EUROPEAN LEADERSHIP NETWORK

UK thinking on AI integration and interaction with nuclear command and control, force structure, and decision-making

Alice Saltini

November 2023

The European Leadership Network (ELN) is an independent, non-partisan, pan-European NGO with a network of over 300 past, present and future European leaders working to provide practical real-world solutions to political and security challenges.

This project is funded by the Bureau of Arms Control, Verification and Compliance (AVC).

This paper is one of four bibliographies commissioned by the ELN on Chinese, French, Russian, and British perspectives on AI integration in nuclear decision-making, from a range of non-governmental experts. It is part of the ELN's project "Examining the impact of artificial intelligence on strategic stability: European and P5 perspectives".

About the author

Alice Saltini
Research coordinator for the European Leadership Network

As the Research Coordinator at the European Leadership Network (ELN), Alice Saltini is actively involved in a range of projects within the Global Security Program that include an examination between the interplay of AI and nuclear risks. With a keen interest in AI regulation at the intersection with nuclear systems, Alice is dedicated to developing policy solutions in this field, contributing constructively to ongoing discussions.

Believing in the importance of transparency in AI, Alice seeks to understand and address challenges presented by the potential implications of advanced AI models for nuclear command and control systems. Her insights have found a platform in a number of multilateral fora, such as the NPT.

Prior to joining the ELN, Alice interned for the Comprehensive Nuclear-Test-Ban Treaty Organization in the External Relations, Protocol and International Cooperation Section, and worked as a Research Assistant at the James Martin Center for Nonproliferation Studies. Alice is also a recent cohort of the CTBTO Youth Group's CTBTO-CENESS Research Fellowship 2022.

She holds a Master's degree in Russian studies and a Post Graduate Certificate (Pg Cert) in Nonproliferation Studies from the Middlebury Institute of International Studies, benefiting from a full merit-based scholarship during her tenure at the Middlebury Institute.

Combining her academic background with practical experience, Alice hopes to foster informed dialogues about the integration of AI in the nuclear domain.

1. Introduction

The integration of artificial intelligence (AI) in the military and defence sector has kicked off a new era of military competition among nuclear-armed states, particularly the P5. The exploration of AI capabilities has emerged as a pathway to national and military power, driving the P5 states to pursue AI for military and defence purposes. As AI applications in the military domain evolve, P5 states are increasingly considering the benefits and risks of application in their nuclear command, control, and communication (NC3) systems.[1]

While AI integration in the military domain offers the potential for enhancing strategic stability, it also presents threats to UK national security.[2] Adversaries, if equipped with AI capabilities, may seek information superiority and exploit AI unethically. Ensuring the ethical and legal use of AI for military and defence purposes, while maintaining human control and accountability over AI-powered weapons systems, has become a priority for the UK Government.[3] However, vulnerabilities stemming from technological limitations inherent to AI, as well as issues underpinning human-machine interaction, raise concerns about the consequences these models may entail.

Given these considerations, this review paper compiles and analyses the British literature on the UK's perception of military and nuclear applications of AI and their impact on strategic stability and NC3. The paper assesses the UK's debates on strategic opportunities and risks, examining the development of AI-enabled systems in defence and NC3. It also explores risk mitigation measures identified by scholars and the UK Ministry of Defence (MOD), with a particular focus on the concept of 'safe and responsible AI.' Additionally, the paper offers recommendations for unilateral measures that the UK can take, as well as multilateral initiatives within the P5 framework, to address the risks associated with AI in nuclear decision-making.

This assessment draws from various sources, including official documents such as the UK's 'Defence AI Strategy' and the 'Ambitious, safe and responsible: our approach to the delivery of AI-enabled capability in Defence' policy paper. Additionally, it incorporates insights from the UK NGO communities, official statements, and other openly available documents and papers related to AI and strategic stability.

In particular, this paper seeks to:

1. Analyse the UK's official stance on AI integration in military and nuclear systems;

2. Collect and analyse open-source UK literature that focuses on the integration of AI in military systems, with a specific emphasis on NC3 and decision-making systems;

3. Examine the UK's role in mitigating risks associated with AI and its military applications;

4. Analyse how the internal debate on AI unfolds within the UK, including how it swings or aligns between official sources and independent experts;

5. Explore additional measures that the UK can adopt to address a broader risk reduction perspective, specifically considering nuclear implications.

Ensuring the ethical and legal use of AI for military and defence purposes, while maintaining human control and accountability over AI-powered weapons systems, has become a priority for the UK Government.

2. Initial observations

AI is defined by the UK MOD as “a family of general-purpose technologies, any of which may enable machines to perform tasks normally requiring human or biological intelligence, especially when the machines learn from data how to do those tasks”.[4]

Since the mid-2010s, AI technologies have gained substantial attention in British politics, with the Government recognising the significance of maintaining a leading position in AI development.[5] Even before a thorough understanding of the potential of AI had been achieved, funding and attention were already being directed towards the field from government and academic sources.[6]
In June 2022, the MOD released the UK’s ‘Defence Artificial Intelligence Strategy’ and a policy paper on the ‘Ambitious, safe and responsible’ use of AI, underscoring the priorities set out by the UK in the development of AI systems in the military and defence sector:

> “Our vision is that, in terms of AI, we will be the world’s most effective, efficient, trusted and influential Defence organisation for our size:
>
> - Effective – through the delivery of battle-winning capability and supporting functions, and in terms of our ability to collaborate and partner with the UK’s allies and AI ecosystem;
> - Efficient – through innovative use of technology to deliver capability, conduct operations, and realise productivity benefits across our organisation;
> - Trusted – by the public, our partners and our people, for the safety and reliability of our AI systems, and our clear commitment to lawful and ethical AI use in line with our core values;
> - Influential – in terms of shaping the global development of AI technologies and managing AI-related issues to positive ends, working collaboratively and leading by example.”[7]

According to the MOD strategy, the interest generated by military applications of AI stems from the potential disruption caused by emerging technologies. According to the ‘Defence AI Strategy’, “These technologies, and the operational tempo they enable, are likely to compress decision times dramatically, tax the limits of human understanding and often require responses at machine speed”.[8]

Based on UK open-source literature, some general observations can be drawn:

1. Although a considerable portion of British literature on AI predominantly focuses on its impact on the military, there is a relatively limited amount of research that considers the implications of AI for strategic stability and nuclear decision-making. However, the most recent documents released by the MOD, namely the ‘Defence AI Strategy’ and the ‘Ambitious, safe and responsible: our approach to the delivery of AI-enabled capability in Defence’ policy paper, provide insights into how AI can be leveraged to enhance UK defence capabilities. These

Although a considerable portion of British literature on AI predominantly focuses on its impact on the military, there is a relatively limited amount of research that considers the implications of AI for strategic stability and nuclear decision-making.

documents outline the current strategy being pursued by the UK in the integration of AI in the defence sector. Experts and scholars within the UK actively contribute to the ongoing debate on the impact of AI on nuclear decision-making and strategic stability by conducting analyses of the associated risks and opportunities, with a particular focus on studying potentially destabilising effects.

2. When it comes to military applications of AI, official documents acknowledge the potential risks associated with its use, but they often lack a thorough analysis of these risks and their broader implications. Instead, the focus is primarily on harnessing the opportunities presented by AI to maintain the security, stability, and democratic nature of the UK.

> "AI has enormous potential to enhance capability, but it is all too often spoken about as a potential threat. 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. We also recognise that the use of AI in many contexts, and especially by the military, raises profound issues. We take these very seriously – but think for a moment about the number of AI-enabled devices you have at home and ask yourself whether we shouldn't make use of the same technology to defend ourselves and our values."[9]

The MOD emphasises the importance of using AI in a responsible, safe, and ethical manner, ensuring human oversight, accountability, and control. The 'Defence AI Strategy' and 'Ambitious, Safe, and Responsible' policy paper provide guidelines for legal, ethical, and responsible usage of AI in military operations.

On the other hand, non-official documents explore the implications of AI in the military domain in greater detail. One particularly prominent issue is the debate surrounding automation, specifically in the context of autonomous weapons, also known as lethal autonomous weapons (LAWS). This topic has sparked fervent public debate, with civilian initiatives calling for a moratorium on the development of weapons systems that can engage targets without direct human control.[10] Numerous initiatives led by both the public and parliamentarians aim to investigate the risks associated with AI in the military and develop mitigation measures. Some notable examples include:

- An All-Party Parliamentary Group on Artificial Intelligence (APPG AI), established in 2017 and led by members of the House of Commons and House of Lords. It specifically focuses on exploring the economic, social, and ethical implications of developing AI, including its military applications.

- The House of Lords Liaison Committee established a 'Special Committee on the Use of Artificial Intelligence in Weapon Systems' in 2023. The committee's objective is to scrutinise the risks and prospects associated with the deployment of autonomous weapons. Nevertheless, the topic of nuclear weapons, especially NC3, was also discussed during at least one oral evidence session, in June 2023. Importantly, experts specialising in the overlap of AI and NC3 were invited to share their insights at the Committee.

- The UK campaign to stop killer robots (part of the global Stop Killer Robots campaign), which is led by UK-based organisations and civil society representatives who aim to promote laws governing autonomy in weapon systems.

3. When it comes to AI applications in the context of nuclear weapons, official documents assert the necessity for human political control over nuclear weapons at all times.[11] However, they do not explore the potential for AI adoption specifically for nuclear weapons systems. The 'Ambitious, Safe and Responsible' policy paper, while discussing AI-enabled weapons in general, provides only a superficial examination of AI implications. Similarly, the 'Defence AI Strategy' touches on the impact of AI on strategic stability and nuclear weapons, but lacks in-depth analysis.

   In contrast, UK-based experts identify risks associated with the use of AI in nuclear decision-making and offer a much greater degree of detail and extensive analysis of the risks and implications.[12] However, it's worth noting that public debate on automation in nuclear decision-making is notably limited compared to the debate on LAWS. Moreover, discussion about the role of AI in nuclear weapons is mostly dominated by a select group of UK experts.

   Representatives from UK-based institutions have significantly contributed to the discourse on this subject. Among them, Dr James Johnson from the University of Aberdeen has published extensively on this topic and is notable for his book 'AI and the Bomb', published in February 2023. This work considers how AI affects nuclear strategy, especially in the context of decision-making. Other academic institutions or NGOs active in the field are King's College London's Centre for Science and Security Studies, the European Leadership Network, and Chatham House. Additionally, other academic institutions in the UK, such as the University of Cambridge's Centre for the Study of Existential Risk, the Oxford Martin School, the Centre for Governance of AI emphasise the risks and security implications arising from AI, including its governance.

   While there are non-UK-based NGOs like SIPRI and the Nuclear Threat Initiative (NTI) contributing valuable English-language insights into the subject, they are outside the scope of this paper due to their non-UK origins.

3. AI applications in the military domain

The UK recognises the strategic opportunities and challenges presented by the emergence of AI as a transformative and potentially disruptive new technology. The UK views AI integration in military systems as a strategic priority and an opportunity to gain an edge over adversaries, but also as a responsibility to use ethically and legally, particularly in the context of automation in weapons systems.[13] The MOD aims to adopt and utilise AI to enable strategic and operational advantages, such as for better-informed decision-making and quicker responses to threats. It seeks to strengthen the defence AI ecosystem through collaborations with various sectors and shape global AI developments to promote security, stability, and democratic values.

Official documents indicate that the UK is actively applying AI technologies in warfare to predict adversaries' behaviour, support decision-making processes, and enhance situational awareness. AI is also seen as particularly valuable in nuclear-related intelligence, surveillance, and reconnaissance (ISR) operations. These documents highlight the advantages of AI in data collection and analysis, as well as its ability to accelerate decision-making. However, there is also an awareness of the risks and vulnerabilities AI poses, particularly in case of system's malfunction, but also due to problems generated by the interaction between humans and machines (such as automation bias).

Overall, while official documents acknowledge the risks of AI integration in military systems, the potential benefits and advantages it offers over adversaries (such as for faster data analysis and collection) appear to be given greater importance. Thus, the MOD strategy aims to address the challenges of AI to maximise its benefits in a competitive context. Non-official documents, on the other hand, tend to adopt a more cautious approach, exploring the inherent risks of AI. Experts and analysts highlight the dual nature of AI, capable of simultaneously enhancing and weakening strategic stability by increasing and reducing decision-making time and situational awareness.[14]

> **While official documents acknowledge the risks of AI integration in military systems, the potential benefits and advantages it offers over adversaries appear to be given greater importance.**

AI benefits

As a result, the opportunities presented by AI underscore the need for the UK to maintain its leadership in the field to promote stability.[15] AI is identified by the MOD as an essential element for future military capabilities, with a broad spectrum of applications that promise to significantly increase efficiency, precision, and effectiveness in defence operations. According to the MOD, AI offers significant enhancements across various domains from administrative processes to combat operations.

The following points are identified by the UK literature as *benefits* provided by AI-military applications.

1. Achieving decision advantage

- Increasing the quality of decision-making and allow for rapid responses to threats.

2. Improving efficiency

- Enhancing the speed of processes and support functions.

3. Unlocking new capabilities

- Improving AI-enabled cyber defence;
- Enhancing the persistence, reach and mass effectiveness of military forces;
- Protecting military personnel from physical harm;
- Contributing to defence intelligence.

4. Empowering the military force

- Focusing human talents on higher value functions requiring ingenuity, contextual thinking and judgement.

“Defence applications for AI stretch from the corporate or business space - the ‘back office’ - to the frontline: helping enhance the speed and efficiency of business processes and support functions; increasing the quality of decision-making and tempo of operations; improving the security and resilience of inter-connected networks; enhancing the mass, persistence, reach and effectiveness of our military forces; and protecting our people from harm by automating ‘dull, dirty and dangerous’ tasks”.[16]

“Even if AI-augmented weapons and systems are unable to produce better decisions than humans, militaries that use AI will doubtless gain significant advantages on the battlefield (e.g. remote-sensing, situational-awareness, battlefield-manoeuvre, and a compressed decision-making loop), compared to those who depend on human judgment alone; in particular, in operating environments that demands endurance and rapid decision-making across multiple combat zones”.[17]

“The trial[18] demonstrated the potential for AI to quickly process vast quantities of data, providing commanders with better information during critical operations and transferring the cognitive burden of processing data from a human to a machine”.[19]

“Integrated multi-disciplinary delivery teams will be at the heart of our approach to AI adoption and the development of effective Human-Machine Teaming, combining human cognition, inventiveness and responsibility with machine-speed analytical capabilities”.[20]

“Existing sensors collect far too much data to sift through manually, especially when operators must make critical decisions quickly, such as offensive counter air and defensive counter air missions. The approach helps to synthesise oceans of data into actionable intelligence and accurate targeting information at speed and on a greater scale. This enables the faster and more accurate collection and synthesisation of data and facilitates more informed command and control decisions”.[21]

“There are a range of other emerging technologies that present opportunities to support compliance and verification regimes, including distributed ledger technology for nuclear materials control, image recognition for verification activities, metadata for geolocation, AI and synthetic environments for improved military planning and wargaming”.[22]

"There are many AI capabilities and applications that the MOD views as being essential for future military capabilities, including the following:

- **Improving coverage of the battlefield and automating information processing and management cycles.** These capabilities include unmanned automated ISR platforms and software that can 'pre-filter, fuse and classify all data flows, eliminate paralysing information overload, and accelerate the observe, orient, decide and act (OODA) loop of decision-makers'.

- **Making the logistics chain more agile and less manpower intensive.** These include self-driving transport vehicles and automated logistic monitoring software.

- **Increasing the persistence, reach, mass and precision of weapon systems.** Notably, these include loitering munitions, an automated technology that could deliver 'step changes in military capability'.

- **Enhancing the capability to fight cyberwar**".[23]

Adversary use of AI

From the MOD perspective, AI is viewed as a potential threat when in the hands of adversaries. This is because adversaries could utilise AI to gain information superiority and a strategic advantage. Moreover, there is concern that adversaries may employ AI in unethical ways, which would run counter to the core values upheld by liberal democracies.[24] The following points outline the challenges associated with AI when used by adversaries in military applications:

1. **Heightening threats above and below the threshold of armed conflicts:**

- Enhancing high-end military capabilities

- Enhancing less sophisticated products

2. **Unethical use of AI from states or non-state actors:**

- Undermining confidence in AI performance

- Enhancing cyber and information warfare

> "The use of AI by adversaries will heighten threats above and below the threshold of armed conflict. AI has potential to enhance both high-end military capabilities and simpler low-cost 'commercial' products available to a wide range of state and non-state actors".[25]
>
> "Artificial Intelligence has the potential to significantly increase the impact of malicious cyber attacks, potentially probing for and exploiting cyber vulnerabilities at a speed and scale that is impossible for human monitored systems to defend against".[26]

From the MOD perspective, AI is viewed as a potential threat when in the hands of adversaries. This is because adversaries could utilise AI to gain information superiority and a strategic advantage... [or] may employ AI in unethical ways.

> “Adversary appetite for risk suggests they are likely to use AI in ways that we would consider unacceptable on legal, ethical or safety grounds. Equally, adversaries will use a range of information, cyber and physical means to attack our AI systems and undermine confidence in their performance, safety and reliability (e.g. by ‘poisoning’ our data, corrupting hardware components in our supply chain, or interfering with communications and commands)”.[27]

> “Potential threats include enhanced Cyber and information warfare, AI-enabled surveillance and population control, accelerated military operations and the use of autonomous physical systems. Non-state actors are seeking to weaponise advanced commercial products to spread terror and hold our forces at risk. As these case studies illustrate, this is not a hypothetical future but the here and now”.[28]

AI integration challenges

The integration of AI into weapons systems entails intrinsic risks. While the MOD acknowledges the potential risks associated with AI integration in weapon systems, official documents often do not provide a comprehensive analysis of these risks and their implications.[29] However, scholars and analysts have extensively examined these risks.

The UK acknowledges that the development and deployment of AI systems have the potential to perpetuate biases. AI’s potential risks, such as unpredictable military applications, necessitate clear accountability and well-defined responsibility across the entire lifecycle of these systems. The unpredictable nature of AI, especially in new environments, increases the risk of unintended consequences.

The following points outline the challenges encountered in AI integration into weapons systems. Some of the risks include:

1. **Introducing algorithmic bias**

- Facilitating discriminatory outcomes;
- Leading to disproportionate harms for certain groups of users.

2. **Obscuring lines of responsibility and accountability**

- Widening the responsibility gap between systems that take decisions or make recommendations, and the human commanders responsible for them.

3. **Furthering unpredictability**

4. **Presenting unintended consequences and incentives:**

- Enabling certain incentives;
- Influencing other systems beyond their intended effect;
- Leading to misunderstanding, miscalculation or uncontrolled escalation.

5. **Introducing doubts on safety and reliability:**

- Ensuring the safety and reliability of AI-enabled weapons systems to prevent unintended consequences or malfunctions.

6. **Maintaining human control:**

- Avoiding the risk of delegating critical decisions to fully autonomous systems without human involvement.

> “AI systems make decisions based on the data they have been trained on. If that data - or the system it is embedded in - is not representative, it risks perpetuating or even cementing new forms of bias in society. It is therefore important that people from diverse backgrounds are included in the development and deployment of AI systems”.[30]
>
> “AI-enabled systems offer significant benefits for Defence. However, the use of AI-enabled systems may also cause harms (beyond those already accepted under existing ethical and legal frameworks) to those using them or affected by their deployment. These may range from harms caused by a lack of suitable privacy for personal data, to unintended military harms due to system unpredictability. Such harms may change over time as systems learn and evolve, or as they are deployed beyond their original setting. Of particular concern is the risk of discriminatory outcomes resulting from algorithmic bias or skewed data sets.”[31]
>
> “Responsibility for each element of an AI-enabled system, and an articulation of risk ownership, must be clearly defined from development, through deployment – including redeployment in new contexts – to decommissioning. This includes cases where systems are complex amalgamations of AI and non-AI components, from multiple different suppliers. In this way, certain aspects of responsibility may reach beyond the team deploying a particular system, to other functions within the MOD, or beyond, to the third parties which build or integrate AI-enabled systems for Defence. Collectively, these articulations of human control, responsibility and risk ownership must enable clear accountability for the outcomes of any AI-enabled system in Defence. There must be no deployment or use without clear lines of responsibility and accountability, which should not be accepted by the designated duty holder unless they are satisfied that they can exercise control commensurate with the various risks”.[32]
>
> “The unpredictability of some AI systems, particularly when applied to new and challenging environments, increase the risks that unforeseen issues may arise with their use. The relative difficulties with interpreting how some forms of AI systems learn and make decisions present new challenges for the testing, evaluation and certification of such systems”.[33]
>
> “The relative difficulties with interpreting how some forms of AI systems learn and make decisions present new challenges for the testing, evaluation and certification of such systems. In addition, the high potential impact of AI-enabled systems

for Defence raises the stakes for potential side effects or unintended consequences, particularly when they could cause harms for those interacting with them".[34]

"We must nevertheless expect increasing numbers of AI-enabled force elements in real or virtual theatres of operations, and for them to operate at increasing speeds. They may be hard to distinguish from more conventional forces, but display unexpected behaviours. This could be intentional, the unexpected consequences of various system interactions, or the result of cyber-attack or some other manipulation. Some capabilities will simply malfunction, especially if safety and reliability standards are compromised in the rush to field new battlefield capability. Increasing numbers of autonomous platforms and reduced human involvement in (or even control over) operations could alter conflict thresholds and create spirals of violence and escalation".[35]

"We will not simplistically assume that AI inherently reduces workforce requirements, even if it does change the activities we need people to undertake. Where staff are affected by the adoption of AI, we will support them and help them find new roles and skills".[36]

4. AI effects on strategic stability and implications for the nuclear domain

Official documents provide only a superficial exploration of the effects of AI on strategic stability, generally highlighting that AI can have both positive and negative effects. However, UK-based experts and scholars consider how AI might impact strategic stability and nuclear decision-making. They recognise that AI creates both new opportunities and vulnerabilities, which can have both stabilising and destabilising effects on strategic stability. As an example, AI for ISR has the potential to strengthen or erode NC3 systems. The reliable functioning of AI for ISR can enhance situational awareness, providing an advantage. However, the spoofing or manipulation of such AI systems can decrease situational awareness, presenting an obstacle and posing a potential risk to crisis stability.[37]

Specifically, AI can *undermine* strategic stability by:

- Weakening deterrence credibility and signalling;
- Compressing decision-making time, allowing for escalation or inadvertent use of nuclear weapons;
- Exacerbating misunderstanding and misperceptions during a crisis;
- Encouraging a premature deployment of insufficiently tested AI.

> "Military AI systems functioning at machine speed could push the pace of combat to a point where the actions of machines eclipse the ability of human decision-makers to control (or even comprehend) events. In extremis, human commanders might lose control of the outbreak, course, and termination of warfare. Were humans to effectively lose (or pre-delegate) control of warfare to machines, inadvertent escalation pathways and crisis instability would increase, potentially with catastrophic results. Compelled by the speed and precision of AI to make decisions in a compressed timeframe, a state might accept higher risks and escalate a conflict with the belief it was in a 'use it or lose it' situation, or a lack of confidence in its ability to guarantee the safety and control of its nuclear arsenals".[38]

> "The integration of AI applications into early-warning (especially nuclear) systems could compress the decision-making timeframe, and accelerate the various stage of the escalate ladder to launch a missile, which would adversely affect crisis stability at a conventional and nuclear level of warfare".[39]

> "Military AI is likely to exacerbate the destabilizing and escalatory effects of an increasingly complex interplay of advanced military technology in a multipolar nuclear world order [sic]. Nuclear-armed states leveraging AI to achieve or sustain first-mover advantages in this multipolar context will likely destabilize this fragile order with uncertain outcomes".[40]

> "The key risk for international security is, therefore, that geopolitical pressures compel states to use AI-enabled autonomous weapon systems before the technology underlining them is sufficiently mature – which would make

Military AI is likely to exacerbate the destabilising and escalatory effects of an increasingly complex interplay of advanced military technology in a multipolar nuclear world order.

> these systems more susceptible to subversion. In extremis, an enemy may believe that AI is more effective than it actually is, leading to erroneous and potentially escalatory decision-making. In an effort to avoid situations such as this states will need to proactively co-ordinate (at military, diplomatic, industry, and academic level) as AI technology matures".[41]

Conversely, AI can *strengthen* strategic stability by:

- Allowing for better-informed decision-making, thus reducing the risk of miscalculation and unintended escalation.

- Allowing for early warning and detection or arms control verification.

> "The major powers' nuclear command-and-control systems increasingly rely on AI programmes, or, more precisely, expert systems and machine-learning algorithms, to enhance information flow, situational awareness and cyber security. Such capabilities can provide such systems with a larger window of opportunity in which to respond in the event of a crisis and thereby support de-escalation".[42]

> "With its myriad possible applications, AI has the potential to disrupt strategic paradigms further, for example by encouraging machine-speed escalation. Conversely, AI could have the ability to improve strategic stability, for example by allowing more complex modelling, better-informed decision-making, and therefore reduced risk of miscalculation and unintended escalation".[43]

The challenges associated with AI integration in weapons systems have raised significant concerns in British literature, particularly regarding the **increased automation of nuclear decision-making functions**. These concerns primarily stem from the limitations of AI itself. Overall, the literature emphasises the crucial role of humans as the ultimate driving force behind decision-making, advocating for a balance of human-machine interaction. AI can serve in decision-making support functions, acting as an 'adviser' to human decision-makers, providing information in a shorter time frame. However, there is recognition of the difficulties in ensuring the correct functioning of AI, including the risks of data manipulation and other technical issues.

Official documents lack detail in their analyses or plans for AI integration with NC3 systems. However, the MOD acknowledges that AI systems are restricted in their ability to engage in contextual thinking, make ethical judgments, and understand intent, which are capabilities possessed by human decision-makers. Therefore, official documents reiterate the commitment to maintain "human political control" over nuclear weapons at all times.[44] While a certain degree of automation is already underway, full automation of nuclear decision-making is not expected to occur. Strict human control is expected to be maintained throughout the process.

A limited number of non-official documents and other literature sources examine the impact of AI on NC3, specifically for the UK, while other researchers and analysts explore the topic across all nuclear-weapons states. Given the ongoing automation efforts,

The literature emphasises the crucial role of humans as the ultimate driving force behind decision-making, advocating for a balance of human-machine interaction.

the UK literature suggests that **AI is not anticipated to have a significant impact on NC3 in the short term**. This is due to the unpredictability and limited understanding of the algorithms used in machine learning and autonomous systems (i.e. deep learning models), making the risks of integration too high. However, as nuclear-weapons states pursue AI 'supremacy', there is a steady increase in the speed at which AI is being applied across military functions. This poses the potential risk of premature deployment of this technology from nuclear-weapons states without adequate consideration of its implications.

> "Our pursuit of AI-enabled capabilities will not change our view that our people are our finest asset. AI has tremendous power to enhance and support their work (e.g. enabling analysts to make sense of ever greater quantities of data), but we understand that some challenges require human creativity and contextual thinking, and that the real-world impact of military action demands applied human judgement and accountability".[45]

> "The types of algorithm underlying machine learning-driven applications and complex autonomous systems remain too unpredictable due to the problems of transparency and explainability. Nuclear command-and-control systems are too safety critical to be left to algorithms that engineers and operators cannot fully understand. Moreover, relatively traditional rule-based algorithms would be sufficient to further automate command and control. There seems to be a general agreement among nuclear-weapon experts that machine learning and autonomy should not be integrated into nuclear command and control, even if technological developments would permit it".[46]

> "We will ensure that – regardless of any use of AI in our strategic systems – human political control of our nuclear weapons is maintained at all times".[47]

> "At the strategic level of decision-making, AI-enabled command and control systems will likely be able to avoid many shortcomings inherent to human strategic decision-making during the "fog of war" such as: the susceptibility to invest in sunk costs, skewed risk judgment, cognitive heuristics, and group-think".[48]

> "Even if they do not revolutionize nuclear command, control and communications (NC3) systems, however, advances in machine learning and autonomous systems could bring some qualitative improvements in the nuclear command-and-control architecture [sic]. They could be used to enhance protection against cyberattacks and jamming attacks. Machine learning could also help planners to more efficiently manage their forces, including their human resources. Similarly, autonomous systems could be used to enhance the resilience of the communications architecture. Long-endurance UAVs could, for example, be used to replace signal rockets in forming an alternative airborne communications network in situations where satellite communication is impossible".[49]

“The inability of AI to understand context (i.e., the rationale and consequences of actions) or empathize (i.e., determine intent) would likely become a liability during wartime, when a degree of flexibility down the chain of command is generally considered positive [sic]. For example, the near catastrophic ICBM test at the US Vandenberg Air Force Base during the 1962 Cuban Missile Crisis was attributed to officers following pre-defined protocols without questioning these guidelines in the context of new information. Human errors typically occur at an individual level and seldom repeat in the same way; by contrast, AI systems may conceivably fail simultaneously and repeat this failure indefinitely’.[50]

5. Safe, responsible, ethical, and legal use of AI in military operations

The MOD has placed significant emphasis on the responsible use of AI in the military sector... To achieve [this], the MOD focuses on three key responsibilities: safety, ethics, and legality.

Given the inherent risks associated with increased automation of nuclear decision-making, including the limitations of AI in ethical judgment, understanding intent, and contextual thinking, the MOD has placed significant emphasis on the responsible use of AI in the military sector. To achieve the goal of safe and responsible use of AI, the MOD focuses on three key responsibilities: safety, ethics, and legality. Safety of AI systems would ensure that any malfunctions or unintended consequences would have no serious implications; ethical considerations are placed to ensure that AI aligns with established ethical principles; and legal compliance would ensure that AI applications in military operations adhere to international laws and norms. By prioritising these three components, while maintaining human political control, the MOD aims to ensure that AI is employed in a manner that upholds the principles of accountability, dependability, and responsible decision-making.

> “However, we recognise that the use of AI in many contexts – and especially by the military – raises profound issues. There are concerns about fairness, bias, reliability, and the nature of human responsibility and accountability. (For example, there are well documented instances of recruiting software demonstrating racial or gender bias). Unintended or unexpected AI-enabled outcomes could clearly have particularly significant consequences in an operational context”.[51]

> “The UK will lead by example, working with partners around the world to make sure international agreements embed our ethical values, and making clear that progress in AI must be achieved responsibly and safely, according to democratic norms and the rule of law”.[52]

> “Human-Machine Teaming will therefore be our default approach to AI adoption, both for ethical and legal reasons and to realise the ‘multiplier effect’ that comes from combining human cognition and inventiveness with machine-speed analytical capabilities”.[53]

Safety

To ensure safety, the MOD is committed to implementing robust safety measures for AI. These measures aim to ensure that AI systems comply with safety rules and regulations. The responsibility for AI safety regulations lies with the Defence Safety Authority (DSA), which is responsible for establishing and enforcing defence regulations to ensure safety across the MOD’s current capabilities. In the event of accidents or incidents, the Defence Accident Investigation Branch (DAIB) operates within the DSA to conduct thorough investigations. The DAIB’s primary focus is to identify the factors that contribute to these accidents, including any factors related to AI capabilities.

> “Head Office, the DAIC, Defence Equipment & Support and the Defence Safety Authority will establish a comprehensive framework for the testing, assurance, certification and regulation of AI-enabled systems – both the human and the technical component of Human Machine Teams. Our approach to AI risk management will be based on the ALARP

> (As Low as Reasonably Practical) principle that is commonplace in Defence for safety-critical and safety-involved systems. This regime will recognise the importance of appropriate testing through the lifetime of systems, reflecting the possibility that AI systems continue to learn and adapt their behaviour after deployment".[54]

To gain insight into the potential risks associated with AI in both the public and private sectors, particularly in relation to national defence and security, the 'National AI Strategy' and the 'Defence AI Strategy' highlight **approaches for mitigating risks**, including evaluating technical expertise within the Government, recognising the value of research infrastructure, and emphasising the importance of appropriate testing. As AI continues to evolve, the UK recognises the need to gain a comprehensive understanding of its strategic implications. To achieve this, the MOD highlights the need to engage with allies, academia, and the private sector. Additionally, wargaming and red teaming exercises are conducted to assess and anticipate potential risks and challenges. Testing AI-enabled systems and understanding their decision-making processes are also recognised as important considerations in ensuring the safe and responsible use of AI in defence and security.

To regulate AI at a strategic level, the UK has introduced various **regulatory frameworks**, such as the National Security and Investment Act, which came into force in 2022 and aims to scrutinise investments in critical sectors to safeguard national security, and the National Resilience Framework, which aims at enhancing the country's preparedness and resilience in the face of potential threats.

> "We will support the National Science & Technology Council and the Office for Science & Technology Strategy to develop broad perspectives on these strategic implications, providing unique Defence expertise, intelligence, analysis and insight – e.g. through wargaming, red-teaming and scenario-based investigations. We will engage closely with allies, partners, academia and civil society to drive forward strategic studies and build the capacity to understand and anticipate the strategic impacts and risks of AI in defence. Recognising AI's profound impact across many sectors, we will seek to learn lessons from areas (e.g. Finance) which have devised protocols to limit shocks despite highly competitive and fast-paced environments".[55]
>
> "The Office for AI will coordinate cross-government processes to accurately assess long term AI safety and risks, which will include activities such as evaluating technical expertise in government and the value of research infrastructure. Given the speed at which AI developments are impacting our world, it is also critical that the government takes a more precise and timely approach to monitoring progress on AI, and the government will work to do so.
>
> The government will support the safe and ethical development of these technologies as well as using powers through the National Security & Investment Act to mitigate risks arising from a small number of potentially concerning actors. At a strategic level, the National Resilience Strategy will review our approach to emerging technologies; the

> Ministry of Defence will set out the details of the approaches by which Defence AI is developed and used; the National AI R&I Programme's emphasis on AI theory will support safety; and central government will work with the national security apparatus to consider narrow and more general AI as a top-level security issue'".[56]

> "AI systems present fundamentally different testing and assurance challenges to traditional physical and software capabilities, not least as it can be technically challenging to explain the basis for a system's decisions. This is a significant risk to delivering our strategic objectives. We must strike the right risk balance, ensuring new AI-enabled capabilities are safe, robust, effective and cyber-secure, while also delivering at the pace of relevance – in hours, in the case of some algorithms".[57]

Fostering trust in AI across all sectors, including defence, is also seen by the MOD as a key aspect of enabling data-driven innovation in the public sector, with a view to support the development of trustworthy, adoptable, and transparent AI technologies through the National AI Research and Innovation Programme. The MOD has particularly focused on ensuring the reliability and trustworthiness of AI-generated data in the defence sector.

> "Data is a critical strategic asset, second only to our people in terms of importance. In recognition of this the Defence data transformation is underway with a central Data Office established within Defence Digital, as well as a Defence Data Framework (2021) to transform Defence's culture, behaviour and data capabilities.
>
> There remains much to do, however, as our vast data resources are too often stove-piped, badly curated, undervalued and even discarded".[58]

Trust in AI varies depending on the domains in which it is used. Surveys are regularly conducted by the UK Government to gauge public opinion on the integration of AI in both businesses and the public sector, in an effort to understand the public's views and to regulate AI systems accordingly. These **surveys have revealed growing awareness of the potential risks and harms associated with AI**, particularly in regard to fairness, bias, and accountability. For example, the Centre for Data Ethics and Innovation (CDEI)'s Public Attitudes to Data and AI (PADAI) Tracker Survey tracks public opinion towards AI and data over time. The second wave of the survey, published in November 2022, revealed increased concerns about data security and privacy and a strong desire for robust governance of AI in high-risk scenarios. However, the survey also highlighted that trust in AI is closely linked to trust in the specific organisation(s) using it, and that people recognise the potential benefits of AI, particularly in the health and economic sectors.[59]

Ethics

AI will be employed in an ethical way, ensuring the protection of UK values and retaining the support of its allies and partners, the British citizens, and key stakeholders. To that end, the MOD developed a set of five ethical principles resulting from consultation with expert stakeholders and partners in the public sector, industry and academia, and created an AI Ethics Advisory Panel.[60] The principles are based on the following parameters:

1. **Human centricity**, whereby AI systems should be designed with a focus on humans, taking into account the full spectrum of their effects on people—both good and bad—throughout the entire duration of the system's use.

2. **Responsibility**, where accountability for AI-enabled systems must be clearly assigned to individuals, who are responsible for the systems' outcomes and maintain control over them throughout their operational life, ensuring that there are defined methods for human oversight.

3. **Understanding**, which dictates that ethical decision-making in Defence must always be supported by a proper understanding of the situation by the decision-makers.

4. **Bias and harm mitigation**, which calls for proactive measures to minimise the risk of damages that may arise from systems' bias. This might entail, for example, setting safeguards and performance thresholds.

5. **Reliability**, which mandates that the AI systems operated by the MOD must consistently meet their design and deployment specifications and function within the bounds of acceptable performance levels. These parameters should be subject to continuous evaluation and verification to maintain assured reliability, especially as AI systems adapt and improve over time.

Legality

The legal framework of AI will be set in compliance with the International Humanitarian Law (IHL) in the context of law of armed conflict, employment law, privacy and procurement.

> "Defence always seeks to abide by its legal obligations across the full range of activities from employment law to privacy and procurement, and the law of armed conflict, also known as International Humanitarian Law (IHL). It has robust practices and processes in place to ensure its activities and its people abide by the law. These practices and processes are being – and will continue to be – applied to AI-enabled capabilities.
>
> Deployment of AI-enabled capabilities in armed conflict needs to comply fully with IHL, satisfying the four core principles of distinction, necessity, humanity and proportionality. We are very clear that use of any system or weapon which does not satisfy these fundamental principles would constitute a breach of international law".[61]

To uphold the value of legality, the issue of autonomous weapons and the challenge of their compliance with the IHL has sparked considerable debate.

To uphold the value of legality, the issue of **autonomous weapons** and the challenge of their compliance with the IHL has sparked considerable debate. LAWS are weapons that can select and attack targets without meaningful human control and, as such, they pose a fundamental challenge to the protection of civilians and to the compliance with the IHL. The UK has consistently emphasised the importance of maintaining human control over the use of force as well as ethical considerations in the development and use of emerging technologies. As a result, the UK strongly rejects fully autonomous weapons that operate without meaningful and context-appropriate human involvement as they would not adhere to the ethical and responsible standards.

The UK has been actively engaged in discussions and negotiations related to autonomous weapons at the international level, particularly within the framework of the Convention on Certain Conventional Weapons (CCW). The Group of Governmental Experts (GGE) on LAWS of the CCW was established to consider proposals and elaborate possible measures related to the normative framework in the area of LAWS. Within the Group, the UK has consistently advocated for a responsible and ethical approach to the development of autonomous weapons, calling for meaningful human control and compliance with legal obligations, including IHL. The country has supported the continuation of the GGE on LAWS and the adoption of 11 guiding principles on LAWS in 2019, and it has advanced proposals on the elaboration of a document with agreed guidelines and best practices 'on how states should approach the development and use of emerging technologies in the area of LAWS at each stage of its lifecycle'.[62] The proposed document has the goal to assess weapons' characteristics that would be in compliance with the IHL and those which would be incompatible.

At the GGE, the UK recognised the challenge of reaching an international agreement on definitions for full or partial autonomy, particularly due to the fact that any approach to this matter should also allow for rapid technological advancement but at the same time ensure that, as technology develops, no circumventions are allowed.[63] For this reason, the UK, along with other states, cautioned against imposing prohibitions or restrictions on LAWS that could hamper innovation or legitimate military applications and thus did not support a legally binding instrument on LAWS.[64] In particular, the UK clarified its preference for non-binding instruments, such as the establishment of a code of conduct or good practices to ensure that autonomous weapons can comply with IHL.

> "We will continue to work closely with international allies and partners to address the opportunities and risks around autonomy in weapons systems. Global governance for such systems is a difficult task. It will be challenging to reach international agreement on definitions for full or partial autonomy on a technical or systems level. It is also important to ensure any approach allows for rapid technological advancement, and doesn't become redundant or isn't able to be circumvented as technology develops. Such international processes must be inclusive, and involve all key actors in this space if they are to be effective.
>
> We believe the best approach is to focus on building norms of use and positive obligations to demonstrate how degrees of autonomy in weapons systems can be used in accordance with international humanitarian law – with suitable levels of human control, accountability and responsibility. Setting

> out those characteristics that would make it inherently impossible for a system to comply with international humanitarian law is key to this, and we will continue to engage actively in the international arena to reach consensus on them. The UN Group of Government Experts on LAWS under the Convention for Certain Conventional Weapons will continue to be our primary avenue for such discussions. Our own approach, driven by the AI Ethical principles, is to build understanding, best practice and codes of conduct through which we can achieve ethical outcomes in our use of AI".[65]

Additionally, debates in the context of autonomous weapons systems have been very lively amongst parliamentarians and civil society. At the APPG AI meeting on 7 September 2022, on 'National security and defence', a group of parliamentarians discussed the ethical, legal, and political challenges of using AI and autonomous weapon systems in warfare. The panel highlighted the risks posed by these technologies to human dignity and security, and expressed concerns about the inadequacy of existing humanitarian and human rights laws in effectively regulating them. The panel also criticised the UK's policy on autonomous weapons as being too ambiguous, calling for transparent governance and monitoring mechanisms, as well as the establishment of an international treaty to ban or regulate autonomous weapons.

In response to these concerns, the House of Lords Liaison Committee established the 'AI in Weapon Systems Committee' in January 2023 with the purpose of examining the use and integration of artificial intelligence in weapons systems. The Committee launched an inquiry and issued a public call for evidence to gather information and perspectives on the effects of autonomy in weapons systems and to assess whether the current framework of IHL is sufficient to regulate autonomous weapons effectively. The Committee also tackled the impact of AI at the intersection with NC3, with experts sharing their insights at the Committee on the risks posed by this technology in critical decision-making scenarios.

Moreover, the UK Stop Killer Robot campaign is a UK-based coalition of non-governmental organisations that seeks to prevent the development and use of LAWS. As a part of the global Campaign to Stop Killer Robots, this initiative urges states to place a national moratorium on LAWS and to participate in international debate and dialogue on this issue. Moreover, the campaign calls on the UK to work internationally to upgrade arms control treaties to ensure that human rights, ethical and moral standards are retained.[66]

6. Conclusion and recommendations

As detailed in the 'Defence AI Strategy' and the 'Ambitious, safe and responsible use of AI' policy papers, the MoD aims to leverage the power of AI to bolster national defence capabilities. Central to this strategy is an emphasis on AI adoption, integration of AI technologies within the defence landscape, and ethical considerations tied to AI usage in defence scenarios. The Strategy acknowledges the critical need for maintaining the accuracy and integrity of datasets. It argues that robust and reliable AI systems play a pivotal role in mitigating the risk of bias. It also underlines the necessity for innovative approaches in the testing and verification of AI systems, exploring opportunities and implications presented by AI-enabled systems.

Additionally, the Strategy aims to improve information sharing and best practices with partners and allies. It highlights the critical role of industry and academia in driving AI innovation, advocating for enhanced collaboration in these sectors and encouraging joint investment to build and maintain the industrial capabilities needed for defence. This includes fostering export and international collaboration opportunities.

Analysing from a policy perspective, the 'Defence AI Strategy' and the 'Ambitious, Safe and Responsible' AI policy paper establish a roadmap to mitigate AI-induced risks by:

- Enhancing mutual trust and security;
- Fostering dialogue on nuclear risk reduction with a view to reducing misunderstanding, miscalculation, or uncontrolled escalation;
- Promoting a safe and responsible use of AI;
- Further studying the effects of AI on the cyber, space and nuclear domain;
- Limit the spread of dual-use technologies through existing non-proliferation, disarmament and export control regimes;
- Ensuring human control of nuclear weapons at all times.

> "The UK is at the forefront of work internationally to reduce the risk of nuclear conflict and enhance mutual trust and security. We will champion strategic risk reduction and seek to create dialogue among states possessing nuclear weapons, and between states possessing nuclear weapons and non-nuclear weapon states, to increase understanding and reduce the risk of error, misinterpretation, and miscalculation. We will study the effects of AI on the inter-linked domains of cyber, space and nuclear, examining AI's potential to accelerate or amplify developments linked to other emerging and strategic technologies. We will promote and engage with international dialogue aimed at identifying and addressing crucial AI-related strategic risks".[67]

> "We must take appropriate steps to limit the possibility of misunderstanding, miscalculation or uncontrolled escalation arising from these factors. At times, it may be crucial to know whether or not a system was AI-enabled or not, and to what degree. This could be particularly important in the event of a

> crisis or flash-point, or in the already murky context of sub-threshold activity. We will engage with allies and partners to understand these issues and develop proposals for codes of conduct and other confidence-building measures which can reduce the risk of accidental engagements, collateral damage, and miscalculations. We will also share best practice internationally on how to conduct TEV&V and weapons reviews, such as practical descriptions and case studies regarding the use and parameters of system control, where appropriate. In this context, it is critical that we engage with potential adversaries and nations whose approach to adopting AI differs from our own, and that we strongly advocate safe and responsible use".[68]

> "We must reinforce, reinvigorate and adapt, balancing the opportunities of new technologies with appropriate controls to constrain access to 'military grade' AI applications. Where appropriate, we will work through existing non-proliferation, disarmament and export control regimes, treaties and organisations to ensure we balance the opportunities of new technologies with appropriate controls".[69]

Overall, the UK strategic response to risks posed by autonomy in weapons systems centres on the formulation and adoption of codes of conduct and rules of the road. These guidelines aim to demonstrate how a limited degree of autonomy in weapons systems, underpinned by human control, can align with international law, as well as to ensure that no weapons systems operate without significant, context-appropriate human involvement throughout their operational lifecycle.

Potential risk reduction measures

However, this strategy necessitates complementary measures addressing a broader risk reduction perspective, specifically considering nuclear implications. One critical aspect not sufficiently covered in the strategy is the integration of AI in nuclear decision-making. The employment of AI in this area carries inherent risks, potentially escalating tensions and, in worse case scenarios, lowering the threshold for nuclear weapon deployment. While the UK has asserted that decision-making will not be fully handed over to AI, the strategy remains unclear on how the integration of AI in nuclear decision-making might lead to inadvertent escalation.

At the multilateral level, specifically within the P5 framework, the UK can actively promote dialogue to address the implications of AI on nuclear decision-making and strategic stability, with a view to make progress in establishing norms or codes of conduct.

At the *multilateral* level, specifically within the P5 framework, the UK can actively promote dialogue to address the implications of AI on nuclear decision-making and strategic stability, with a view to make progress in establishing norms or codes of conduct. For instance, a crucial first step could be the P5 arriving at a shared understanding of what constitutes strategic stability in the age of AI, delineating clear red lines that must not be crossed. Such an understanding would ensure all parties comprehend the boundaries and consequences, thus reducing the likelihood of miscalculations or inadvertent escalations. As the former UK National Security Adviser Stephen Lovegrove has argued:

> "We need to establish new norms for behavior in the context of hybrid and tech-enabled conflict, setting red lines for the gray zone as it emerges as a new arena for strategic competition [*sic*]".[70]

Expanding on the joint statement released by France, the UK, and the US at the Tenth NPT Review Conference[71], the P5 could **formulate a collective pledge to preserve human involvement in critical decision-making processes concerning nuclear weapons systems**. This need for human oversight is underscored by a common understanding amongst all P5 states that, particularly in the context of nuclear weapons, human judgment should always be a key element in the decision-making process.[72]

However most-advanced AI systems (such as advanced deep learning-based models) are too premature to be integrated in nuclear decision-making due to their opacity and unpredictability. An understanding among all P5 states should be put in place on the extent to which an integration of such systems should not take place. This shared stance is primarily motivated by the need to mitigate the risks associated with integrating AI into nuclear decision-making, some of which all P5 states similarly identify in their respective internal debates.

Ultimately, the P5 states should progress beyond merely issuing high-level principles about the military applications of AI and concentrate on their **practical implementation**. A joint effort to devise tangible norms and guidelines regarding the safe and responsible use of AI in NC3 systems could be a strategic move. Key objectives could include ensuring nuclear decision-makers have a thorough understanding of AI-enhanced tools and creating guidelines to prevent tampering in sensitive NC3 processes. This would involve addressing concerns linked to data integrity, biases, and potential vulnerabilities introduced by AI.

Additionally, an emphasis on cultivating critical thinking skills to counteract automation bias and empowering military personnel to evaluate AI system functioning and outcomes effectively could be integral. The P5 states should prioritise **investments in AI education and training programs** for operators and decision-makers in the defence realm. This education initiative would not only foster critical thinking skills to counter automation bias but also enable personnel to effectively assess the functioning and outputs of AI systems. The training should underscore the significance of human supervision and decision-making in NC3 processes.

From a *unilateral* perspective, as one of the world's top three most advanced nations in AI, the UK should continue to **invest in R&D**, especially in understanding the interpretability of AI models. The inherent complexity of numerous AI models often leads to an opacity issue, colloquially known as the 'black box' dilemma. This poses a significant challenge for decision-makers, who need to be able to understand, trust, and verify the decisions or recommendations generated by AI models. If advanced AI is integrated into NC3, even for support functions, enhancing the interpretability of AI models becomes a critical objective. Current research initiatives exploring ways to **improve model interpretability, especially those focusing on mechanistic interpretability** (essentially attempting to understand and predict the behaviour of neural networks by reverse engineering them and elucidating the algorithms they use), are showing significant promise.

As one of the world's top three most advanced nations in AI, the UK should continue to invest in R&D, especially in understanding the interpretability of AI models.

Moreover, as AI continues to advance, there is an increasing need for comprehensive legal guidelines around its use in warfare. As one of the leading countries in AI, the UK is uniquely positioned to influence global norms and standards. It could benefit the UK to re-evaluate its stance and take a more active role in shaping the rules and regulations concerning the use of AI in autonomous weapons systems. While codes of conduct and rules of the road are useful for managing autonomous weaponry, **legally binding measures are pivotal for preventing the proliferation of dual-use systems** and ensuring that no state, including potential adversaries, integrates a level of autonomy into decision-making that threatens global safety and strategic equilibrium. Extensive research and cooperation across the AI landscape are required to reconcile the UK's commitment to binding legal instruments with its pursuit of AI technological advancements.

As part of this approach, the UK should continue to **promote public-private partnerships** to address the AI skills deficit within the public sector. Official documents acknowledge that collaboration with private enterprises and academic institutions can leverage their expertise and resources, thus addressing the challenges of integrating AI into weaponry and military systems. However, to collaborate with the MOD, private companies and academia demand that the UK acts responsibly with AI integration in weapons systems.[73] This means firstly engaging in dialogue to dispel misconceptions and concerns about weapon system automation, fostering a greater understanding and willingness to collaborate. Secondly, the UK should establish frameworks for productive collaboration, for instance ensuring alignment of project objectives. In parallel, collaborations between academic institutions and the public sector could spur innovation and bolster AI capabilities in defence sectors.

> **The UK is uniquely positioned to influence global norms and standards. It could benefit the UK to re-evaluate its stance and take a more active role in shaping the rules and regulations concerning the use of AI in autonomous weapons systems.**

References

[1] Vincent Boulanin, Lora Saalman, Petr Topychkanov, Fei Su, & Moa Peldán Carlsson, “Artificial Intelligence, Strategic Stability and Nuclear Risk”, *SIPRI* (June 2020), https://www.sipri.org/sites/default/ files/2020-06/artificial_intelligence_strategic_stability_and_nuclear_risk.pdf; Jill Hruby & M. Nina Miller, “Assessing and Managing the Benefits and Risks of Artificial Intelligence in Nuclear-Weapon Systems”, *NTI Paper* (August 2021), https://www.nti.org/wp-content/uploads/2021/09/NTI_Paper_AI_r4.pdf; James Johnson, *AI and the Bomb: Nuclear strategy and risk in the digital age* (Oxford: Oxford University Press, 2023) ; Tim McDonnell, Mary Chesnut, Tim Ditter, Anya Fink & Larry Lewis, “Artificial Intelligence in Nuclear Operations: Challenges, opportunities, and impacts”, *Center for Naval Analyses* (April 2023), https://www.cna.org/reports/2023/04/Artificial-Intelligence-in-Nuclear-Operations.pdf; Alexa Wehsener, Andrew W. Reddie, Leah Walker, & Philip J. Reiner, “AI-NC3 Integration in an Adversarial Context: Strategic Stability Risks and Confidence Building Measures”, *Institute for Security and Technology* (February 2023), https://securityandtechnology.org/wp-content/uploads/2023/02/AI-NC3-Integration-in-an-Adversarial-Context.pdf; Michael T. Klare, “Skynet Revisited: The Dangerous Allure of Nuclear Command Automation”, *Arms Control Today*, (April 2020) https://www.armscontrol.org/act/2020-04/features/skynet-revisited-dangerous-allure-nuclear-command-automation. Peter Hayes, “Nuclear Command, Control, and Communications (NC3): Is There a Ghost in the Machine?,” *Nautilus Institute* (9 April 2018).

[2] *Defence Artificial Intelligence Strategy* (London: UK Ministry of Defence, June 2022), p. 59, https://www.gov.uk/government/publications/defence-artificial-intelligence-strategy/defence-artificial-intelligence-strategy.

[3] *Defence Artificial Intelligence Strategy* (London: UK Ministry of Defence, June 2022), https://www.gov.uk/government/publications/defence-artificial-intelligence-strategy/defence-artificial-intelligence-strategy; *Global Britain in a competitive age: the Integrated Review of Security, Defence, Development and Foreign Policy* (London: HM Government, March 2021), https://assets.publishing.service.gov.uk/media/60644e4bd3bf7f0c91eababd/Global_Britain_in_a_Competitive_Age-_the_Integrated_Review_of_Security__Defence__Development_and_Foreign_Policy.pdf.

[4] *Defence Artificial Intelligence Strategy* (London: UK Ministry of Defence, June 2022), p. 4, https://www.gov.uk/government/publications/defence-artificial-intelligence-strategy/defence-artificial-intelligence-strategy.

[5] *Global Britain in a competitive age: the Integrated Review of Security, Defence, Development and Foreign Policy* (London: HM Government, March 2021), p, 7, https://assets.publishing.service.gov.uk/media/60644e4bd3bf7f0c91eababd/Global_Britain_in_a_Competitive_Age-_the_Integrated_Review_of_Security__Defence__Development_and_Foreign_Policy.pdf.

[6] Vincent Boulanin, Lora Saalman, Petr Topychkanov, Fei Su, and Moa Peldán Carlsson, “Artificial intelligence, strategic stability and nuclear risk”, *SIPRI* (June 2020), p. 52, https://www.sipri.org/sites/default/files/2020-06/artificial_intelligence_strategic_stability_and_nuclear_risk.pdf.

[7] *Defence Artificial Intelligence Strategy* (London: UK Ministry of Defence, June 2022), p. 5, https://www.gov.uk/government/publications/defence-artificial-intelligence-strategy/defence-artificial-intelligence-strategy.

[8] Ibid.

[9] Ibid, p. 1.

[10] *See Moratorium on fully autonomous robotic weapons needed to allow the UN to consider fully their far-reaching implications and protect human rights: Amnesty International written statement to the 23rd session of the UN Human Rights Council* (23 May 2013), https://www.amnesty.org/en/wp-content/uploads/2021/05/ACT300382013ENGLISH.pdf.

[11] *Defence Artificial Intelligence Strategy* (London: UK Ministry of Defence, June 2022), p. 7, https://www.gov.uk/government/publications/defence-artificial-intelligence-strategy/defence-artificial-intelligence-strategy.

[12] See James Johnson, *AI and the Bomb: Nuclear strategy and risk in the digital age* (Oxford: Oxford University Press, 2023); Marina Favaro, “Weapons of Mass Distortion’, Centre for science & security studies (May 2021), https://www.kcl.ac.uk/csss/assets/weapons-of-mass-distortion.pdf; James Johnson, “Artificial intelligence & future warfare: implications for international security”, *Defense & Security Analysis* (24 April 2019) pp. 147-169, https://www.tandfonline.com/doi/abs/10.1080/14751798.2019.1600800; James Johnson, “Artificial Intelligence in Nuclear Warfare: A Perfect Storm of Instability?”, *The Washington Quarterly*, 43:2 (2020), https://www.tandfonline.com/doi/abs/10.1080/0163660X.2020.1770968; Mark Fitzpatrick, “Artificial Intelligence and Nuclear Command and Control”, *Survival* (21 May 2019), https://doi.org/10.1080/00396338.2019.1614782; Marina Favaro, Heather Williams, “Written evidence submitted by Marina Favaro and Heather Williams *Centre for science & security studies* (12 May 2021), https://committees.parliament.uk/writtenevidence/35529/html/.

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

[14] Marina Favaro, "Weapons of Mass Distortion', *Centre for science & security studies* (May 2021), https://www.kcl.ac.uk/csss/assets/weapons-of-mass-distortion.pdf.

[15] *Defence Artificial Intelligence Strategy* (London: UK Ministry of Defence, June 2022), p. 5, https://www.gov.uk/government/publications/defence-artificial-intelligence-strategy/defence-artificial-intelligence-strategy.

[16] *Defence Artificial Intelligence Strategy*, (London: UK Ministry of Defence, June 2022), p. 9, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1082416/Defence_Artificial_Intelligence_Strategy.pdf.

[17] James Johnson, "Artificial intelligence & future warfare: implications for international security", *Defense & Security Analysis* (24 April 2019), p. 4, https://doi.org/10.1080/14751798.2019.1600800.

[18] Reference to the Spring Storm exercise, conducted in May 2021, where AI was employed in a British military operation for the first time.

[19] *Defence Artificial Intelligence Strategy, (London: UK Ministry of Defence*, June 2022), p. 10, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1082416/Defence_Artificial_Intelligence_Strategy.pdf.

[20] Ibid, p.7.

[21] Marina Favaro, "Weapons of Mass Distortion', *Centre for science & security studies* (May 2021), p. 21, https://www.kcl.ac.uk/csss/assets/weapons-of-mass-distortion.pdf.

[22] Ibid, p. 23.

[23] Vincent Boulanin, Lora Saalman, Petr Topychkanov, Fei Su, and Moa Peldán Carlsson, "Artificial intelligence, strategic stability and nuclear risk", *SIPRI* (June 2020), p. 55, https://www.sipri.org/sites/default/files/2020-06/artificial_intelligence_strategic_stability_and_nuclear_risk.pdf.

[24] *Defence Artificial Intelligence Strategy*, (London: UK Ministry of Defence, June 2022), p. 11-13, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1082416/Defence_Artificial_Intelligence_Strategy.pdf

[25] Ibid, p. 11.

[26] Ibid, p. 13.

[27] Ibid, p. 11.

[28] Ibid, p. 12.

[29] See: *Defence Artificial Intelligence Strategy*, (London: UK Ministry of Defence, June 2022), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1082416/Defence_Artificial_Intelligence_Strategy.pdf; *Ambitious, Safe, Responsible: Our approach to the delivery of AI-enabled capability in Defence*, (London: UK Ministry of Defence, June 2022), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1082991/20220614-Ambitious_Safe_and_Responsible.pdf.

[30] *National AI Strategy*, HM Government (September 2021), p. 16, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1020402/National_AI_Strategy_-_PDF_version.pdf.

[31] *Ambitious, Safe, Responsible: Our approach to the delivery of AI-enabled capability in Defence*, (London: UK Ministry of Defence, June 2022), p. 11, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1082991/20220614-Ambitious_Safe_and_Responsible.pdf.

[32] Ibid, p. 10.

[33] Ibid, p. 5.

[34] Ibid.

[35] *Defence Artificial Intelligence Strategy*, (London: UK Ministry of Defence, June 2022), p. 56, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1082416/Defence_Artificial_Intelligence_Strategy.pdf.

[36] Ibid, p. 15.

[37] Marina Favaro, "Weapons of Mass Distortion', *Centre for science & security studies* (May 2021), p. 20, https://www.kcl.ac.uk/csss/assets/weapons-of-mass-distortion.pdf

[38] James Johnson, "Artificial Intelligence in Nuclear Warfare: A Perfect Storm of Instability?", *The Washington Quarterly*, 43:2 (2020), p. 4, https://www.tandfonline.com/doi/abs/10.1080/0163660X.2020.1770968.

[39] James Johnson, "Artificial intelligence & future warfare: implications for international security", *Defense & Security Analysis*, (24 April 2019), p. 6, https://doi.org/10.1080/14751798.2019.1600800.

[40] James Johnson, 'Artificial Intelligence in Nuclear Warfare: A Perfect Storm of Instability?', *The Washington Quarterly*, 43:2 (2020), p. 15, https://www.tandfonline.com/doi/abs/10.1080/0163660X.2020.1770968.

[41] James Johnson, "Artificial intelligence & future warfare: implications for international security", *Defense & Security Analysis*, (24 April 2019), p. 6, https://doi.org/10.1080/14751798.2019.1600800.

[42] Mark Fitzpatrick, "Artificial Intelligence and Nuclear Command and Control", *Survival* (21 May 2019) https://doi.org/10.1080/00396338.2019.1614782

[43] Marina Favaro, Heather Williams, "Written evidence submitted by Marina Favaro and Heather Williams", *King's College London* (12 May 2021), https://committees.parliament.uk/writtenevidence/35529/html/.

44 *Defence Artificial Intelligence Strategy*, (London: UK Ministry of Defence, June 2022), p. 7, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1082416/Defence_Artificial_Intelligence_Strategy.pdf.

45 *Defence Artificial Intelligence Strategy*, (London: UK Ministry of Defence, June 2022), p. 15, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1082416/Defence_Artificial_Intelligence_Strategy.pdf.

46 Vincent Boulanin, Lora Saalman, Petr Topychkanov, Fei Su, and Moa Peldán Carlsson, "Artificial intelligence, strategic stability and nuclear risk", *SIPRI* (June 2020), p. 24, https://www.sipri.org/sites/default/files/2020-06/artificial_intelligence_strategic_stability_and_nuclear_risk.pdf.

47 *Defence Artificial Intelligence Strategy*, (London: UK Ministry of Defence, June 2022), p. 7, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1082416/Defence_Artificial_Intelligence_Strategy.pdf.

48 James Johnson, "Artificial intelligence & future warfare: implications for international security", *Defense & Security Analysis* (24 April 2019), p. 4, https://doi.org/10.1080/14751798.2019.1600800.

49 Vincent Boulanin, Lora Saalman, Petr Topychkanov, Fei Su, and Moa Peldán Carlsson, "Artificial intelligence, strategic stability and nuclear risk", *SIPRI* (June 2020), p. 25, https://www.sipri.org/sites/default/files/2020-06/artificial_intelligence_strategic_stability_and_nuclear_risk.pdf.

50 James Johnson, 'Artificial Intelligence in Nuclear Warfare: A Perfect Storm of Instability?', *The Washington Quarterly*, 43:2 (2020), pp. 7-8, https://www.tandfonline.com/doi/abs/10.1080/0163660X.2020.1770968

51 *Defence Artificial Intelligence Strategy*, (London: UK Ministry of Defence, June 2022), p. 14, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1082416/Defence_Artificial_Intelligence_Strategy.pdf.

52 Ibid, p. 11.

53 Ibid, p. 15.

54 Ibid, p. 27.

55 Ibid, p. 58.

56 *National AI Strategy*, HM Government, September 2021, p. 60, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1020402/National_AI_Strategy_-_PDF_version.pdf.

57 *Defence Artificial Intelligence Strategy*, (London: UK Ministry of Defence, June 2022), p. 27, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1082416/Defence_Artificial_Intelligence_Strategy.pdf.

58 Ibid, p. 24.

59 "Public attitudes to data and AI: Tracker survey (Wave 2)", *Centre for Data Ethics and Innovation* (2 November 2022), https://www.gov.uk/government/publications/public-attitudes-to-data-and-ai-tracker-survey-wave-2/public-attitudes-to-data-and-ai-tracker-survey-wave-2.

60 *Ambitious, Safe, Responsible: Our approach to the delivery of AI-enabled capability in Defence*, (London: UK Ministry of Defence, June 2022), p. 9-11, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1082991/20220614-Ambitious_Safe_and_Responsible.pdf

61 Ibid, p. 6.

62 "Meeting of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects" (13 December 2019), https://digitallibrary.un.org/record/3856241?ln=en; "Proposal for a GGE document on the application of International Humanitarian Law to Emerging Technologies in the Area of Lethal Autonomous Weapons Systems", March 2022, https://view.officeapps.live.com/op/view.aspx?src=https%3A%2F%2Fdocuments.unoda.org%2Fwp-content%2Fuploads%2F2022%2F05%2F03032022-UK-Proposal-for-Mar-2022-LAWS-GGE.docx&wdOrigin=BROWSELINK

63 *Ambitious, Safe, Responsible: Our approach to the delivery of AI-enabled capability in Defence*, (London: UK Ministry of Defence, June 2022), p. 13, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1082991/20220614-Ambitious_Safe_and_Responsible.pdf

64 Ray Acheson, Allison Pytlak. Rep. "Autonomous Weapons and Questions of Ethics, Control, and Accountability" (3 June 2022), https://www.reachingcriticalwill.org/disarmament-fora/ccw/2022/laws/ccwreport/16277-ccw-report-vol-10-no-4.

65 *Ambitious, Safe, Responsible: Our approach to the delivery of AI-enabled capability in Defence*, (London: UK Ministry of Defence, June 2022), p. 13, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1082991/20220614-Ambitious_Safe_and_Responsible.pdf.

66 "Campaign to Stop Killer Robots reaction to the Integrated Review", *UNA-UK* (12 May 2021), https://una.org.uk/news/campaign-stop-killer-robots-reaction-integrated-review.

67 *Defence Artificial Intelligence Strategy*, (London: UK Ministry of Defence, June 2022), p. 59, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1082416/Defence_Artificial_Intelligence_Strategy.pdf.

68 Ibid, p. 56.

69 Ibid, p. 54.

[70] Sir Stephen Lovegrove, "The Future of Arms Control, Strategic Stability, and the Global Order", *CSIS* (28 July 2022), https://www.csis.org/analysis/future-arms-control-strategic-stability-and-global-order.

[71] "Principles and responsible practices for Nuclear Weapon States" Working paper submitted by France, the United Kingdom of Great Britain and Northern Ireland and the United States of America (29 July 2022), https://www.un.org/sites/un2.un.org/files/npt_conf.2020_e_wp.70.pdf.

[72] Maximilian Hoell & Sylvia Mishra, "Artificial Intelligence in Nuclear Command, Control, and Communications: Implications for the Nuclear Non-Proliferation Treaty" in: The Implications of Emerging Technologies in the Euro-Atlantic Space: Views from the Younger Generation Leaders Network, ed. by Julia Berghofer, Andrew Futter, Clemens Häusler, Maximilian Hoell & Juraj Nosál (Cham: Palgrave Macmillan, 2023), pp. 123 – 142 (p. 140).

[73] *Defence Artificial Intelligence Strategy*, (London: UK Ministry of Defence, June 2022), p. 42, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1082416/Defence_Artificial_Intelligence_Strategy.pdf.

The European Leadership Network (ELN) is an independent, non-partisan, pan-European NGO with a network of over 300 past, present and future European leaders working to provide practical real-world solutions to political and security challenges.

Contact

Published by the European Leadership Network, November 2023

European Leadership Network (ELN)
8 St James's Square
London, UK, SE1Y 4JU

@theELN | europeanleadershipnetwork.org

The opinions articulated in this report represent the views of the author, and do not necessarily reflect the position of the European Leadership Network or any of its members. The ELN's aim is to encourage debates that will help develop Europe's capacity to address pressing foreign, defence, and security challenges.

EUROPEAN
LEADERSHIP
NETWORK

European Leadership Network
8 St James's Square
London, SE1Y 4JU
United Kingdom

Email: secretariat@europeanleadershipnetwork.org
Tel: 0203 176 2555

Follow us

europeanleadershipnetwork.org