Proceedings of the 3rd International Conference on Global Politics and Socio-Humanities
DOI: 10.54254/2753-7048/2025.LD31954

The Legal and Moral Responsibility of AI Systems in Future Warfare

Xinyao Jia

New Channel-UIBE Qingdao A-level Centre, Qingdao, China
jxy080712@icloud.com

Abstract. As the development of artificial intelligence, it has been rapidly deployed in military operations, especially lethal autonomous weapon systems, which has sparked intense debates about the attribution of responsibility for AI's killing on the battlefield. Existing research mostly focuses on attributing responsibility to a single entity (such as developers or commanders), but this article argues that these views oversimplify the multi-level accountability framework required for governing the ethics and legal responsibilities of military AI. By integrating ethical frameworks and case studies, as well as legal considerations, this article critiques two major flaws in the current discourse: the neglect of contextual distinctions among various factors such as developers, operators, and commanders, and the lack of clarity regarding the delineation between legal and moral responsibilities. This article proposes contextualized responsibility: in different situations, different individuals (developer, operator, commander) should be held accountable, and in specific circumstances, this extends to legal liability. The contribution of this article lies in addressing the shortcomings of existing literature and providing a refined framework for the allocation of responsibility in military AI, which also ensures the core supervisory role of humans in lethal decision-making.

Keywords: artificial intelligence, military conduct, human rights, civilian casualty

1. Introduction

During the 2021 Gaza conflict, the Israeli military used an AI system called 'The Gospel' to rapidly generate airstrike targets through algorithms. Reports indicate that the system misidentified civilian structures (such as residential buildings and media offices) as Hamas bases, resulting in the deaths of over a hundred civilians. However, the Israeli military claimed the system was only used for 'decision support', refused to disclose the algorithm' s logic or target selection criteria, nor did it acknowledge responsibility for the misjudgments.

Thus it can be seen, with the rapid advancement of artificial intelligence, military forces around the world are increasingly integrating AI into their operations. In the future, armies may become fully automated, relying on AI to make critical decisions on the battlefield. However, this raises significant ethical and legal questions: if an AI system kills humans during war, who should be held responsible—the developers (The people who writes code to design AI), the commanders (High-level decision-makers who do not directly control AI), or the operators (The direct executor of the system, the person who monitors and deploys AI actions throughout the process). The urgency of this inquiry is underscored by unresolved real-world incidents.

This article summarizes various contextualized responsibilities: developers are morally and legally accountable for issues such as algorithm opacity, willful misconduct, code defects, and failure to promptly improve (in certain circumstances, this extends to legal liability); operators are legally and morally responsible for foreseeable harm caused by failure to conduct pre-deployment evaluations, abuse during system failures, or failure to intervene in a timely manner, but the boundary of their obligations depends on their actual supervisory capabilities; if commanders deploy AI in violation of the principles of distinction and proportionality in international humanitarian law, they must bear both legal and moral responsibilities.

This article makes three contributions. First, it systematically distinguishes moral responsibility (who should be condemned by conscience) from legal responsibility (who should be prosecuted according to the international law) in AI military operations. Second, it reveals the rules for responsibility allocation in different scenarios and clarifies the accountability standards for developers, military authorities, and commanders at different levels. Third, This achievement provides an accountability operation guideline for AI military applications in international wars, preventing autonomous weapons from becoming a gray area of liability exemption.

The rest of this paper begins with a literature review and their shortcomings exposing. Next, different scenarios are distinguished, followed by an analysis of the legal and moral responsibilities of various actors in each scenario. Finally, the paper concludes by summarizing the key points and dicussing the implications of the research.

2. Literature review

Before embarking on the research, I will first sort out the existing diverse viewpoints on the three responsible parties: developers, operators and commanders.

Regarding developers, flawed design makes them responsibility: intentional evasion of legal regulation [1], unpredictable AI behavior [2], inability to distinguish legal and illegal targets [3], and limited human control [4].

Scholars also talk about operators who need to assume responsibility.

1. When the system violates ethical norms (such as: the laws of war), operators are responsible [5].

2. Operators are the sole moral subject. When AI causes harm, since only humans possess moral judgment ability, the responsibility lies with the operators [6].

3. Operators are the final link in the decision-making chain, and their responsibility for initiating and monitoring the system cannot be excused [4].

Finally, some scholars also talk about commanders' responsibility: deploying AI despite knowing its insecurity [7], need to bear legal responsibility even without direct control [8].

In summary, although there have been studies on related issues, there are still two major structural flaws:

First, scholars have confused moral responsibility (subjective fault based on ethical obligations) with legal responsibility (objective attribution based on causal relationships), and have failed to establish a hierarchical determination standard;

Second, there is a lack of a dynamic analysis framework based on the operational context for the allocation of responsibilities among developers, the military, and commanders - what level of technological involvement, algorithm transparency, or battlefield urgency will trigger accountability from different entities?

3. Different scenarios

Here, I develop a framework to analyze responsibilities for different actors at different stages (see Figure 1). Four stages can be identified in the use of AI weapons: development, deployment, application, and feedback. At first three stages, developers, commanders, and operators are the main actor in charge, respectively, while the last stage involves all three actors.

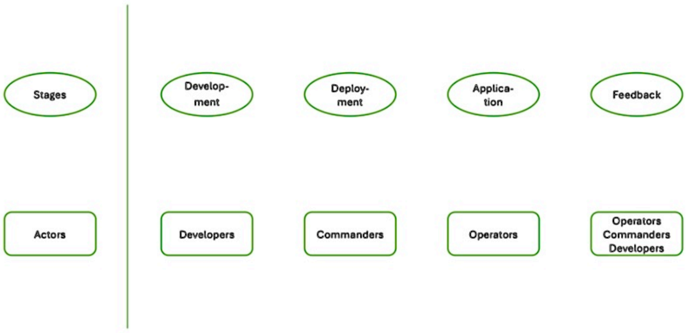

Figure 1. The stages and corresponding actors

3.1. Developers

3.1.1. Algorithm opacity

If the developer deliberately conceals the system logic, it directly violates the ethical obligation of technical transparency. According to Kantian ethics, technical design must comply with the absolute moral law that 'human rationality can review', and concealing behavior is equivalent to deception.

3.1.2. Willful violation and intentional setting

If the developer knowingly proceeds with the research and development despite knowing that the system may indiscriminately kill civilians (for example: setting the code to attack all moving targets), it violates the 'distinction principle' of the Geneva Convention (distinguishing combatants from civilians), and constitutes an 'intentional design' that actively violates international law, and is liable for war crimes.

3.1.3. Code issues leading to actual harm

If the system causes large-scale civilian casualties due to design flaws (for example: misidentifying hospitals as military targets), it should be held accountable based on the consequences.

3.1.4. Negligence

If the developer fails to test the system risks (such as not simulating attack scenarios), resulting in subsequent operators being unable to control the harm, they shall be responsible for the foreseeable catastrophic consequences.

3.1.5. Response lag

Failure to properly improve after problems occurred during use

3.2. Operators

3.2.1. Failure to control issues in time

If an operator fails to promptly shut down the system when he/she notices that it has gone out of control (for example, AI starts attacking civilians), it violates the professional obligation of 'Protecting Lives'. According to the Duty Theory, the operator has the responsibility to ensure that the use of technology complies with humanitarian principles. Failure to act is considered dereliction of duty.

3.2.2. Failure to predict harm

If the operator fails to assess the scope of harm before deployment (for example, does not set fire control parameters), it violates the "Prudence Obligation". The Duty Theory emphasizes that the operator, as the executor of technology, must actively avoid the violation of rules.

3.2.3. Responsibility for expanding harm

If the operator could have reduced casualties by 50% through control but fails to act, then the operator's choice leads to 'additional avoidable deaths', and the operator should bear responsibility for these consequences.

3.2.4. Weighing risks and benefits

If the operator allows the system to mistakenly kill civilians due to the desire to quickly end the battle, it can be regarded as prioritizing "military efficiency" over "value of human life", and is considered immoral because the overall harm far exceeds the benefits.

3.3. Commanders

3.3.1. Legality of pre-deployment assessment

If the commander uses the system to attack only clearly military targets, his/her behavior is legal. However, if the system itself is illegally designed (such as indiscriminate bombing), even if the commander is unaware, he/she still needs to bear the joint liability for "using illegal weapons".

3.3.2. Indulging civilian casualties

If civilian deaths far outweigh the military value and violate the "proportionality principle" of the Geneva Conventions (imbalance between military gains and civilian costs), it constitutes willful war crimes. If there are alternative options with lower casualties (such as special forces' raids instead of AI bombings), and the commander still opts for the high-risk system, he/she shall be held responsible for the avoidable deaths.

4. Legal and moral responsibility

The above content has discussed under what circumstances each of the three responsible parties should bear responsibility. Then, how to distinguish the legal liability and moral condemnation situations of the three responsible entities?

4.1. Developers

4.1.1. Legal liability

If there are design flaws in the AI system, the developer shall bear relevant legal responsibilities. (For example: identifying civilians as targets, identifying hospitals as military bases). If the system is deliberately developed to violate international laws of war (such as autonomously attacking hospitals), it may constitute war crimes.

4.1.2. Legal basis

All countries' "Product Liability Laws" stipulate that products cannot have dangerous defects. International law also prohibits the development or deployment of fully autonomous lethal weapon systems.

4.1.3. Ethical responsibility

When developing weapon-grade AI, should safety measures be set in advance (for example: installing an emergency stop switch for the AI)? Should the risk of technological abuse be considered? If the development continues despite knowing that the technology may be abused, it is like selling knives to terrorists.

4.1.4. Examples

A company that wrote code for drones deliberately deleted the instruction 'prohibit attacking schools', resulting in civilian casualties. The developer shall bear legal responsibility.

4.2. Operators

4.2.1. Legal liability

Using an AI weapon without testing will lead to accidents. The operator shall bear legal responsibility (for example: bypassing safety checks and deploying directly).

4.2.2. Legal basis

International treaties require that new weapons must undergo humanitarian reviews, just like conducting experiments before the launch of new drugs.

4.2.3. Ethical responsibility

It depends on whether the operator ensures that humans can take over the AI at any time (for example: setting a 'manual confirmation is required every three attacks'). If letting the AI kill by itself is done for convenience, it is equivalent to shirking the responsibility for human judgment.

4.2.4. Examples

If the operator fails to check whether the combat machine can distinguish between soldiers and civilians, this is a systemic failure and the operator shall bear legal responsibility.

4.3. Commander

4.3.1. Legal liability

When discovering that AI is committing indiscriminate killing without giving orders to stop, the commander must bear legal liability.

4.3.2. Legal basis

The Geneva Conventions clearly stipulate that the commander must prevent war crimes.

4.3.3. Moral responsibility

The commander should not overly rely on AI decisions. Even if AI suggests 'bombing the entire building', the commander should prioritize protecting civilians (for example: switch to a smaller-scale attack).

4.3.4. Examples

When a drone suggests bombing a house with civilians inside, and the commander approves it without consideration, this is a typical dereliction of duty and should bear legal liability.

5. Conclusion

5.1. Summary

This article has proposed a contextualized responsibility framework for allocating legal and moral accountability in the use of military AI systems. By systematically distinguishing between developers, operators, and commanders, and analyzing their respective responsibilities under different scenarios, this framework addresses the shortcomings of existing literature, which often oversimplifies accountability or conflates legal and moral obligations. The refined framework ensures that human oversight remains central to lethal decision-making, preventing autonomous weapons from becoming a gray area of liability exemption.

5.2. Refinement of the framework

To further enhance this framework, future research should:

1. Empirical Validation: Test the framework against real-world cases of AI military deployments to assess its applicability and identify potential gaps.
2. Dynamic Contexts: Explore how varying levels of battlefield urgency, technological transparency, and human-machine interaction influence responsibility allocation.

3. International Consensus: Investigate pathways for global legal standards to harmonize accountability mechanisms across nations.

References

[1] Calo, R. (2017). Legal Liability for AI Decisions. University of Washington Law Review, 94(3), 1-28.

[2] Scharre, P. (2018). Army of None: Autonomous Weapons and the Future of War. W.W. Norton & Company.

[3] International Committee of the Red Cross (ICRC). (2021). Autonomous Weapon Systems under International Humanitarian Law. Geneva: ICRC Publications.

[4] Ekelhof, M. (2019). Moving Beyond Semantics on Autonomous Weapons: Meaningful Human Control in Operation. Global Policy, 10(3), 343-348.

[5] Sharkey, N. (2012). The Evitability of Autonomous Robot Warfare. International Review of the Red Cross, 94(886), 787-799.

[6] Sparrow, R. (2007). Killer Robots. Journal of Applied Philosophy, 24(1), 62-77.

[7] Walsh, T. (2022). *Machines Behaving Badly: The Morality of AI*. MIT Press.

[8] Schmitt, M. N., & Thurnher, J. S. (2013). Out of the Loop: Autonomous Weapons and the Law of Armed Conflict. Harvard National Security Journal, 4(2), 231-281.