

The background of the cover is a composite image of Earth from space. The left side shows a bright, curved horizon of the planet with swirling white and grey cloud patterns over a dark blue ocean. The right side shows a dark, starry space with a dense, glowing spiral of golden-yellow city lights, representing a global network or risk. The overall color palette is dark with highlights of orange, yellow, and white.

AN ANTHOLOGY OF GLOBAL RISK

EDITED BY
SJ BEARD AND TOM HOBSON



<https://www.openbookpublishers.com>

©2024 SJ Beard and Tom Hobson

Copyright of individual chapters is maintained by the chapter's authors



This work is licensed under an Attribution-NonCommercial 4.0 International (CC BY-NC 4.0). This license allows you to share, copy, distribute and transmit the text; to adapt the text for non-commercial purposes of the text providing attribution is made to the authors (but not in any way that suggests that they endorse you or your use of the work). Attribution should include the following information:

SJ Beard and Tom Hobson (eds), *An Anthology of Global Risk*. Cambridge, UK: Open Book Publishers, 2024, <https://doi.org/10.11647/OBP.0360>

Copyright and permissions for the reuse of many of the images included in this publication differ from the above. This information is provided in the captions and in the list of illustrations. Every effort has been made to identify and contact copyright holders and any omission or error will be corrected if notification is made to the publisher.

Further details about CC BY-NC licenses are available at <http://creativecommons.org/licenses/by-nc/4.0/>

All external links were active at the time of publication unless otherwise stated and have been archived via the Internet Archive Wayback Machine at <https://archive.org/web>

Digital material and resources associated with this volume are available at <https://doi.org/10.11647/OBP.0360#resources>

ISBN Paperback: 978-1-80511-114-6

ISBN Hardback: 978-1-80511-115-3

ISBN Digital (PDF): 978-1-80511-116-0

ISBN Digital eBook (EPUB): 978-1-80511-117-7

ISBN XML: 978-1-80511-119-1

ISBN HTML: 978-1-80511-120-7

DOI: 10.11647/OBP.0360

Cover image: Javier Miranda, Alien planet, June 18, 2022, <https://unsplash.com/photos/nc1zsYGkLFA>

Cover design: Jeevanjot Kaur Nagpal

21. It Takes a Village: The Shared Responsibility of “Raising” an Autonomous Weapon

Amritha Jayanti and Shahar Avin

Highlights:

- Expectations around future capabilities of Lethal Autonomous Weapons Systems (LAWS) have raised concerns for military risks, ethics, and accountability. The UK’s position has attempted to address these concerns through a focused look at the weapons review process, human-machine teaming (or “meaningful human control”), and the ability of autonomous systems to adhere to the Rules of Engagement.
- Further, the UK has stated that the existing governance structures — both domestic and international — around weapons systems are sufficient to deal with concerns around the development, deployment, and accountability for emerging LAWS, with no need for novel agreements on the control of these weapons systems.
- In an effort to better understand and test the UK’s position on LAWS, the CSER ran a research project that interviewed experts in multiple relevant organisations, structured around a mock parliamentary inquiry of a hypothetical LAWS-related civilian death.

- The responses highlighted different conceptions of future systems, which were sometimes complementary but sometimes contradictory, as well as challenges and accountability measures. They have provided rich “on the ground” perspectives and highlight the very wide range of intervention points where humans are expected, and should be supported, to make decisions that enable legal, safe, and ethical weapon systems. These all need to be considered by any military that is considering acquisition and deployment of autonomous and semi-autonomous weapon systems.

This chapter was initially presented as a workshop paper in 2020. Using expert interviews and scenarios, the chapter provides an empirically informed account of the multiple points at which meaningful human oversight and control of autonomous weapons ought to be exercised. Similar methodological approaches are presented in several chapters of this volume, including 8, 16 and 14.

1. Introduction

With the increasing integration of digital capabilities in military technologies, many spheres of the public — from academics to policy-makers to legal experts to nonprofit organisations — have voiced concerns about the governance of more “autonomous” weapons systems. The question of whether autonomous weapons systems pose novel risks to the integrity of governance, especially as it depends so heavily on the concept of human control, responsibility, and accountability, has become central to the conversations.

The United Kingdom (UK) has posited that lethal autonomous weapons (LAWS), in their current and foreseeable form, do not introduce weaknesses in governance; existing governance and accountability systems are sufficient to manage the research, development, and deployment of such systems and the most important thing we can do is focus on improving our human-machine teaming. Our research project seeks to test this theory by asking: with the introduction of increasingly autonomous agents in war (lethal autonomous weapons/LAWS), are the current governance structures (legal, organisational, social) in fact sufficient for retaining appropriate governance and accountability in

the UK MoD? By attempting to confront strengths and weaknesses of existing governance systems as they apply to LAWS through a mock parliamentary inquiry, the project uncovers opportunities for governance improvements within Western military systems, such as the UK.

2. Background

Computers and algorithms are playing a larger and larger role in modern warfare. Starting around 2007 with writings by Noel Sharkey, a roboticist who heavily discusses the reality of robot war, members of the research community have argued that the transition in military technology research, development, and acquisition to more autonomous systems has significant, yet largely ignored, moral implications for how effectively states can implement the laws of war.¹ Segments of this community are concerned with the ethics of decision-making by autonomous systems, while other segments believe the key concern is regarding accountability: how responsibility for mistakes is to be allocated and punished. Other concerns raised in this context, e.g. the effects of autonomous weapon systems on the likelihood of war, proliferation to non-state actors, and strategic stability, are beyond the scope of this brief, though they also merit attention.

2.1 UK position on LAWS

The United Kingdom's representatives at the UN Group of Governmental Experts (GGE) on Lethal Autonomous Weapon Systems (LAWS) have stated that the UK believes the discussions should "continue to focus on the need for human control over weapon systems and that the GGE should seek agreement on what elements of control over weapon systems should be retained by humans".² The UK, along with other actors, such as the United States, believe that a full ban on LAWS could be counterproductive, and that there are existing governance structures in place to provide appropriate oversight over the research, development, and deployment of automated weapons systems:

...[T]he UK already operates a robust framework for ensuring that any new weapon or weapon system can be used legally under IHL. New weapons and weapons systems are conceived and created to fulfil a

specific requirement and are tested for compliance with international law obligations at several stages of development.³

The UK is also interested in a “technology-agnostic” focus on human control because it believes that it will “enable particular attention to be paid to the key elements influencing legal, ethical and technical considerations of LAWS”, as opposed to “debated definitions and characteristics” which, ultimately, may “never reach consensus”. The position emphasises that taking a “human-centric, through-life” approach would enable human control to be considered at various stages and from multiple perspectives. This includes across all Defense Lines of Development, the acquisition of weapons systems, and their deployment and operation. It is the UK’s position that the existing institutional infrastructure builds-in accountability measures throughout the weapon system lifecycle.

3. Methodology

In order to stress-test the governance and accountability structures that exist for UK weapon systems, and how they would apply to LAWS, we developed a hypothetical future scenario in which a UK LAWS kills a civilian during an extraction mission in Egypt. In order to ensure a level of feasibility and accuracy of construction, the scenario was built based on a wargaming scenario publicly published by RAND.⁴ We then ran a facilitated role-play exercise based on our modified scenario with an initial group of Cambridge-based experts. With their feedback and the lessons from the role-play, we developed the final version of the scenario which we then used in the research study (see Appendix).

This final iteration of the LAWS scenario was used to run a mock UK parliamentary inquiry through which we interviewed 18 experts across various areas of expertise, including (but not limited to) UK military strategy, military procurement, weapons development, international humanitarian law, domestic military law, military ethics, and robotics.

The interviews ranged from 45 to 120 minutes and explored a variety of questions regarding the case. The main objective of the interviews was to catalyse a meaningful discussion around what information the experts deemed important and necessary in order to decide who should

be held accountable in the aftermath of this scenario. A sample of the questions asked include:

- Who holds the burden of accountability and responsibility?
- What explanations and justifications for actions are needed?
- What information is necessary to come to a conclusion about the burden of accountability?
- Are there any foreseeable gaps because of the autonomy of the weapons systems?

The responses and dialogue of these 18 interviews were then reviewed and synthesized in order to develop a landscape of strengths and weaknesses of the current governance and accountability schemes related to UK institutions as they relate to LAWS, as well as recommendations on addressing any identified weaknesses. The full report is under preparation, but we are happy to share our preliminary key findings and recommendations below.

4. Key Findings

The main takeaway from the “inquiry”, from both a legal and organisational standpoint, was that assessing accountability is in the details. This contrasts with what we perceive as a dominant narrative of “meaningful human control”, which focuses mainly on human control, and the design of that interaction, at the point of final targeting action. The disconnect between the accountability across a weapon’s lifetime and the focus on final targeting decision was observed throughout the various expert interviews. “Meaningful human control” has become the *idée fixe* of domestic and international conversations for regulation of LAWS but it disadvantageously provides a limited lens through which most experts and relevant personnel think about accountability.

To contrast this heavily focused narrative, the interviews have highlighted a whole range of intervention points, where humans are expected to, and should be supported in making decisions that enable legal, safe, and ethical weapon systems. These are arguably points that should be considered in “meaningful human control”. These include, but are not limited to:

Establishment of military need:

- defining military necessity for research, development, and/or procurement; and
- choice of technological approach based on political and strategic motivations.

(Main related stakeholders: UK MoD; UK Defense Equipment and Support (DE&S); private military contracting companies, such as BAE Systems, Qinetiq, General Dynamics)

Technical capabilities and design:

- trade-offs between general applicability and tailored, specific solutions with high efficacy and guarantees on performance;
- awareness, training, and foreseeability of contextual factors about intended use situations that may affect the performance of the weapon system; and
- documentation and communication of known limitations and failure modes of the system design.

(Main related stakeholders: private military contracting companies, such as BAE Systems, Qinetiq, General Dynamics; UK Defense Science and Technology, UK Defense and Security Analysis Division)

Human-computer interaction design:

- choices of what data to include and what data to exclude;
- trade-offs between clarity and comprehensiveness;
- level of technical information communicated; and
- parallel communication channels: to operator in/on the loop, to command centres further from the field, and to logs for future technical analysis or legal investigation.

(Main related stakeholders: Private military contracting companies, such as BAE Systems, Qinetiq, General Dynamics; UK Defense Science and Technology; UK Defense and Security Analysis Division; UK MoD military personnel — human operators)

Weapons testing:

- choice of parameters to be evaluated, frequency of evaluation, and conditions under which to evaluate;
- simulation of adversaries and unexpected situations in the evaluation phase;
- evaluation of HCI in extreme conditions; and
- evaluation of the human-machine team.

(Main related stakeholders: private military contracting companies, such as BAE Systems, Qinetiq, General Dynamics; UK DE&S; UK MoD military personnel — human operators)

Procurement:

- robust Article 36 review;
- assessment of operational gaps, and trading-off operational capability with risks;
- trade-off between cost effectiveness and performance of weapons systems;
- documentation and communication of trade-offs so they can be re-evaluated as the context or technology changes;
- number and type of systems;
- provisioning of training and guidance; and
- provisioning for maintenance.

(Main related stakeholders: UK DE&S; Article 36 convened expert assessment group; private military contracting companies, such as BAE Systems, Qinetiq, General Dynamics)

Weapons deployment:

- informing commanders about the capabilities and limitations of the system, of their track record in similar situations, and of novel parameters of the new situation;
- establishing and training for appropriate pre-deployment testing schemes to capture any vulnerabilities or “bugs” with specific weapons system;

- checking for readiness of troops to operate and maintain systems in the arena; and
- expected response of non-combatants to the presence of the weapon system.

(Main related stakeholders: UK MoD commanding officers; UK MoD military personnel — human operators)

Weapons engagement:

- awareness of limiting contextual factors, need to maintain operator awareness, and contextual knowledge; and
- handover of control between operators during an operation.

(Main related stakeholders: UK MoD military personnel — human operators)

Performance feedback:

- ensuring a meaningful feedback process to guarantee process improvement, reporting of faulty actions, communicating sub-par human-machine techniques and capabilities, and more.

(Main related stakeholders: UK MoD military personnel — human operators; UK MoD commanding officers; UK DE&S; private military contracting companies, such as BAE Systems, Qinetiq, General Dynamics)

5. Recommendations

5.1 Dialogue shift: Emphasising control chain and shared responsibility

The prioritisation of “meaningful human control” for LAWS-related risk mitigation and governance anchors the scope of control points around final targeting decisions. The narrative implies that this is the main area of control that we want to manage, focus, and improve on in order to ensure that the weapons systems we are deploying are still acting with the intent and direction of human operators. Although this is an important component of ensuring thoughtful and safe autonomous weapons systems, this is only a fraction of the scope of control points. In order for us to acknowledge the other points of control throughout the research,

development, procurement, and deployment of LAWS, we need to be inclusive in our dialogue about these other points of human control.

5.2 Distribution of knowledge: Personnel training

Training everyone who touches the research, development, deployment, etc. of LAWS on international humanitarian law, robot ethics, legality of development, responsibility schemes, and more, would contribute to a more holistic approach to responsibility and accountability, and, at its best, can contribute to a culture that actively seeks to minimise and eliminate responsibility gaps through a collaborative governance system.⁵ This distribution of understanding around governance could provide a better landscape for accountability through heightened understanding of how to contextualise technical decisions. Further, it can provide an effective, granular method for protecting against various levels of procedural deterioration. With shifting geopolitical pressures, as well as various financial incentives, there could easily be a deterioration of standards and best practices. A collaborative governance scheme that is based on a distributed understanding of standards, military scope, international norms, and more, can provide components of a meaningful and robust governance plan for LAWS. This distribution of knowledge, though, must be coupled with techniques for reporting and transparency of procedure to be effective.

5.3 Acknowledging the politics of technical decision-making/ design specifications

“Meaningful human control”, through its dialogue anchoring, also puts a heavy burden on the technical components of design decisions, such as best practices for human-computer interactions. The politics of quantification in technical decision systems for autonomous systems should not be undervalued. The way any autonomous system decides what actions to take and what information to show is a highly political decision, especially in the context of war. It is important to understand which parts of the design process are more political than they are technical, who should be involved in those decisions, and how to account for those decisions in the scope of research and development (to inform a proper, comprehensive collective responsibility scheme).



Appendix available online at <https://doi.org/10.11647/OBP.0360#resources>

Notes and References

- 1 Carpenter, C. 'From "stop the robot wars!" to "ban killer robots"', *Lost Causes* (2014), pp. 88–121. <https://doi.org/10.7591/cornell/9780801448850.003.0005>
- 2 Human Machine Touchpoints. *The United Kingdom's Perspective on Human Control Over Weapon Development and Targeting Cycles* (2018).
- 3 Human Machine Touchpoints (2018).
- 4 Khalilzad, Z. and I. O. Lesser. *Selected Scenarios from Sources of Conflict in the 21st Century*. RAND (1998), pp. 317–18.
- 5 Ansell, C. *Collaborative Governance* (2012). <https://oxfordindex.oup.com/view/10.1093/oxfordhb/9780199560530.013.0035>