The Ethics of Engineering Autonomous Systems in Defence

Jason Alexander Chan a1867657

Computer Science 7237: Concepts in Artificial Intelligence and Machine Learning,
University of Adelaide, South Australia.

Contributing authors: a1867657@adelaide.edu.au;

Abstract

The surge in Autonomous Systems (AS) technology in the defence sector will change the nature of warfare. Today, countries are pouring billions into their defence budgets to develop next generation AS for the future of war. This investment is precipitated by the highest levels of geopolitical tensions since the Cuban Missile Crisis. More specifically, there are concerns about the proliferation of a subclass of AS, which are Autonomous Weapon Systems (AWS). Meanwhile, global regulations on AWS have not kept up despite public sentiment and the huge body of academic literature on topic calling for a comprehensive ban. Thus, AS in defence are a practically inevitable. Attention and energy should instead turn to ensuring that defence companies develop AS with a strong understanding of the ethical implications of their work. This paper conducts a literature review on the ethics of AS in defence by examining academic journals, Australian Government policy, and white papers by engineering professional bodies. This paper concludes that a significant gap exists between the discussion in academia and the public domain vs. engineering practice. A framework to bridge this gap is proposed to aid defence companies with assessing the ethical rigour of their business and engineering practices with respect to AS.

1 Introduction

Between 2017-21 the United States, China, Russia, South Korea and the European Union spent a combined estimate of USD 35.8 billion on military drones [1]. In 2022, the United States alone will spend USD 8.2 billion on Uncrewed Vehicle (UxV) systems [2]. The United Nations has failed to introduce regulations on AWS despite successive attempts in the last eight years. Countries disagree about the definitions of an AS vs. AWS, what regulation to apply, and the impossibility of the ban itself since the technology is dual purpose (civilian and military) [3]. Meanwhile, the first lethal autonomous attack has already occurred in 2020 in Libya. Turkish Kargu 2 loitering munition attack drones *"were programmed to attack targets without requiring data connectivity between the operator and the munition: in effect a true fire, forget and find capability."* [4].

This paper discusses the ethics of AS and AWS in defence in four sections.

- Section 2: Summary of the Ethical Risks
- Section 3: Summary of the Current Solutions
- Section 4: Discussion about the Effectiveness of the Current Solutions
- Section 5: Recommendations to Defence companies working on AS

1.1 Definitions in this Paper

This paper relies on the following definitions.

- System: the combination of elements that functions together to produce the capability required to meet a need [5].

- Autonomy: the ability of a system to achieve goals while operating independently of external control. Requires self-directedness (to achieve goals). Requires self-sufficiency (operate independently). Autonomy is not Artificial Intelligence but may use it. Autonomy is not Automation but often relies on Automation. [5].
- Automation: Automation is not self-directed but instead requires command and control (e.g. set of pre-planned set of instructions). A System can be automated without being autonomous [5].
- Artificial Intelligence (AI): Software techniques such as machine learning, perception, search, probabilistic methods, classification, neural networks etc. [5].
- AS (AS): Any robotic or software system that exhibits autonomy.
- Autonomous Weapon System (AWS): Robotic or software systems that selects and attack targets without human intervention. Subset of AS.

2 Ethical Risks

This section outlines the ethical risks of AS and AWS in defence.

2.1 Lack of Meaningful Human Control

The human needs to be responsible for the outcomes of AS systems. The United Nations states that "Accountability cannot be transferred to machines." [6]. The concept of meaningful human control can be classified into five levels [7].

a. Humans deliberate about specific targets before initiating an attack
b. Humans choose from a list of targets suggested by a program
c. Programs select the calculated targets and needs human approval before attack
d. Programs select calculated targets and allocate humans a time-restricted veto before attack
e. Programs select calculated targets and initiate attacks **without human involvement**

Having humans in the loop is a necessary but insufficient condition for realising meaningful control. Consider the cases of Malaysia Airlines Flight 17 in 2014 and Ukrainian International Airlines Flight 752 in 2020, which were both shot down due to human error [8]. A human in the loop is insufficient - the boundary of interactions between a human operator and any system needs careful design to ethically minimise inadvertent harm.

2.2 Conflict Escalation

Automated reactions raise the risk of escalating conflicts. Machine Systems can perform calculations at the rate of machine speeds and this may be beyond the reaction time for a human to intervene. Consider the case of two combatant AS, each of which automatically reacting to the other in an escalation feedback loop. Even "when a single System is predictable, or even deterministic, which such Systems interact with other Systems or ... in large swarms... their collective behaviours can become intrinsically unpredictable" [9]. Learning Systems further "compound the problem of predictable use".

2.3 Obfuscated and Non-attributable Attacks

Obfuscated and non-attributable attacks are another ethical risk of AS. "States have legal obligations to make attacks practically attributable"[9]. Arguably, cyber warfare attacks may be considered AS. Consider the 2011 revelation about the Stuxnet computer worm which damaged an Iranian nuclear facility. Stuxnet autonomously spread to hundreds of thousands of computers around the world to reach its target. Such attacks "do carry a risk of collateral damage, with a risk of political blowback if the attacking parties are identified" [10]. Consider, on the other hand, non-kinetic attacks such as Russia's political interference in the 2016 United States election and again in 2020 with automated social media bots. Even though no physical harm was suffered it is arguable that misinformation campaigns intended to sow discord are unethical use of automation technologies.

2.4 Unregulated AWS Proliferation

AS technology is already proliferating. Commercial off-the-shelf (COTS) drones can easily be re-purposed for military uses. COTS drones, for example, are currently used by both sides in the

2022 war in Ukraine [11]. The components for AS themselves are also commercially available and cheap to obtain. As at 2020, Australia, China, Israel, Russia, South Korea, Turkey, the United Kingdom, the United States are investing in AS for defence [12]. More countries will follow. AWS systems will inevitably develop as a consequence of AS investment.

2.5 Human Rights Abuse

There is an alarming risk that AS are used by "Tyrants and despots...to gain or retain control over a population which would not otherwise support them. AS might be turned against peaceful demonstrators when human law enforcement might not do the same."[9]. Military technology can be exported, stolen or sold to unaligned actors, authoritarian states, private militaries, domestic police, terrorist groups.

3 Current Solutions

3.1 Introduce AWS Global Regulations

There are three positions on global regulation of AWS. The first is do nothing because International Humanitarian Law is sufficient, the second is a comprehensive ban on AWS, and the third is to develop legal norms requiring and defining "meaningful human control" [13]. The body of literature overwhelmingly calls for the second or third option.

The latest attempt for a comprehensive ban of AWS stalled yet again in December 2021 at the United Nations because it was blocked by the United States and Russia. According to Scharre (2019), a ban on AWS is impossible [14]. Since a comprehensive ban is still unreachable, "a majority of states pushed for an international treaty to be developed" which suggests that after eight years of stalled discussions, regulation on AWS may be approaching an inflection point. An international treaty, however still needs defined terms.

3.2 Standardise Definitions to Support Regulation

A standardised definition of AWS and AWS is imperative if regulation is to be passed and enforced. In 2014, the International Committee of the Red Cross (ICRC) (2014) concluded that there was no international agreed definition of an AWS and proposed that AWS is defined as "weapons that can independently select and attack targets" [13].

In 2016, the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems asserted that "confusions about definitions regarding... artificial intelligence, AS, and autonomous weapons systems stymie more substantive discussions about crucial issues" [9]. They expanded ICRC's AWS definition to "any weapon system with autonomy in its critical functions...that can select and attack targets without human intervention".

But as at 2020, definitions are still problematic. The IEEE lament that there is no "precision in definitions about autonomous weapons and human control". They recommend that these definitions should be "grounded in technical realities of today... not limited by disagreements over what technologies may or may not be... farther into the future" [6].

Bradshaw (2013) on the other hand, strongly condemns the effort of standardising definitions because the fundamental problem for lawmakers is their lack of understanding of systems theory in general.

> "The attempt to define autonomy has resulted in a waste of both time and money spent debating and reconciling different terms and may be contributing to fears of unbounded autonomy. The definitions have been unsatisfactory because they typically try to express autonomy as a widget or discrete component, rather than a capability of the larger system enabled by the integration of human and machine abilities"[3].

3.3 Increase Public Awareness

Generating public awareness creates public discussion and elevates the topic of AS in defence. Launched in 2013, the public movement stopkillerrobots.org, now undersigned by 26 Nobel Peace Laureates and counting 180 organisations as its members, seeks to " rejects the automation of killing and instead promotes the principle of human control over emerging technologies."

In 2015 an open letter on artificial intelligence was signed by physicist Stephen Hawking,

CEO Elon Musk and over leading 150 academics in artificial intelligence. It asked whether AWS can be made to "comply with humanitarian law", whether an agreed definition for AWS can actually be developed to facilitate regulation, how AWS could be integrated with existing military systems and finally, outlined a list of ethical risks [15].

The problem with stoking public pressure, however, is that AS in defence naturally provokes fear due to inherent biases in the public imagination about dystopian robotic futures. Rosendorf (2022), for example surveyed 1006 US citizens on the question of the ethics of military strikes resulting in collateral damage and found that 'increasing weapon autonomy is associated with ... lower perception of ethicality' than 'conventional inhabited and remote-controlled systems'. The authors note that the public may equate 'unethical' with the novelty of the technology and having limited experience with autonomous machines [16].

Interestingly, Devitt conducted a survey of 2500 Australian respondents and found that "Australians have low trust in AI systems but generally 'accept' or 'tolerate' AI. They found that Australians trust research institutions and Defence organisations the most to use AI and trusted commercial organisations the least." [17].

3.4 Introduce National AWS Regulation

Even if multi-lateral regulation are unsuccessful, countries can unilaterally propose and adopt their own AWS regulations. Australia is a laggard. In 2017, Australia at the United Nations stated that it considers a sweeping prohibition of AWS to be premature [17]. In 2021, Australia at the United Nations acknowledged that International Humanitarian Law applies to AWS and that "states...[must] undertake necessary weapons reviews during the study, development and adoption of such systems". Interestingly, "Australia also believes that AWS have the potential to uphold International Humanitarian Law compliance". Australia proposed an AWS code of conduct or good practice guideline. [18].

Australia doesn't yet have an overarching governance framework for Artificial Intelligence in Defence. However, it is taking the first steps toward one with its allies. [17].

> "Australia is a founding partner in the US's AI Partnership for Defense (PfD) that includes Canada, Denmark, Estonia, France, Finland, Germany, Israel, Japan, the Republic of Korea, Norway, the Netherlands, Singapore, Sweden, the United Kingdom, and the United States. In doing so, Australia has aligned its AI partnerships with AUKUS, five-eyes (minus New Zealand), the Quad (minus India) and ASEAN via Singapore. In particular Australia is seeking to increase AI collaboration with the US and UK through AUKUS."

3.5 Self-Regulate AS in the Defence Industry

In the absence of global and national regulation, the Australian Defence industry has proactively sought to develop its own guidelines for the ethical development of Defence AI. In 2021, the Australian Department of Defence released its report titled "A method for Ethical AI in Defence". It proposes an Ethical AI checklist and risk matrix for defence companies working on AI products in defence. The report also recommended the inclusion of a "Legal and Ethical Assurance Program Plan" for any future defence contracts involving complex Defence AI systems. The report's limitation, however, is that it is only "a report on the outcomes of a workshop ... and doesn't represent and official position of Defence" [19].

Recommendations and standards from the civilian domain could be adopted by the defence industry. However, even civilian standards are lacking. Standards Australia, which is the "peak non-government, not-for-profit standards organisation" released its 2020 Artificial Intelligence Standards Roadmap. It recommends that Australia needs to increase its representation on global AI standards and acknowledged that "we need to promote Australia's security interests,...trade and investment agenda [and] Australia's evolving values". This highlights the impossibility of divorcing defence and civilian AI technologies. [20].

3.6 Uplift Ethical Engineering Design Techniques

Despite the lack of global and national regulations and even industry guidelines, Engineers can bear the responsibility to proactively apply ethical engineering design techniques when developing AS for defence. Verdiesen et al. (2018) ran an

online survey to solicit impressions on the importance of six variables (type of mission, outcome, type of weapon, type of character, number of characters and location) for the requirements of a Moral Machine for AWS. Their paper proposed this survey to evaluate a "grounded view on the perception of the moral acceptability of AWS of the wider public" [21].

Umbrello (2019) proposed a Value Sensitive Design framework for AWS that "intends to embed stakeholder values into design, encourage stakeholder cooperation and coordination, and promote social acceptance of [AWS] as a preferable future fact of war". The approach requires "conceptual, empirical and technical investigations' to determine values of the stakeholders, envision how those values can be construed as design requirements, and evaluate how those design requirements can be supported" [22]. Umbrello asserts that programming AWS with existing military conventions that define Rules of Engagements and Laws of War are "are more than sufficient to govern battlefield action" [22]. Umbrello clearly demonstrates that they do not understand the challenge software robustness and monumental effort of assuring that an AS will comply with that encoding in a highly unstructured and dynamic battlefield environment.

The IEEE 7000-2021 *Standard Model for Addressing Ethical Concerns During System Design* could be considered by the Defence Industry. It claims to be a " highly practical approach to minimising potential value harms associated with product or systems design". It is a Value-Based Engineering approach which is broadly equivalent to Umbrello's Value Sensitive Design framework.

3.7 Review and Revise the Company's Code of Ethics

Modern businesses have a code of ethics that guide their business activities. Similarly, engineering professional bodies have a ethical code of ethics. Companies ought to review and revise their Business and Engineering Code of Ethics because the technology and responsibilities for AS in defence are evolving at speed. Companies need to stay current.

> ... existing codes of ethics may fail to properly address ethical responsibility for AS, or clarify ethical obligation of engineers with respect to AWS. [9]

Finally, individual engineers working on AS for defence may not even know their ethical obligations. Thus it is incumbent on their companies to ensure they do.

4 Discussion

4.1 The Gap between Academia and Engineering Practice

This paper concludes that the effort poured into the academic discussions on the topic of AWS and AS ethics aren't yet adopted by defence engineering practice. Furthermore, the development of AS and AWS in defence is "accelerating largely outside of public and academic attention in the discipline of International Relations" [23]. There is currently no public data on the AS and AWS ethics policies of defence companies. Thus it remains beyond the scrutiny of academics.

Countries ultimately need to mandate AS and AWS ethical obligations on the defence companies working for them to drive practical outcomes. Australia's proposal to introduce "Legal and Ethical Assurance Program Plan" for any future defence contracts involving complex Defence AI system would be supremely effective because it imposes reporting obligations its Defence companies [19].

There is no discussion about when ethical reviews should be introduced in the engineering lifecycle, which governs the product development process from ideation to operation and disposal. The Technology Readiness Level (TRL) in Figure 1 is a useful framework to generalise the engineering lifecycle. Ethical design review gates should be introduced at TRL5 at a minimum and conducted at each subsequent TRL level. Any earlier is impractical because the technology is too immature to exist beyond the lab. Any later and its too late - at TRL8 the product already exists at scale with its deficiencies hard-wired into the design. It is too late at TRL9 when the system is in-service.

Finally, there is no discussion about how defence companies should translate the recommendations made by the literature review into action. This is an enormous gap. To actually drive

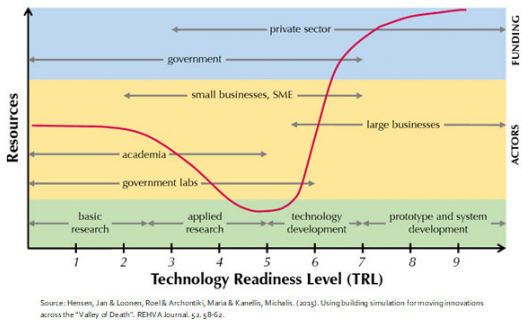

Fig. 1 The Technology Readiness Level

practical change, requires an understanding of how engineering companies operate. The key steps are:

1. Assign a company board member to be made accountable for ethical AS and AWS development.
2. Develop a policy on what ethical AS and AWS engineering means to that company.
3. Publish that policy to the engineering workforce and ensure they understand their ethical obligations.
4. Re-baseline that ethical policy at least annually.
5. Inject ethical review gates into the engineering life cycle.
6. Introduce mechanisms for engineers to report ethical issues. Two important mechanisms are independent ethics officers and a centralised reporting system that catalogues when ethical concerns are closed.

4.2 Definitions for AS and AWS are Desirable

Despite the views of Bradshaw (2013) claiming that "[attempting] to define autonomy has resulted in a waste of both time and money", standardised definitions for AS and AWS are desirable.

At a minimum, defence companies working on AS and AWS need to adopt and make definitions consistent internally within their companies as the first step. This is essential to communicate effectively about AS and AWS design. Definitions can even be adopted from civilian standards such as IEEE 7007-2021: Ontological Standard for Ethically Driven Robotics and Automation Systems.

It is pragmatic to adopt the IEEE's recommendation that definitions are "grounded in technical realities of today... not limited by disagreements over what technologies may or may not be... farther into the future". This will help dialogue on the ethics of AS and AWS avoid the quagmire of speculative scenarios beyond the capabilities of today's technologies.

As robotics becomes more widely available in civilian industries, the exercise of standardising definitions will become easier. As at 2022, there are dozens of consumer facing robotic systems that will increasingly feature in our daily experiences. Segments include as robo-taxis (Waymo, Zoox, Cruise), home automation (iRobot's Roomba), industrial inspections (Boston Dynamics), agriculture (John Deere) to name a few. Discussions between regulators, technologies and end-users will be improved as a consequence.

4.3 Engineering Techniques are not Practical Enough

The Australian Department of Defence's AI ethical risk matrix is a sound approach. But the provided template is not practically useful because it is effectively a blank table. Three modifications are proposed to make the risk matrix more effective for assessing ethical risks in AS and AWS design.

1. Pre-fill risk matrix with the ethical risks outlined in section two of this paper. This takes the guesswork out of the exercise because its difficult for working engineers to conceive them. These baseline risks are unlikely to change. Ensure new risks can be added.
2. Add columns for risk score (before mitigation), a column for the proposed mitigation, and a column for estimated residual risk score (after mitigation).This aligns with conventional product safety engineering analysis.
3. Add columns to assign names to the risk owner, mitigation owner, and independent reviewer

In a first attempt to derive design requirements for a moral machine for autonomous weapons, Verdiesen et. al. (2018) conducted a large-scale study of the moral judgement of people regard are AWS deployment but no design requirements were actually derived. Instead their work only proposed

a survey to gather public opinion on the matter to inform requirements design.

Arguably, rather than refer to academic papers, Defence engineering companies are better off adopting practices outlined in IEEE 7000-2021: Standard Model Process for Addressing Ethical Concerns During System Design. This one of the world's first industry standards to offer guidelines on how to implement ethical concerns during system design. The standard is 82 pages and is reasonably comprehensive but would require investment in personnel to be comprehensively implemented. This would be impractical for start ups and small companies.

4.4 Keep Driving Public Awareness

Public awareness can influence the decision makers in the defence industry and the engineers working on its projects. Consider Google's Project Maven, whose aim was to help the Pentagon process the sheer volume of drone video data. Its functionalities included object detection, classification and alerts. More than 3100 Google employees signed an open letter to its CEO outlining their concerns [24]. Project Maven was not an AWS system [25].

5 Recommendations

At a minimum, Defence companies working on AS and AWS should adopt the following recommendations.

1. Assign a company board member to be accountable for ethical AS and AWS development
2. Publish an internal engineering policy on ethical AS and AWS development with a consistent set of internal definitions (see section 4.2). This policy must acknowledge the ethical risks of AS and AWS work for defence (See section 2.
3. Re-baseline that ethical policy at least annually with the evolving solutions proposed in AS and AWS global and national regulations, industry recommendations, academic journals, and military analysis (See section 3).
4. Inject ethical review gates into the engineering life cycle (See section 4.1).
5. Introduce a modified AI Risk matrix (See section 4.3) that is used at project kick off and at each subsequent ethical design review gate.
6. Introduce mechanisms for engineers to report ethical issues. Two important mechanisms are independent ethics officers and a centralised reporting system that catalogues when ethical concerns are closed.

6 Conclusion

Defence companies working on AS and AWS need to be aware of the ethical risks. A global race to develop AS and AWS in defence is already underway while geopolitical tensions are at their highest since the Cuban Missile Crisis. Thus, AS and AWS in defence are a practically inevitable. Despite the fact that global regulations are bogged down and national regulations absent, there is enough publicly available content on the topic for Defence companies working on AS and AWS to be proactive about ethical engineering. Ethical engineering of AS and AWS is possible. This paper summarised the AS and AWS ethical risks, the current solutions, discussed their advantages and disadvantages, and finally proposed a recommendation for defence companies can ethically engineer AS and AWS. Defence companies developing AS and AWS need a strong understanding of the ethical implications of their work.

Declarations

This article represents my own views and not that of my company.

References

[1] J. Haner and D. Garcia. The artificial intelligence arms race: Trends and world leaders in autonomous weapons development. *Global Policy*, 10:331–337, 2019.

[2] Association for Unmanned Vehicle Systems International. Department of defence unmanned systems budget report fiscal year 2022. 2022.

[3] Robert R.R. Johnson M. Woods D.D. Bradshaw, J.M. The seven deadly myths of "autonomous systems". *IEEE Intelligent Systems*, 2013.

[4] United Nations. Letter dated 8 march 2021 from the panel of experts on libya established pursuant to resolution 1973 (2011) addressed to the president of the security council. 2021.

[5] NASA Jet Propulsion Lab and Carnegie Mellon University Robotics. Workshop on autonomy for future nasa science missions. 2018.

[6] Conn A. Garcia D. Gill A. Llorens A. Noorma M. Heather Roff Bloch, E. Ethical and technical challenges in the development, use, and governance of autonomous weapons systems. *IEEE Standards Association*, 2020.

[7] N. Sharkey. Staying in the loop: Human supervisory control of weapons. *Autonomous Weapons Systems: Law, Ethics, Policy*, 2016.

[8] I. Bode and T. Watts. Meaningless human control: Lessons from air defence systems on meaningful human control for the debate on aws. *Centre for War Studies, University of Southern Denmark*, 2021.

[9] IEEE. Reframing autonomous weapons systems. *The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems*, 2016.

[10] J.P. Farwell and Rohozinski R. Stuxnet and the future of cyber war. *Global Politics and Strategy*, 53, 2011.

[11] T. Brewster. China's dji and its billionaire chief put in an awkward spot as both sides in ukraine war use its drones. *Forbes*, 2022.

[12] B. Stauffer. Country positions on banning fully autonomous weapons and retaining human control. *Human Rights Watch*, 2020.

[13] Red Cross Expert Meeting in Geneva. Autonomous weapon systems technical, military, legal and humanitarian aspects. 2014.

[14] P. Scharre. Army of none: Autonomous weapons and the future of war. 2019.

[15] Dewey D. Russell, S. and M. Tegmark. Research priorities for robust and beneficial artificial intelligence. *Open letter*, 2015.

[16] Smetana M. Rosendorf, O. and M. Vranka. Autonomous weapons and ethical judgments: Experimental evidence on attitudes toward the military use of "killer robots. *Journal of Peace Psychology*, 2022.

[17] S.K Devitt and D. Copeland. Australia's approach to ai governance in security defence. 2021.

[18] Australian Government. Australia's national statement. *United Nations 2021 6th Review Meeting on the Convention on Certain Conventional Weapons (CCW)*, 2021.

[19] Australian Department of Defence. A method for ethical ai in defence. 2021.

[20] Standards Australia. An artificial intelligence standards roadmap: Making australia's voice heard. 2021.

[21] Dignum V. Verdiesen, I. and I. Rahwan. Design requirements for a moral machine for autonomous weapons. *SAFECOMP Workshop 2018: International Conference on Computer Safety, Reliability, and Security*, 2018.

[22] S. Umbrello. Designing war machines with values. *Delphi: Interdisciplinary Review of Emerging Technologies 1 (2)*, pages 30–34, 2018.

[23] I. Bode and H. Huelss. Autonomous weapons systems and chaning norms in international relations. *Review of International Studies*, 44, part 3:393–413, 2018.

[24] P. Croft and H. Rijswijk. Negotiating 'evil': Google, project maven and the corporate form. *Law Technology and Humans*, 2 (1):75–90, 2020.

[25] R.H. Shultz and R.D. Clarke. Big data at war: Special operations forces, project maven and twenty first century warfare. *Modern War Institute, West Point*, 2020.