Balancing Power and Ethics: A Framework for Addressing Human Rights Concerns in Military AI

Mst Rafia Islam[1], Azmine Toushik Wasi[2]
[1]Independent University, Bangladesh [2]Shahjalal University of Science and Technology, Bangladesh
2030391@iub.edu.bd, azmine32@student.sust.edu
Both authors contributed equally to the project

1 Introduction

AI has advanced significantly in recent years, resulting in a wide range of applications in both the civilian and military sectors. The military, which is continually seeking innovations for more effective, faster, and stronger technology or weapons, sees AI as a perfect answer to address these needs [8, 11]. AI technologies offer many advantages, such as increased operational efficiency, precision targeting, and reduced human casualties [18, 14]. However, these advancements come with profound ethical and legal concerns, particularly regarding potential human rights violations [3, 14, 6].

Use of AI in military applications raises serious concerns [12, 1, 11]. Autonomous weapons systems, capable of making life-or-death decisions without human intervention, threaten the right to life and violate international humanitarian law. While AI-enhanced surveillance systems improve data collection, they can lead to widespread privacy violations and unnecessary monitoring of individuals [11]. Additionally, inherent biases in AI systems may result in discriminatory practices, exacerbating existing disparities and violating principles of equality, non-discrimination, and human rights [3, 14].

Motivated by these concerns, in this work, we propose a novel three stage framework to address human rights issues in the design, deployment, and use of military AI, with each aspect consisting of multiple components. Our work aims to propose a comprehensive framework for evaluating human rights violations by military AI, addressing both technical and statutory aspects. By examining the ethical implications and regulatory challenges, this framework seeks to provide a balanced approach to harnessing the benefits of AI in military operations while safeguarding human rights.

2 Framework

As shown in Figure 1, this framework addresses human rights concerns in the Military AI lifecycle across three key phases: **Design, Deployment, and Use**, with each phase having several components. Below, we discuss each phase and its components in detail.

2.1 Design Phase

Design phase involves creating military AI systems, focusing on algorithms and functionalities. It sets the foundation for AI behavior, ensuring it operates correctly, without bias and discrimination.

Targeting Errors of Autonomous Weapons. AI-powered autonomous weapons systems can make life-and-death decisions without human intervention [4]. AI systems can make errors in identifying and targeting individuals, leading to unintended casualties and injuries among civilians [2]. Implementing ethical guidelines in the design and development of military AI systems can ensure compliance with human rights standards. Moreover, incorporating human oversight in AI decision-making processes ensures accountability and ethical compliance. Human-in-the-loop (HITL) [13] systems allow human operators to intervene and override AI decisions when necessary.

Bias and Discrimination. AI systems can perpetuate or even exacerbate existing biases, leading to discriminatory practices in targeting or profiling based on race, ethnicity, or other characteristics [10]. By developing algorithms that detect and mitigate biases in AI systems can prevent discriminatory outcomes. This not only enhances the fairness of AI applications but also helps in building public trust and ensuring compliance with ethical standards.

2.2 In Deployment Phase

In Deployment phase involves integrating military AI systems into specific environments, focusing on their usage while emphasizing ethical considerations and legal compliance.

2024 Harms and Risks of AI in the Military (HRAIM 2024).

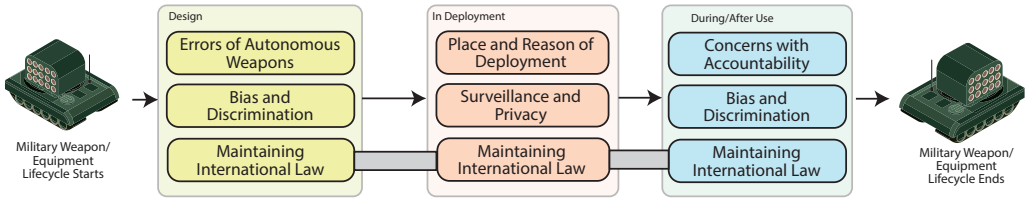

Figure 1: Our framework for addressing human rights concerns in military AI

Surveillance and Privacy. AI-driven surveillance systems can infringe on individuals' privacy rights, leading to unauthorized monitoring and data collection [7]. Implementing strong data privacy and security measures to protect sensitive information collected and processed by military AI systems. This includes encryption, access controls, and regular security assessments.

Deployment of the weapon. To ensure maximum efficacy while minimizing collateral damage [19], the automated weapon system must be deliberately positioned in a location that maximizes operational range and precision. In order to prevent unintended harm, this entails a detailed investigation of the terrain, potential targets, and civilian areas [5]. Continuous monitoring and changes are required to ensure its efficacy and safety. Additionally, real-time data analysis and feedback loops should be integrated to adapt to dynamic battlefield conditions, ensuring that the system remains responsive and minimizes risks to non-combatants [2].

2.3 During/After Use Phase

During/After Use phase encompasses the actual operation of military AI systems during conflicts or missions. It assesses the AI's performance in real-time, monitoring for accountability and compliance with human rights and legal standards.

Concerns with Accountability. Transparency of AI decision-making processes complicates accountability, making it difficult to attribute responsibility for human rights violations [15]. When AI systems make decisions, it creates a gap in justice and responsibility, as it becomes challenging to hold any individual or entity accountable for wrongful actions [16]. By ensuring AI systems are transparent and their decision-making processes are explainable, accountability issues can be resolved. Ethical guidelines, such as the Asilomar AI Principles [9], emphasize the importance of transparency, accountability, and human oversight in the development and deployment of military AI. Adhering to these principles can help mitigate the violations of human rights.

Bias and Discrimination. AI systems can perpetuate or worsen existing biases, leading to discriminatory practices in targeting or profiling based on race, ethnicity, or other characteristics [7]. To mitigate these biases post-deployment, continuous monitoring and auditing of AI systems are essential. Implementing feedback loops to identify and correct biased outcomes can ensure fairness [17]. This approach addresses biases and helps build public trust in AI technologies by demonstrating a commitment to ethical standards and accountability.

2.4 Violation of International Humanitarian Law

Deployment of AI in military operations poses serious risks to International Humanitarian Law (IHL), which aims to protect non-combatants. Autonomous weapons can make decisions without human oversight, leading to indiscriminate attacks that violate the principle of distinction [2]. AI targeting decisions may also compromise proportionality, as these systems might not effectively balance military advantage with civilian harm [5]. The lack of transparency and accountability in AI decision-making complicates compliance with IHL, hindering the investigation and attribution of unlawful actions. Adherence to IHL requires strong ethical guidelines, oversight mechanisms, and international cooperation to prevent violations and uphold the laws of war.

3 Challenges and Future Work

Several challenges may hinder the proper implementation of our framework. There is no global agreement on the ethical use of military AI. Groups like the Campaign to Stop Killer Robots push for a ban, but leading AI nations resist binding commitments. Accountability is also complex, as it's hard to determine who is responsible when AI systems cause harm—creators, commanders, or the AI itself. Additionally, AI evolves quickly, outpacing legal systems and creating oversight gaps that could be exploited. Future research should focus on establishing international agreements and fostering collaboration between engineers, ethicists, and legal experts to keep pace with AI advancements.

References

[1] David Adam. Lethal ai weapons are here: how can we control them? *Nature*, 629(8012):521–523, April 2024.

[2] Peter Asaro. On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making. *International Review of the Red Cross*, 94(886):687–709, June 2012.

[3] Charles Bradley, Richard Wingfield, and Megan Metzger. National artificial intelligence strategies and human rights: A review. Technical report, Global Partners Digital and Global Digital Policy Incubator, Stanford Cyber Policy Center, April 2020. Research assistance from Madeline Libbey and Amy MacKinnon.

[4] Edward Hunter Christie, Amy Ertan, Laurynas Adomaitis, and Matthias Klaus. Regulating lethal autonomous weapon systems: exploring the challenges of explainability and traceability. *AI and Ethics*, 4(2):229–245, February 2023.

[5] Rebecca Crootof. The killer robots are here: Legal and policy implications. *Cardozo law review*, 36:1837, 2014.

[6] EU. Eu artificial intelligence act. Technical report, European Union, 2024.

[7] Steven Feldstein. *The global expansion of AI surveillance*, volume 17. Carnegie Endowment for International Peace Washington, DC, 2019.

[8] Sophie-Charlotte Fischer. *Military AI Applications: A Cross-Country Comparison of Emerging Capabilities*, page 39–55. Springer International Publishing, 2022.

[9] Future of Life Institute. Asilomar ai principles, August 2017. Coordinated by FLI and developed at the Beneficial AI 2017 conference.

[10] Faith Gordon. Virginia eubanks (2018) automating inequality: How high-tech tools profile, police, and punish the poor. new york: Picador, st martin's press. *Law, Technology and Humans*, page 162–164, November 2019.

[11] Sarah Grand-Clément. Artificial intelligence beyond weapons: Application and impact of ai in the military domain, 2023.

[12] Tshilidzi Marwala. Militarization of ai has severe implications for global security and warfare, July 2023.

[13] Eduardo Mosqueira-Rey, Elena Hernández-Pereira, David Alonso-Ríos, José Bobes-Bascarán, and Ángel Fernández-Leal. Human-in-the-loop machine learning: a state of the art. *Artif. Intell. Rev.*, 56(4):3005–3054, April 2023.

[14] Daragh Murray. Adapting a human rights-based framework to inform militaries' artificial intelligence decision-making processes. *St. Louis University Law Journal*, 68, 2024.

[15] Claudio Novelli, Mariarosaria Taddeo, and Luciano Floridi. Accountability in artificial intelligence: what it is and how it works. *AI Soc.*, 39(4):1871–1882, August 2024.

[16] Frank Pasquale. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press, January 2015.

[17] Inioluwa Deborah Raji, Andrew Smart, Rebecca N White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, and Parker Barnes. Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. 2020.

[18] Adib Bin Rashid, Ashfakul Karim Kausik, Ahamed Al Hassan Sunny, and Mehedy Hassan Bappy. Artificial intelligence in the military: An overview of the capabilities, applications, and challenges. *International Journal of Intelligent Systems*, 2023:1–31, November 2023.

[19] Torsten Stein. Collateral damage, proportionality and individual international criminal responsibility. In *International Humanitarian Law Facing New Challenges*, pages 157–161. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.