Ministry of Defence

The Government Response to the Report by the House of Lords AI in Weapon Systems Committee: ‘Proceed with Caution: Artificial Intelligence in Weapon Systems’

(Session 2023–24 HL Paper 16)

February 2024

The Government Response to the Report by the House of Lords AI in Weapon Systems Committee: 'Proceed with Caution: Artificial Intelligence in Weapon Systems'

(Session 2023–24 HL Paper 16)

Presented to Parliament
by the Secretary of State for Defence
by Command of His Majesty

February 2024

CP 1023

This publication is available at www.gov.uk/official-documents.

ISBN 978-1-5286-4669-7

E03059416 02/24

Printed on paper containing 40% recycled fibre content minimum

Printed in the UK by HH Associates Ltd. on behalf of the Controller of His Majesty's Stationery Office

Contents

Introduction

1. The Government welcomes the report by the House of Lords AI in Weapon Systems Committee entitled 'Proceed with Caution: Artificial Intelligence in Weapon Systems', dated 1 December 2023 (HL Paper 16), and is grateful to all who gave evidence in the preparation of the report.

2. Artificial Intelligence (AI) technologies have the potential to transform every aspect of Defence fundamentally. It is essential that the UK's Armed Forces are able to embrace these technologies to maintain our technological edge within a competitive, volatile and challenging international security environment. At the same time, we recognise that the adoption of these general-purpose enabling technologies poses significant challenges in a high-impact Defence context. We welcome the Committee's thorough and thought-provoking analysis. These issues will be of increasing importance in the years to come as AI-enabled military systems and capabilities become more common.

3. The Ministry of Defence (MOD) published the Defence AI Strategy[1] in June 2022 alongside the 'Ambitious, Safe, Responsible'[2] policy statement, which includes our Defence AI Ethical Principles. These documents set out our overall approach to the development and adoption of these transformative technologies in line with our AI Ethical Principles and the values and standards of the society we protect.

4. The MOD is actively engaging with a very wide range of experts (including technologists, ethicists, legal advisers and civil society stakeholders) to understand the issues and concerns associated with the use of AI in weapons, and to develop appropriate policies and control frameworks. This is not a new challenge for Defence – we have extensive experience of adapting to embrace new technologies and capabilities. We are currently assessing the appropriate ways in which our existing, robust and effective legal, safety and regulatory compliance regimes may need to evolve to tackle new challenges posed by AI technologies.

5. We are committed to safe and responsible use of AI in the military domain. We are clear that we will use AI to augment the capabilities of our service personnel and derive military advantage through effective human-machine teaming; that accountability for military effects can never be delegated to a machine; and that we will always comply with our national and international legal obligations. At the same time, we know some adversaries may seek to misuse advanced AI technologies, deploying them in a manner which is malign, unsafe and unethical. We are working with allies and partners through international forums to develop norms and standards for military AI and to ensure that any illegal, unsafe or unethical use of these technologies is identified, attributed and held to account.

---

[1] Ministry of Defence. (2022). *Defence Artificial Intelligence Strategy.* https://www.gov.uk/government/publications/defence-artificial-intelligence-strategy

[2] Ministry of Defence. (2022). *Ambitious, Safe, Responsible: Our approach to the delivery of AI-enabled capability in Defence.* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1082991/20220614-Ambitious_Safe_and_Responsible.pdf

6. This Command Paper responds to the conclusions and recommendations set out in the Committee's report. For ease of reading, the conclusions and recommendations have been grouped thematically into four sections: General Principles; Ethics, Legal and Governance; Safeguarding against risks; and Enablers. In each section, related conclusions (in italics) and recommendations (in bold) are highlighted in separate text boxes, with the Government response following in plain text.

Section 1 – General Principles

7. This section addresses six general conclusions and recommendations from the Committee's report, including points about definitions, the publication of statistics, and engagement with Parliament and the general public.

Conclusion and Recommendation (C&R) 4: *The UK's lack of an operational definition of [Autonomous Weapons Systems (AWS)] is a challenge to its ability to make meaningful policy on AWS and engage fully in discussions in international fora. Other states and organisations have adopted flexible, technology-agnostic definitions and we see no good reason why the UK cannot do the same.* (Paragraph 53)

C&R 5: **In acknowledgement that autonomy exists on a spectrum and can be present in certain critical functions and not others, the Government should without further delay adopt operational definitions of 'fully' and 'partially' autonomous weapon systems as follows:**

- **'Fully' autonomous weapon systems: Systems that, once activated, can identify, select, and engage targets with lethal force without further intervention by an operator.**
- **'Partially' autonomous weapon systems: Systems featuring varying degrees of decision-making autonomy in critical functions such as identification, classification, interception and engagement. (Paragraph 54)**

8. **The Government respects the argument put forward by the Committee but does not agree with these conclusions and recommendations and does not intend to adopt an official or operative definition of AWS at this time.**

9. Our priority is to maximise our military capability in the face of growing threats, mindful at all times that we must uphold our obligations under International Humanitarian Law and abide by the ethical principles we have outlined. We must be ambitious for our AI-enabled military capability; but we have also set out very strong commitments that we will exploit and use AI in ways that are safe, legal, ethical and responsible. Our strong stance on the applicability of International Humanitarian Law (IHL) to AI-enabled capability means that our position is that the irresponsible and unethical behaviours and outcomes about which the Committee is rightly concerned are already prohibited under existing legal mechanisms. We are concerned, however, about the implications of formally adopting any AWS definitions. While acknowledging the argument that definitions are an aid to policymaking and discussion, there is a strong tendency in the ongoing debate about autonomous weapons to assert that any official AWS definition should serve as the starting point for a new legal instrument prohibiting certain types of systems (we do not suggest that this is the Committee's intent). This represents a threat to UK Defence interests, and at the worst possible time, given Russia's action in Ukraine and a general increase in bellicosity from potential adversaries.

10. The Government rejects any suggestion that a definition of the kind proposed by the Committee, would, on its own, be sufficient to determine questions about the compliance or non-compliance of particular AWS with IHL or to support a presumption towards a ban. We recognise that the proposed definitions focus on the *potential* performance of a system in a particular part of its lifecycle, regardless of the controls that might be applied or take effect at different points in that lifecycle, and therefore the outcomes that might arise. Nevertheless, even systems prevented from displaying any autonomous behaviour in practice could be caught up in a definition-driven ban argument because of underlying system potential. This would clearly be an unacceptable constraint.

11. We maintain that meaningful human control, exercised through context-appropriate human involvement, must always be considered across a system's full lifecycle. We and our key partners have been clear that we oppose the creation and use of AWS that would operate in any other way, but we face potential adversaries who have not made similar commitments and are unlikely to be as responsible. Rather, as we have seen in other domains, adversaries will seek to use international pressure and legal instruments to constrain legitimate research and development while actively pursuing unsafe and irresponsible use cases. It is important that the UK maintains the freedom of action to develop legal and responsible defensive capabilities to protect our people and our society against such hostile activities.

C&R 1: *The lack of available statistics on the UK's spending on AI in defence means that it is difficult to determine whether the level of spending is appropriate and to compare it internationally.* (Paragraph 17)

C&R 2: **The Government must publish annual spending on AI in defence as part of the Ministry of Defence's Finance and Economics Statistics Bulletin series. (Paragraph 17)**

12. The Government recognises that the publication of official statistics on spending levels can have widespread benefits in terms of tracking investment over time, benchmarking national performance and ensuring balanced investment. We are committed to providing as much transparency as possible around Defence AI investment to aid public and parliamentary scrutiny, including through the publication of statistics.

13. However, it is extremely challenging to provide authoritative figures on the MOD's overall investment in AI owing to its nature as a general-purpose enabling technology. AI is not typically a capability in and of itself, but rather enables different types of functionality as part of a much broader system or programme. It is therefore difficult to calculate the spend on particular AI components as cost data is typically integrated within broader programme costs. Drawing a comparison with civilian systems, it would be difficult to assign values to the AI contained in a typical smartphone, but AI is increasingly a key element integrated within a wide range of applications. The MOD is exploring medium-term solutions that may give a better picture of overall AI spending across Defence.

C&R 34: The Government must allow sufficient space in the Parliamentary timetable and provide enough information for Parliament, including its select committees, to scrutinise its policy on AI effectively. We naturally understand that elements of policy development may be highly sensitive; but there are established ways of dealing with such information. Arguments of secrecy must not be used to sidestep accountability. (Paragraph 250)

14. Effective Parliamentary accountability is of paramount importance to the Government. For example, throughout this Inquiry the MOD has made every effort to be transparent with the Committee about our approach and intent for the responsible adoption of AI, including through submission of oral and written evidence and through closed briefings. The MOD remains committed to engaging with Parliament and its Select Committees on these important subjects. Where national security considerations prevail, such as in relation to classified programmes, we will ensure suitable information is shared via appropriate mechanisms with Parliament to enable effective democratic scrutiny.

C&R 35: The Government must ensure that it engages with the public on AI-enabled AWS. It must also ensure that ethics are at the heart of its policy. (Paragraph 251)

15. Scrutiny and challenge are essential parts of the policy-making process and the MOD has a range of established and effective mechanisms to engage with external subject experts and stakeholders on potentially contentious issues. However, the nature of Defence and security work does impose certain constraints on external engagement beyond those that may apply to government departments more generally. It is not widespread practice within Defence to engage with the public at large (i.e. by way of public consultation or polling). We do make significant efforts to understand different perspectives on issues around AWS, including through forums such as the Defence AI Ethics Advisory Panel, public engagement at events (e.g. the annual AI Summit during UK Tech week), outreach to academic partners and civil society stakeholders, and assessing research undertaken externally or by other government departments.[3]

16. Questions about the MOD's approach to embedding AI ethics considerations within our policy, processes and culture, including through publication of a comprehensive "Whole Force" policy document (known internally as a Joint Service Publication – JSP) in early 2024, are answered in the next section. That JSP will give direction and guidance on understanding and implementing our AI ethical principles in practice, helping to ensure (alongside other controls and assurance processes)

[3] For example, the Centre for Data Ethics and Innovation (CDEI) conducts an annual Public Attitudes Tracker Survey to understand attitudes towards data and data-driven technologies, including artificial intelligence (AI), and how these attitudes change over time. (Centre for Data Ethics and Innovation. (2023). *Public attitudes to data and AI: Tracker survey (Wave 3)*. Department for Science, Innovation & Technology. https://www.gov.uk/government/publications/public-attitudes-to-data-and-ai-tracker-survey-wave-3/public-attitudes-to-data-and-ai-tracker-survey-wave-3#foreword)

that teams and decision-makers across the department and Armed Forces make appropriate, balanced and responsible decisions about the adoption and exploitation of AI to achieve our strategic objectives.

Section 2 – Ethics, Legal, and Governance

17. This section addresses recommendations and conclusions regarding the domestic and international implementation of the MOD's AI Ethical Principles in the context of legal, safety and governance frameworks, including methods for ensuring meaningful human control and the purpose and future of the AI Ethics Advisory Panel. Thirteen conclusions and recommendations are addressed in this section.

C&R 10: *Context-appropriate human control is a difficult concept to define, presenting challenges to the development of policy on AWS. Determining whether human control has been satisfied and setting a minimum level of human involvement in a system involves considering many nuanced factors such as the complexity and transparency of the system, the training of the operator, and physical factors such as when, where and for how long a system is deployed. (Paragraph 102)*

C&R 11: **We note the Ministry of Defence's definition of "context-appropriate" and "human involvement". The Government must ensure that human control is consistently embedded at all stages of a system's lifecycle, from design to deployment. This is particularly important for the selection and attacking of targets. (Paragraph 103)**

C&R 23: *Human decision-making is central to legal accountability for the use of AWS. Accountability cannot be transferred to machines.* (Paragraph 188)

C&R 24: **The Government must commit to integrating meaningful human control into all AI-enabled AWS which it deploys so that human accountability can clearly be assigned for use of AWS on the battlefield. (Paragraph 189)**

18. The Government has clearly set out its commitment to ensuring meaningful human control (and therefore human accountability) through context-appropriate human involvement throughout the lifecycle of AI-enabled military systems. This commitment derives from the UK's obligations under International Law and is further enshrined through the Defence AI Ethical Principles. Any UK military capability must only be used in a manner that achieves lawful, ethical and effective ends, though the controls that are applied will vary depending upon the nature of the capability and context of use. Nevertheless, the requirement of lawful and ethical effect is immutable.

19. The term 'context-appropriate' is important as there are a vast range of potential applications for AI across defence, with each being subject to specific contextual factors. These include the purpose of use, physical and digital environment, nature of possible threats, risks associated with system behaviour, regulatory environment, and so on. These contextual factors will shape the type and timing of human involvement to ensure that it is best tailored to meet military, safety, legal and ethical objectives. For example, an at-sea engagement, or use of a high-speed defensive air system, might warrant a greater degree of autonomous functioning than would be appropriate in a complex urban environment.

20. ‘Human involvement’ is demonstrated at the numerous points throughout the system lifecycle at which authorised, suitably qualified and experienced people exercise judgement to influence, direct or limit the behaviour of an AI-enabled system and its effects. Some of those decisions reach beyond those individuals who deploy the system, across multiple parts of the MOD and beyond. This lifecycle approach to human involvement and control of AI-enabled systems is described in the 2018 and 2020 UK working papers published at the UN GGE on LAWS. The MOD continues to conduct research and policy work to determine the most effective ways to implement context-appropriate human involvement, and ultimately to put the MOD ethical principles for AI in Defence into practice. This includes active engagement and collaborative research alongside our closest international allies including NATO and 5-Eyes partners.

21. We entirely agree with the Committee that human responsibility and accountability for decisions on the use of weapons systems cannot be transferred to machines. Defence has significant expertise in understanding how accountability works on a systemic basis and how to assign accountability to the right level. We apply accountabilities through duty holding in a safety context, clear governance frameworks for the conduct of military operations, and through the chain of command and application of rules of engagement to apply military personnel and capabilities in a manner that achieves strategic military effect. New technological capabilities are adopted within that system of accountabilities.

C&R 12: The Government must ensure that any personnel required to use AWS have been provided with the training to ensure they have sufficient technical knowledge of how the system operates and its limitations, enabling operators to have confidence and capacity to override decisions where necessary. Such training needs to encompass the technical characteristics of systems, but also the exercise of human agency and legal compliance in controlling them. (Paragraph 104)

22. We agree with this recommendation, with the slight qualification that operators are unlikely to require the same knowledge of the technical characteristics of systems as, for example, a systems or software engineer. Exactly what constitutes ‘sufficient technical knowledge’ will depend on the precise nature of a given system, and potentially also wider factors. Overall, however, we agree that proper training is essential before personnel operate or are given operational command of AWS, to ensure that meaningful human control and accountability is maintained at all times. In the Defence AI Strategy, the Department committed to exploring options to establish and manage a distinctive ‘AI-enabled military operator’ skill set or trade, and to institute clear and auditable processes for the licensing and routine re-certification of military AI operators where appropriate. The MOD agrees that such training and licencing regimes should encompass both technical skills and wider legal & ethical compliance and will continue to work to ensure such items are included within training provided to personnel using AWS.

C&R 20: *We have heard significant concerns about the ability of AWS which use AI technology in the targeting process to be used in compliance with IHL. The Government also acknowledges that there must be "context-appropriate" human control over any AWS which can identify, select and attack targets* (Paragraph 181)

C&R 21: **The Government must demonstrate that AI-enabled AWS which it develops or deploys will function under sufficient levels of human control to be compliant with IHL on the battlefield. (Paragraph 182)**

C&R 22: **The Government must demonstrate to Parliament that it has in place an effective system to perform Article 36 weapons reviews for AI-enabled AWS, particularly AWS which continue to learn and modify their behaviour after they have been deployed, including setting thresholds for triggering a new review. (Paragraph 183)**

23. The Government wholly agrees that the weapon systems it deploys must be used in a manner which is compliant with IHL. As set out above, there are various established mechanisms through which the MOD will ensure that context-appropriate human involvement is exercised throughout the lifecycle of AI-enabled military systems, including weapons[4]. This includes governance of the system before the use of a military capability (e.g. policy frameworks, risk management processes, system test and evaluation and operator training); during use (e.g. targeting processes, setting system parameters and monitoring performance, battlespace management); and after use (e.g. 'after action' reporting and investigations, system updates). This layered approach will ensure that commanders and operators understand the capabilities and limitations of any AI-enabled military systems under their authority, and that there are effective controls or limits in place to enable them to fulfil their ethical and legal obligations.

24. Article 36 Weapons Reviews are an important element of this overall governance framework. Weapons Reviews support the MOD in meeting its obligation to determine whether any 'new weapon, means or method of warfare' developed or deployed is capable of properly functioning for its designed purpose to achieve a defined outcome or else would, in some or all circumstances, be prohibited by international law. Weapons Reviews do not, however, sit alone in governing the use of weapons in light of legal obligations, and are not simple 'review and release' events. Systems developers can expect an iterative process to review and – should a system update result in material changes to the nature of a particular weapon system or capability – re-review would be appropriate.

25. The MOD recognises that the nature of AI technologies may pose particular challenges in the context of military systems (e.g. the potential for evolved behaviour by 'learning systems'), and that our extant Article 36 Legal Review processes will need to be adaptable in order to be 'future-proof' against such challenges. We are

[4] Ministry of Defence. (2022). *Ambitious, Safe, Responsible: Our approach to the delivery of AI-enabled capability in Defence.* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1082991/20220614-Ambitious_Safe_and_Responsible.pdf

actively working with expert stakeholders and international partners to understand and tackle these challenges, and to support international conversations about compliance with IHL by sharing lessons and best practice.

26. The MOD will continue to make every effort to be transparent with Parliament and the public about our governance processes, as demonstrated in the evidence and private briefings we have provided to this Inquiry.

C&R 25: **We call for a swift agreement of an effective international instrument on lethal AWS. It is crucial to develop an international consensus on what criteria should be met for a system to be compliant with IHL. Central to this is the retention of human moral agency. Non-compliant systems should be prohibited. Consistent with its ambitions to promote the safe and responsible development of AI around the world, the Government should be a leader in this effort. (Paragraph 204)**

27. We set out to the Committee in written and oral evidence our belief that IHL provides an effective, proven and technology-agnostic framework to regulate the development and use of new military systems, including those enabled by AI. It is prohibited to deploy a system in a manner that is inconsistent with the principles of IHL. The UK intends to remain an active and influential participant in international dialogues to regulate so called Lethal Autonomous Weapons Systems and considers the Group of Governmental Experts to the Convention on Certain Conventional Weapons (LAWS GGE) to be the most appropriate international forum to advance negotiations on these issues.

28. Since the inception of the LAWS GGE, the UK has worked with partners to build international understanding of the issues involved and to develop an international consensus on IHL compliance. We have put forward (singularly or in partnership with other states) various initiatives to this end:

- In May 2023, the UK joined a Joint Statement with 51 other states which highlighted the approach that acknowledged that some weapons would be automatically prohibited under International Humanitarian Law (IHL). The Statement also welcomes the focus on the role of humans in the context of autonomy in weapon systems by ensuring an appropriate level of human involvement throughout the life-cycle of the weapon system and preserving human responsibility and accountability.
- In March 2023, the UK co-sponsored *"Draft articles on autonomous weapon systems" proposal which inter alia includes Articles on "Preventing Autonomous Weapon Systems That, By Their Nature, Are Incapable of Use in Accordance With IHL".* This includes suggested *"Regulatory Measures to Ensure Accountability"*. We have also proposed that the LAWS GGE develop a document that would constitute an authoritative and comprehensive statement of the application of International Humanitarian Law and agreed best practice with regard to LAWS.

- Most recently, the UK joined consensus at the recent CCW Meeting of High-Contracting Parties to agree a new strengthened Mandate for the LAWS Group of Government Experts which directs the group to *"formulate a set of elements of an instrument"* by 2025 *"without prejudging its nature"*. The UK will continue to play a leading role in fulfilling this new mandate.

C&R 26: **The Government should make explicit how it intends to implement domestically the five principles outlined in Ambitious, safe and responsible and the draft articles submitted to the 2023 Group of Governmental Experts. (Paragraph 212)**

C&R 27: **The Government should set out its plans to become a leader in setting responsible standards at every stage of the lifecycle of AWS, including responsible development and governance of military AI. These standards should refer to the Ministry of Defence's Five Ethical Principles for AI in Defence. (Paragraph 213)**

29. The Government entirely agrees with these recommendations and is already taking concrete steps to deliver these outcomes. We are determined to adopt AI safely and responsibly because no other approach would be in line with the values of the British public; meet the demands of our existing rigorous approach around safety and legal compliance; or allow us to develop the AI-enabled capability we require.

30. The MOD is currently working to adapt legal, safety and regulatory policies, processes and compliance regimes to embed the Defence AI Ethical Principles at each stage of the System Lifecycle. As an example of this, we will shortly publish the 'Dependable AI in Defence' Joint Service Publication (JSP) setting out the governance, accountabilities, processes and reporting mechanisms that will need to be put in place across Defence to operationalise the MOD's 'Ambitious, Safe and Responsible' policy. However, it is important to be clear that the adoption and integration of novel technologies and capabilities is not a new challenge for Defence. We have established and effective risk management systems in place with clear lines of accountability and assurance and controls frameworks embedded throughout the lifecycle of any military capability. Where unique requirements are required, owing to the nature or functionality of AI, we will review and augment these approaches, working with established frameworks wherever possible.

31. The MOD is also engaging with a wide range of international bodies, partners, and stakeholders to promote our approach to responsible military AI and champion global norms and standards for the safe development and use of these technologies. This includes bilateral and multilateral dialogues with key partners (particularly the U.S., 5-Eyes and NATO) to drive policy and technical alignment, as well as broader outreach to build communities of interest and champion core values. As examples:

- In February 2023, the UK played a prominent role in the first 'Responsible Use of AI in the Military Domain' (REAIM) summit in the Hague, co-hosted by the Netherlands and the Republic of Korea (RoK) and featuring high-level representation from 100 countries.

- The UK is a founding member of the 'AI Partnership for Defense', which comprises 16 likeminded nations (Australia, Canada, Denmark, Estonia, France, Finland, Germany, Israel, Japan, the RoK, Norway, the Netherlands, Singapore, Sweden, the UK, and the US).
- The MOD played a leading role in the development of the NATO AI Strategy and its principles-based ethical approach. We actively support NATO's Data and Artificial Intelligence Review Board (DARB), which is developing a Responsible AI (RAI) Certification Standard and best practice risk management approaches.

C&R 32: **The Government has asserted that transparency and challenge are central to its approach. From the evidence we have taken, we have not found this yet to be the case. The Government should increase the transparency of advice provided by the AI Ethics Advisory Panel by publishing its Terms of Reference, membership, agendas, and minutes, as well as an annual transparency report. (Paragraph 247)**

C&R 33: **The Government should immediately expand the remit of the AI Ethics Advisory Panel to review the practical application of ethical principles in armed conflict and to cover ethics in relation to the development and use of AI in AWS. (Paragraph 248)**

32. The Government accepts that we have more to do, noting that work is already underway to review the role of the Defence AI Ethics Advisory Panel (EAP) and increase the transparency of advice. The UK is a leading nation in terms of transparency about our approach to military AI and our proactive approach to encouraging external challenge, and in having established an independent expert panel to advise on AI Ethics in Defence. We intend to build on these strengths.

33. The MOD's EAP was established in 2021 to '*provide expert advice, scrutiny and challenge across the full span of principles, policies and frameworks relevant to the delivery of ethical AI outcomes within Defence*'. It is chaired by the MOD's 2nd Permanent Secretary, bringing together experts (including critical perspectives) from Defence, academia, industry, and civil society. The panel has now met six times. It was instrumental in the development of the MOD's AI Ethical Principles, advised on the development of the Defence AI Strategy and has provided constructive challenge on our developing approach to operationalising AI ethics, including by reviewing draft policy, proposing external best practice, and reviewing real life case studies. For example, at its most recent meeting in September 2023, the panel considered a report by the Government's Centre for Data Ethics and Innovation (CDEI) into how three current MOD R&D projects have understood and worked to apply the AI Ethical Principles.

34. The MOD will shortly publish the minutes of the previous 6 AI Ethics Advisory Panel (EAP) meetings on the gov.uk website, alongside the EAP Terms of Reference and Membership. We are currently examining options for the EAP's future role, including the extent to which an independent expert function could consider sensitive Defence use cases, and options for delivering regular transparency reporting. The Committee's recommendation that the MOD expand the remit of the EAP to review

the practical application of ethical principles in armed conflict and to ethics in relation to the development and use of AI in AWS aligns with our thinking about the role of the EAP. We believe these are already within scope of the EAP Terms of Reference, which refer to ‘the full span of principles, policies and frameworks relevant to the delivery of ethical AI outcomes within Defence.’

Section 3 – Safeguarding against risks

35. This section addresses recommendations and conclusions regarding AI safety from the perspective of potential escalation in conflicts, use against the UK, and integration within strategic systems. Six conclusions and recommendations are addressed in this section.

> **C&R 3**: **The Bletchley Declaration of November 2023 is, inevitably, aspirational, but it is a start. We commend the contents of the Declaration and encourage the Government to apply its principles to AI in defence. (Paragraph 24)**

36. The Government believes that the principles set out in the Bletchley Declaration are well aligned with our approach to Responsible AI in Defence, as encapsulated in the Defence AI Ethical Principles. Both documents focus on the need to develop human-centric, trustworthy, safe and responsible AI. Similarly, the MOD actively works with allies and through international forums (e.g. the LAWS GGE and the Responsible AI in the Military Domain (REAIM) initiative) to shape global AI developments to promote security, stability and democratic values. In refreshing our AI policy, we will examine options for applying the principles more explicitly.

37. The MOD will continue to work closely with the Department for Science, Innovation and Technology (DSIT) and the newly formed UK AI Safety Institute to understand potential threats arising from 'Frontier AI' capabilities and to support wider initiatives aimed at mitigating national security challenges.

> **C&R 15**: **We note with concern that at the moment there is not enough being done to protect UK systems from interference or attack, or to develop methods to counter the use of AWS by adversaries. It is one thing to deploy a system without challenge, but quite another to cope not only with enemy action but with the realities of the battlefield. The Government must recognise the risk posed to our own side by enemy AWS, avoiding a "sole-ownership fallacy", and must take action to ensure the resilience, as far as possible, of the UK's own systems. (Paragraph 123)**

38. The Defence AI Strategy clearly highlights the risks posed by the adoption of AI-enabled military capabilities by adversaries, along with the risk that certain nations and groups may misuse these technologies in ways that are illegal, unsafe or unethical. Significant work is underway across the MOD – and in partnership with other government departments and key allies – to understand potential threats, ensure that our own systems are resilient to attack, and develop appropriate countermeasures.

C&R 16: *The proliferation of commercially available drones, coupled with the widening availability of AI software, including open-source software, could enable non-state actors to produce AWS from widely available civilian technologies.* (Paragraph 134)

C&R 17: **The Government must demonstrate to Parliament that it is committed to ensuring 'deterrence by denial' to defend its own citizens from the use of AWS by nonstate actors, as well as methods to limit the proliferation of the precursors of AWS. (Paragraph 135)**

C&R 18: **The development of AI capabilities, including AWS, has the potential to bring significant strategic benefits to the UK and its allies, for example enhanced conventional deterrence. However, the Government must not use AI-enabled AWS in a way that could result in unintended increases in escalatory risk. (Paragraph 149)**

39. The Government recognises that AI technologies present both opportunities and significant strategic challenges for the UK. This includes the risk that adversaries and non-state actors may use AI to exacerbate existing threats (e.g. cyber-attacks and misinformation), along with the potential for new challenges arising from misuse or misadventure.

40. Significant work is underway across government to monitor technology trends, understand the most concerning national security risks presented by AI and develop appropriate mitigation strategies. This includes, as examples: collaborations with civil society and the technology sector – and in particular Frontier AI companies – to develop safeguards and regulatory frameworks; research and expert consultations to understand potential risks of miscommunication and inadvertent military escalation at times of tension; and action with international partners to address potentially destabilising effects. Where appropriate, we will work through existing non-proliferation, disarmament and export control regimes, treaties and organisations to limit the spread of key strategic or sensitive technologies and deter the deliberate transfer by States to non-state or proxy actors.

41. On domestic security, protection of the public from non-state actors using drones is carried out by the Home Office Counter-Drones Unit, which provides cross-government leadership on domestic drone security. They are responsible for understanding the rapidly evolving threat from drones to people and places in the UK, ensuring that new technological developments in both drone and counter-drone technology are factored into risk assessment, capability development and innovation investment, and supporting operational partners to build effective countermeasures.

42. The Counter Drone Unit takes a threat-agnostic approach and has focussed on building comprehensive Detect, Track and Identify, and deterrence capabilities (both active, with effector usage, and more passive, such as with protective security precepts and advice to CNI) with police forces so that when someone decides to use a drone in a dangerous, negligent or illegal manner, potentially with the intent to harm or cause damage, there is a far greater chance of prevention. This is further bolstered with forensic and investigatory capabilities.

C&R 19: **The risks inherent in current AI systems, combined with their enhanced escalatory risk, are of particular concern in the context of nuclear command, control and communications. The Government should lead international efforts to achieve a prohibition on the use of AI in nuclear command, control and communications. (Paragraph 161)**

43. The UK Government has publicly committed to ensure that – regardless of any use of AI in our strategic systems – human political control of our nuclear weapons is maintained at all times. The UK is actively encouraging other nuclear states to make a similar commitment. The UK is at the forefront of work internationally to reduce the risk of nuclear conflict and enhance mutual trust and security, and will continue to promote and engage with international dialogue aimed at identifying and addressing crucial AI-related strategic risks.

Section 4 – Enablers

44. This section considers conclusions and recommendations relating to the core enabling capabilities that will be required to accelerate the adoption of AI in Defence. This includes skills and talent, high-quality data, reforms to the MOD procurement system and challenges around testing and verification. Eleven conclusions and recommendations are addressed in this section.

C&R 30: *Issues of pay and ethical concerns act as barriers to recruitment. AI is highly complex and requires a very high degree of knowledge and qualifications in order to develop it. This requires officials to be the "brightest and the best". But the Ministry of Defence is hamstrung by the Government's requirement that all staff should be paid using existing Civil Service paygrades. This has resulted in salaries offered by the Ministry of Defence being around 50 percent of those offered by commercial enterprises. This situation cannot be allowed to continue.* (Paragraph 237)

C&R 31: **The Government must solve this problem. It must be able to deploy sufficient qualified staff to work on AI and to deliver demanding scrutiny of procurement offers from private developers and manufacturers. This might be achieved by establishing new pay scales, or by bringing in private sector staff on secondment. Either way, it will be challenging but absolutely necessary if we are to have the ability to compete on the international stage and safeguard our country. (Paragraph 238)**

45. The Government recognises the significant challenges in recruiting, training and retaining AI talent amid intense national and global competition for skilled AI professionals. Whilst the public sector may not be able to compete directly with the private sector on pay, an AI career in Defence does offer unique opportunities including working on some of the most challenging and impactful problem sets and the sense of purpose in contributing to national priorities. As set out in the Defence AI Strategy, we are taking a range of steps to make Defence AI an attractive and aspirational choice, working with partners across government to explore freedoms and flexibility around recruitment and retention allowances, developing new mechanisms to identify and incubate AI talent within Defence (including in the reservist community) and developing new partnerships with the technology sector (industry and academia) to encourage greater talent interchange. We will shortly appoint a Capability Lead for AI Talent & Skills to drive this work forward in partnership with the Front Line Commands and our Enabling Organisations.

46. In parallel, the MOD is focused on the need to upskill on AI at all levels of the organisation, from senior decision-makers to procurement officers and front-line operators. Our people must have enough understanding of the capabilities and limitations of AI-enabled systems to work confidently, effectively and responsibly with these tools and capabilities. We have effective processes to ensure that users are closely involved throughout R&D activities, from specifying requirements through to exercising with new AI capabilities in 'real world' situations – an example of this was Exercise Spring Storm in Estonia in 2021 when soldiers from the 20th Armoured Infantry Brigade trialled a novel machine learning engine to process complex data

and accelerate planning and decision-making. We are exploring options for a distinct 'AI enabled military operator' skill set or trade, potentially including licensing and certification requirements that would encompass both technical skills and wider legal and ethical understanding. Further information on our approach to accelerated procurement is provided below.

C&R 6: *In addition to implementing appropriate human input and control in the design phase, high-quality training data, where any bias can be identified and accounted for, is crucial to the development of robust AI models. However, real-world data to train AWS is limited in quantity and quality, and models and tools may be third party, in which case the training data and processes may not be available for inspection.* (Paragraph 79)

C&R 7: **We welcome the Government's commitment to ensuring the gathering and processing of high-quality data sets. In order to achieve this aim, the Government must dedicate sufficient resources to projects which further this goal, including the arrangement of data-sharing agreements with allied partners, and the continuous audit and independent certification of datasets as appropriate. (Paragraph 80)**

47. The Government recognises data as a critical strategic asset and the essential foundation for AI-enabled military capabilities. The MOD's Chief Data Officer is leading the drive to transform Defence into a data-driven organisation, in line with the Digital Strategy for Defence[5] and the Data Strategy for Defence[6]. This includes people, process and technology elements. The Data Strategy in particular emphasises the importance of Defence retaining ultimate ownership of its data, and we are working to ensure this is explicit in our commercial arrangements with all Defence suppliers.

48. We are actively pursuing data-sharing partnerships with key allies, both bilaterally and through multilateral forums including the AUKUS security partnership, the 'Five-Eyes' community and NATO. This requires focus on policy and process as well as technical enablers. Our Defence Data Fabric already provides the tools required to enable AUKUS to implement AI for acoustics, critical for maritime security. As we continually expand the capabilities of the Data Fabric at Secret, a key requirement is to support data and model sharing between allies and industry partners. Central to this will be providing the tools to enable suitably skilled professionals to both use and assure in-house and third-party developed datasets and models.

49. Relevant teams within Defence will examine options to ensure that commercial arrangements with suppliers require appropriate levels of model transparency (including the model, the data, and details of how it was trained), along with robust

[5] Ministry of Defence. (2021). *Digital Strategy for Defence: Delivering the Digital Backbone and unleashing the power of Defence's data*. https://assets.publishing.service.gov.uk/media/60afae56d3bf7f435f43c7af/20210421_-_MOD_Digital_Strategy_-_Update_-_Final.pdf

[6] Ministry of Defence. (2021). *Data strategy for Defence*. https://www.gov.uk/government/publications/data-strategy-for-defence/data-strategy-for-defence

inspection and review processes. This will include clear ownership of intellectual property and address the need for stakeholders to comply with the Defence AI Ethical Principles.

C&R 28: *We heard widespread concern about the Ministry of Defence's procurement processes. While we appreciate the complexities, this is all the more concerning given the additional challenges of creating effective processes for AI in defence.* (Paragraph 235)

C&R 29: **The Government should set up an independent committee of experienced executives to overhaul its defence AI procurement system. The committee should in particular recommend the best way for the Government to specify objectives for systems in advance with clear criteria and how these criteria should be continually monitored and enforced post-deployment, including regular independent AI auditing. As part of this, the Government should require that software developers and manufacturers provide effective through-life support to address any issues. (Paragraph 236)**

50. As outlined above and in the Defence AI Strategy, the Government recognises that significant change is required across the MOD to transform into an 'AI Ready' organisation. This includes accelerating ongoing work to streamline our procurement processes, moving from traditional 'waterfall' programmes to systems that reflect the challenges posed by rapidly evolving technologies.

51. The MOD's 'Commercial X' function was established to drive this change, disrupting outdated practices and processes while incentivising and rewarding commercial individuals based on their ability to deliver quickly and then scale necessary technologies. This new approach lends itself well to working with small and medium sized suppliers, which are prominent in the AI space. In addition, Commercial X and the Defence AI Centre are intrinsically linked to ensure that Defence Digital can set up 'AI' as a subcategory of spend aligned with the new category management approach to Defence acquisition policy. This will help to ensure a specific focus on the requirements for speed and agility associated with AI-enabled capability projects, and support transparency on Defence AI spending (as outlined in para 12). In conjunction with updated procurement regulations, it offers a once-in-a-generation opportunity to ensure the newly unlocked freedoms are applied openly and allow for the rapid exploitation of a constantly developing technology.

52. The Government recognises the value of independent expertise and challenge in helping overhaul Defence acquisition processes to ensure we can adopt new technologies at pace. Commercial X has already started to work with a spectrum of suppliers – varying from emerging AI companies to established 'Big Tech' firms, as well as existing industry forums such as Defence Suppliers Forum – to co-create new ways of working for digital capabilities. We have established cross-cutting executive governance within Defence to direct and oversee our AI transformation (including commercial) and, going forward, will examine options to augment this AI governance with external expertise, potentially including an independent body as recommended.

C&R 8: *Testing AWS properly against all possible scenarios which may arise after deployment is extremely challenging and indeed may be impossible. However, it is vital that only systems which meet sufficient, context appropriate standards of reliability and predictability make their way into use.* (Paragraph 90)

C&R 9: **The Government must develop standards for use in the testing, verification, and validation of autonomous weapon systems. These standards should cover but not be limited to aspects of data quality and sufficiency, human-machine interaction and appropriate transparency and resilience. (Paragraph 91)**

53. The Government agrees that AI-enhanced capabilities must undergo rigorous testing to ensure that operators can be confident in their use and effect. The MOD already has robust and effective processes and procedures to ensure that new or novel military capabilities are safe, secure and operate as intended. We will ensure that our Testing, Evaluation, Validation and Verification (TEV&V) policies, processes and standards are reviewed – and as necessary updated – as we mature our understanding of the particular requirements for safe and responsible deployment of AI-enabled military capabilities and back office services. This includes socio-technical factors relating to assuring and assessing whole human-AI teams together with surrounding training, processes and precautions. We also note that changes to the context of use or operational environment may require models to be substantially updated or re-trained, and therefore re-tested.

C&R 13: *AI-enabled AWS could offer step changes in defence capability including increased speed, efficiency and accuracy. These capabilities, if realised, have the potential to change the nature of warfare and reduce casualties.* (Paragraph 122)

C&R 14: **The Government must ensure that there is sufficient research and resources to realise this potential and it must be realistic about the capabilities and limitations of AI systems, benchmarking the performance of AWS against the operation and fallibility of non-AI-enabled and human-operated systems. (Paragraph 122)**

54. The Government recognises that significant R&D is required to understand the most impactful applications for AI technologies across the span of Defence capabilities and in every operating domain. As highlighted above, the MOD is reviewing and updating our policies and processes to ensure that we are identifying the right opportunities and capability concepts, developing the right evidence base to inform future force planning and balance of investment exercises, and investing in the right tools and processes to test and evaluate the safety, effectiveness and impact of future systems and capabilities.

55. The MOD's approach to AI-enabled capability is based on human-machine teaming by default, combining human cognition and inventiveness with machine-speed analytical capabilities. Therefore, while our research will explore where AI can comparatively outperform a human at a given task, our priority is to understand how to optimise the human-machine team. This research, together with the adoption of a

human-centred design approach, will help to underpin implementation of the MOD AI Ethical Principles of human centricity, responsibility and understanding; as well as being an enabler of operational effectiveness, system safety and resilience. In addition to metrics associated with task performance, other benchmarks – such as the impact on human workload, personnel and training implications – are likely to be key factors in determining the effectiveness, performance and responsible use of AI systems.

C&R 36: *Overall, we welcome the fact that the Government has recognised the role of responsible AI in its future defence capability. AI has the potential to provide key battlefield and strategic benefits. However, in doing so, the Government must embed ethical and legal principles at all stages of design, development and deployment. Technology should be used when advantageous, but not at unacceptable cost to the UK's moral principles. (Paragraph 252)*

56. The Government is grateful to the members of the House of Lords Committee and to those who gave evidence for their time, expertise, and engagement in producing this report. We welcome the Committee's overarching conclusions on the importance of securing future military edge through the adoption of AI-enabled capabilities, and agree with the Committee that we must ensure that ethical and legal principles are embedded at all stages of the capability lifecycle.

57. Defence has a longstanding and proud tradition of adapting to integrate new technologies into our military capabilities across the UK Armed Forces. While we recognise that the adoption of AI may pose some novel risks and issues, including where AWS are concerned, we seize the challenge and are actively working at pace to develop the right policies, processes and standards to develop and deploy AI-enabled military capabilities safely and responsibly.

58. "Proceed with caution", the overall message of this report, mirrors the MOD's approach to AI adoption. We are developing our capabilities iteratively and learning lessons as our understanding of legal, ethical and technical challenges increases. The MOD will continue to engage constructively with Parliament as we do.

E03059416
978-1-5286-4669-7