 2016.
9. Wojtczak, S. "Endowing Artificial Intelligence with Legal Subjectivity." AI & Society, Vol. 37, 2022.
10. Walz A., Firth-Butterfield K., Implementing ethics into artificial intelligence: a contribution, from a legal perspective, to the development of an AI governance regime. USA: Duke Law & Technology Review. 2019. – 56 p.
11. Solaiman S. Legal personality of robots, corporations, idols and chimpanzees: a quest for legitimacy // Artificial Intelligence and Law. – 2017. – Vol. 25. – № 2. – P. 155–179.



-
12. Russell S., Norvig P. (2020). Artificial Intelligence: A Modern Approach (4th ed.). Pearson Education. 1166 p.
 13. Мосечкин И.Н. Искусственный интеллект в уголовном праве: перспективы совершенствования охраны и регулирования: [монография]. – Киров: Вятский государственный университет, 2020. – 111 с.
 14. Hubig C. Benefits and Limits of Autonomous Systems in Public Security // European Journal for Security Research. 2019. – P. 1–13.
 15. Goodfellow I. Generative adversarial nets //Advances in neural information processing systems. – 2014. – T. 27.
 16. Bostrom N. Superintelligence: Paths, Dangers, Strategies. – USA: Oxford, Oxford Univ. Press, 2014. – 390 p.
 17. Abbosov M.M. Jinoyat qonunchiligida sun'iy intellekt tushunchasi va ijtimoiy xavfliligi. Ijtimoiy-gumanitar fanlarning dolzarb muammolari №4 (5) – Toshkent, 2025. – B. 375–382.