TITLE: INTERNATIONAL RESPONSIBILITY FOR AUTONOMOUS ARTIFICIAL INTELLIGENCE DECISIONS IN ARMED CONFLICT: BRIDGING LEGAL GAPS IN THE AGE OF ALGORITHMIC WARFARE

AUTHOR: Dr. Mohamed Kamal Arafa Elrakhawi

DEDICATION
To the defenders of international humanitarian principles who refuse to let technological advancement outpace moral accountability. May this work contribute to a legal architecture that ensures human dignity remains non-negotiable, even in the most autonomous of battlefields.

TABLE OF CONTENTS
Title
Author Name
Dedication
Table of Contents
Chapter One: The Evolution of Autonomous Systems in Modern Warfare and the Legal Void
Chapter Two: State Responsibility and the Attribution Challenge Under International Law
Chapter Three: The Liability Nexus: Developers, Operators, and Command Structures
Chapter Four: Reconstructing Accountability: Integrating the Martens Clause with Emerging AI Norms
Chapter Five: Victim Compensation, Remedial Mechanisms, and the Future of Humanitarian Oversight
Conclusion
References
Author Name
Intellectual Property and Copyright Notice

CHAPTER ONE: THE EVOLUTION OF AUTONOMOUS SYSTEMS IN MODERN WARFARE AND THE LEGAL VOID
The integration of artificial intelligence into military operations marks a paradigm shift in the conduct of hostilities. For the purposes of this study, autonomous weapons systems are defined in accordance with established defense taxonomies as platforms capable of identifying, selecting, and engaging targets without real-time human intervention, operating across varying levels of autonomy from automated assistance to fully independent decision-making. This technological trajectory challenges the foundational premises of international humanitarian law, which historically presumed human judgment as the ultimate arbiter of proportionality, distinction, and military necessity. Current legal discourse remains heavily concentrated on the prohibition of lethal autonomous weapons, yet the pressing doctrinal gap lies in the allocation of international responsibility when autonomous systems commit violations. The law of armed conflict was drafted for human actors operating within hierarchical command structures. Algorithms, however, function through probabilistic reasoning, adaptive learning, and opaque decision-making pathways. When a machine misidentifies a civilian structure as a military objective due to training data bias or environmental interference, existing attribution

mechanisms falter. The legal void is not merely technological but conceptual. International law requires a clear causal chain between conduct and consequence, yet autonomous decision-making fractures this chain across multiple nodes: software architecture, data curation, operational deployment, and post-deployment environmental variables. This chapter establishes the doctrinal baseline by examining how the principles of distinction and proportionality translate into algorithmic parameters, and why the absence of a direct human trigger does not absolve states of their obligations under the Geneva Conventions and their Additional Protocols. It further analyzes the jurisprudential tension between technological determinism and legal adaptability, arguing that the international legal system must evolve from a human-centric liability model to a systems-based accountability framework.

CHAPTER TWO: STATE RESPONSIBILITY AND THE ATTRIBUTION CHALLENGE UNDER INTERNATIONAL LAW

State responsibility constitutes the cornerstone of international accountability. The International Law Commission's Articles on Responsibility of States for Internationally Wrongful Acts require that conduct be attributable to the state and constitute a breach of an international obligation. Autonomous artificial intelligence complicates both prongs. Attribution traditionally hinges on the actions of state organs or entities exercising effective control. While the International Court of Justice's jurisprudence in Nicaragua (1986) anchors attribution in effective control over non-state actors, and Bosnia v. Serbia (2007) affirms a state's due diligence obligation to prevent violations, autonomous systems destabilize both paradigms by introducing algorithmic unpredictability that defies traditional foreseeability standards. When an autonomous system operates beyond predefined parameters due to emergent behavior or adversarial machine learning manipulation, the question of effective control becomes profoundly ambiguous. This chapter dissects the attribution dilemma by evaluating three doctrinal approaches: the strict liability model, which holds deploying states accountable regardless of fault; the fault-based model, which requires proof of negligence in system design, testing, or deployment protocols; and the hybrid responsibility model, which allocates liability proportionally based on the degree of human oversight, system autonomy level, and operational context. The analysis demonstrates that existing international law does not adequately address algorithmic black boxes, and proposes a modified attribution standard grounded in the principle of reasonable foreseeability in technological contexts. It further examines how state practice and opinio juris are gradually shifting toward recognizing autonomous deployment as an inherently state-controlled activity, thereby triggering direct responsibility for violations irrespective of intermediate technological failures.

CHAPTER THREE: THE LIABILITY NEXUS: DEVELOPERS, OPERATORS, AND COMMAND STRUCTURES

The operationalization of autonomous weapons disperses responsibility across a fragmented network of actors. Traditional military command responsibility, codified in Additional Protocol I and customary international law, requires superiors to prevent, repress, or report violations committed by subordinates. Autonomous systems disrupt the subordinate-superior relationship by introducing non-human decision-makers. This chapter maps the liability nexus across three critical nodes: defense contractors and software developers, military operators and

commanders, and political decision-makers authorizing deployment. Developers bear responsibility for algorithmic integrity, training data validation, and safety certification. When commercial entities supply dual-use artificial intelligence to state militaries, questions of corporate complicity and extraterritorial human rights obligations arise under frameworks such as the United Nations Guiding Principles on Business and Human Rights. Operators and commanders retain responsibility for mission parameters, rules of engagement, and continuous monitoring. The shift from human-in-the-loop to human-on-the-loop architectures directly engages the ongoing doctrinal debate on meaningful human control, a standard increasingly endorsed by intergovernmental bodies as the minimum threshold for lawful autonomous deployment. This shift dilutes direct intervention capacity but does not eliminate the duty to exercise legally sufficient oversight over critical combat functions. Political authorization carries strategic accountability, particularly when states deploy untested autonomous systems in densely populated areas without adequate risk mitigation protocols. This chapter argues for a layered liability architecture that recognizes concurrent responsibilities without allowing any single actor to deflect accountability through technological opacity. It proposes mandatory pre-deployment certification standards, transparent algorithmic auditing requirements, and clear delineation of command protocols that maintain meaningful human control over lethal decision pathways.

CHAPTER FOUR: RECONSTRUCTING ACCOUNTABILITY: INTEGRATING THE MARTENS CLAUSE WITH EMERGING AI NORMS

The Martens Clause, originating in the 1899 Hague Conventions, affirms that civilians and combatants remain under the protection of the principles of international law derived from established custom, humanity, and the dictates of public conscience. This provision serves as a dynamic interpretive tool, particularly in technological contexts where treaty law lags behind innovation. This chapter examines how the Martens Clause can bridge the accountability gap in autonomous warfare by invoking humanitarian principles that transcend specific weapon classifications. The clause mandates that the absence of explicit treaty prohibition does not permit unfettered technological deployment. Public conscience, shaped by growing civil society advocacy, academic consensus, and intergovernmental deliberations, increasingly views fully autonomous lethal decision-making as incompatible with the principles of humanity. This chapter analyzes soft law instruments, including United Nations Group of Governmental Experts reports, International Committee of the Red Cross position papers, and regional declarations, as normative precursors to binding regulation. It evaluates the jurisprudential potential of treating algorithmic violations as breaches of customary international law, particularly when they systematically undermine distinction or cause disproportionate civilian harm. The analysis further explores how the Martens Clause imposes a proactive duty on states to conduct ethical-legal impact assessments prior to autonomous system deployment. By anchoring accountability in customary humanitarian principles and general principles of law recognized under Article 38(1)(c) of the Statute of the International Court of Justice, the international community can develop interim accountability mechanisms while formal treaty negotiations progress. This chapter concludes that normative reconstruction must prioritize human dignity over technological efficiency, establishing a legal presumption against autonomy in critical combat functions until verifiable compliance with international humanitarian law can be guaranteed.

CHAPTER FIVE: VICTIM COMPENSATION, REMEDIAL MECHANISMS, AND THE FUTURE OF HUMANITARIAN OVERSIGHT

Accountability remains hollow without effective remediation for victims. Traditional war claims commissions and domestic litigation pathways are ill-equipped to address algorithmic violations due to evidentiary barriers, jurisdictional fragmentation, and the complexity of tracing harm to specific technical failures. This chapter proposes a comprehensive remedial architecture tailored to autonomous warfare. It advocates for the establishment of an international autonomous systems compensation fund, financed through mandatory state contributions and defense industry levies, to provide rapid relief to civilian victims regardless of fault attribution. Drawing institutional lessons from established mechanisms such as the United Nations Compensation Commission for Gulf War claims and the International Criminal Court Trust Fund for Victims, the proposed fund would operate with streamlined evidentiary standards that recognize the inherent opacity of artificial intelligence decision-making. This chapter further examines the role of international judicial bodies in adjudicating algorithmic violations, proposing specialized expert panels within existing tribunals to evaluate technical evidence, audit deployment protocols, and assign proportional liability. Transparency emerges as a non-negotiable prerequisite for accountability. The analysis recommends mandatory incident reporting frameworks, independent international monitoring missions, and standardized data preservation requirements for all autonomous deployments. Humanitarian oversight must evolve from reactive investigation to proactive governance, incorporating algorithmic impact assessments, continuous compliance monitoring, and public disclosure of operational parameters. This chapter concludes that a robust remedial ecosystem requires institutional innovation, cross-jurisdictional cooperation, and the recognition that technological advancement cannot supersede the right to redress for victims of armed conflict.

CONCLUSION

The autonomousization of warfare represents a profound challenge to the coherence of international responsibility. Existing legal frameworks were constructed for human actors operating within predictable command structures, and they struggle to accommodate machine learning systems that adapt, evolve, and occasionally malfunction beyond human anticipation. This study demonstrates that accountability cannot be abandoned in the pursuit of technological efficiency. State responsibility must be anchored in a modified attribution standard that emphasizes reasonable foreseeability, rigorous pre-deployment testing, and meaningful human control over lethal functions. Liability must be distributed across developers, operators, and political decision-makers through a layered architecture that prevents accountability dilution. The Martens Clause provides a vital normative bridge, invoking humanity and public conscience to constrain autonomous deployment until compliance with international humanitarian law can be verifiably ensured. Victim compensation mechanisms must be modernized to address evidentiary complexity and jurisdictional fragmentation, establishing rapid relief pathways that operate independently of prolonged fault determination. The international community stands at a critical juncture. Regulatory inaction risks normalizing algorithmic violations, while premature prohibition could stifle legitimate defensive applications. The path forward requires a balanced, evidence-driven framework that prioritizes human dignity, mandates transparency, and enforces

proportional accountability. Only through deliberate legal adaptation can the international system ensure that artificial intelligence serves humanitarian imperatives rather than undermining them.

REFERENCES

International Law Commission. Articles on Responsibility of States for Internationally Wrongful Acts. United Nations General Assembly, 2001.
International Committee of the Red Cross. Position on Artificial Intelligence and International Humanitarian Law. Geneva: ICRC, 2021.
United Nations. Report of the Group of Governmental Experts on Lethal Autonomous Weapons Systems. UN Doc. CCW/GGE.1/2023/3, 2023.
Schmitt, Michael N. (Ed.). Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations. Cambridge: Cambridge University Press, 2017.
Tallinn Manual 3.0 Editorial Board. Tallinn Manual 3.0 on International Law Applicable to Artificial Intelligence and Cyber Operations. Cambridge: Cambridge University Press, 2024.
United Nations Institute for Disarmament Research. Accountability in the Age of Autonomous Weapons. Geneva: UNIDIR, 2022.
Hague Conventions. Regulations Respecting the Laws and Customs of War on Land, with Martens Clause. The Hague, 1899.
Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts. Geneva: International Committee of the Red Cross, 1977.
United Nations Human Rights Council. United Nations Guiding Principles on Business and Human Rights. New York: UN, 2011.
International Court of Justice. Military and Paramilitary Activities in and against Nicaragua (Nicaragua v. United States of America). Merits, Judgment. ICJ Reports 1986.
International Court of Justice. Application of the Convention on the Prevention and Punishment of the Crime of Genocide (Bosnia and Herzegovina v. Serbia and Montenegro). Merits, Judgment. ICJ Reports 2007.
Boulanin, Vincent, and Maaike Verbruggen. Mapping the Development of Autonomy in Weapon Systems. Stockholm: Stockholm International Peace Research Institute, 2017.
Sparrow, Robert. Killer Robots. Journal of Applied Philosophy, Vol. 24, No. 1, 2007, pp. 62-77.
Asaro, Peter. On Banning Autonomous Weapon Systems. International Review of the Red Cross, Vol. 94, No. 886, 2012, pp. 687-709.
Crawford, James. State Responsibility: The General Part. Cambridge: Cambridge University Press, 2013.
United Nations Compensation Commission. Report and Recommendations Made by the Panel of Commissioners Concerning the First Installment of Claims for Damages to the Natural Environment and Depletion of Natural Resources. UN Doc. S/AC.26/2001/18, 2001.
International Criminal Court. Trust Fund for Victims. Strategic Plan 2014-2018. The Hague: ICC, 2014.

AUTHOR: Dr. Mohamed Kamal Arafa Elrakhawi