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

Mst Rafia Islam, Azmine Toushik Wasi

Introduction

AI advancements can bring improved **efficiency, precision**, and **reduced casualties** in military operations.

But, AI in military use risks human rights violations through **autonomous decision-making, privacy breaches**, and **discriminatory biases** in systems.

Motivated by these concerns, we propose a **three-stage** framework to address human rights issues in **the design, deployment**, and **use** of military AI.

(Murray, 2024; Rashid et al., 2023)

Some parts of the presentation may include graphics of weapons or military equipment, intended solely for clarity and context.

Thank you for your understanding.

Motivation

Rapid AI adoption in the military raises concerns about **ethics** and **international law** compliance.

AI systems may **perpetuate biases and violate human rights**, including the **right to life and privacy.**

Lack of **transparency** and **accountability** complicates justice and oversight.

Autonomous Systems in Military

AI-based autonomous military systems function by processing data through ML algorithms to **make decisions or perform tasks without direct human control**. They rely on sensors and real-time data inputs to **perceive their environment, analyze the situation**, and **determine optimal actions**, typically using deep learning, reinforcement learning, and computer vision.

AI risks in autonomous systems include **decision errors** like **misidentifying targets, biases from flawed training data, lack of accountability due to opacity,** and **vulnerabilities to hacking or adversarial attacks**.

Our Framework

Design

Focuses on **developing algorithms** and **functionalities** to ensure military AI operates correctly, without bias or discrimination.

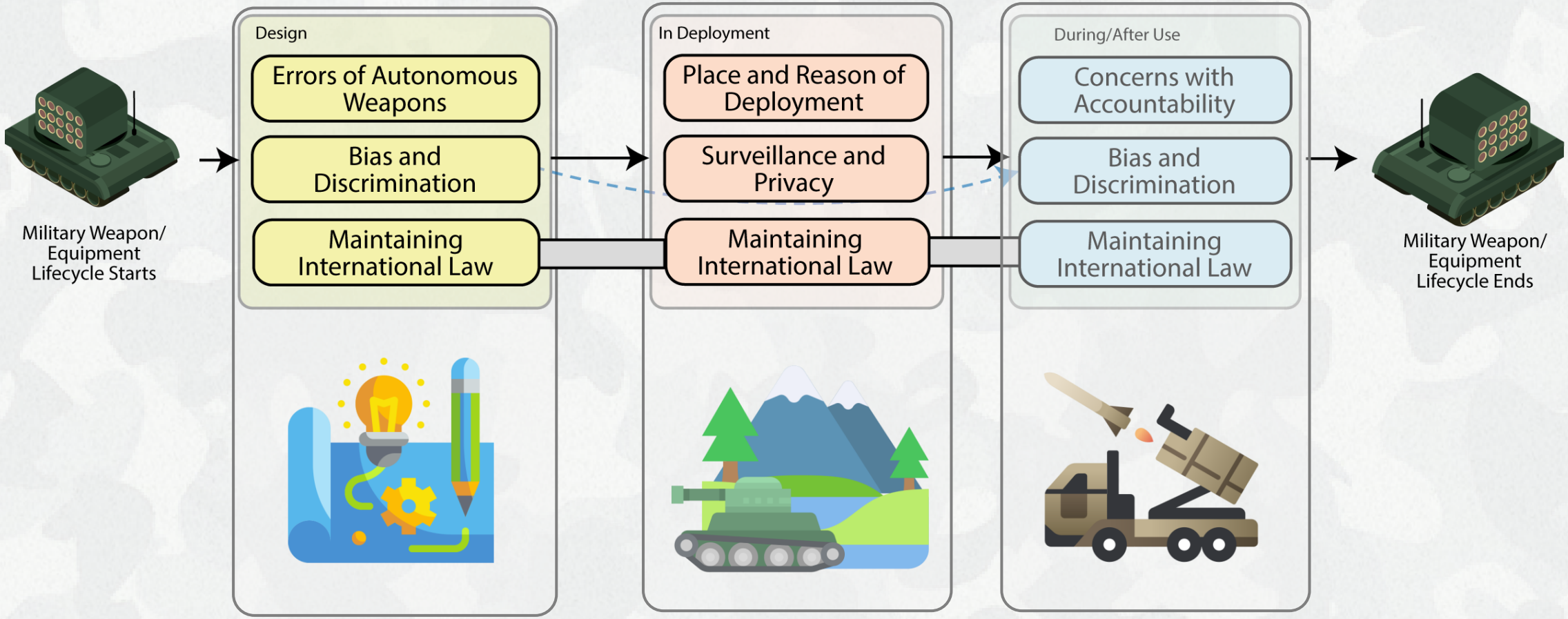
In Deployment

Checks if military AI systems are **deployed correctly** in specific environments, maintaining **ethical** considerations and **legal** compliance.

During/After Use

Involves the use of military AI systems during conflicts or missions, ensuring **accountability**, **human rights**, and **legal** compliance.

Our Framework



Design Phase

Design Phase

Design Phase ensures military AI systems are **unbiased, accountable**, and **ethical** by integrating **fairness, transparency**, and **human oversight.** This requires diverse training data, robust ethical frameworks, and transparent processes to mitigate bias, **uphold humane values**, prevent harm, and build trust.

Errors of Autonomous Systems

Errors in autonomous systems arise from **flawed perception**, **decision-making**, or execution due to sensor inaccuracies, biased data, or unforeseen conditions. Solutions include algorithms like **YOLO for perception**, **RL and MCTS for decision-making**, and **MPC for execution.**

Bias and Discrimination

Bias in autonomous military systems arises from **skewed data or flawed algorithms**, causing unfair outcomes. Solutions include **adversarial debiasing, fairness metrics, data balancing**, and XAI tools like SHAP for transparent and equitable decision-making.

(Wray, 2021; Ali et al. 2024)

Design Phase

Example:

"They wanted to allow us to attack [the junior operatives of hamas] automatically. That's the Holy Grail. Once you go automatic, target generation goes crazy."
– referring to an IDF Officer

The IDF's AI system, Lavender, identified 37,000 potential targets in Gaza and was linked to 15,000 deaths during the war's early weeks, as reported by +972 Magazine, Local Call, and The Guardian. This raises ethical concerns over its use in conflict.

Yuval Abraham, +972 Magazine (formed by four Israeli writers in Tel Aviv), 2024

Deployment Phase

Deployment Phase

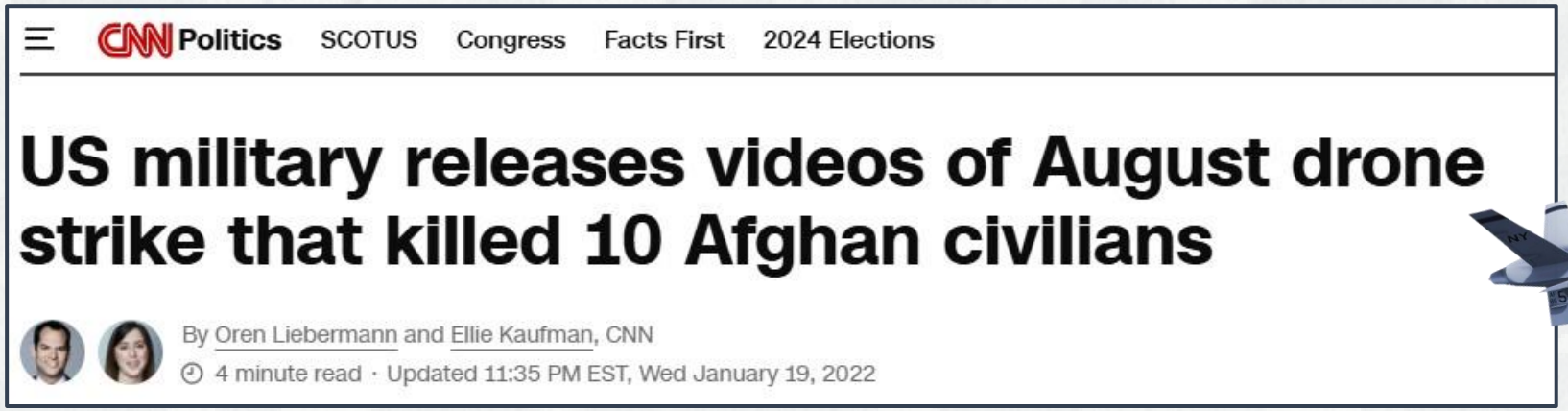
Deployment phase ensures that the AI-based military system is deployed in **proper ground** (with no possible issues of bias and harms by AI integration), **ethical AI use** and **legal compliance , which** is crucial for **responsible**, **humane applications**, such as reducing collateral damage in critical operations.

Place and Reason of Deployment	AI systems should be deployed in **environments** suited to their **capabilities and legal compliances** and there is **no possible ground for discrimination** against any specific group. Also, should **avoid high-stake grounds** to avoid hard consequences.
Surveillance and Privacy	AI-powered surveillance systems must **balance security needs** with individual **privacy** rights. Ethical deployment ensures that surveillance is conducted transparently, minimizing intrusive monitoring while safeguarding personal freedoms and confidentiality.

(Putro et al. 2024)

Deployment Phase

Example:



U.S. intelligence **mis-identified** Zemari Ahmadi, an Afghan aid worker, as an **ISIS-K operative**, misinterpreting his routine activities, like loading water canisters into a Toyota Corolla, as preparations for an attack. **Rushed verification and faulty analysis** led to a drone (MQ-9 Reaper) strike that killed Ahmadi and nine others, all civilians, near a suspected ISIS compound.

CNN, New York Times - 2021, 2022

During/After Use

During/After Use

Use phase requires **continuous oversight** and **monitoring** of military AI to ensure ethical decision-making and prevent misuse, which is essential to **safeguard human lives**, **maintain accountability**, and ensure that AI operates within legal and moral boundaries, especially in high-stakes situations.

Concerns with Accountability	Concerns with accountability in AI arise when it's unclear **who is responsible for decisions made by autonomous systems**. Ensuring clear chains of responsibility, transparency in decision-making, and the ability to trace actions back to human oversight is must to maintain trust and prevent unintended harm.
Bias and Discrimination	While using military AI, biases can lead to **unfair targeting or decisions**, harming vulnerable groups. Regular audits and updates are essential to **prevent discrimination** and ensure ethical outcomes.

(Maathuis, 2024)

Use Phase

Example:

The New York Times

A.I. Drone May Have Acted on Its Own in Attacking Fighters, U.N. Says

A United Nations report suggested that a drone, used against militia fighters in Libya's civil war, may have selected a target autonomously.

The drone, a Kargu-2, was used as soldiers tried to flee, the report said.
"Once in retreat, they were subject to continual harassment from the unmanned combat aerial vehicles and lethal autonomous weapons systems," according to the report, which was written by the U.N. Panel of Experts on Libya and released in March.

New York Times - 2021

International Humanitarian Law

Military AI systems pose significant risks to the protection of civilians under **International Humanitarian Law (IHL)** and the **Geneva Conventions**.

How existing IHL principles apply to military AI:

Distinction

AI systems must be capable of distinguishing between **combatants and civilians**, and between **military objectives and civilian objects**.

Proportionality

AI systems must be designed to ensure attacks **don't cause excessive civilian casualties** or damage.

Necessity

The use of force, even if AI-controlled, **must be necessary** for achieving a **legitimate military objective**.

(Putro et al. 2024)

Challenges

- ✓ Accountability, Control, and Explain ability
- ✓ Lack of Willingness
- ✓ Security and Reliability Challenges
- ✓ Ethical and Legal Issues
- ✓ Over-Reliance on AI
- ✓ Need for International Frameworks

(Schraagen, 2023; Putro et al. 2024; Cîrdei, 2024)

EU AI Act

Not enough!

- × **Regulatory Void:** While the **EU AI Act** imposes stringent requirements on civilian AI, it **explicitly excludes military AI used for defense**, leaving a regulatory gap that raises concerns about their ethical and legal use.
- × **Dual-Use Technology Risks:** Many AI systems with **civilian applications can also serve military purposes**, but their military deployment remains unchecked under the current framework.
- × **Challenges in Governance:** National-level regulation may be **insufficient** given the transnational impact of military AI, highlighting the need for coordinated international oversight.
- × **Need for Unified Guidelines:** There is **lack of a comprehensive framework** which connects the AI Act's principles to dual-use and military AI, ensuring ethical and responsible development in defense contexts.

(Fanni, 2023; Csernatoni, 2024; EDRi, 2024)

How Our Framework Can Help?

Why this framework is effective?

- ✓ Life-cycle Coverage
- ✓ Avoiding Discrimination in All Stages
- ✓ Maintaining IHL
- ✓ Ethical Deployment
- ✓ Operational Accountability
- ✓ Continuous Oversight

Computational Intelligence and Operations Laboratory

Thank You!

If you have any questions, please ask!

Scan to read the full paper!

```
@misc{islam2024balancingpowerethicsframework,
    title={Balancing Power and Ethics: A Framework for Addressing Human Rights Concerns in Military AI},
    author={Mst Rafia Islam and Azmine Toushik Wasi},
    year={2024},
    eprint={2411.06336},
    archivePrefix={arXiv},
    primaryClass={cs.CY},
    url={https://arxiv.org/abs/2411.06336}, }
```