Academic Year	Module	Assessment Number	Assessment Type
2024	Concepts and Technologies of AI	2	Report

“ AI: Balancing Innovation with Ethical Integrity: Opportunites and Challenges across various Fields”.

Student name: Diwash Adhikari

Section: L5CG1

Student ID: 2407736

Module Leader: Siman Giri

Tutor: Siman Giri

Abstract

The swift adoption of AI in military applications has substantially altered the landscape of modern warfare, raising critical ethical concerns. This report investigates the moral implications of AI in autonomous weapons systems, focusing on accountability for predictions made by AI in conflict zones and the risk of unintended harm. It also explores national and international initiatives, including the ethical guidelines developed by the European Commission for trustworthy AI and the recommendations issued by UNESCO on AI governance, aimed at establishing solid ethical frameworks to govern AI use in military cases. The research also highlights the difficulties introduced by AI-driven supervision, intelligence gathering, and the potential growth of conflicts due to autonomous decision-making. A generic framework is proposed to address these issues, emphasizing human control, ethical design principles, and global cooperation to align AI technologies per societal norms and human rights. This report concludes with a generic framework and discussion aimed at securing AI ethical creation and use in military systems.

Table of Contents

1. Introduction

Ethics in AI refers to the principles that direct the behavior of artificial intelligence to reflect and uphold human values (SAP, 2024). However, the swift integration of AI into various industries has introduced numerous ethical challenges to ensure its beneficial and responsible use. Some primary ethical issues include bias, privacy breaches, job displacement, deepfakes, and exploitation of intellectual property (Watters, 2023). These difficulties highlight the necessity of strong ethical frameworks to direct AI development and deployment.

To tackle and mitigate the ethical challenges associated with AI, several key principles and guidelines have been established globally. Core ethical principles encompass fairness, aiming to ensure that AI systems do not reinforce biases, and transparency, which enables a clear understanding of AI operations (European Commission, 2019); accountability, ensuring AI-driven decisions, and privacy in safeguarding individual data. Guidelines from bodies like the European Council framework for reliable AI and UNESCO's directives provide actionable frameworks for developing AI models that respect human rights and societal values (UNESCO, 2021).

Ethical AI should possess essential features like fairness, transparency, and privacy. Fairness guarantees that AI systems operate without bias, while transparency allows users to understand AI models. Accountability ensures that developers are responsible for their AI system's actions, a point emphasized by the European Commission (2019). Privacy is another critical feature of ethical AI, which safeguards user data to ensure it is used responsibly, as highlighted by UNESCO (2021). Developers should

incorporate ethical principles into AI design to align the system with human and societal values. This involves comprehensive testing, detailed documentation, and compliance with legal standards. Additionally, involving diverse teams and consistently evaluating AI systems after deployment is essential for upholding ethical standards.

2. AI and Modern Warfare:

The deployment of AI in Autonomous Weapons Systems poses critical ethical challenges, especially concerning responsibility and the potential for unintended harm. Autonomous weapons often claim to enhance civilian protection through precision targeting, but they fall short in practice. Autonomous weapons systems face prominent challenges in accurately distinguishing between combatants and civilians, precisely in fluid and complex environments of conflict zones. This inability derives from the AI algorithm's limitations in interpreting complex human behaviors and context-specific scenarios, increasing the risk of targeting errors and collateral damage (Marijan, 2022). This raises questions about transparency when an AI-driven system causes harm. The uncertainty around who bears responsibility – the developers, military operators, or commanding authorities complicates legal and moral considerations. The potential for unindented harm is high, as autonomous systems may malfunction or misinterpret data, causing civilian damage and raising conflicts. Marijan emphasizes the need for strict regulations and human oversight to prevent misuse and diminish risks. To ensure ethical use, nations must prioritize international cooperation, establish robust guidelines, and hold stakeholders accountable for developing and deploying these systems.

AI-enabled surveillance and intelligence gathering raise serious problems in balancing national security with individual privacy. While these technologies can enhance security by improving data preprocessing and threat detection, they also suggest significant privacy risks, including mass surveillance and potential civil freedoms violations. Marijan highlights that nations have an ethical responsibility to deploy such systems transparently, with strict monitoring and safeguards to protect privacy. This includes ensuring legal compliance and minimizing damage by limiting surveillance scope and restricting access to sensitive data. To ensure ethical use, nations must establish legal frameworks, hold developers and users accountable, and promote international cooperation to prevent misuse. A balanced approach is needed to safeguard security and individual freedoms.

AI has the potential to heighten conflict, especially in military contexts, due to the increased use of autonomous weapons systems. One ethical dilemma surrounding these technologies is the possibility of Autonomous weapons systems acting independently, without human intervention, in high-stakes military situations.These systems are designed to select and engage target autonomously, may create decisions that are not alinged with human ethical standards, leading to unintended consequences such as targeting error. The risk of miscalculation increases when AI-driven systems are delployed in dynamic combat environments,where the uncertainity of human behavioiur and rapidly changing circumstances can challenge the decision-maling capabilites of machines.In this context, the absence of huamn judgment and emotional intelligence could result in catastrophic mistakes, escalating conflicts and causing unnecessary destruction. To overcome these risks, it is important to implement strict oversight, regulation, and clear accountability

mechanisms to guarantee that AI technologies are deployed responsibly and ethically in military settings.

3. Generic Framework

To guarantee that AI models are developed ethically, they must operate under human supervision and avoid causing harm in any context. AI systems should be designed to act responsibly in public spaces, respecting societal norms and human well-being. Developers are ethically obligated to be transparent about their AI's purpose, processes, and limitations, ensuring the technology is accessible and understandable to shareholders. Transparency builds trust and mitigates potential biases in predictions. Moreover, even in private organizations, AI research work should be shared with the public to foster collaborative improvements and avoid ethical setbacks.

Following guidelines like transparency, robustness, and accountability is paramount. Developers should ensure users can comprehend how AI models make decisions, maintaining clarity in their processes. The system must be secure and safe, minimizing the risk of malfunction. Finally, accountability lies within developers and users, who must take responsibility for the outcomes and consequences of the AI system they create and deploy.

4. Discussion

Building ethical AI is essential to ensure technology serves humanity responsibly, especially in high-stakes applications like autonomous weapons and surveillance, where the consequences of errors and misuse can be dangerous. AI models often struggle to interpret complex human behavior and adapt to increasingly changing environments, highlighting the irreplaceable value of human supervision and judgment in preventing unintended harm. Establishing a transparent legal framework, implementing strict accountability mechanisms, and fostering international cooperation are crucial to addressing risks such as privacy violations, system biases, and potential misuse. Developers, policymakers, and stakeholders carry a profound ethical responsibility to inset robustness, fairness, and transparency principles into AI systems, ensuring they prioritize societal well-being over mere efficiency. By fostering trust, protecting human rights, and aligning technological advancements with core ethical values, ethical AI enables sustainable progress, where innovation supports safety, justice, and the shared benefit of all humankind.

5. Bibliography

SAP. (2024, August 9). *What is AI ethics?* From SAP: https://www.sap.com/resources/what-is-ai-ethics

Watters, A. (2023, November 16). 11 Common Ethical Issues in Artificial Intelligence.

European Commission. (2019). *Ethics guidelines for trustworthy AI*. European Commission.

UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. UNESCO.

Marijan, B. (2022, November 28). *Autonomous Weapons: The False Promise of Civilian Protection*. From CIGI: https://www.cigionline.org/articles/autonomous-weapons-the-false-promise-of-civilian-protection/