Fine words, Few assurances

Assessing new Ministry of Defence policy on the military use of Artificial Intelligence

DRONE WARS

Drone Wars UK,
Unit 34, The Wincombe Centre,
Shaftesbury,
SP7 9QJ

www.dronewars.net
info@dronewars.net

Written by Peter Burt
December 2022

Introduction

This short paper analyses the UK's approach to the use of artificial intelligence (AI) for military purposes[1] as set out in two recently policy documents. Part one reviews and critiques the Ministry of Defence's (MoD's) Defence Artificial Intelligence Strategy[2], published in June 2022, while the second part considers the UK's commitment to 'responsible' military artificial intelligence capabilities, presented in the document 'Ambitious, Safe, Responsible'[3] published alongside the strategy document.

What was once the realm of science fiction, the technology needed to build autonomous weapon systems is currently under development by in a number of nations, including the United Kingdom. Due to recent advances in unmanned aircraft technology, it is likely that the first autonomous weapons will be a drone-based system.

Drone Wars UK believes that the development and deployment of AI-enabled autonomous weapons would give rise to a number of grave risks, primarily the loss of human values on the battlefield. Giving machines the ability to take life crosses a key ethical and legal Rubicon. Lethal autonomous drones would simply lack human judgment and other qualities that are necessary to make complex ethical choices on a dynamic battlefield, to distinguish adequately between soldiers and civilians, and to evaluate the proportionality of an attack.

In the short term it is likely that the military applications of autonomous technology will be in low risk areas, such logistics and the supply chain, where, proponents argue, there are cost advantages and minimal implications for combat situations. These systems are likely to be closely supervised by human operators. In the longer term, as technology advances and AI becomes

[1] For a summary of the various military applications of artificial intelligence please see the Drone Wars UK publication 'None too clever? Military applications of artificial intelligence', 7 December 2021: https://dronewars.net/2021/12/07/none-too-clever/

[2] Ministry of Defence: 'Defence Artificial Intelligence Strategy'. June 2022. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1082416/Defence_Artificial_Intelligence_Strategy.pdf

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

more sophisticated, autonomous technology is increasingly likely to become weaponised and the degree of human supervision can be expected to drop.

The real issue perhaps is not the development of autonomy itself but the way in which this milestone in technological development is controlled and used by humans. Autonomy raises a wide range of ethical, legal, moral and political issues relating to human judgement, intentions, and responsibilities. These questions remain largely unresolved and there should therefore be deep disquiet about the rapid advance towards developing autonomous weapons systems.

Despite the seeming inevitability of autonomous weapon systems there are a range of measures which could be used to prevent their development, such as establishing international treaties and norms, developing confidence-building measures, introducing international legal instruments, and adopting unilateral control measures. Drone Wars UK takes the view that the UK should be fully involved in developing these measures on the international stage.

However, at this point in time, the government seemingly wishes to keep its options open, often arguing that it does not want to create barriers which might hinder underlying research into AI and robotics. Nonetheless, plenty of controlled technologies, such as encryption, or in the area of nuclear, biological and chemical science, can be used for civil or military purposes and are controlled without stifling underlying research.

With this in mid, we turn to the two MoD policy documents.

> “Drone Wars UK believes that the development and deployment of AI-enabled autonomous weapons would give rise to a number of grave risks, primarily the loss of human values on the battlefield. Giving machines the ability to take life crosses a key ethical and legal Rubicon.”

Part 1

Unpicking the UK's aspirations for military artificial intelligence

The Defence AI Strategy basically sets out four aims: to adapt and exploit AI "at pace and scale" for military advantage; to transform the MoD into an 'AI ready' organisation; to increase collaboration on defence and security with the UK's AI industry, and to collaborate internationally to shape global AI developments "to promote security, stability, and democratic values". A pledge to continue to work with the UN, other states, civil society, industry and academia to develop and promote best practice in the use of AI and autonomy in weapons systems is important and welcome.

Defence Secretary Ben Wallace's foreword to the strategy neatly sets out the MoD's world-view underpinning the strategy. The foreword asserts that "AI-enabled systems do indeed pose a threat to our security, in the hands of our adversaries, and it is imperative that we do not cede them a vital advantage." This is another way of saying if we don't do this, then others will, and we mustn't be left behind. It also reinforces "Defence's place at the heart of Government's drive for strategic advantage through science & technology" – the view that the science and technology sector should be driven by military needs.

The ambitions ...

Let's start by looking at what the MoD means by 'artificial intelligence'. An answer to a recent Parliamentary Question amplifies on the definition given in the strategy document, characterising AI as "a family of general-purpose technologies with ubiquitous potential applications from the back office to the battlespace".[4] Although MoD plays down the war-fighting applications of AI, with the same answer stating that "in most cases AI will be an enabler for a broader system or capability (e.g. supporting more informed logistics planning) not a capability programme in itself", the strategy emphasises that "AI must be the essential future technology for almost everything that we do".

The strategy proposes various measures to transform the MoD into an 'AI ready' organisation, based on training, cultural change, and the way in which data is

[4] 'Defence: Artificial Intelligence'. Question for Ministry of Defence. UIN HL1997, tabled on 21 July 2022. https://questions-statements.parliament.uk/written-questions/detail/2022-07-21/HL1997

used. Institutionally, the Defence AI and Autonomy Unit will be strengthened, a new 'Digital Backbone' will upgrade IT infrastructure, and a Defence AI Centre has been set up to bring together personnel from across MoD's functions to champion, enable, and innovate for the adoption of AI within the department. The Defence AI Skills Framework will provide training and support career development in sector.

In near term, the MoD's intention is to take a dual track approach to the adoption of AI: firstly by rolling out existing AI and data technology to improve effectiveness, efficiency, and productivity, while at the same time and investing in research and development for next generation AI systems for adoption in the future. The strategy notes that over 200 AI-related research and development programmes are already underway in MoD, ranging from machine learning applications to 'generation after next' research. More information is to be published to help industry and academia understand the areas of research and development that the MoD will be prioritising.

... and the pitfalls

It must be noted, however, that government hopes for the potential of AI are served with a heavy slug of optimism. Regardless of the hype, the reality is that, despite decades of research, AI systems are still too fragile and error-prone to be relied upon in safety-critical applications such as driverless cars, let alone applications relating to the use of military force.

The strategy document certainly succeeds on its own terms: in showing that the MoD is serious about putting AI on the military front line. Indeed, defence officials' enthusiasm for AI jumps out of every page. However, this will not be a simple matter and a range of technical obstacles – such as the need to overcome unexpected system interactions and behaviours, maintain high safety and reliability standards, and minimise the risks of misunderstanding, miscalculation, and uncontrolled escalation on the battlefield – will vastly complicate the task. The strategy recognises this, and acknowledges the "extreme and even existential risks" which AI may pose to humanity. The 'Ambitious, Safe, Responsible' document looks at some of these issues in more detail (see part 2 below). However, even that document only suggests high level approaches to mitigating risks posed by AI and lacks specific proposals for safeguards, and so the strategy can generate only limited confidence that the MoD's processes are robust enough to tackle the challenges. Details and timescales on safeguarding arrangements are in short supply, but MoD does mention that it plans to prepare an AI Technical Strategy, which will presumably provide more information. It has also stated that it is developing frameworks to assess risk and ensure compliance of systems across the full range of AI functionality.[5] Time will tell how effective these are likely to be.

Indeed, the aspiration for an 'ambitious delivery of capability', intended to "enable – rather than constrain – the adoption of AI-enabled solutions and capabilities" and avoid "self-imposed limitations which would risk being arbitrary, constraining, and habitually out-dated", is a clear indication that MoD will be focusing on developing AI capabilities first and foremost, with safeguarding taking second place. Throughout the document, commitments to address ethical concerns and questions of trust are hedged with caveats that they must not impede AI development or collaboration on AI. This seems to

[5] 'Defence: Artificial Intelligence'. Question for Ministry of Defence. UIN HL2088, tabled on 5 September 2022. https://questions-statements.parliament.uk/written-questions/detail/2022-09-05/HL2088

be based on the mistaken premise that safeguards would necessarily constrain technological advances in AI, and the hostility of Conservative ministers to any kind of regulation is apparent throughout the document.

The strategy states that the MoD's approach to AI risk management will be based on the ALARP (As Low As Reasonably Practical) principle that is commonly used for safety-critical and safety-involved systems.[6] We would argue that, given the novel features of AI technology, a more rigorous, precautionary approach will be necessary, particularly with high-risk AI applications which are involved in warfighting and are in the early stages of development, which will require MoD to look beyond the 'reasonably practical' safety standard.

Statements such as: "We must maintain a broad perspective on implications and threats, considering extreme and even existential risks which may arise, and ensuring our risk management practices acknowledge and are suitably adapted to the uncertainty" are clearly intended to provide reassurance about the consequences of the widespread adoption of military AI systems. Yet such assurances have no practical value unless they are backed up by deeds. We have seen how governments have failed to acknowledge or adapt to the clear and present danger of climate change posed by carbon-based technologies – also an existential risk for humanity – in any meaningful way, and there are no grounds to believe that that things will be any different with AI and the technologies of the future.

Perhaps as notable as the intention to introduce AI widely into defence processes is the government's aim to use AI to "transform the UK's economic landscape, requiring a whole-of-society, all-of-government effort that will span the next decade". This raises profound questions about democracy and how decisions are made in society. Ministers have made no effort to have a national conversation about such significant changes and the public has neither debated, nor is ready for, the fundamental changes that the widespread adoption of AI will bring. We would argue that these changes will not necessarily benefit humanity in general but will merely serve the ends of the corporations and governments who are promoting the use of AI. No-one has asked us whether we want any of this.

The absence of public debate in itself represents a problem for the MoD, and even some senior military figures argue that there needs to be greater public discussion on AI and new technologies (albeit that they hope such discussion will result in public support for AI). Air Marshall Johnny Stringer, Deputy Commander of NATO Allied Air Command has warned that if the public and media are not engaged early on with regard to new technology such as military AI, the West could be running a real risk of acquiring capabilities with restrictions.[7] The Defence Artificial Intelligence Strategy rightly notes that "trust is a fundamental, cross cutting enabler of any large-scale use of AI. It cannot be assumed – it must be earned". However, as yet MoD and wider government have made no attempt to earn public trust – let alone consent – on the uses of AI.

[6] 'ALARP "at a glance"'. Health and Safety Executive. **https://www.hse.gov.uk/managing/theory/alarpglance.htm**

[7] Tweet from Tim Robinson, 14 July 2022. **https://twitter.com/RAeSTimR/status/1547526268266102785**

AI, the military, and the tech sector

The strategy also reflects the views of Conservative Ministers that the military should be elevated to the heart of all aspects of society. Defence has "a critical role to play in preparing and pioneering the use of AI to support both national security and economic prosperity", and "should be a natural partner for the UK AI sector". The AI strategy sets out steps to stimulate and support the UK's defence and security 'AI ecosystem' by developing a new Defence and National Security AI Network which will encourage civil sector investment in defence-relevant AI research and development. In part, the strategy is intended to reassure academics and potential civil tech sector partners that the MoD will be a responsible collaborator. As Drone Wars UK has already shown, the MoD already works closely with a number of UK universities, the Alan Turing Institute, and small businesses on the development of applied AI and robotics technologies.[8] One of the roles of the Defence AI Centre will be to develop closer links with AI training institutes, universities and professional bodies to "highlight the opportunities and impact of working in Defence; dispel myths, and catalyse a two-way flow of talent."

Given the government's aim of stimulating the UK's AI sector across the board, there are clear risks associated with attempting to militarise the sector. The approach bears close parallels to China's military-civil fusion strategy,[9] which has drawn criticism from Western governments as "an attempt to deliberately erase the line between China's military and civilian sectors" and for "abusing the access that China's brightest scholars and researchers have earned to universities around the world, or the access successful Chinese tech companies have to the global economy".[10] The AI strategy specifically mentions China's military-civil fusion strategy, warning that research partnerships and dual-use technology agreements with China may contribute directly to the Chinese national security apparatus. At best it is ironic, and at worst perverse, that the MoD is blurring the military and civilian applications of advanced technologies while at the same time discouraging tech sector investment in China on the same grounds. As the White House website points out, "Even if the Chinese Communist Party gives assurances about your technology being confined to peaceful uses, you should know there is enormous risk to America's national security".[11] Why should tech sector innovators or investors concerned about human rights be any more inclined to accept the UK government's assurances that it will only use technology for non-military purposes?

Presumably the government's answer to this question would be that, unlike China, the UK intends to "shape global AI developments to promote security, stability and democratic values", and that it will "shape the development of AI in line with UK goals and values". But what exactly are the UK's values in post-Brexit Britain, when the government is prepared to throw aside international conventions on the treatment of refugees,[12] grant UK military forces exemptions

[8] Drone Wars UK: 'Off The Leash: The development of autonomous military drones in the UK'. 10 November 2018. https://dronewars.net/2018/11/10/off-the-leash-autonomous-drones/

[9] Elsa B. Kania and Lorand Laskai: 'Myths and Realities of China's Military-Civil Fusion Strategy'. Center for a New American Security, 28 January 2021. https://www.cnas.org/publications/reports/myths-and-realities-of-chinas-military-civil-fusion-strategy

[10] Jenny Bavisotto: ' China's Military-Civil Fusion Strategy Poses a Risk to National Security'. DipNote: Military and Security, US State Department, 30 January 2020. https://2017-2021.state.gov/chinas-military-civil-fusion-strategy-poses-a-risk-to-national-security/index.html

[11] US State Department: 'The Chinese Communist Party's Military-Civil Fusion Policy' (undated). https://2017-2021.state.gov/military-civil-fusion/index.html

[12] Full Fact: 'The facts about people seeking asylum in the UK'. 30 November 2021. https://fullfact.org/immigration/asylum-seeker-november-2021/

from international humanitarian law,[13] and dismantle the protections provided by the Human Rights Act?[14] States of all political shades pay lip service to the notion of promoting peace and democracy, but the proof of their faith is in their actions on the international stage, not in fine sounding policy statements. The Environmental and Social Governance panels of tech sector companies, charged with assessing the ethics and impacts of projects that the company might become involved in, can be expected to take a more down-to-earth view of the UK government's commitment to human rights and democratic values than the MoD policy officers who write such platitudes.

Conclusions

Ultimately, the Defence Artificial Intelligence Strategy gives an enlighting but depressing insight into the perspectives and understandings of strategists and ministers at the Ministry of Defence. The justification for the UK's use of AI for military purposes is based on a narrative of maintaining Western dominance – technical, geopolitical, and in warfighting – cloaked in the guise of maintaining the UK's values because, as the Defence Artificial Intelligence Strategy says, "we know that adversaries will use technology in ways that we would consider unethical and unsafe". Autonomous systems are "increasingly the key to the successful generation of overwhelming force in the battle space", and so the military strategy of the future is reduced to a single question about technical force, where machine-system meets machine-system and the largest, fastest, most technologically advanced system will win. The second part of this paper will examine the conflict between MoD's aspirations to move forward with military AI and its stated intention to take an ethical approach to the development of AI.

> **"Regardless of the hype, the reality is that, despite decades of research, AI systems are still too fragile and error-prone to be relied upon in safety-critical applications such as driverless cars, let alone applications relating to the use of military force."**

[13] Freedom From Torture: 'The Overseas Operations Act: What you need to know' (undated). https://www.freedomfromtorture.org/the-overseas-operations-act-what-you-need-to-know

[14] Rajeev Syal: 'Raab urged to let parliament scrutinise Human Rights Act replacement'. Guardian, 21 June 2022. https://www.theguardian.com/law/2022/jun/21/dominic-raab-bill-of-rights-human-rights-act-replacement-letter

Part 2

Ethical, safe, and responsible?

The BBC Reith Lectures, broadcast every Christmas, are a popular series of annual radio lectures given by leading figures of the day to advance public understanding and debate about issues of contemporary interest. In December 2021 audiences around the world tuned in to listen to Dr Stuart Russell, professor of Computer Science at the University of California, Berkeley,[15] discuss how artificial intelligence (AI) might affect how humanity will live in the future.

In the second of four lectures Dr Russell discussed the risks of employing AI systems in warfare. A prominent contributor to the subsequent discussion was Dr Steven Meers, Head of the AI Lab at the Defence Science and Technology Laboratory (DSTL).[16] Responding to a question on how significant moral issues are when looking for AI solutions to military problems, Dr Meers said:

> *"When we are trying to develop future concepts or future countermeasures to autonomous systems within defence, our kind of ethical and responsible approach really is at the forefront of what we do. We have ethicists that we work with who help us guide and develop our approach. In particular, we think very hard about the vulnerabilities and the kind of the misuses of the technologies that we develop, so we focus very hard on kind of responsible and ethical application that really saves lives and tries to reduce harm. So, absolutely it is front and centre of our approach".*[17]

This sounds reassuring – as indeed it is intended to. But to what extent are ethical concerns and risks really at the 'front and centre' of the MoD's artificial intelligence programmes? This section of the paper will outline the measures the MoD is taking to address such issues and critique their effectiveness.

[15] Stuart J. Russell. Berkeley EECS. https://www2.eecs.berkeley.edu/Faculty/Homepages/russell.html

[16] Steven Meers, Linkedin. https://uk.linkedin.com/in/steven-meers-06a71b89?original_referer=

[17] 'BBC Reith Lectures 2021 – Living with Artificial Intelligence'. Lecture 2: The Future Role of AI in Warfare. BBC Radio 4 Transcript, p19. https://downloads.bbc.co.uk/radio4/reith2021/BBC_Reith_Lectures_2021_2.pdf

The context

MoD clearly recognises that ethical issues must be recognised in developing and implementing AI systems, and has made some positive, if unambitious, moves in this respect. The government's stated opposition to the "creation and use" of AI-enabled weapon systems "which operate without meaningful and context-appropriate human involvement throughout their lifecycle" is welcome.[18] MoD evidently accepts that there is a line that should not be crossed in terms of machines making decisions in combat, and that ethical principles are important in deciding where to draw this line. However, as yet the department has given no indication what exactly "context appropriate human involvement" means and how it would be guaranteed in practice when operating at machine speeds. Political scientist Antoine Bousquet has pointed out that "by collapsing the decision-action cycle to fractions of a second, hyperwar will be a type of conflict where human decision-making is almost entirely absent from the observe-orient-decide-act (OODA) loop"[19]. Under these circumstances, how could it be possible to ensure a decision to kill is always made by a human?

On the international front, the UK has recently given cautious support to calls to ensure that weapons system always remain under meaningful human control. In October the UK, along with 69 other states, endorsed a joint statement on Lethal Autonomous Weapons Systems at the UN General Assembly First Committee on Disarmament and International Security.[20] The statement emphasises the necessity for human beings to exert appropriate control, judgement and involvement in relation to the use of weapons systems in order to ensure that any use is in compliance with international law. Along with the other members of the UK Human Rights Council (including all other permanent members of the Security Council) the UK also supported a resolution recognising "the imperative of a human remaining central in the use of force", and asking the Human Rights Council's advisory committee to prepare a study examining the human rights implications of new and emerging technologies in the military domain.[21] While these small steps in the right direction are to be welcomed, they do not make up for the vagueness of the government's position on human involvement in the use of force, which is tantamount to saying "just trust us" to decide what will eventually happen.

MoD's key statement on how it intends to address the maze of ethical issues surrounding the adoption of AI for ethical issues is set out in the policy document 'Ambitious, Safe, Responsible', which was published alongside the Defence Artificial Intelligence Strategy earlier this year. Like Dr Meers' comments at the Reith Lectures, its aim is to reassure and the document outlines a number of aspirations for delivering AI systems which operate in a principled and human way. The document announces that "MoD will develop and deploy AI-enabled systems for purposes that are demonstrably beneficial", including "upholding human rights and democratic values". This is a big claim to make, and the measures laid out in the paper do not give us confidence that the MoD will

[18] 'Autonomous Weapons.' Question for Ministry of Defence. UIN HL2032, tabled on 21 July 2022. https://questions-statements.parliament.uk/written-questions/detail/2022-07-21/HL2032/

[19] Antoine Bousquet: 'The Scientific Way of Warfare: Order and Vhaos on the Battlefields of Modernity'. C. Hurst & Co., 2022. P232

[20] Ousman Noor: '70 states deliver joint statement on autonomous weapons systems at UN General Assembly'. Stop Killer Robots, 21 October 2022. https://www.stopkillerrobots.org/news/70-states-deliver-joint-statement-on-autonomous-weapons-systems-at-un-general-assembly/

[21] United Nations General Assembly: 'Resolution adopted by the Human Rights Council on 7 October 2022. 51/22. Human rights implications of new and emerging technologies in the military domain' https://documents-dds-ny.un.org/doc/UNDOC/GEN/G22/520/64/PDF/G2252064.pdf

succeed in doing this. Rather than providing convincing answers, a lack of detail in the document raises questions about how MoD will actually meet this aim.

MoD's Ethical Principles for AI in Defence

Human-Centricity The impact of AI-enabled systems on humans must be assessed and considered, for a full range of effects both positive and negative across the entire system lifecycle.

Responsibility Human responsibility for AI-enabled systems must be clearly established, ensuring accountability for their outcomes, with clearly defined means by which human control is exercised throughout their lifecycles.

Understanding AI-enabled systems, and their outputs, must be appropriately understood by relevant individuals, with mechanisms to enable this understanding made an explicit part of system design.

Bias and Harm Mitigation Those responsible for AI-enabled systems must proactively mitigate the risk of unexpected or unintended biases or harms resulting from these systems, whether through their original rollout, or as they learn, change or are redeployed.

Reliability AI-enabled systems must be demonstrably reliable, robust and secure.

Principles and panels

The two policy centrepieces of the 'Ambitious, Safe, Responsible' document are a set of ethical principles for the use of AI in defence and the appointment of an Ethics Advisory Panel for MoD to provide independent scrutiny and challenge. The ethical principles (see box) are intended to set the ethical framework which will guide MoD's approach to adopting AI, and will apply to all cases where AI systems are used – "from battlespace to back office" – and across the entire lifecycle of the systems.

According to MoD, these principles will form the core of the UK's approach to creating agreed norms for AI in defence internationally and in "working with partners and allies to shape the global development of AI in the direction of freedom, openness and democracy". They were developed in partnership with the government's Centre for Data Ethics and Innovation following 18 months of consultation with expert stakeholders (although not the broader public or civil society). The MoD's principles are similar to principles adopted by the US Department of Defence[22] and NATO[23] for the ethical use of AI.

Stuart Russell has pointed out that at least 300 sets of AI principles have been drawn up by governments, international organizations, corporations, and professional societies, and that these are "mostly bland platitudes lacking in precision and teeth".[24] While welcome, and not without value, the MoD's ethical

[22] Defense Innovation Board: 'AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense'. 31 October 2019. https://media.defense.gov/2019/Oct/31/2002204458/-1/-1/0/DIB_AI_PRINCIPLES_PRIMARY_DOCUMENT.PDF

[23] NATO: 'Summary of the NATO Artificial Intelligence Strategy'. 22 October 2021. https://www.nato.int/cps/en/natohq/official_texts_187617.htm

[24] Stuart Russell: 'Politicians must prepare for AI or face the consequences'. The House, 25 November 2022. https://www.politicshome.com/thehouse/article/politicians-must-prepare-for-ai-or-face-the-consequences

AI principles as yet have no coherent means of implementation or enforcement. A lack of examples showing how the principles would apply also makes it harder to envisage how they would be put to everyday use. Their successful adoption will depend upon leadership and political and military culture in determining how AI systems are used 'on the ground', and on diplomatic skills in coaxing other nations and tech sector corporates to ensure that their applications of AI are ethically grounded.

Unfortunately the lack of open consultation during the drafting of these principles suggests that they have been prepared on the basis of a limited range of perspectives, and at worst only represent the perceptions of a relatively narrow range of predominantly white male technocrats and military planners. Ordinary people, and most likely marginalised groups and those in the Global South, will ultimately face the consequences and impacts of the MoD's decisions on AI systems, yet the public have not been given a voice in determining any element of the government's AI strategy.

Despite the proliferation of corporate statements on ethical principles, those working in the tech sector are sceptical about the prospect that ethical AI design will be adopted as a norm over the next decade.[25] A non-random poll conducted by the Pew Research Centre in early 2021 found that 68% of experts in the field thought that ethical principles focused primarily on the public good will not be employed in most AI systems by 2030.[26] Their concerns recognised that the main developers and deployers of AI are focused on profit-seeking and social control, and that global competition will matter more to the development of AI than any ethical issues. MoD's aspiration to ensure that its development and use of AI remains ethical will face powerful challenges over the years ahead.

The MoD AI Ethics Advisory Panel, while also to be welcomed, will only be effective if it has teeth. The panel's remit does not extend beyond an advisory role and it has no formal decision-making powers. It appears that there are limits to the areas it will be able to focus on: the 'Ambitious, Safe, Responsible' document states that "the panel has not been involved in the creation of policy relating to Lethal Autonomous Weapons Systems, nor the department's policy on AI safety". There are therefore many questions about the panel's scope and likely effectiveness. Will its advisory role extend to the armed forces as well as the MoD's civilian elements? How will the panel's advice be integrated into day-to-day AI research, development, and application? Can it consult more broadly on particularly sensitive issues? And what happens if MoD's military leadership or ministers disagree with the panel's advice?

MoD has published details of the AI Ethics Advisory Panel's membership. It would be too cynical to say that the panel has been packed with industry and government representatives, but while there are good people and strong voices on the panel, its members are largely 'government friendly' and drawn from within the AI and defence establishments – 'safe pairs of hands', in other words. Radical and critical voices are few and far between. The panel should insist upon openness and transparency not just on membership but also other elements of its work – for example, meetings should be open to non-members to attend and ask questions, and terms of reference and records of meetings should be published.

[25] Sebastian Klovig Skelton: 'AI experts question tech industry's ethical commitments'. ComputerWeekly.com, 31 October 2022. https://www.computerweekly.com/feature/AI-experts-question-tech-industrys-ethical-commitments

[26] Lee Rainie, Janna Anderson and Emily A. Vogels: 'Experts Doubt Ethical AI Design Will Be Broadly Adopted as the Norm Within the Next Decade'. Pew Research Center, 16 June 2021. https://www.pewresearch.org/internet/2021/06/16/experts-doubt-ethical-ai-design-will-be-broadly-adopted-as-the-norm-within-the-next-decade/

Ducking and diving

Despite the good intentions of 'Ambitious, Safe, Responsible', its thrust is considerably blunted by a significant number of hedges and qualifications throughout the text and statements which undermine the importance of an ethical framework. Comments such as "it is imperative that we deliver maximum effectiveness and efficiency using AI across the spectrum of Defence activities" – on the opening page of the document – do not give the reader confidence that ethics will be a priority in delivering AI systems. In places the qualifications are sufficient to completely negate the original good intentions: for example, the claim that:

> *We can exert satisfactory and rigorous human control over AI-enabled systems without always requiring some form of real-time human supervision. Indeed doing so may act as an unnecessary and inappropriate constraint on operational performance. For example, to defend a maritime platform against hypersonic weapons we may need defensive systems which can detect incoming threats and open fire faster than a human could react.*

This statement underplays the high risks already associated with automated air defence systems, where failings have led to a number of tragic incidents when the civilian airliners Iran Air Flight 655 (1988), Malaysian Airlines MH 17 (2014), and Ukrainian Airlines PS752 (2020) have been shot down.[27] At times it feels as if a senior civil servant, worried about policies that might impede the implementation of military AI programmes, has gone through the document with a red pen at a late stage in the editing process and added a set of restrictions to make sure that nothing in it might stop the MoD from doing whatever it wants with AI systems.

Despite recognising – and making much of – the risks arising from the use of AI systems for military applications, the MoD's proposals for addressing these risks are less than satisfactory, and are surprisingly complacent. 'Ambitious, Safe, Responsible' asserts that "this is not a new challenge for the Department. Defence is bound by UK law and has a robust regime for compliance. Defence activities also include those that are inherently dangerous and require additional risk management beyond that of our statutory obligations". This ignores the facts that regulation of AI systems will be intrinsically different to anything that MoD regulates at present, and will raise a host of new issues, and that MoD's current performance at regulating high hazard activities not particularly good.[28] Proposals to appoint 'Accountable Officers' responsible for oversight of AI activity, to provide technical oversight from the Defence Artificial Intelligence Centre, and to explore how the Defence Safety Authority will consider AI are all very general and aspirational.[29] The bottom line is that the MoD's current AI activities are currently unregulated and the department as yet has no concrete proposals for regulating them.

There are parallels here with the government's position on the use of other new high-risk technologies. Police forces in several areas are implementing the use of controversial live facial recognition technology for public surveillance despite

[27] Ingvild Bode and Tom Watts: 'Meaning-less human control: Lessons from air defence systems for lethal autonomous weapons'. Drone Wars UK and University of southern Denmark, 19 February 2021. https://dronewars.net/2021/02/19/meaning-less-human-control-lessons-from-air-defence-systems-for-lethal-autonomous-weapons/

[28] See, for example: The Ferret: 'Defence Nuclear Safety Regulator'. https://theferret.scot/tag/defence-nuclear-safety-regulator/

[29] 'Defence: Artificial Intelligence'. Question for Ministry of Defence. UIN HL2088, tabled on 5 September 2022. https://questions-statements.parliament.uk/written-questions/detail/2022-09-05/HL2088

a 2020 Court of Appeal ruling that such technology breached privacy rights and broke equalities laws.[30] No legislation exists which explicitly authorises police use of live facial recognition technology and data regulators do not have explicit authorisation to limit use of the technology.[31] Despite a 2019 recommendation from the House of Commons Science and Technology Committee that a legislative framework to control use of such technology should be introduced none of the governments in power since then have done anything to introduce such protections.[32] Government and regulators are either unwilling or unable to stop the proliferation of facial recognition technology. In the same way, the uses of AI, including for warfighting, are developing without any proper framework of scrutiny or 'red lines', and Parliament is powerless to do anything about it.

To explain the MoD's approach to the development of AI, 'Ambitious, Safe, Responsible' presents a bizarre analogy likening the development of AI to electrical appliances. It asserts that "Initial planning would not normally hinge on discussions about whether the likely incorporation of electricity meant that the system could be delivered safely and responsibly," and stating that this is similar to the way the MoD thinks about AI. This is another way of saying that opaque, error-prone, and unsafe technology should be tolerated as part of the glorious crusade onwards to a bright AI-driven future. Rather than starting from the position of seeking to defend people and their personal data from the risks posed by AI, the MoD is starting from the position that 'we're determined to develop AI, so how can we do that?'. The hazards and accidents that result from this approach[33] are not inevitable, but are the predictable and avoidable consequences of choices made by tech sector companies who roll out immature and dangerous products and governments who fail to regulate the sector.

Autonomous weapon systems: policy missing in action

Perhaps the place where 'Ambitious, Safe, Responsible' shows its weaknesses most glaringly is in its position statement on lethal autonomous weapon systems – emerging weapons systems which would be able to identify and destroy targets without human involvement. The use of autonomous weapon systems would raise major ethical concerns. Ministers are keen on the adoption of autonomous technologies by the armed forces, with James Heappey, Minister of State at the Ministry of Defence, recently telling the House of Commons that autonomy is key to generating "overwhelming force on the battlefield" and "a more lethal force – even a bigger force" without necessarily requiring more personnel.[34] The government has to date refused to support calls for a ban on autonomous weapon systems citing two reasons. It claims that "International Humanitarian Law already provides a robust, principle-based framework for the regulation of development and use of all weapons systems including weapons that contain autonomous functions".[35] It also argues that "Without international

[30] Liberty: 'Legal Challenge: Ed Bridges v South Wales Police', undated. https://www.libertyhumanrights.org.uk/issue/legal-challenge-ed-bridges-v-south-wales-police/

[31] Peter Fussey and Daragh Murray: 'Policing Uses of Live Facial Recognition in the United Kingdom'. In 'Regulating Biometrics: Global Approaches and Urgent Questions', Ed. Amba Kak, p78-85. AI Now Institute, 2020. https://ainowinstitute.org/regulatingbiometrics-fussey-murray.pdf

[32] House of Commons Science and Technology Committee: 'The work of the Biometrics Commissioner and the Forensic Science Regulator. 18 July 2019. P21, para 8. https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/1970/1970.pdf

[33] See the AI Incident Database, https://incidentdatabase.ai/

[34] 'Armed Forces: Size Targets'. House of Commons, 18 July 2022. https://www.theyworkforyou.com/debates/?id=2022-07-18c.688.2

[35] 'Autonomous Weapons: Treaties'. Question for Ministry of Defence. UIN HL2033, tabled on 21 July 2022. https://questions-statements.parliament.uk/written-questions/detail/2022-07-21/HL2033

consensus on the definitions or characteristics of weapons with levels of autonomy, a legal instrument would have to ban undefined systems, which would present difficulties in the application of any such ban and which could severely impact legitimate research and development of AI or autonomous technologies".

This position reflects a familiar hostility within the current government to regulation, which should come as no surprise to observers of the UK's performance in international forums. However, the government offers no measures for controlling this potentially dangerous, rapidly emerging technology other than aspirations to work with partners to build norms of use and positive obligations on how autonomy in weapons systems should be developed, and to build understanding, best practice, and codes of conduct on their use. Alongside this is a whine that "global governance for such systems is a difficult task". Global governance is always a difficult task, but particularly so if you have no commitment to it and no positive vision for the future. The government's position on lethal autonomous weapons is akin that of the US gun lobby to proposals for controlling firearms: that weapons don't need control, and it is OK to allow 'responsible' humans to use them (where we define who is 'responsible').

Although the position statement on lethal autonomous weapons systems mentions that the UK hopes to use codes of conduct to control the use of such weapons, it is notable that the statement makes no mention of the eleven guiding principles on emerging technologies in the area of lethal autonomous weapons which were agreed by the UN Group of Government Experts on Certain Conventional Weapons at the end of 2019.[36] Adoption of these principles was strongly supported by the UK, and at the time regarded as a significant diplomatic achievement. That the 'Ambitious, Safe, Responsible' document does not even mention the guiding principles tells us all we need to know about the effectiveness of voluntary codes of conduct.

[36] High Contracting Parties to the Convention on Certain Conventional Weapons: 'Guiding Principles affirmed by the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System'. CCW/MSP/2019/9, 2019. https://www.ccdcoe.org/uploads/2020/02/UN-191213_CCW-MSP-Final-report-Annex-III_Guiding-Principles-affirmed-by-GGE.pdf

Beyond 'Ambitious, Safe, Responsible'

In order to find out how MoD's ethical appraisal process works in practice, Drone Wars UK submitted a Freedom of Information Act request[37] to DSTL asking for copies of their procedures and guidelines for conducting ethical assessments of artificial intelligence research projects. The reply stated that:

- DSTL follows the Data Ethics Framework[38] regarding appropriate and responsible data use in government;
- Any AI research involving human participants complies with Joint Service Publication 536;[39]
- DSTL conducts and funds research to deepen understanding of the ethical issues and approaches associated with Defence applications of AI, for example this recent paper co-authored by DSTL staff and researchers from the Alan Turing Institute.[40] Please note this is a research paper and does not necessarily represent MoD policy;
- DSTL has supported MoD in developing a set of approaches for responsible AI adoption. As outlined in the National AI Strategy,[41] the MoD will shortly set out the details of the approaches by which Defence AI is developed and used. These will be shortly published via gov.uk but currently represent policy under development and given the imminent publication and the increased risk of uncertainty when they are launched, we assess release would not be in the public interest at this stage. [This presumably related to the Defence AI Strategy and 'Ambitious, Safe, Responsible' paper, which at the time of DSTL's reply had yet to be published.]

While helpful, this does not represent a clear protocol for applying ethical standards to research projects and is vague as to how any appraisal works in practice. We asked for a 'worked example' from a real-life DSTL project: the 'intelligent ship' project which DSTL has supported to develop a network of multiple AI systems to revolutionise decision-making on a warship.[42] In reply, DSTL told us:

> *With reference to your request for: "A copy of the ethical assessment conducted for DSTL's 'intelligent ship' project", we can confirm no information is held. Under Section 16 of the Act (Advice and Assistance) we can advise that the research within DSTL's Intelligent Ship Project has not to date required a formal ethical assessment, but has conformed to the policies and principles outlined above.*

[37] Defence Science and Technology Laboratory: Response to Freedom of Information Act request FOI 2021/15482, 25 March 2022: **https://dronewars.net/wp-content/uploads/2022/11/DSTL-FOI-letter-250322.pdf** Following publication of the Defence AI strategy, DSTL confirmed that its ethical framework remained unchanged in Response to Freedom of Information request FOI 2022/12671, 22 November 2022: **https://dronewars.net/wp-content/uploads/2022/11/FOI-2022-12671_Redacted.pdf**

[38] Central Digital and Data Office: 'Data Ethics Framework'. 13 June 2018. **https://www.gov.uk/government/publications/data-ethics-framework**

[39] Ministry of Defence: 'Defence research involving human participants (JSP 536)'. 15 September 2016. **https://www.gov.uk/government/publications/defence-research-involving-human-participants-jsp-536**

[40] Mariarosaria Taddeo, David McNeish, Alexander Blanchard, and Elizabeth Edgar: 'Ethical Principles for Artificial Intelligence in National Defence'. Philosophy & Technology 34:1707 1729, 13 October 2021. **https://link.springer.com/content/pdf/10.1007/s13347-021-00482-3.pdf**

[41] Office for Artificial Intelligence, Department for Digital, Culture, Media & Sport, and Department for Business, Energy & Industrial Strategy: 'National AI Strategy'. 22 September 2021. **https://www.gov.uk/government/publications/national-ai-strategy**

[42] Defence Science and Technology Laboratory, Defence and Security Accelerator, and Ministry of Defence: 'Competition document: Intelligent Ship Phase 2'. 29 June 2020. **https://www.gov.uk/government/publications/competition-intelligent-ship-phase-2/competition-document-intelligent-ship-phase-2**

This answer came as some surprise, as the project is a high-profile initiative which has been described as technologically high risk and which, dealing with topics such as mission execution and power and propulsion systems, raises safety issues.[43] The project is at a relatively early stage, but to be effectively planned for and dealt with, ethical issues need to be considered at an early stage in development programmes. DSTL apparently holds no documentation relating to how ethical factors have been addressed during the 'intelligent ship' process, yet documenting is a key function of the ethical appraisal process: recording why decisions are made, what data has been used and for what, and what procedures have been followed for approving decisions, rules, or instructions that determine how an algorithm functions.

The approach taken to the ethics of the intelligent ship project certainly did not seem to match up to Steven Meers' assertions that DSTL "focus very hard on kind of responsible and ethical application that really saves lives and tries to reduce harm" and that an ethical and responsible approach "really is at the forefront of what we do". As Kate Crawford, author of the 'Atlas of AI', has observed: "The great majority of university-based AI research is done without any ethical review process".[44] Despite the hype, it is not clear how the MoD's ethical assessments are any improvement on this.

What needs to change?

While recognising the ethical and practical hazards of AI, the MoD's approach to tackling these hazards is conservative, unambitious, and lacking in commitment. MoD apparently seems to think it has now 'ticked a box' on its path towards implementing AI technologies: it can now say that it has an ethical approach to AI, and use this as a get-out-of-jail-free card when it is challenged on its applications of AI. However, close scrutiny of the MoD's AI strategy documents raises serious questions about its approach to the governance of AI and digital technologies. It also points to a deep conflict between the government's stated democratic values rooted in human rights on the one hand, and a technocratic impulse to race forward with AI at all costs on the other.

Given that the government is not going to change direction on its AI strategy and military aspirations for AI, what might be done to hold MoD to account on its pledges that it will operate a responsible AI programme? One possibility might be to set up an Parliamentary oversight committee, akin to the Intelligence and Security Committee, to scrutinise and challenge the use of AI by government departments, including the military. To be effective, this would require Parliamentarians to demonstrate independence and insistence to a degree which has hitherto been absent when reviewing the activities of the armed forces. Likewise, sharper critique and coverage from journalists and the media would also be welcome, but is unlikely to happen. In the absence of broader interest, civil society organisations such as the Campaign to Stop Killer Robots and the Ada Lovelace Institute will doubtless continue their excellent work to provide the necessary scrutiny, despite the meagre resources available to them.

[43] Harry Lye: 'Machine learning for the future fleet: Dstl's Intelligent Ship'. Global Defence Technology, February 2021. https://defence.nridigital.com/global_defence_technology_feb21/ai_dstl_intelligent_ship

[44] Kate Carwford: 'Atlas of AI'. Yale University Press, 2021.

Looking more broadly, further steps will be needed to protect the public if AI is to become the technology which shapes our future. These could include the following proposals, none of which will find favour with the government, tech sector, and research interests which are driving the AI juggernaut.

- Introduce an Artificial Intelligence Act to regulate the technology and ban the most harmful applications, as the European Union is currently doing.[45] There should be no exemptions for military applications – potentially the most harmful and high-risk applications of all.
- Establish an Artificial Intelligence regulatory agency at arms-length from government. This should have real teeth and should cover all sectors, including the military. MoD's current internal regulators operate to lower standards than their civilian cousins, are less transparent, and are not independent.
- Clearly state that the UK will ensure that its weapon systems will always be under meaningful human control and that the UK supports a binding international treaty to incorporate such a requirement into international humanitarian law.
- Establish a Parliamentary committee to investigate the implications of the development of autonomous weapon systems and make recommendations.
- Establish a 'digital hippocratic oath', to be administered by professional institutes and universities, committing computer scientists to uphold ethical principles.

There are even more serious issues to consider when thinking about the ethics of AI. Ultimately, the widespread application of AI raises huge questions for humanity. Carbon emissions, for example, from training large machine-learning systems are substantial, while AI systems demonstrate intrinsic racial, gender and ethnic biases. The motives for developing AI – societal control, warfighting, and pursuit of monopoly profit – represent the worst of human traits. As autonomous systems increase in complexity, we can expect a corresponding decrease in our ability to predict and control such systems. In the light of these questions, can AI ever really be considered ethical, safe, or responsible?

> **"The document announces that "MoD will develop and deploy AI-enabled systems for purposes that are demonstrably beneficial", including "upholding human rights and democratic values". This is a big claim to make, and the measures laid out in the paper do not give us confidence that the MoD will succeed in doing this."**

[45] For more information about the EU Artificial Intelligence Act see the independent website 'The Artificial Intelligence Act'. https://artificialintelligenceact.eu/

DRONE WARS

Shining a spotlight
on military drones