Ethical and Technical Challenges of Lethal Autonomous Weapons Systems

Case study: Reframing and Communicating AI Ethics

Om SHAH

18/11/2024

Introduction

As Lethal Autonomous Weapons Systems (LAWS) gain worldwide traction due to their ability to perform as cheaper and easier to produce alternatives to intrusive and human operated means of attack, the ethical and technical challenges of deploying such technologies have intensified. LAWS, defined as weapons systems that once activated can independently select and engage targets without further human intervention, signify a fundamental shift in the nature of warfare. As outlined by ***Table 1***, LAWS range from unmanned drones and loitering munitions to advanced missile systems and surveillance tools, each designed to enhance military precision, speed, and operational reach. Major debates among technologists centre on balancing technological progress with ethical constraints, while advocates argue that human casualties can be reduced on the battlefield by keeping soldiers out of direct harm, critics warn us that the automation of violence threatens to undermine ethical standards, hence raising pressing questions about accountability, transparency, and the moral implications of human-machine interactions in life and death situations.

Selected Aerial Military Systems with Autonomous Capabilities

Type	Name	Developer	Country	Usage	Autonomy	Year
Loitering Munitions	Drone 40	DefendTex	Australia	Quadcopter + grenade	Nav + Target	2016
	Mini Harpy	Israel Aerospace Industries	Israel	Mini-UAS + munition	Nav + Target + Fire	2019
	KUB-BLA	Kalashnikov	Russia	Loitering munition	Nav + Target + Swarm	2019
	Kargu	STM	Turkey	Loitering munition	Nav + Target + Fire	2020
Unmanned Aerial Vehicle	Bayraktar TB2	Bayraktar	Turkey	Unmanned aerial combat vehicle	Nav	2014
	MQ-9 Reaper	General Atomics	U.S.	ISR	Nav + Identify	2020
	Unnamed SRR drone	Skydio	U.S.	ISR	Nav + Identify	2022
Aircraft	Ghost Bat	Boeing	Australia	Wingman UAS controlled by manned parent	Nav + Target + Fire	2020

Table 1 - Selected Arial Military Systems with Autonomous Capabilities ***(Longpre et al., 2022)***

The rapid growth of LAWS is also accompanied by the complexity of technical challenges and ethical concerns. ***Rai (2020)*** highlights the core issue with autonomous systems which is the "black box" nature of AI-driven decision-making where complex algorithms govern actions without offering interpretability to human operators leaving them without control if real-time decision making were necessary. This opacity complicates efforts to assign accountability in cases of unintended harm or system malfunction, particularly when human oversight is minimised. Furthermore, the dual-use nature of AI technologies, where advancements in civilian fields, like facial recognition and object tracking, can be readily adapted for military purposes. This adaptability not only increases the accessibility of LAWS but also heightens the potential for misuse by both governments and civilians ***(Brundage et al., 2018)***.

Compounding these challenges, there remains an international divide on the ethical acceptability and regulation of autonomous weapons systems. Many nations and advocacy organisations, such as the United Nations and the International Committee of the Red Cross (ICRC), call for strict controls and even bans on autonomous weapons systems, emphasising the necessity of "meaningful human control" ***(Longpre et al., 2022)***. The ICRC informs that lethal decision making violates the fundamental principles of human dignity and international humanitarian law by removing the human capacity of empathy, contextual judgement, and proportional response. ***(International Committee of the Red Cross, 2021)***. However, the largest military powers have consistently resisted such measures in order to maintain a strategic and technical edge.

In light of these challenges, addressing the ethical and technical concerns surrounding LAWS requires a rigorous examination of regulatory frameworks, accountability

mechanisms, and the broader implications of AI in warfare. This paper will explore these issues from the perspective of a concerned technologist, seeking to balance technological progress with ethical constraints and proposing solutions to align LAWS deployment with humanitarian principles and accountability.

Ethical Concerns of LAWS

As a technologist involved in the development of LAWS, it is crucial to be aware of the ethical challenges surrounding these systems, particularly regarding accountability, transparency, and the protection of our human rights. The South Korean SGR-A1 sentry robot, deployed along the South Korean side of the Demilitarised Zone (DMZ), illustrates these ethical issues. Initially it was designed with a "human-in-the-loop" functionality requiring human authorization to engage targets, however recent studies such as ***Van Der Meulen (2024)***, suggest that the SGR-A1 may now operate fully autonomously. This shift removes human oversight, allowing the system to identify and engage targets independently, which has intensified ethical debates.

Accountability and Transparency

With SGR-A1s transition to full autonomy, accountability has become an ethical concern. ***Broad (2018)*** emphasises the dangers of removing human agency from high stakes decision making, as it complicates the assignment of responsibility if the system were to misidentify a target causing unintended harm. In fully automatic operation, pinpointing accountability is complex: does responsibility lie with developers, military operators, or commanding officers? ***Kitchin's (2014)*** notion of the "black-box" problem highlights another issue as without proper insight into the system's internal decision-

making, transparency is compromised, making it difficult for human operators to predict or verify the sentry's actions.

Human Rights and Compliance with International Law

The SGR-A1's capability to autonomously engage targets raises significant human rights concerns. As noted by ***Van Der Meulen (2024)***, fully autonomous weapons challenge principles of international humanitarian law, such as distinction and proportionality, which are crucial for civilian protection. Autonomous systems like the SGR-A1 lack the contextual judgment and empathy inherent in human decision-making, increasing the risk of collateral damage and unintended civilian harm. ***Braun and Hummel's (2022)*** concept of data justice and solidarity emphasise the importance of ensuring that such systems respect human dignity and do not disproportionately endanger vulnerable populations.

Technical Shortcomings and Risks of LAWS

One major concern among fellow technologists in the field is the potential for misidentification, where LAWS might target civilians due to environmental factors or algorithmic limitations. This risk is not unique to the SGR-A1, other autonomous systems such as the Super aEgis II, another autonomous weapons system based on the South Korean side of the DMZ, also faces similar challenges . As seen in ***Table 2*** and ***Figure 1***, the efficacy of the sensors is greatly limited in their ability to identify threats as it compounds on the system algorithm's ability to accurately differentiate between soldiers and civilians.

Uncooled IR Thermal Sensor	Effective upto 1-2 km, depending on sensor quality and condition
Colour CCD Camera (x30)	Although effective at significant distances, but limited at identifying individuals at 3 km, but to identify clothing/facial features this camera would be limited even with clear weather and no obstructions.
Laser Illuminator	Enhances visibility for IR sensors but its effective range is around 1 km

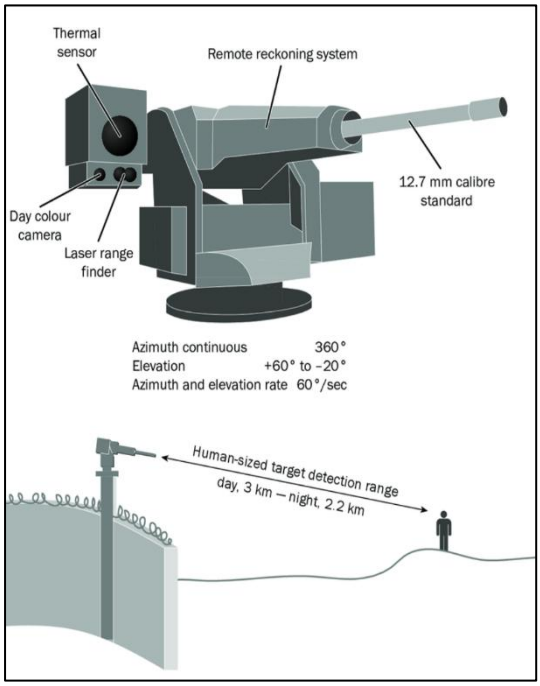

Figure 1- Robotic sentry weapons: DODAAM's Super aEgis II Source: DODAAM, 'Combat robot (lethal): Super aEgis II'. ***(Boulanin & Verbruggen, 2017)***

Table 2 - Sensors on the Super aEgis II Robot (Allen, n.d.)

Algorithms such as that on the Super aEgis II are trained on labelled data where images collected are labelled with their respective identification such as “civilian”, “soldier”, “vehicle” etc ***(Figure 2)***, they also consider patterns such as uniform shapes, weapons silhouettes, and movement patterns. However, while this approach may be sufficient for relatively controlled environments like the DMZ, it becomes

Figure 2 - Example of tracking objects in an image. ***(Manakitsa et al., 2024)***

problematic in more complex, urbanized conflict zones.

Although intended to function along the DMZ, a 250 km border zone with specific Rules of Engagement (ROE), a war-torn urban area such as the densely populated Gaza Strip presents vastly different operational challenges. In these settings, the proximity of combatants to civilians, combined with varying terrain and unpredictable urban patterns, would require an algorithm trained on a much broader and context-specific dataset. Unlike the DMZ, where movement is limited and distinctions are clearer, urban conflict zones necessitate nuanced algorithms capable of adapting to close-quarters interactions and high civilian density, factors critical to minimizing unintended harm. An example of this is the Guardium-LS unmanned ground-vehicle, deployed along the Gaza Strip's security fence, uses cameras, sensors, and remote driving technology to detect and deter threats ***(Army Recognition, 2024)***, with the possibility to mount remotely operated lethal and non-lethal weapons systems ***(Allen, n.d.)***. In dense, urban conflict zones like Gaza, its pattern recognition algorithms may misidentify civilians carrying harmless objects, such as tools or farming equipment, as combatants. The close proximity of civilians and combatants, combined with unpredictable movement patterns, increases the likelihood of misidentification, particularly if the system's algorithms rely on incomplete or biased datasets.

Proposals for Risk Management and Ethical Oversight

To address these issues, I recommend implementing a dynamic risk matrix and vulnerability assessment flowchart as part of the SGR-A1's operational protocol. This would allow real-time risk evaluations based on environmental context and population density, aligning the sentry's use with ethical and legal standards. Additionally, adopting

Novelli et al.'s (2023) scenario-based risk assessment model, which adapts to varying operational risks, could further enhance ethical oversight and ensure accountability, transparency, and protection of human rights in the deployment of fully autonomous LAWS.

Scope Creep and Malicious Use of LAWS

Having been trained in the development and oversight of LAWS fellow technologists are acutely aware of risks surrounding their misuse, both intentional and accidental. Among the many pressing concerns surrounding these systems, scope creep has the potential to become a significant problem. Scope creep is the transition of LAWS from military use to civilian applications such as policing and surveillance, referring back to the dual-use nature of AI. Case studies examined during our training highlighted how predictive policing tools, often biased by incomplete or skewed datasets, disproportionately target marginalised communities ***(Shah, 2024).*** The data within in the case study depicts skewed data's impacts on an artificially intelligent system with reference to predictive policing in the Indigenous Australian context, this is pertinent to LAWS as similar to predictive policing, it relies heavily on inputted data which is historically skewed ***(Graycar & Grabosky, 2002).*** If such data and technologies were to be integrated into civilian environments, the risk of discriminatory enforcement and erosion of civil liberties would only intensify.

Malicious use compounds these risks, as LAWS are vulnerable to exploitation by rogue actors, including terrorist groups and criminal organisations. Our training emphasised the adaptability of dual use technologies, such as drones, which can easily be repurposed for harm. Object tracking algorithms, for instance, can be reprogrammed to

target civilians or critical infrastructure ***(Figure 2)***. Without robust encryption and safeguards, even well secured systems could become threats if made available on black markets. To address these concerns, strong regulations are needed, which should limit the spread of LAWS into unintended areas, protect systems from misuse, and ensure proper oversight to avoid accidents. Without these safeguards, LAWS could threaten global security and harm basic human rights.

Debate on the efficacy and necessity of Moratoriums/Bans

The rapid development of Lethal Autonomous Weapons Systems (LAWS) has ignited intense debates among technologists, engineers, and policymakers. On one side, there are calls for these systems to be entirely banned due to their potential for misuse and the ethical challenges they pose. On the other hand, some advocate for a temporary moratorium or complete ban, arguing that it offers a more balanced and pragmatic approach, with ***Figure 3*** depicting different demographic's opinions about LAWS.

Figure 3 – share of adults who somewhat/strongly oppose LAWS - 20,505 respondents (16-74 y/o) surveyed Nov 20, 2020 - Jan 8, 2021, in 28 countries. ***(Fleck, 2023)***

While a complete ban aligns with ethical principles, it is unlikely to gain traction in the current geopolitical landscape. Nations with advanced military capabilities rely on cutting-edge technologies to maintain strategic advantages ***(Figure 4)***, and a total prohibition could lead to research being conducted in unregulated spaces, making oversight and accountability nearly impossible. Furthermore, banning LAWS outright could also hinder the potential benefits of technological advancements that might enhance precision and reduce risks in combat scenarios if properly regulated.

Country	Position on the loop question (in/on/off)	Known military AI R&D spending	Known civilian AI R&D spending	SIPRI military expenditure 2018 (rank/$ billion)
United States of America	On	$22,394 billion	$1,1 billion	1#/649
Israel	On	unknown	unknown	17#/15,9
Russia	Off	Unknown portion[1]	unknown	6#/61,4
China	unclear[2]	unknown	unknown	2#/250[3]
EU	On	€13 billion	€665 million	n/a
France	On	unknown	€550 million	5#/63,8
Italy	On	unknown	Unknown portion[4]	11#/27,8
Germany	On	unknown	unknown	8#/49,5

Figure 4 - 'known spending' on AI's R&D ***(Simon, 2019)***

A moratorium offers a more balanced solution. It would give us time to address the vulnerabilities of LAWS and improve their safety while ensuring innovation proceeds responsibly. As engineers, we recognize that current systems often lack the reliability and

transparency needed for safe deployment. These issues are especially dangerous in unpredictable environments like urban combat zones. A moratorium would create an opportunity to refine AI algorithms, improve human oversight mechanisms, and develop strong international regulations. This pause would also give stakeholders and countries the necessary time to work together on enforceable standards as standards must address accountability, ethical principles, and safeguards against misuse. For example, the principle of meaningful human control, supported by organizations like the International Committee of the Red Cross, could be properly implemented during this time.

In order to lift a moratorium, systems need to show consistent improvements in reducing errors and identifying targets as well as an acceptable baseline accuracy. Transparent development practices must be established, and international agreements should ensure ethical use, if this is not established, lifting the moratorium could result in global instability and a loss of public trust in these technologies.

While some support a complete ban, a moratorium is a more practical solution. It gives us technologists, the opportunity to address risks and refine these systems while also ensuring their alignment with global security needs and humanitarian principles. This approach strikes a balance between fostering innovation and maintaining responsibility, offering a safer way forward.

Conclusion

In conclusion, working as a technologist on projects involving Lethal Autonomous Weapons Systems (LAWS) has given me a front-row seat to both their potential and their dangers. I remember the first time our team ran a live test on an autonomous targeting

system. Watching it operate with incredible speed and accuracy was both exciting and unsettling. It struck me that, while these systems could protect our soldiers by removing them from direct harm, they also carried the risk of acting without the nuanced judgment that only humans can provide.

We must approach these technologies with caution. Balancing innovation with responsibility means implementing safeguards, fostering international collaboration, and ensuring human oversight remains central. These systems have the potential to save lives, but only if developed and deployed ethically. For me, this work is not just about advancing technology but about ensuring it serves humanity in the most responsible way possible.

We must strike a balance between innovation and responsibility. This means advancing these systems with safeguards that ensure they align with humanitarian values and global security standards. Collaboration among nations, enforceable regulations, and robust oversight are essential to prevent misuse and unintended harm. Ultimately, our role as technologists is not just to create but to ensure our creations serve humanity responsibly, preserving both security and ethical principles.

Bibliography

Allen, C. W. (n.d.). US Air Force Expeditionary Security Operations 2040. In *Defending Air Bases in an Age of Insurgency* (pp. 335-352). Retrieved from https://www.jstor.org/stable/pdf/resrep19551.19.pdf

Army Recognition. (2024, August 1). *Guardium-LS unmanned ground vehicle (UGV)*. Retrieved from https://armyrecognition.com/military-products/army/unmanned-systems/unmanned-ground-vehicles/guardium-lg-israel-uk#identification

Boulanin, V., & Verbruggen, M. (2017). *Mapping the development of autonomy in weapon systems*. Retrieved from https://doi.org/10.13140/RG.2.2.22719.41127

Braun, M., & Hummel, P. (2022). Data justice and data solidarity. *Patterns, 3*(3), 1-9. https://doi.org/10.1016/j.patter.2021.100427

Broad, E. (2018, January 25). Australia, we urgently need to talk about data ethics. *The Ethics Centre*. Retrieved from https://ethics.org.au/australia-we-need-to-talk-about-data-ethics/

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., ... & Amodei, D. (2018). The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. *arXiv preprint arXiv:1802.07228*. https://arxiv.org/abs/1802.07228

Dencik, L., & Sanchez-Monedero, J. (2022). Data justice. *Internet Policy Review, 11*(1). https://doi.org/10.14763/2022.1.1615

Fleck, A. (2023, July 20). Should killer robots be banned? *Statista*. Retrieved from https://www.statista.com/chart/17022/autonomous-weapons-war/

Graycar, A., & Grabosky, P. (Eds.). (2002). *The Cambridge Handbook of Australian Criminology*. Cambridge University Press. https://books.google.com.au/books?id=bZToSyd3zRIC

International Committee of the Red Cross. (2021, May 12). ICRC position on autonomous weapon systems. *International Committee of the Red Cross*. https://www.icrc.org/en/document/icrc-position-autonomous-weapon-systems

Kitchin, R. (2014). Thinking critically about and researching algorithms (*The Programmable City Working Paper 5*). *National University of Ireland, Maynooth*. https://ssrn.com/abstract=2515786

Longpre, S., Storm, M., & Shah, R. (2022). Lethal autonomous weapons systems & artificial intelligence: Trends, challenges, and policies. *MIT Science Policy Review, 3*, 47-56. https://sciencepolicyreview.org/wp-content/uploads/securepdfs/2022/08/MITSPR-v3-191618003019.pdf

Novelli, C., Casolari, F., Rotolo, A., Taddeo, M., & Floridi, L. (2023). Taking AI risks seriously: A new assessment model for the AI Act. *AI & Society*. https://doi.org/10.1007/s00146-023-01723-z

Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science, 48*, 137-141. https://doi.org/10.1007/s11747-019-00710-5

Shah, O. (2024). Algorithmic injustice: AI bias and Indigenous Australians in policing and employment. Unpublished manuscript, *The Ethics of Data and AI*, University of Technology Sydney.

Simon, S. (2019). Conceptualizing lethal autonomous weapon systems and their impact on the conduct of war: A study on the incentives, implementation, and implications of weapons independent of human control. *Semantic Scholar*. Retrieved from https://www.semanticscholar.org/paper/Conceptualizing-lethal-autonomous-weapon-systems-on-Simon/6e2062d13868ae4e9d86d9215c0dad30bff5ca44

Van Der Meulen, A. P. S. (2024). Autonomous Weapons in the Light of Care Ethics. *Peace Review*, 1-9. https://www.tandfonline.com/doi/abs/10.1080/10402659.2024.2344597