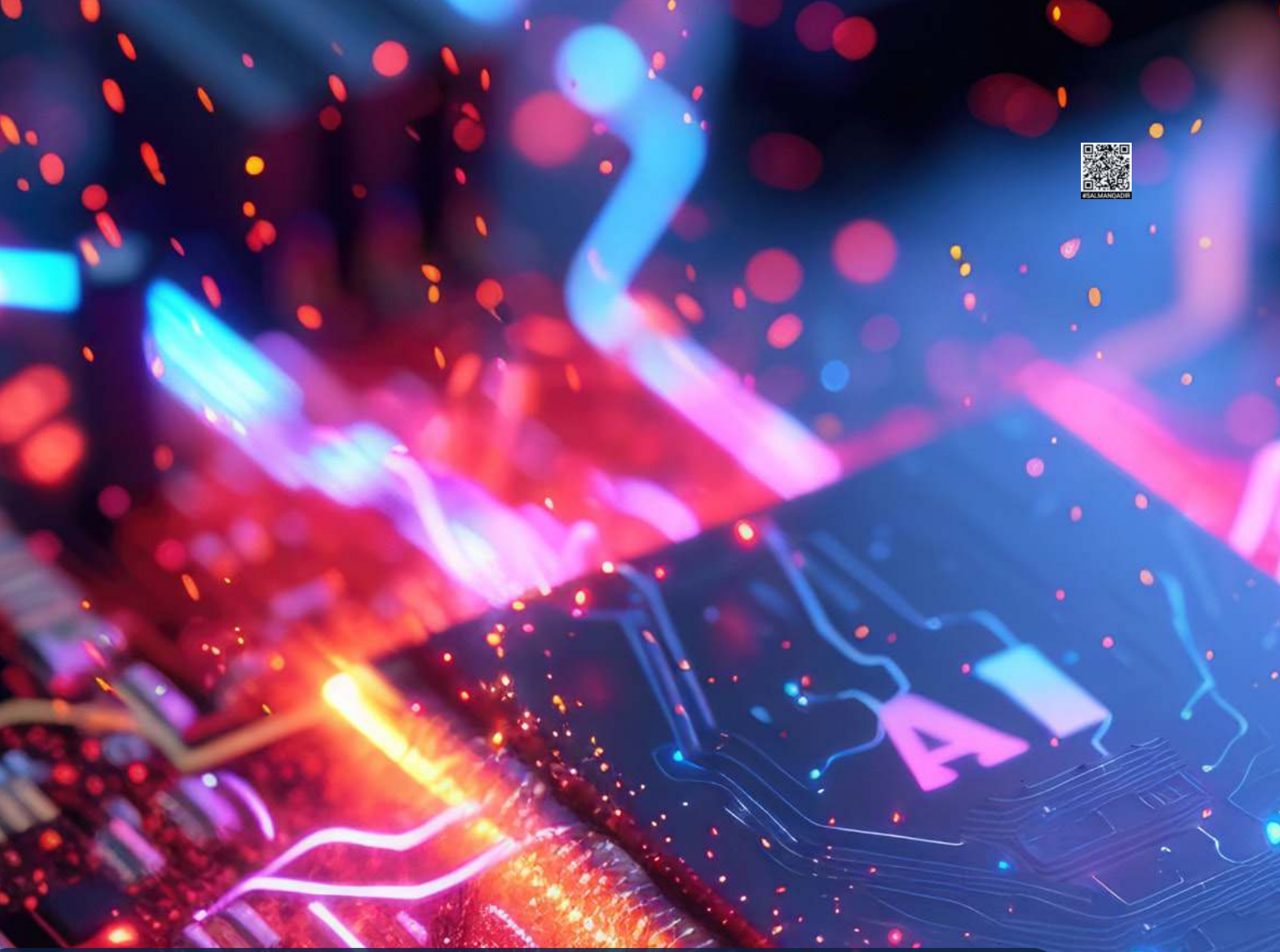


RAND EUROPE

# Strategic competition in the age of AI

## Emerging risks and opportunities from military use of artificial intelligence

James Black, Mattias Eken, Jacob Parakilas, Stuart Dee, Conlan Ellis, Kiran Suman-Chauhan, Ryan Bain, Harper Fine, Maria Chiara Aquilino, Mélusine Lebret, Ondrej Palicka

# Preface

## Background to this study

Artificial intelligence (AI) holds the potential to usher in transformative changes across all aspects of society, economy, and policy, including defence and security. The UK aspires to be a leading player in the rollout of AI for civil and commercial applications, and in the responsible development of defence AI. This necessitates a clear and nuanced understanding of the emerging risks and opportunities associated with the military use of AI, as well as how the UK can best work with others to mitigate or exploit these.

In March 2024, the Defence AI & Autonomy Unit (DAU) of the UK Ministry of Defence (MOD), and the Foreign, Commonwealth and Development Office (FCDO) jointly commissioned a short scoping study from RAND Europe. The goal was to provide an initial exploration of ways in which military use of AI might generate risks and opportunities at the strategic level – conscious that much of the research to date has focused on the tactical level or on non-military topics (e.g. AI safety). Follow-on work will then explore these issues in more detail to inform the UK strategy for international engagement on these issues.

This technical report aims to set a baseline of understanding of strategic risks and opportunities emerging from military use of AI. A standalone summary report focuses on high-level findings for decision makers.

## About RAND

This study was conducted by a mix of RAND staff on both sides of the Atlantic. With offices in the UK, Belgium, and the Netherlands, RAND Europe is the European arm of RAND, a non-profit research institute and the largest policy research organisation in the world. RAND's mission is to help improve public policy and decision making through objective research and analysis, having delivered over 75 years of classified and unclassified studies for UK, US and other allied governments.

RAND has been involved in research into the military and strategic implications of AI since the 1950s, having also played a vital role in developing game theory, deterrence theory, and nuclear strategy.

For more information on the study, this report or RAND, please contact:

James Black
Assistant Director
Defence and Security Research Group
RAND Europe
e. jblack@randeurope.org

# Summary

## The advent of AI is ushering in profound changes to competition and conflict

AI is best understood as a dual-use set of **general-purpose technologies,** hardware-enabled but software-based. Unlike traditional military technologies, they are highly democratised and proliferating fast. Innovation is driven by the private sector for commercial uses, not by government or defence.

Collective understanding of military applications and implications is **improving, but from a low base.** Too often, debate prioritises **certain high-profile issues** – e.g. lethal autonomous weapon systems (LAWS) or artificial general intelligence (AGI) – at the expense of other topics. It focuses on the tactical at the expense of the strategic; risks at the expense of opportunities; or the immediate consequences of military AI at the expense of the **second- and third-order effects** that might be most impactful in the long run.

To address this, the MOD and FCDO commissioned this study to develop a **conceptual framework** mapping the strategic risks and opportunities arising from military AI.

**Figure 0.1 Framework: strategic risks and opportunities of military use of AI**


*Source: RAND Europe analysis (2024).*

## AI poses complex, significant and underappreciated risks to defence and security

Of the many risks and opportunities explored in detail in this report, the most pressing include:

- **Information manipulation**, such as AI deepfakes, which could not only drive political, economic and social problems but also skew military decision making in times of crisis.
- **Empowerment of non-state actors** with asymmetric capabilities that challenge the dominance of state militaries or, in the worst-case scenario, new tools of mass destruction (e.g. bioweapons).
- The interlinked impacts of AI on the **offence–defence balance** between adversaries, on **escalation dynamics** towards warfighting, and on the **stability of nuclear deterrence**. These issues are especially concerning amidst intensifying superpower rivalries and in a world already grappling with other drivers of insecurity (e.g. Ukraine, Israel–Iran, Taiwan, migration, climate change).
- The potential catastrophic safety and security risks associated with any future **advent of AGI**.

**Table 0.1 Priority risks and opportunities for action**

| SELECTED PRIORITY RISKS AND OPPORTUNITIES | | | INITIAL ASSESSMENT OF IMPACT | | |
|---|---|---|---|---|---|
| **Framework category** | | **Issue** | **Significant** i.e. potential to disadvantage in sub-threshold | **Severe** i.e. potential to disadvantage in conventional war | **Catastrophic** i.e. potential for catastrophe or existential threat |
| **National** | | Economic disruption and warfare | ● | | |
| | | Information-manipulation (e.g. deepfakes) | ● | ● | ● |
| | | Changes to defence productivity, mass and lethality | ● | ● | |
| **International** | By actor type | Erosion of RBIO and governance institutions | ● | ● | |
| | | AI-enabled repression (and export thereof) | ● | | |
| | | Empowerment of non-state actors (e.g. bioweapons) | ● | ● | ● |
| | By conflict type | Changes to military offence-defence balance | ● | ● | |
| | | Impact on escalation dynamics | ● | ● | ● |
| | | Impact on nuclear | ● | ● | ● |
| **Macro-trends** | | Prospects for AGI and non-alignment | ● | ● | ● |

*Source: RAND Europe analysis (2024).*

At home, there are also major issues to contend with in terms of disruptive impacts on **domestic politics** and the **economy**. These shape the ends and means available to Defence. Abroad, AI could similarly have profound implications for the health of the **rules-based international order**, depending on whether and how effectively nations, industry and civil society work together to manage its effects. There is significant concern among AI experts about the extent to which AI could tip the balance in favour of repressive and **authoritarian modes of governance** in many parts of the world, while simultaneously threatening to subvert democratic politics, pollute the **information environment** and undermine societies' **will-to-fight**.

## Equally, leadership on military AI could deliver outsized benefits

Many of these potential risks could also manifest as opportunities for strategic advantage. The balance of pros and cons from the rollout of AI hinges on how quickly and effectively nations are **able to adapt** institutions such as their Armed Forces to exploit AI's benefits. Similarly, it depends on how well governments can exert influence internationally to shape global behaviours on military AI in a direction that suits their interests and values. This means being willing to deliver the significant investments, organisational reforms and cultural changes needed to transform Defence's approach to new technologies.

## Urgent action is needed to mitigate emerging risks and exploit opportunities

To address these challenges, nations must urgently develop a **comprehensive action plan** that considers the complex interplay of technical advances *in* AI, geopolitical competition *over* and *through* AI, and evolving norms *around* AI in the international system. This should draw upon a **toolkit of mechanisms** to influence **different audiences**, employing all diplomatic, information, military and economic (DIME) levers to bring together a proactive set of:

- Efforts to boost the responsible uptake of AI and maximise its benefits to Defence.
- Efforts to limit the adoption of military AI by non-state and terrorist actors, or hostile / rogue states, while also imposing costs on them to influence their actions.
- Efforts to shape global, minilateral and bilateral governance arrangements for military AI.

**Table 0.2 Toolkit of mechanisms for shaping global defence AI developments**

| CATEGORY OF TOOLKIT | PRIORITY ACTIONS |
|---|---|
| Mechanisms to boost AI adoption and benefits for UK Defence |  Accelerate investment in and adoption of AI across Defence, while increasing resilience against hostile or accidental misuse of AI |
| Mechanisms to restrict AI adoption and benefits for adversaries |  Adopt a campaigning approach to restrict, slow, or increase the costs to adversaries (state or non-state) of deploying military AI |
| Mechanisms to shape emerging governance arrangements for military AI |  Play a leading role in awareness raising, problem finding, and sharing learning about military AI risks |
| |  Develop transparency and confidence building measures with key allies (e.g. US) and competitors (e.g. China) to reduce escalation risks |
| |  Promote an inclusive, participatory approach to build an emerging global consensus on norms of responsible behaviour around military AI, as a prelude to more robust binding agreements in future |
| |  Promote parallel development of minilateral mechanisms for reducing urgent nuclear- and bio-related AI risks |
| |  Investigate ways to incorporate AI into verification and compliance mechanisms, and vice versa |
| |  Over time, consolidate the current fragmented landscape of AI governance initiatives into a more concrete architecture |

*Source: RAND Europe analysis.*

This should also build on **lessons from other domains** – as examined in this report – and the momentum of **recent high-level initiatives** on AI. Prominent examples include the Bletchley Summit, the Responsible AI in the Military Domain (REAIM) summit, and the Political Declaration on Military AI.

# Table of contents

# Tables

# Figures

# Boxes

# Abbreviations

| | |
|---|---|
| AI | artificial intelligence |
| AIPfD | AI Partnership for Defence |
| AGI | artificial general intelligence |
| ASI | artificial super intelligence |
| AUKUS | Australia–United Kingdom–United States |
| C2 | Command and Control |
| C4ISTAR | Command, Control, Communications, Computers and Intelligence, Surveillance, Target Acquisition and Reconnaissance |
| DAU | Defence AI & Autonomy Unit |
| DIME | Diplomatic, Information, Military, Economic |
| DoD | US Department of Defense |
| EU | European Union |
| FCDO | Foreign, Commonwealth and Development Office |
| FVEY | Five Eyes |
| GC-REAIM | Global Commission on Responsible AI in the Military Domain |
| GGE | Group of Government Experts |
| GPT | general-purpose technology |
| IHL | international humanitarian law |
| LAWS | lethal autonomous weapons systems |
| LLM | large language model |
| LOAC | Law of Armed Conflict |
| MCF | military-civil fusion |
| ML | machine learning |
| MOD | Ministry of Defence |
| NATO | North Atlantic Treaty Organization |

| | |
|---|---|
| NGO | non-governmental organisation |
| PAG | Partners Across Government |
| REAIM | Responsible AI in the Military Domain |
| TCBM | transparency- and confidence-building measures |
| TTC | EU–US Trade and Technology Council |
| UK | United Kingdom |
| UN | United Nations |
| US | United States |

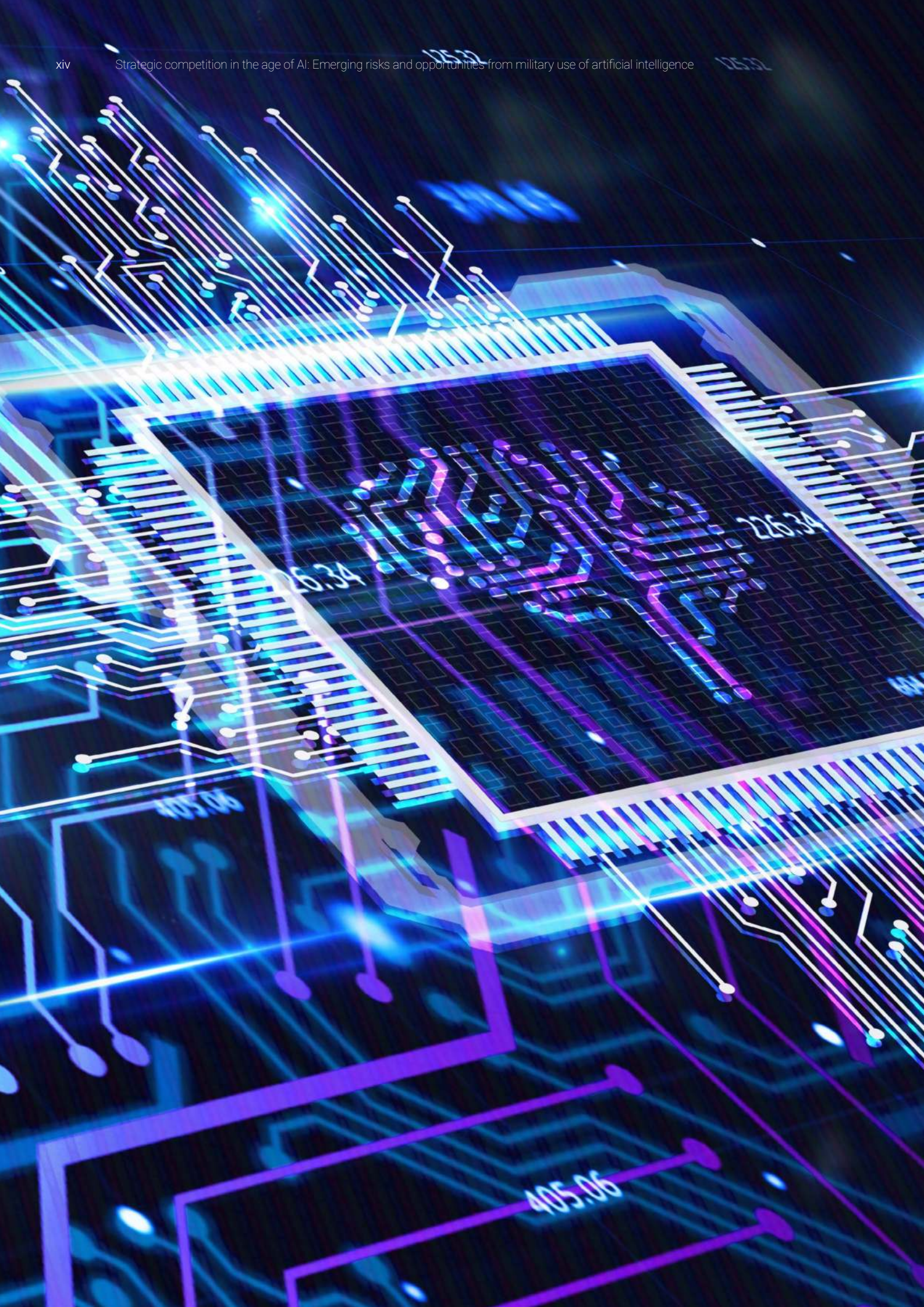
# Acknowledgements

This short study would not have been possible without contributions from several institutions and individuals. The authors are especially grateful to Mike Gibson, Sam Phillips and Oliver Mahoney of the DAU in the UK MOD, and to Anthony Stanley and Sian Evans of the FCDO, for their sponsorship of the research and their feedback throughout the study.

Within RAND, thanks are owed to Kiera Addair from Knowledge Services for her support with the extensive literature review, to Jess Plumridge and Hannah Beelam for their graphic design on the final report; to Ben Plumridge and Abi Saffrey for their copy editing; and to our quality assurance reviewers, Erik Silfversten and Luke Huxtable, for providing comments and feedback on drafts.

Crucially, too, the RAND team are grateful to the insightful contributions made by the 54 government stakeholders and academic, think tank or industry experts who took part in research interviews within the tight timeframes for this study.

Notwithstanding all these valued inputs, all errors or omissions in this draft report remain the sole responsibility of the authors.


26.34
405.06

# Chapter 1. **Introduction**

## 1.1. Research scope, objectives and methodology

### 1.1.1. Adoption of AI by state militaries and armed non-state groups is ushering in significant changes to the character of competition and conflict

Development, integration and use of AI for military purposes could have profound implications for the future of warfare and for international peace and security more generally. This presents both opportunities and risks, whether at the tactical, operational or strategic level. Already, uncrewed robotic systems are being deployed in unprecedented numbers and with increasing levels of autonomy across the battlespaces of Ukraine, Israel–Gaza and the Red Sea. AI is similarly being integrated into intelligence analysis, command and control (C2), targeting, fires, training, simulation, equipment monitoring, and logistics.

Such trends have driven debates over the impact and possible trajectories of algorithmic warfare, known as 'intelligentised' warfare by China's People's Liberation Army (PLA). These discussions do not only concern the immediate military, ethical and legal impacts. There is also an increasing recognition of broader strategic implications. Examples include concerns about the effects of AI on state competition or conflict escalation, extending to the risk of nuclear warfare.[1]

Faced with a nascent understanding of these cascading implications, governments must work with industry, academia and civil society to improve understanding of how to accentuate the benefits of military AI while reducing potential risks and hazards at the strategic level. For its part, the UK Government published a Defence AI Strategy in June 2022, with a focus on responsible development and deployment of AI in a defence context. This entails not only accelerating adoption of AI by the MOD and Armed Forces, but also proactively 'shaping global AI developments' to manage both benefits and risks. The UK thereby aims to play a leading role in shaping governance of this fast-moving technology by collaborating with allies and partners, engaging neutral countries and, where interests align, working with adversaries to build a shared understanding of the possible strategic implications and how to manage AI's global impact.

### 1.1.2. The MOD and FCDO commissioned RAND to build a framework for thinking about these emerging strategic risks and opportunities from military use of AI

Against this backdrop, RAND was asked to deliver an exploratory study into the potential impacts, both beneficial and detrimental, of military AI on the strategic level. Specifically, this one-month initial study aimed to identify components of a potential conceptual framework that could aid in comprehending the strategic effects of AI to help facilitate an informed response from UK Government, including Defence.

As discussed in Chapter 2, the UK's Defence AI Strategy defines AI as a collection of

1 Kania (2019); Liu & Maas (2021).

general-purpose technologies, each of which has the potential to empower machines to execute tasks that would typically necessitate biological intelligence. Machines can also learn from large datasets how to perform these tasks, such as identifying patterns, acquiring knowledge from experiences or making predictions. Currently, the state-of-the-art in AI is advancing rapidly, with much uncertainty over its future directions. As such, any framework for conceptualising the risks and benefits of military AI needs to consider the wide range of current applications and possible future developments. This does not mean making predictions about the most likely trajectories for technological progress, as these are likely to be wrong. Rather, it means mapping out different potential strategic risks and opportunities that could come to pass depending on how both technology and governance mechanisms evolve. These should be grouped based on common features to enable a systematic approach to thinking about issues such as deterrence, strategic stability or proliferation.

Armed with this mapping of possible impacts of military AI at the strategic level, this exploratory study was then tasked with high-level identification of possible ways to influence these risks or opportunities. This included incorporating any insights emerging from other sectors (e.g. nuclear arms control, space, biotech) or the approaches and thinking of other countries around the world.

### 1.1.3. The research team employed a multi-method approach, combining literature review and interviews with government, industry, think tanks and academia

To inform this initial exploratory study, the research team drew on:

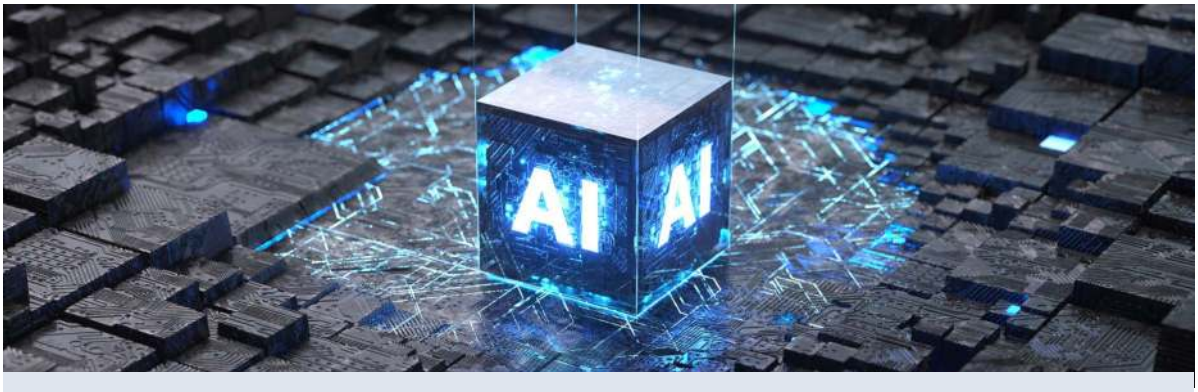
- A narrative literature review of ~200 academic or 'grey' sources, derived from a long-list of 1,500.
- Semi-structured interviews with over 50 stakeholders and experts from across government, UN, NATO, defence industry, AI firms, academia, think tanks and non-governmental organisations.
- Seven external workshops or webinars and two parliamentary inquiries held alongside the study.
- Iterative development of a conceptual framework in consultation with the MOD and FCDO.

More information on the methodology and interviewees can be found in Annexes A and B respectively. The findings presented below are not intended as definitive, but rather as the basis for further research and discussion – conscious especially that RAND's study was undertaken in only four weeks, placing tight constraints on the time available both for data gathering and for framework development or testing.

This technical report details the study's findings, beginning with the logic behind the proposed conceptual framework (Chapter 2), before moving to different categories of impact from military AI (Chapters 3–6), priority issues for action (Chapter 7), lessons from other domains (Chapter 8), and a toolkit of measures to address them (Chapter 9), before concluding with next steps (Chapter 10).

# Chapter 2. Towards a framework for strategic impacts from military AI

This chapter outlines the case for a more structured and multidisciplinary approach to mapping potential strategic risks and opportunities arising from military use of AI. In doing so, it delves into fundamental aspects of this technology, including its dual-use nature encompassing both civil and defence applications.




**Box 2.1 Summary of findings: Chapter 2**

AI should be understood as a set of general-purpose technologies (GPTs), hardware-enabled but software-based. Unlike traditional military technologies, they are highly democratised and proliferating fast; innovation is being driven by the private sector for civil and commercial uses, rather than by governments or defence establishments.

Our understanding of the applications and implications of these technologies is improving, but from a low base. Despite a lot of hype around AI, there are significant gaps in both our theoretical understanding and our empirical data on the potential benefits, drawbacks and risks of different use cases for AI, including in a military setting.

This has prompted intense and at times highly ideological debates among global AI experts, and left policy makers grappling with high levels of uncertainty around the likely pace and direction of future advances.

Crucially, this uncertainty not only exists in relation to the technical dimension of AI, but also its human element. As well as a set of GPTs, AI needs to be understood as a complex socio-technical system. Military applications will be shaped as much by the operational, organisational and cultural context in which AI technologies are developed and deployed as by the underlying characteristics of the technology itself.

Despite the deficiencies in current evidence and understanding, policy makers cannot afford to wait for perfect clarity before acting on AI rollout and governance. Given this urgent need to bring structure to thinking about military AI, this chapter proposes an initial and high-level categorisation of different types of strategic risk and opportunity, with different sub-categories then elaborated upon in subsequent sections of this report.

*Source: RAND Europe analysis.*

## 2.1. Problems with current understanding of the impacts of military AI

### 2.1.1. AI is a family of GPTs, a broad scope that complicates discussions of impact given the diversity of AI techniques and applications

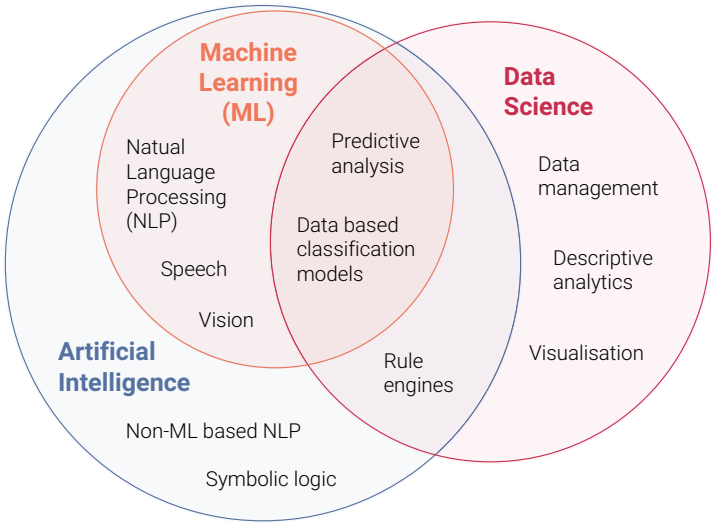
AI is increasingly recognised as a set of GPTs, in some ways akin to the combustion engine, electricity or the Internet. GPTs are technologies that have the potential to drastically impact productivity across many sectors, including defence, and significantly transform societal structures and individual lifestyles. AI is therefore much more than any single given technology (for example, large language models [LLMs] such as ChatGPT). Rather, AI is a heterogenous group of different systems, methods and applications, each with their own developmental trajectories and implications. There are also strong overlaps with fields such as data science and links to hardware technologies such as computing or robotics.

**Box 2.2 Definition of AI in the UK Defence AI Strategy**

'… 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.'

*Source: UK MOD (2022).*

**Figure 2.1 Depiction of overlaps between AI, ML and data science**




**Overlapping technologies**

**AI:** Machines that perform tasks normally requiring human intelligence, specially when the machines learn from data how to do those tasks.
*– UK National AI Strategy*

**ML:** Computer algorithms that can 'learn' by finding patterns in sample data and then apply this to new data to produce useful outputs, often using neural networks.
*– Alan Turing Institute*

**Data Science:** Research that involves the processing of large amounts of data in order to provide insights into real-world problems.
*– Alan Turing Institute*

*Source: UK MOD (2022).*

At a high level, AI can be further differentiated into Narrow, Broad and Strong AI:

- **Narrow AI, also sometimes known as Weak AI,** refers to AI systems that are designed to perform a narrow task (e.g. facial recognition or Internet searches) and can only operate under a limited predefined range.[2] They are specialised systems that excel in their specific tasks but lack the ability to understand or apply knowledge beyond their programming.
- **Broad AI** refers to an approach to AI that focuses on creating systems capable of generalising knowledge and skills across multiple tasks and domains. These systems would be able to adapt to tasks, but not at the level of sentience or comparable to human performance.
- **Strong AI**, such as artificial general intelligence (AGI), refers to those systems able to understand, learn, adapt, and implement knowledge across a broad range of tasks at a level equal to or beyond human capabilities. AGI, or the related concept of artificial superintelligence (ASI), is a long-term goal of many research programmes, but largely theoretical at this point.

Problematically, debates over the impact of AI suffer all too often from imprecision and conflation between these different types of system, or misunderstandings between technical and policy communities who come at the issue with different perspectives, assumptions and skillsets.[3]

### 2.1.2. AI forms part of a complex socio-technical system, with deep uncertainty about both the trajectory and pace of development and real-world adoption

AI should not be seen in isolation but as a complex socio-technical system with a significant human element across the lifecycle, from initial research and development (R&D) through to commercialisation and real-world deployment.[4] Humans directly and indirectly affect the strengths, weaknesses and biases of AI systems, for example through initial engineering, training data for algorithms or the quality of prompts used when tasking LLMs. In turn, the impact of AI is not solely determined by the technology itself, but also by how it is integrated with other systems, how it is used by humans, and how it is perceived and governed by institutions, regulators and societies.[5] This combination of AI with other technologies – whether legacy systems or other novel technologies such as biotech, quantum or robotics – may result in some of the most significant impacts.[6] Yet understanding the nuances of technology convergence requires systems thinking and a diverse mix of interdisciplinary expertise that can be hard to achieve.

Uncertainty extends not only to technological progress but also to the absorptive capacity of organisations, such as Defence, to translate new technologies into applications and innovation. The pace of change will be influenced by a wide range of factors. These include advances in the technology itself, but also policy decisions, societal acceptance and economic conditions. Many national or defence AI strategies focus

2 Sheikh et al. (2023).

3 David Galbreath, interview by the authors, 19 March 2024; Giacomo Persi Paoli, interview by the authors, 2 April 2024.

4 Schaefer et al. (2021).

5 Anonymous, interview by the authors, 5 April 2024; Chris Spedding, interview by the authors, 15 March 2024.

6 Mouton et al. (2023).

as much, if not more, on overcoming barriers to AI adoption as on progressing the technical state-of-the-art.[7] These include a lack of workforce skills, fragmented approaches to data management or sharing, and antiquated procurement systems. The latter struggle to deal with the software-driven nature, iterative development cycles or sheer pace of change in AI technologies.[8]

Cultural barriers compound these uncertainties.[9] Only a minority of AI technical specialists understand the policy process, or the myriad ways in which AI may affect different organisational contexts, including the military.[10] Other issues include a lack of security clearances for many AI experts and, as demonstrated by the high-profile backlash within Google against its work for the Pentagon on Project Maven, an ethical aversion of some tech firms and employees to working with Defence. In turn, few political leaders, civil servants or military personnel have deep technical knowledge in AI or related fields such as data science. This means it can be hard to bridge these interdisciplinary, organisational and cultural divides.[11]

### 2.1.3. Discussions of AI in a military context often focus narrowly on the ethics and tactical impact of lethal autonomous weapons systems, or existential threats

For the reasons outlined above, and more, the future direction and pace of change in AI technologies is uncertain. Similarly unclear are what the cascading second- or third-order effects of AI may be, beyond the more obvious direct impacts.[12] Much of the defence debate focuses on two extremes:

- On the one hand, the tactical military impacts and associated ethical, legal and policy dilemmas associated with the combination of AI and autonomy with advances in robotics – most notably in relation to **lethal autonomous weapons systems** (LAWS) or so-called 'killer robots'.[13]
- On the other hand, the potential existential or **global catastrophic risks** (GCRs) arising from AI, either in terms of **AGI** (which may not be directly military in nature, but could nonetheless pose a threat to the human species if its goals and values did not align with our own survival) or in terms of AI's interplay with weapons of mass destruction (with the bulk of literature in this area focusing on nuclear escalation risks and on the convergence of AI with bioweapons).[14]

Both issues are important, but the focus on these topics to the exclusion of other risks and opportunities may serve as a barrier to, and distraction from, development of a more holistic and nuanced understanding of the strategic impacts of military AI. And while many of the contributors to debates on both LAWS and GCRs bring deep technical knowledge on the realities of AI, popular views are often shaped

7 UK MOD (2022).

8 Andrew van der Lem, interview by the authors, 22 March; Heather Roff, interview by the authors, 27 March 2024.

9 FCDO official, interview by the authors, 19 March 2024.

10 Horowitz & Kahn (2023).

11 Scharre (2023).

12 Anonymous, interview by the authors, 8 April 2024; Anonymous, interview by the authors, 22 March 2024.

13 Meerveld et al. (2023).

14 Scharre & Lamberth (2022); Mouton et al. (2023).

by reductive preconceptions emerging from science fiction (e.g. Terminators, Skynet, etc.).

The focus of many policy makers, analysts and civil society organisations on the ethical implications of LAWS has usefully served to underscore the need for governments to develop, integrate and roll out military AI in a responsible matter, and to build social acceptability and political legitimacy for increasingly autonomous systems.[15] Equally, being responsible also means adopting military AI at a sufficient pace to deter and defeat aggression from adversaries, state or non-state, who might threaten international peace and security if they themselves exploit AI to gain a decisive battlefield advantage.

Such debates can overlook the more prosaic but no less significant impacts that AI can have away from any fighting, for example through integration into procurement, logistics or personnel management systems across Defence.[16] Discussions about whether the human should be 'in', 'on' or 'out of the loop' for decision making are important, especially when considering lethal force.[17] So too, though, are deeper conversations about how to build the best human–machine teams given the strengths, weaknesses and cognitive biases of both human and machine agents, as well as the demands of different specific tasks or situations (e.g. the speed and nature of the decisions to be made).[18] This goes far beyond LAWS or AI's use in targeting and affects the technology's adoption across all aspects of the wider defence enterprise.

Similarly, research and action on GCRs is essential. Even if low-probability, any AI-related developments that could pose an existential risk merit proper analysis, modelling and proactive risk mitigation measures, given the global scale and dire consequences of their potential impacts.[19] Equally, though, the fierce debate that has emerged within the AI community between those focused on existential risk and those focused on nearer-term risks (e.g. concerns around bias, privacy, inequality, etc.) poses a false dichotomy to policy makers. It is imperative to address both types of risk, and this should be feasible with the collective resources and political bandwidth of major governments and tech firms. This means iteratively developing solutions to the immediate practical challenges posed by AI adoption (e.g. developing governance arrangements to mitigate concerns around safety and bias and accentuate the technology's benefits) while also being mindful about any longer-term trends and path dependencies that could lead to GCRs (see Chapter 7).

## 2.2. Contours of a possible conceptual framework

### 2.2.1. The initial framework presented in this report is informed by a set of design criteria, assumptions and caveats

Given the shortcomings in understanding outlined in Section 2.1, the MOD and FCDO asked RAND to develop a structured way of thinking about and categorising the strategic risks and opportunities (collectively,

15 Hoadley & Lucas (2018).

16 Joe Wang, interview by the authors, 21 March 2024.

17 Wong et al. (2020).

18 Schaefer et al. (2021).

19 Scharre & Lamberth (2022).

**Box 2.3 Design considerations for a possible conceptual framework**

- Need to recognise AI as a set of GPTs and a complex socio-technical system with a crucial human dimension.
- Need to reflect the deep uncertainty that exists around the future trajectories and pace of progress in AI technologies, as well as around organisational, cultural, financial and other barriers to real-world adoption.
- Need to focus on the under-scrutinised strategic level of defence, while acknowledging that tactical and operational-level impacts from military AI may have aggregate effects on the strategic level.
- Need to move beyond the important but potentially distracting high-profile debates on LAWS or GCRs to consider a wider range of possible impacts arising from military use of AI.
- Need to be flexible and future-proof, with the framework able to accommodate rapid changes in AI technologies rather than being tied to near-term priorities (e.g. LLMs such as ChatGPT) and thus soon rendered obsolete.
- Need to be coherent with theory (e.g. around strategy, deterrence or warfare), accessible to a non-technical audience, precise with language and, where possible, orthogonal in categorisation of types of impact.
- Need to accommodate not only direct impacts from military AI but also potential second- and third-order effects, and the feedback loops that may occur across levels or areas (e.g. military and non-military levers).

*Source: RAND Europe analysis (2024).*

'impacts') arising from military use of AI. To this end, the research team undertook a review of existing conceptual frameworks and typologies within academic literature, a search which emphasised the lack of any such comprehensive or universally agreed tools. To help develop a new framework, the literature review and interviews with officials and experts emphasised several design considerations.

The remainder of this chapter provides a brief description and visual depiction of a possible framework that aims to meet these design considerations. Chapters 3 to 5 then elaborate on each of the main levels of the framework, and the categories and sub-categories therein.

This initial framework is intended as a guide to further research, analysis and policy discussions. It is certainly not intended to be definitive, given both substantive issues (i.e., the complexity, uncertainty and rapid pace of change in AI as already mentioned) and practical considerations (i.e., the fact that this research was undertaken in a very short period). Instead, the framework is designed as the basis for further debate, iteration and refinement to incorporate further insights and learning over time – especially in terms of prioritising one strategic risk or opportunity over another or improving understanding of timelines to certain breakthroughs in AI – which is beyond the scope of this short exploratory study.

### 2.2.2. The highest level of the framework addresses different types of AI-related risk and opportunity, of which the impacts of military AI are a subset

Figure 2.2 depicts different categories of strategic impact, conscious of the dual-use nature and characteristics of AI (e.g. as a set of GPTs) and uncertainty about the pace and direction of future change in the technological state-of-the-art. At the highest level of abstraction, risks/opportunities can be:

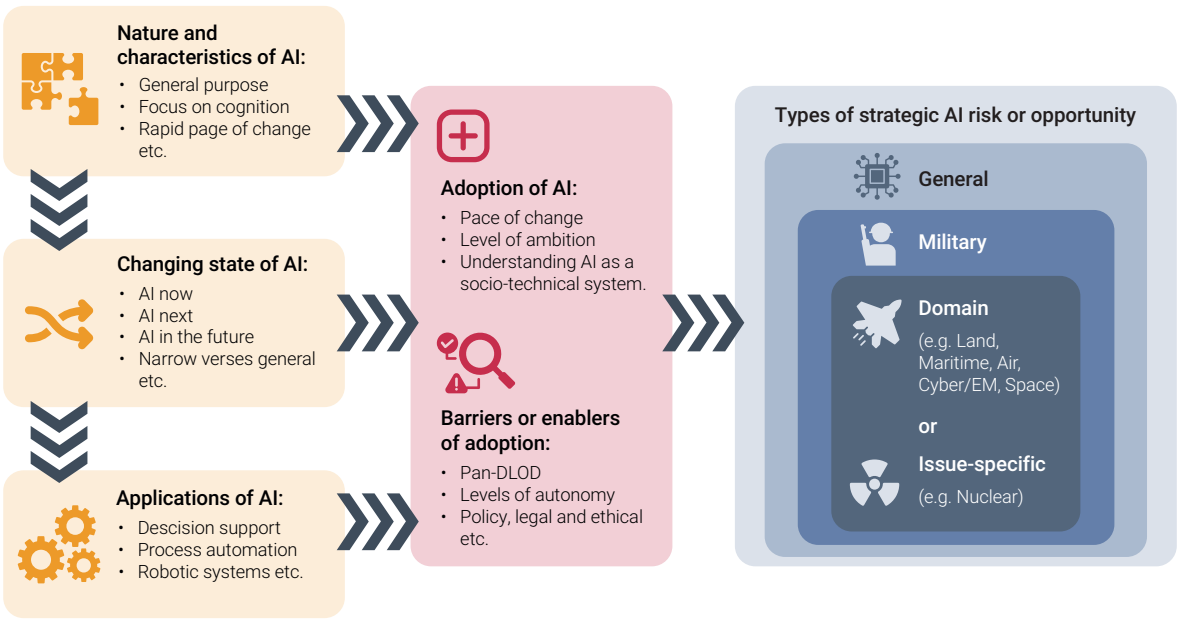
- General (e.g. AI safety);
- Military-specific; or
- Domain- or issue-specific (e.g. nuclear deterrence).

Crucially, as noted above, AI does not exist in a vacuum. Rather, it is best understood as a complex adaptive socio-technical system, with an important human component. As such, the framework also emphasises the need to think about how various barriers/enablers could shape how quickly and in what ways new AI technologies are absorbed into ministries of defence and military organisations.

### 2.2.3. Subsequent levels of the framework address the interplay of impacts at both global and actor level, across the full continuum of competition and conflict

Figure 2.3 moves into more detail on strategic risks and opportunities specifically relating to military AI and breaks this down further across several dimensions. It builds on prior conceptual work for both the US Government (the Department of Defense, and Office for Net Assessment) and the UK MOD (the Secretary of State's Office for Net Assessment and Challenge, and Defence Science and Technology) around the dynamics of strategic competition and how new technologies can affect a given actor's strategic advantage or

**Figure 2.2 Conceptualising the AI lifecycle from development to adoption to impact**




*Source: RAND Europe analysis (2024).*

lack thereof.[20] The framework thereby aims to capture how military use of AI could impact:

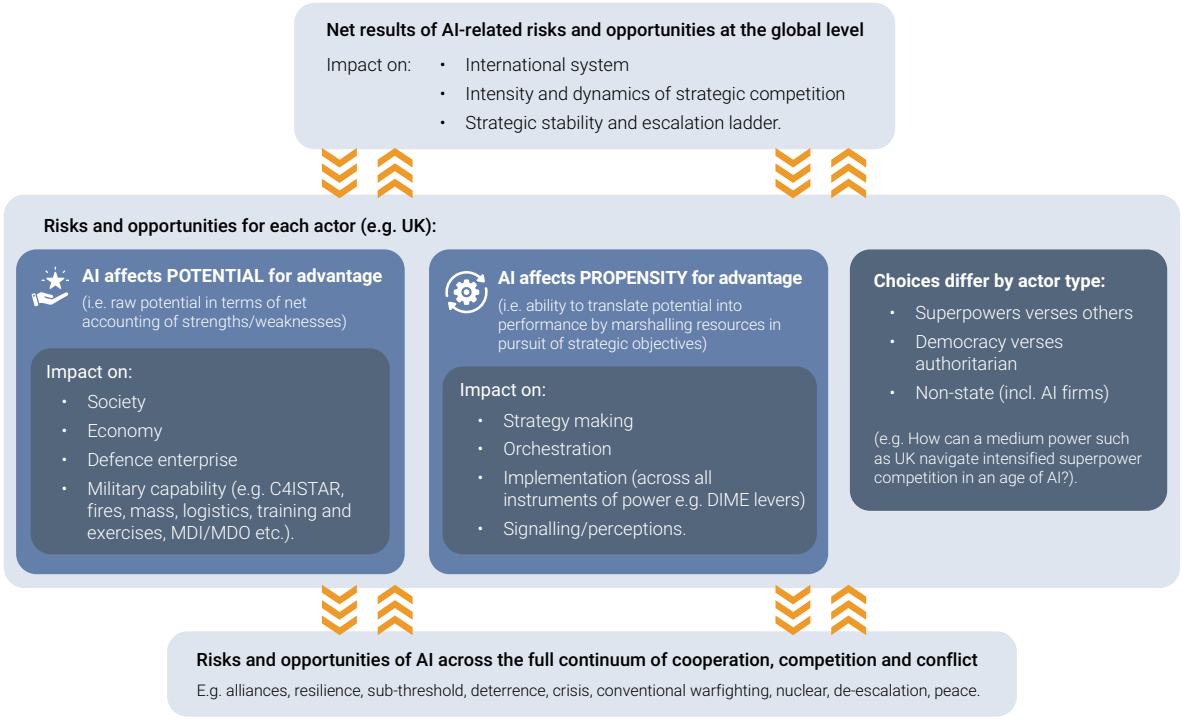
- The international system;
- The intensity and dynamics of strategic competition or collaboration within that system; and
- The potential and propensity of individual actors to achieve strategic advantage within that persistent global competition.

The framework further differentiates between how AI-related impacts manifest differently:

- Between actor type: differing between superpowers (e.g. US and China), medium powers (e.g. UK) and small states; democracies and authoritarian regimes; or state and non-state actors.
- Across the full continuum of cooperation, competition and conflict: from alliance-building through to deterrence, crisis management, conventional warfighting or even nuclear exchanges.

Crucially, the non-linear and relational nature of strategic competition means that the above continuously interact with and impact each other, with a series of feedback loops between impacts at different levels.

**Figure 2.3 Framework: strategic risks and opportunities of military use of AI**




*Source: RAND Europe analysis (2024).*

20 Mazarr, Rhoades et al. (2022); Black et al. (2023); Heath et al. (2024).

For example, AI-related impacts on the stability, polarity, institutions, norms and dynamics of the international system have cascading effects on what nation states compete or collaborate over, how, where, why and with whom. In turn, the impacts of military AI on the asymmetric strengths or weaknesses of a given actor affect their ability to influence strategic outcomes and thus to reshape the international system in their favour, including through war.[21] This reflects the dialectical nature of strategy (a contest between opposing wills and intelligences) and the nature of geopolitics as a complex adaptive system with emergent properties that arise from the interplay of different competing actors and other factors (e.g. environmental or technological change), as well as fog, friction and an element of chance.[22]

## 2.3. Summary

There is a pressing need for a more structured, nuanced, empirically based and interdisciplinary debate over the potential strategic risks and opportunities arising from the growing military use of AI. Currently, there are significant pockets of research, analysis and discussion. But, all too often, different scholarly communities are siloed off from one another. There are consequently substantial knowledge gaps, as well as misunderstandings between technical and policy specialists. Furthermore, much of the debate focuses on certain high-profile issues (e.g. LAWS or GCRs) at the expense of other topics; on the tactical at the expense of the strategic; on the risks at the expense of the opportunities; or on the immediate consequences at the expense of the second- and third-order effects that might be most impactful.

This report proposes a conceptual framework as the basis for mapping the full breadth of potential strategic risks and opportunities emerging from current or future use of AI in a military context, based on a set of design criteria derived from expert interviews and an extensive literature review. It is intended as the basis for iteration, not least as the technical feasibility and real-world impacts of different applications become better known through further theoretical work, empirical research or lived experience.

The following chapters delve into different levels of the framework, providing more detailed discussion of sub-categories of impact within each, and illustrating possible strategic risks and opportunities of note.

21 Anonymous stakeholder, interview by the authors, 25 March 2024.

22 Black et al. (2023).

# Chapter 3. **Impact at the national level**

This chapter considers the first category of the framework: the risks and opportunities that AI poses at the level of individual strategic actors (typically, but not exclusively, nation states). Given the focus of this study on military rather than civil applications of AI, the discussion is framed through the lens of the constant competition among actors for strategic advantage. This in turn influences the balance of power and the degree of peace, prosperity, and stability at the international level – topics covered in Chapter 4.

**Box 3.1 Summary of findings: Chapter 3**

First, this chapter explores how AI could affect the **potential** of actors for advantage (i.e. their raw potential in terms of a net accounting of their strengths and weaknesses). This discussion moves from the higher-level impacts on society and the economy through to those on the defence enterprise. Key issues that emerge include:

- The potential for AI to drive **sweeping economic disruption**, or be weaponised to wage economic warfare, with knock-on effects for prosperity, stability and security, as well as the financial settlement for Defence.
- The potential for AI to be used for **information manipulation** (e.g. highly sophisticated deepfakes), with consequences for everything from political warfare, subversion, electoral interference, crime and public trust.
- The potential for AI to transform the **productivity of the Defence enterprise** and support development and fielding of **military capabilities** with increased mass, survivability and lethality.

A recurring theme, across all these sorts of impacts, is the potential for AI to energise those governance systems (i.e. societies, governments, militaries) that manage to **adapt and integrate AI** into their day-to-day functions, while exacerbating the already acute pressures on those nations or organisations that are left behind.

Second, the chapter then considers how AI might impact the **propensity** of individual actors for advantage (i.e. their ability to translate that raw potential into actual beneficial outcomes depending on how efficiently they marshal their available resources in pursuit of strategic objectives). Here, the evidence suggests:

- AI could bring substantial benefits across the **strategy cycle**, from intelligence gathering and analysis, through to decision support and consideration of alternative courses of action, as well as enabling more effective collaboration across government and with allies and partners to implement an agreed strategy.
- Conversely, though, the literature and interviews emphasise concern about the potential for **AI bias, brittleness and failures**; a lack of proper understanding of how to get the most of both sides of human–machine teams; and limited appreciation of the **limitations, vulnerabilities, dependencies** or through-life support needs of military AI systems – not least given acute shortages of AI expertise within government.
- There is similarly a growing body of work on the risks associated with military AI in terms of the potential unintended consequences for **strategic signalling and perceptions** (or misperceptions) among different actors. AI hype and rhetoric do not help in this regard. If not properly addressed, such issues could drive an 'arms race' narrative around military AI, as well as increasing the chance of unintentional escalation in a crisis.

*Source: RAND Europe analysis.*

## 3.1. Understanding strategic advantage

This level of the framework focuses on the impacts of military AI on a given actor (e.g. the UK or another country) in terms of their ability to exert an influence on the international system discussed in Chapter 4. Here, the categorisation of AI-related risks and opportunities is built around the concept of strategic advantage – one of the central concepts of the UK's Integrated Review in 2021, and its Refresh in 2023.[23] Building on prior RAND research for the UK MOD and US DoD, strategic advantage can be understood in terms of both an actor's *potential* and *propensity* for advantage, as outlined below.

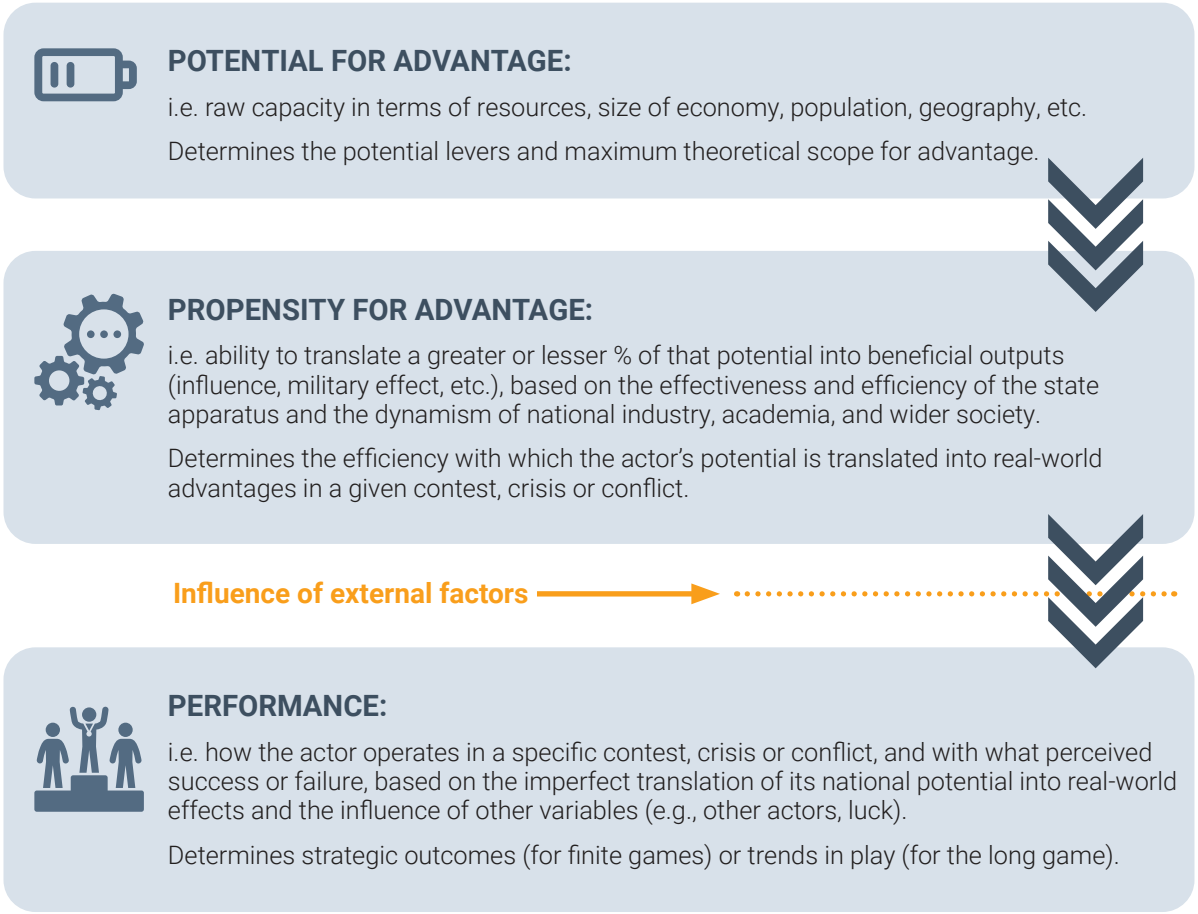
**Box 3.2 Definition of strategic advantage**

'A position of strategic advantage is one in which an actor is more likely than others to achieve their objectives in a given contest, crisis or conflict, having influenced the dynamics of competition in their favour and maximised the relevance of their own areas of asymmetric advantage across all levers of powers.'

*Source: Black et al. (2023), adapted from SONAC (n.d.).*

23 UK Government (2021); (2023).

**Figure 3.1 Relationship between an actor's potential and propensity for strategic advantage**




*Source: Black et al. (2023).*

The following sections examine AI-related impacts across each of these sub-categories, beginning with how strategic risks and opportunities arising from AI could affect an actor's raw potential for advantage.

## 3.2. AI impact: potential for advantage

This sub-category of the conceptual framework address AI-related impacts in terms of:

- Impact on society
- Impact on the economy (including the technology base)
- Impact on the Defence enterprise (including the MOD, Armed Forces and defence industry)
- Impact on available military capability (in terms of Defence Lines of Development[24]).

Collectively, these different themes aim to capture the strategic implications of AI in terms of boosting or undermining an actor's (e.g. a

24 By the UK MOD's definition, the Defence Lines of Development that make up any military capability are training, equipment, personnel, infrastructure, concepts and doctrine, organisation, information, logistics and interoperability. The US DoD and NATO both have their own equivalents, if using slightly different terminology.

nation's) underlying capacity for both hard (i.e. military, coercive) and soft (e.g. diplomatic, persuasive) power, as well as how favourable or volatile their domestic position is socially, politically and economically. This forms the raw potential that can then be translated into useful outputs, and influence over the outcomes of a given competition or conflict, depending on an actor's propensity for advantage (i.e. the effectiveness of their governance systems and strategy making and implementation).

### 3.2.1. AI is projected to have profound if unpredictable effects on society, shaping the infosphere, social attitudes and the effectiveness of governance systems

The literature review and interviews conducted for this study emphasised that, given AI's role as a set of inherently dual-use GPTs, it is impossible to segregate the pure military applications of AI from the wider impacts of civilian AI systems that could have cascading effects on defence and security.[25]

This includes the potential impacts on society itself. Prominent risks found by the research team include AI enabling an unprecedented spread of disinformation, causing social upheaval and atomisation, and undermining trust in facts, institutions or democratic politics.[26] Conversely, AI holds the prospect of enhancing the quality of public services and offering novel solutions to systemic challenges such as climate change, with the balance of risks and opportunities seen as dependent on how AI governance unfolds.[27]

**Table 3.1 National level: AI impacts on society**

| | Description |
|---|---|
| + | • Use of AI and big data helps to boost productivity and the quality of public services, leading to improvements in health, education, social care, transport and other policy goals<br>• AI provides new tools for managing the green energy transition and combating the effects of climate change and natural disasters (alleviating some of demand for military aid to civilian authorities/humanitarian assistance and disaster relief)<br>• AI supports more efficient R&D, enhancing innovation in other areas of science and technology (S&T)<br>• Impacts from AI on the national economy (see Section 3.2.2) and on international security and stability influence patterns of migration and demographic change<br>• Improvement in prosperity (see Section 3.2.2) drives wider social benefits e.g. mental health<br>• Improvements in policy outcomes due to use of AI in turn boost public trust in governance |

25 Johnson (2021b).

26 Tate Nurkin, interview by the authors, 20 March 2024; Joe Wang, interview by the authors, 21 March 2024.

27 Andrew van der Lem, interview by the authors, 22 March 2024.

| | Description |
|---|---|
| – | • AI-enabled deepfakes and disinformation campaigns fuel truth decay and governance crisis<br>• Extremist organisations or hostile states exploit AI to support recruitment and propaganda<br>• Division between AI haves and have-nots leads to increasing social tension<br>• Potential for backlash against AI on ethical, privacy and other grounds<br>• Social upheaval from AI (e.g. impact of automation on jobs) undermines public order, trust in institutions (incl. the military), and national or alliance will-to-fight<br>• Increased challenges to open, democratic societies as AI risks outstrip capacity to respond<br>• Use of AI for pervasive surveillance in more authoritarian states (see Section 3.4) |
| ? | • Uncertain outcomes from efforts to find technical and policy solutions to AI bias<br>• AI and its impacts on the infosphere drive unpredictable changes in cultural identities<br>• Uncertain impacts from AI on party politics and election interference<br>• Uncertain impacts from AI on the social contract<br>• Uncertain impacts from AI on public attitudes, values, and legal and ethical norms, including attitudes towards the use of human vs machine intelligence to perform different tasks<br>• Longer term, vast but uncertain implications from AGI or ASI on understandings of what it is to be human or how to reorganise society in an age of machine superintelligence |

*Source: RAND Europe analysis (2024).*

### 3.2.2. AI is expected to affect all sectors of the economy, if at differing paces, with potential for both winners and losers from any AI-fuelled economic disruption

AI is seen as a central component in the so-called fourth industrial revolution, with machine intelligence and automation affecting all aspects of value chains and touching all sectors of the economy.[28] On the one hand, AI proponents argue that AI and related digital technologies such as robotics, novel compute, telecommunications or big data, could bring substantial improvements in productivity. More efficient use of both capital and labour would then deliver better products and services and drive economic growth. This may advantage those knowledge-based economies that are best able to innovate, attract AI-related talent and develop new AI applications, as well as to extract value from data – a vital commodity, with analogies often drawn to natural resources (i.e. 'data is the new oil'). Conversely, there are risks that AI could fuel unemployment and disruption in labour markets, increased volatility in financial markets (given the rise of

28 Wright (2019).

algorithmic trading), and heightened inequality between those countries, regions, companies or individuals able to embrace the opportunities offered by AI and those excluded from partaking in its economic benefits.[29] Equally, there are concerns that AI could be purposefully weaponised as a tool of economic warfare, causing volatility or exerting coercive influence.[30]

**Table 3.2 National level: AI impacts on the economy**

| | Description |
|---|---|
| + | • AI drives economic growth and regional development<br>• AI boosts productivity across various economic sectors, boosting competitiveness and exports<br>• AI boosts role of data and analysis in optimising supply chains, financial trades or sanctions<br>• Automation reduces need for some jobs, but also creates others, shifting focus onto those activities where humans add most value (e.g. using soft skills)<br>• AI drives better economic/fiscal/monetary policy making backed by economic modelling and foresight, helping to bolster economic resilience and predict, absorb and recover from shocks<br>• Globally, AI supports development of emerging economies, lifting millions out of poverty |
| − | • Automation drives mass unemployment in some industries (esp. white collar)<br>• AI exacerbates skills shortages (e.g. in STEM) and brain drain (e.g. to Silicon Valley)<br>• Competition for advantage in AI leads to a race to the bottom on regulatory standards on issues such as data protections, algorithmic bias, harm prevention or privacy<br>• AI intensifies wealth and income inequality, concentrating the benefits of AI technologies in the hands of a few powerful countries, companies or super-wealthy individuals<br>• Weaponisation of AI drives new forms of economic warfare (e.g. deepfakes or memetic engineering to disrupt financial markets, attacking models behind algorithmic trading, etc.)<br>• Abuse of AI for fraud and other economic crimes, terrorist financing or evading sanctions<br>• Public sector and government finances are stretched by dealing with economic disruptions and the negative externalities of poorly regulated AI, even as private sector actors reap rewards<br>• Conversely, AI, coupled with automation, could potentially hinder the advancement of emerging economies, as the reliance on inexpensive labour diminishes, thereby eliminating a key component of their development trajectory |

29 Sigfrids et al. (2023).

30 Futter (2022).

| | Description |
|---|---|
| ? | • Uncertainty over which countries best seize the economic benefits of AI (beyond US, China)<br>• Uncertain prospects for new forms of AI-related wealth distribution e.g. universal basic income |

*Source: RAND Europe analysis (2024).*

### 3.2.3. AI is similarly expected to have profound impacts on the Defence enterprise and on the military capability development, with strategic consequences

Just as AI is expected to reshape civil and commercial organisations, or non-defence industries, so too is it expected to transform the Defence enterprise. This sub-category of the framework focuses on the risks and opportunities arising from AI's adoption across a ministry of defence as a military-strategic headquarters; across military commands, other top level budget holders and procurement agencies responsible for developing and acquiring new capabilities; and across the defence technological and industrial base. Here, the literature and interviews emphasised that AI could have – and in some cases is already having – profound implications not only on the military tasks, forces and capabilities that Defence needs to provide, but also on the oft-overlooked 'back office' functions that support those requirements.[31] Collectively, these could either transform the strategic bandwidth and competence of a ministry of defence, and the productivity, efficiency, resilience and value-for-money of defence industry, or see them left behind by more agile and innovative competitors – with knock-on effects on military and strategic competition.[32]

**Table 3.3 National level: AI impacts on the Defence enterprise**

| | Description |
|---|---|
| + | • Advanced AI decision support tools help to improve quality of MOD decision making as a military-strategic headquarters and management of Defence's resources and portfolio<br>• Further advances in AI enhance understanding via better intelligence analysis and prediction<br>• Process optimisation using AI tools transforms the efficiency of finance, procurement, logistics, personnel management, maintenance, infrastructure management and other IT systems etc.<br>• AI boosts productivity across the Defence workforce, increasing strategic bandwidth<br>• Applications of AI and other Industry 4.0 technologies to defence industry boosts productivity, industrial competitiveness, exports, supply chain resilience and value for money of defence programmes<br>• Applications of AI to defence R&D help to identify and absorb new S&T at greater pace |

31 Joe Wang, interview by the authors, 21 March 2024.

32 Anonymous, interview by the authors, 25 March 2024.

| | Description |
|---|---|
| − | • Defence has much less significant buying power or ability to shape markets for dual-use AI technologies compared to traditional defence industry – leaving it a rule-taker, not rule-maker<br>• Skills shortages restrict Defence's ability to adopt and exploit AI at pace<br>• A rushed rollout of AI creates new dependencies (e.g. on foreign AI companies), vendor lock-in and other unintended consequences (e.g. impacts on staff morale or retention)<br>• Conversely, rollout of AI in Defence proves too slow, risking Defence being left further and further behind by more innovative competitors or by private sector organisations<br>• Some AI specialists in tech sector are reticent to work with Defence (e.g. as on Project Maven)<br>• Defence industry faces increasing competition for AI-related talent from other sectors |
| ? | • Unique challenges for Defence in navigating ethical and legal sensitivities of AI and autonomy<br>• Uncertain fiscal implications of economic impacts of AI for defence budgets |

*Source: RAND Europe analysis (2024).*

At the next level down, AI is expected to affect not only how military capability is delivered, but also what that capability looks like across all its constituent parts, known as the Defence Lines of Development.

**Table 3.4 National level: AI impacts on military capability**

| Defence Lines of Development | |
|---|---|
| **Training:**<br>• Combination of AI and synthetic environments drives advances in wargaming, training, education, exercises and mission rehearsal<br>• Use of AI and autonomous systems reduces need for live training and associated strain on platforms<br>• Use of AI drives requirement for new training and education packages, e.g. in AI bias<br>• Use of AI demands pipeline of new training data to maintain algorithms, not just training humans | **Equipment:**<br>• Shift from hardware-centric model to emphasis on software as key enabler of capability<br>• Shift from linear to spiral development models, open architectures, data, MLOps pipelines etc.<br>• AI enables increased mass through autonomy and automation across all domains<br>• AI tools (e.g. predictive analytics and equipment health monitoring) support efforts to bolster availability and readiness of military platforms |

| Defence Lines of Development | |
|---|---|
| **Personnel:**<br>• Adoption of AI enables exploitation of asymmetric strengths of human–machine teams<br>• AI supports new people/career management tools<br>• Automation of dull, repetitive tasks boosts morale<br>• AI drives new skills requirements from workforce, as well as competition from the private sector | **Infrastructure:**<br>• Rollout of AI demands new focus on accessing and securing compute, as well as datasets<br>• AI tools and autonomous systems help optimise infrastructure monitoring, prediction of faults, etc.<br>• Increased AI-enabled cyber threats to critical infrastructure, but also AI tools to aid the defence |
| **Concepts and doctrine:**<br>• AI informs better, faster concept and force development, experimentation and testing<br>• AI supports operational analysis and lessons processes by aiding data fusion, analysis and insights into causation<br>• Continuing need for legal reviews to assess new AI and autonomous systems against IHL/ LOAC | **Organisation:**<br>• Use of AI drives changes in process and accompanying organisational structures<br>• Rollout of AI requires overcoming bureaucratic and organisational cultural barriers to innovation<br>• AI places new demands on both leadership and followership at all levels |
| **Information:**<br>• Adoption of AI requires – and reinforces – the drive for transformation in Defence's data strategy<br>• Hostile actors seek to poison data and algorithms<br>• Increasing use of synthetic data and secure clouds supports sharing with allies, partners, industry | **Logistics:**<br>• AI tools help predict demand patterns, optimising supply chain and stockpile management<br>• AI and robotics boost warehouse efficiency<br>• Autonomous systems for 'last mile' logistics in contested environments cut force protection need |

**Defence Lines of Development**

**Interoperability:**

- Opportunities for countries to position themselves as an AI leader and framework nation for others
- Combination of AI and technologies such as cloud and edge computing, connectivity, etc., supports rollout of digital architecture for multi-domain operations and bolsters interoperability with allies and partners
- Conversely, divergent national approaches to AI, autonomy and cross-border data sharing (incl. standards, policy, ethical or legal differences) or varying speeds of AI adoption could undermine alliance cohesion

*Source: RAND Europe analysis (2024).*

The uses to which such capabilities might be put, and the implications of AI-enhanced operations across the continuum of competition and conflict, are examined in more detail in Chapter 5.

## 3.3. AI impact: propensity for advantage

This sub-category of the conceptual framework addresses AI-related impacts on:

- The full strategy cycle, from strategy making to orchestration and implementation;
- All levers of power: diplomatic, information, military, economic (DIME); and
- Signalling to, and perceptions of or by, other actors.

Collectively, these focus on the ways in which AI might help an actor (i.e. a government) to be more efficient and effective in mobilising their national resources (i.e. their potential for advantage) in pursuit of security goals in a competitive environment. This section concludes by exploring the temporal dimension and the debate over whether AI will provide enduring first-mover advantages to those who are fastest to adopt it – or least constrained ethically – or whether such advantages will be fleeting as AI technologies proliferate to other actors.

### 3.3.1. Decision support tools could first enhance, then reimagine, strategy making

Given its focus on machine intelligence, including via gamified learning, it is unsurprising that much of the literature around the military applications of AI focuses on potential uses in decision making, including at the strategic level. AI has long since surpassed human players at certain strategic games (e.g. chess, Go or real-time strategy videogames). In doing so, it has also demonstrated a capacity to develop winning strategies that no human had previously employed.[33] Equally, current limitations to technical capabilities (esp. Narrow AI) means that AI systems – especially those reliant on ML techniques – are not yet able to replicate humans' ability to generalise, adapt and make

33 Hunter et al. (2023).

strategic decisions in the face of unfamiliar, uncertain or complex circumstances.[34]

In the near to medium term, then, the focus is on use of AI alongside other technologies (e.g. data science, modelling, synthetic environments, etc.) as a decision support tool, harnessing the strengths of both human and machine intelligences to make more timely, better informed and higher quality decisions:

- At the beginning of the decision cycle, AI can assist with **intelligence gathering and analysis**.[35]
- AI tools can then assist with **cleaning, fusing, processing and analysing** vast amounts of data from diverse inputs (be they covert or open source), making sense of complexity, and then prioritising and visualising the most pertinent information to human decision makers to help them avoid cognitive overload.[36]
- AI tools then offer means of **wargaming or Red Teaming** potential strategies, helping depict how other actors may react, serving to improve the robustness of the chosen course of action (COA).[37]
- As they become more sophisticated, **AI decision support tools** will also be able to suggest alternative COAs of their own.[38] These can then be modelled, e.g. by running many different iterations of the same scenarios in faster-than-real-time, to build out a refined picture of causal relationships and what does and does not work stochastically, enabling selection of the most promising strategies. These models can then be improved over time with real-world data and ML.
- At the tactical level, there are many situations in which speed of decision making is the driving consideration in apportionment of roles to human vs machine intelligences (e.g. for an uncrewed combat aircraft in a dogfight, the speed of reaction and the fact that comms links back to a human overseer are likely to be jammed may suggest a need for higher levels of autonomy).[39] At the strategic level, though, speed is important, but it is also about **tempo**, i.e. **making high-quality decisions** at the right time. As such, the somewhat slower pace of decision making involved in many cases (with notable exceptions, such as urgently responding to a possible nuclear attack – see Chapter 5) presents greater scope for humans to be in, not merely on, the loop. Examples include strategic-level decisions about defence investment and capability development priorities, about deterrence posture or about military campaign planning.

The literature and interviews reveal mixed opinions as to the pace at which Defence will likely be able to field more sophisticated AI decision support tools, as well as the desired levels of autonomy versus human control. Certainly, machine intelligence brings certain advantages, e.g. the ability to quickly absorb vast amounts of data that would overwhelm human analysts or to make decisions without

34 Rob Solly, interview by the authors, 12 April 2024.

35 Futter (2022).

36 Anonymous, interview with the authors, 15 March 2024.

37 Geist et al. (2024).

38 Andrew Sharpe, interview with the authors, 19 March 2024.

39 Slapakova et al. (2022).

emotion.[40] Newer language models have demonstrated 'theory of mind' (i.e., the capacity to derive insights about the likely perspectives of other agents) and, relatedly, an ability to bluff or deceive – both important traits for a strategist.[41] Conversely, it is important to stress that AI systems remain subject to biases and brittleness; despite recent progress, models are all too susceptible to mistakes, hallucinations or adversarial attacks.[42]

There is currently significant interest in improving the flexibility of AI systems to deal with a wider range of unfamiliar contexts, tasks and decisions, as manifest in large private and governmental investments in large-scale frontier models. Given the importance of maintaining legitimacy and accountability in decision making, there has been considerable investment in recent years in so-called Explainable AI. These are systems that do away with old 'black box' techniques and present users with justifications behind the machine's decisions, articulated in terms that humans can understand.[43] There is a significant discussion surrounding the potential of various AI methods to enhance decision making processes and foster an appropriate level of human trust. This aims to avoid scenarios where humans rely on their own biases and heuristics due to insufficient trust in AI, or conversely place excessive confidence in algorithmic outputs without a full grasp of the underlying assumptions and constraints.[44] Equally, there is hope in the literature that the growing use of AI and related advances in decision science could force human strategists to also get better at articulating their own mental models and justifications for decisions. This would lead to human–machine teams that are more cognisant of biases and more focused on iterative learning.[45]

In the longer term, AGI could then far surpass the cognitive limitations of humans. This would open a wide range of unprecedented strategies for influencing, deterring or defeating adversaries. Equally, though, we return here to the macro risks around AI safety, alignment and meaningful human control.

### 3.3.2. Crucially, any use of AI in strategy making will be hotly contested

It is important to remember the dialectical nature of strategy making. It can be understood as a contest of opposing wills and intelligences in which 'the enemy gets a vote'. For those concerned with strategic theory, there is debate within the literature and interview data as to whether AI fundamentally changes the *nature*, rather than merely the *character*, of strategy – and each of the influential precepts handed down by Clausewitz, Sun Tzu and many others.[46] More practically, the adoption of AI into different stages of the strategy making and decision cycle also reflects an intensifying contest for decision advantage underway between the UK and its allies (esp. the US), and competitors such as Russia, China or Iran.[47]

---

40 Tate Nurkin, interview by the authors, 20 March 2024.

41 Payne (2024).

42 Liu & Maas (2021).

43 Reinhold & Reuter (2022).

44 Johnson (2022); Hughes et al. (2024).

45 Meerveld et al. (2023).

46 David Galbreath, interview by the authors, 19 March 2024.

47 Robles & Mallinson (2023).

Concepts of decision and information advantage have emerged as central to Western thinking about how to achieve advantage. They are prominent, for example, in the UK's Integrated Operating Concept (IOpC).[48] These ideas build on the manoeuvrist approach and the legacy of 'mission command' within NATO militaries; emphasising getting inside adversaries' observe-orient-decide-act (OODA) loops to out-think them, and then presenting that adversary with multiple dilemmas (i.e. decisions where all the options are bad) so as to shape their behaviours in directions that favour NATO's interests and preferred outcomes.[49]

Both Russian and Chinese doctrines emphasise similar concepts, if framed through the lens of their own cultures and historical experiences[50]:

- **Russia's armed forces** stress the uses of AI in operationalising concepts such as 'reflexive control' (influencing adversary's perceptions, access to information and thinking) or 'disorganisation' (seeking to disrupt and paralyse adversary's C2 structures, especially in the initial period of war).
- **China's People's Liberation Army** (PLA) is modernising fast in preparation for 'systems destruction warfare' (the idea that advantage comes not from destroying the enemy's forces in detail, but rather from targeting key nodes and linkages in their C2 systems to confuse, paralyse and ultimately out-think them) and for 'informatised' and 'intelligentised' future wars using AI.

What this translates into, then, is a competition for advantage between opposing AI-enabled systems for what the military calls Command, Control, Communications, Computers and Intelligence, Surveillance, Target Acquisition and Reconnaissance (C4ISTAR), as well as opposing counter-C4ISTAR capabilities:

- Here, AI becomes a **target** (e.g. with each side seeking to poison training data for opponents' AI algorithms, or to exploit the limitations of AI systems), a **defensive aid** (e.g. with AI supporting automated cyber defence) and a **tool supporting offensive action** through a mix of kinetic and non-kinetic effects (e.g. via use of AI in electronic warfare, or for detection and targeting of concealed adversary C2 nodes, such as mobile headquarters, for attack with long-range fires).[51]
- This entails a need for any integration of AI into strategic decision making processes to consider **possible threat vectors** through which adversaries might seek to poison, exploit or degrade that AI system. It also means having technical and procedural **redundancies and reversionary modes** so that humans can fall back on non-AI-dependent systems if their AI systems are disrupted.

It is ultimately unclear what the net results will be of this contest between measures and countermeasures, e.g. in terms of the overall offence–defence balance, the transparency or opacity of the battlespace, or the influence of fog and friction on the decision making of different actors. Competitors will be seeking to bolster the robustness of their own AI-enabled strategy making functions while undermining

48 UK Ministry of Defence (2020).

49 UK Ministry of Defence (2022b); NATO (2023).

50 Black, Lynch et al. (2022).

51 Black et al. (2024); Lucas et al. (2024).

or attacking their adversaries' AI-enabled systems at the same time.[52]

### 3.3.3. AI also affects orchestration and communication of a strategy once one has been formulated

Literature and interviews also suggest that AI tools could have implications for the next stage of the strategy cycle, namely orchestration and communication of that strategy. Possible impacts include:

- AI, along with rollout of other digital technologies, could support realisation of new ways of **collaborating** with PAGs, allies, partners, industry, academia, civil society, and NGOs.
- The nature of AI as a set of GPTs, and as a software-based capability that requires continuous updating, means that use of AI necessarily entails a closer partnership with commercial AI firms. They will become vital to **strategy implementation** and thus must be engaged in something deeper than a transactional customer–supplier relationship.[53]
- AI tools could also play a significant role in **strategic communications**, as well as related activities such as information and influence operations. This includes everything from content generation through to translation or understanding and automating engagement with different target audiences, through to measures of effect to refine comms strategies.[54]

Here, the emphasis is primarily on opportunities, with AI supporting better collaboration across institutional, cultural or language boundaries, and communication of strategies to different audiences. There are nonetheless several associated risks, for example arising from a potential backlash against the legitimacy of strategies developed using AI, or from divergent approaches by different partner organisations to the rollout and regulation of AI tools, including in a military or national security context. These are examined further in Chapter 5's discussion of implications for alliances.

### 3.3.4. AI introduces new risks and opportunities to strategy implementation, affecting all instruments of power (DIME)

AI is similarly expected to affect the last stage of the strategy cycle, namely implementation. Here, literature and interviews emphasise the potential opportunities arising from AI's role as an enabler and force multiplier. Proponents hope AI will generate new efficiencies that increase the reach and likelihood of success when states use instruments of power (i.e. DIME levers) in pursuit of strategic goals.[55]

Of note, research and historical case studies suggest a nation's propensity for advantage – i.e. their ability to realise a greater or lesser portion of their raw potential by mobilising and coordinating all parts of government, the private sector and wider society in pursuit of strategic goals – hinges on a range of factors. These include the underlying characteristics of the society in question (e.g. national identity and will, levels of openness to free thinking,

52 Maurice Chiodo, interview by the authors, 28 March 2024.

53 There are parallels here to the cyber/electromagnetic and space domains, where private tech firms have been integrated more directly into supporting headquarters (e.g. Commercial Integration Cell at the National Space Operations Centre).

54 Joe Wang, interview by the authors, 21 March 2024.

55 Anonymous, interview by the authors, 25 March 2024.

learning and innovation, or of competitive diversity and pluralism). They also, crucially, hinge on the effectiveness (both real and perceived) of governance institutions at responding to the dominant competitive paradigm of the age, including the strategic implications of disruptive technologies such as AI.[56] With this in mind, Table 3.5 outlines strategic risks and opportunities of AI across each of the DIME levers.

**Table 3.5 National level: AI impacts on strategy implementation using DIME levers**

| | Risks | Opportunities |
|---|---|---|
|  | • AI impacts undermine rules-based order, state sovereignty, and institutions (e.g. UN)<br>• Shift to machine intelligence undermines value of diplomacy as a human art<br>• Backlash against stance on military AI | • Magnification of soft power with AI<br>• Participatory approaches to tech governance that build ties with diverse new partners<br>• AI helps develop new mechanisms for building trust, ensuring compliance with treaties etc. |
|  | • Reduced ability to exert global influence in an AI-degraded infosphere<br>• Increased control of algorithms or private firms over infosphere at state's expense | • New AI-enabled means to understand and influence target audiences globally<br>• Improved measures of effect |
|  | • See Section 3.2 on development of the Defence enterprise and military capability<br>• See Chapter 5 on competition and conflict | • See Section 3.2 on development of the Defence enterprise and military capability<br>• See Chapter 5 on competition and conflict |
|  | • Diminishing human influence on economy<br>• AI-enabled economic warfare undermines stability of global markets<br>• AI frustrates enforcement of sanctions, or fight against terrorist financing and crime | • Improvements to government finances (and thus discretionary spending) from AI boom<br>• New AI-enabled economic statecraft tools<br>• New insights into economic data and capacity to better predict effects of policy interventions |

*Source: RAND Europe analysis (2024).*

56 Mazarr (2022); Mazarr et al. (2024).

### 3.3.5. Experts raise concerns over AI's potential unintended consequences for strategic signalling and perceptions

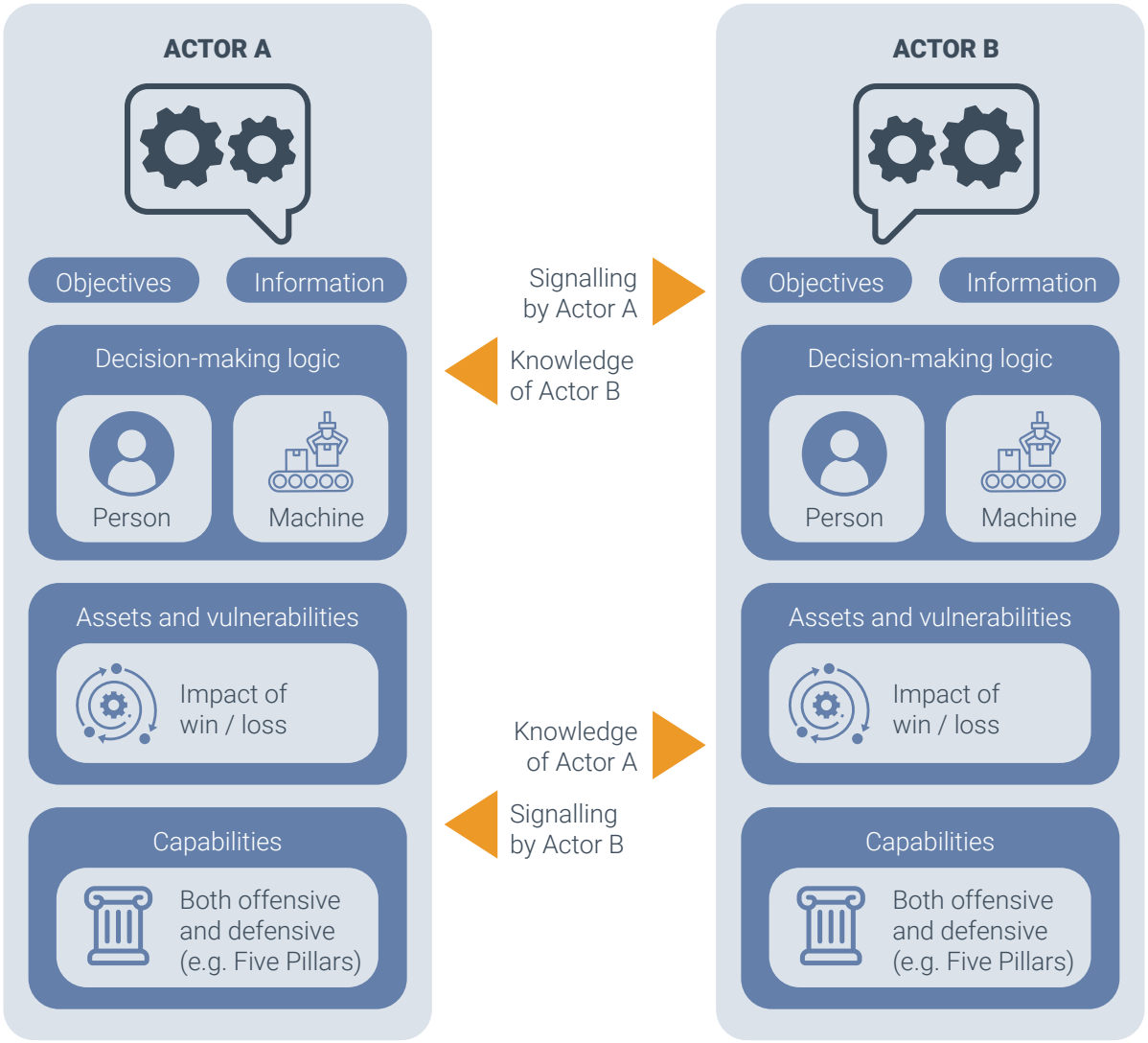
As stressed in preceding sections, strategy is a cycle, with the rollout of AI bringing positive and negative effects across it. As a cycle, it is recursive, rather than linear. Strategy implementation is accompanied by monitoring of changes in the strategic environment to determine if a change in strategy is required.

In part, this means adapting in response to contextual factors (e.g. changes in technology, the economy, climate, etc.). But it also means interpreting the signals that other strategic actors are sending, whether consciously or inadvertently, and making sense of what these might mean about their perceptions, plans or decision calculus. Game theory has long provided explanations for how two or more (human) actors interact, sending and receiving signals from each other, and making inferences and judgements that then inform adjustments to their respective strategies, postures and behaviours. As depicted in Figure 3.2, this is based on imperfect self-knowledge about one's own objectives, information, decision making logic, assets, vulnerabilities and capabilities – and even more imperfect knowledge about those of other actors. This fuels the risk of misperceptions and unintentional escalation, especially under the pressure of a crisis.

AI tools bring new opportunities for addressing such issues, e.g. by improving intelligence analysis and thus understanding of other actors' strategic culture and decision making. But the literature and interviews also show substantial concern among experts that AI's integration into strategic decision making (above all, nuclear command and control) could have unintended consequences.[57] Wargames by RAND and others have emphasised the potential for rapid escalation of crises in which actors fear, wrongly or rightly, that AI may give their adversaries a decisive advantage (e.g. the ability to launch a first strike), or where they believe hostile AI systems may target vital infrastructure (e.g. nuclear C2 systems).

57 Tate Nurkin, interview by the authors, 20 March 2024.

**Figure 3.2 Strategic signalling and (mis)perception with AI through a game-theoretic lens**




*Source: RAND Europe analysis (2024). Note: the Five Pillars refer to defence capabilities and activities in terms of non-proliferation, deterrence, counter force, active defence and passive defence.*

These risks are explored in more detail in specific relation to deterrence, crisis management and nuclear in Chapter 5. But it is important to flag the broader issue of misperception at this juncture, as it is an essential consideration when looking at the impact of AI on any given actor's propensity for advantage. That actor will not be able to achieve their desired strategic outcomes if they send the wrong signals to others and contribute wittingly or unwittingly to escalation dynamics that work against their own national interests, let alone wider peace and security.[58] Relatedly, literature and interviews raise concerns about

58 Scharre & Lamberth (2022).

the rhetoric of an 'arms race' in military AI. It can be argued that such terminology is reductive; more polemical than an accurate description of competition over AI, not least given the deep cross-border linkages between national AI sectors (even the US and China). Equally, some experts express concern that such combative or hyped language may fuel misperceptions that create a security dilemma.[59]

For example, it is noteworthy that the open-source literature in the US, China and Russia often presents divergent narratives on military AI: expressing concern that one's own nation is falling behind others (even as other countries say the same about themselves); employing rhetoric about goals for military AI that conveys strength domestically but might raise unwanted alarm abroad; and at times taking the more bellicose language or technology hype from other countries' own military AI programmes at face value.[60] Intelligence analyses within government should present a more nuanced view on other actors' objectives, capabilities, posture and policies (e.g. around levels of autonomy). Still, heightened geopolitical tensions and the immaturity of transparency and confidence-building mechanisms (TCBMs) around military AI increase the risk of misperceptions that undermine efforts to build a global governance architecture for AI or avoid unintended conflict more broadly.[61]

### 3.3.6. While actors pursue first-mover advantage in AI, some advantages are likely to prove more fleeting than others, given the diffusion of AI as a set of GPTs

This chapter has focused on the theme of advantage, and how AI might impact the potential and propensity of a given actor (typically a nation) to achieve advantage, prompting both strategic risks and opportunities. There is debate within the literature and interviews, however, over how decisive and lasting those advantages may be. Here, there are four main topics of contention:

- First, the importance of **technological breakthroughs** (as in AI) to long-term strategic outcomes, either in warfare or competition more broadly. Here, it can be observed that Western strategic theorists or practitioners have typically placed more stock in the decisive battlefield impact of technology and supposed 'revolutions in military affairs' (RMAs). This brings accusations of techno-determinism, underplaying non-material factors such as will-to-fight.[62] Conversely, Soviet/Russian, Maoist/Chinese and Iranian traditions have typically placed greater emphasis on factors such as political will, mobilisation of popular support and indirect approaches.
- Second, the extent to which **RMAs at the tactical level**, even if they exist, prove decisive at the strategic level. Here, it can be observed that the US and UK have had clear technological superiority in every war they have fought since 1945, winning most battles but not all their wars.

59 Cave & Ó hÉigeartaigh (2019); Roff (2019); Anonymous, interview by authors, 20 March 2024.

60 Hunter et al. (2023); Nabibaidze (2024).

61 Nadibaidze & Miotto (2023).

62 Rossiter (2021).

- Third, the degree to which AI offers **decisive first-mover advantage**. Historically, the first actor to field a disruptive new technology is often not the one to perfect and most benefit from it. Others quickly learn lessons from the first-mover's mistakes and/or develop better concepts of employment or countermeasures. Furthermore, AI is a set of GPTs, and thus inherently more dual-use than most previous disruptive military technologies, with innovation driven by the private sector and thus much more proliferated and democratised.[63] As such, there may be different dynamics for bespoke military AI (e.g. battle management systems). First-mover advantage may be more enduring here as the algorithms remain classified and hard to discern from external observation, as compared to dual-use AI systems, where technologies and concepts of operations can be quickly emulated by others (as, for example, with Russian and Ukrainian emulation of each other's tactics for commercial drones on the battlefield since February 2022).
- Fourth, the extent to which **differing ethical stances** will generate long-term advantage when it comes to AI and, relatedly, autonomous systems. While it is often presented as a truism that Western actors approach deployment of military AI 'with one hand held behind their back', due to greater ethical, policy and legal restrictions than found in authoritarian regimes, it is not clear a) that adversaries necessarily will deploy such systems in fundamentally different ways, b) that lower ethical standards necessarily bring a decisive advantage (especially in the long term, where demonstrating responsible use of military AI is important to maintain political legitimacy both at home and abroad) or c) that the adversary's less ethical uses of AI cannot be countered by other asymmetric means (e.g. counter-C4ISTAR capabilities).[64]

As such, while this chapter has demonstrated that military AI is likely to create a wide range of strategic risks and opportunities, it is unclear how decisive or long-lasting the resultant advantages will be, given the fluidity of strategic competition and the impact of factors besides AI. This merits more analysis.

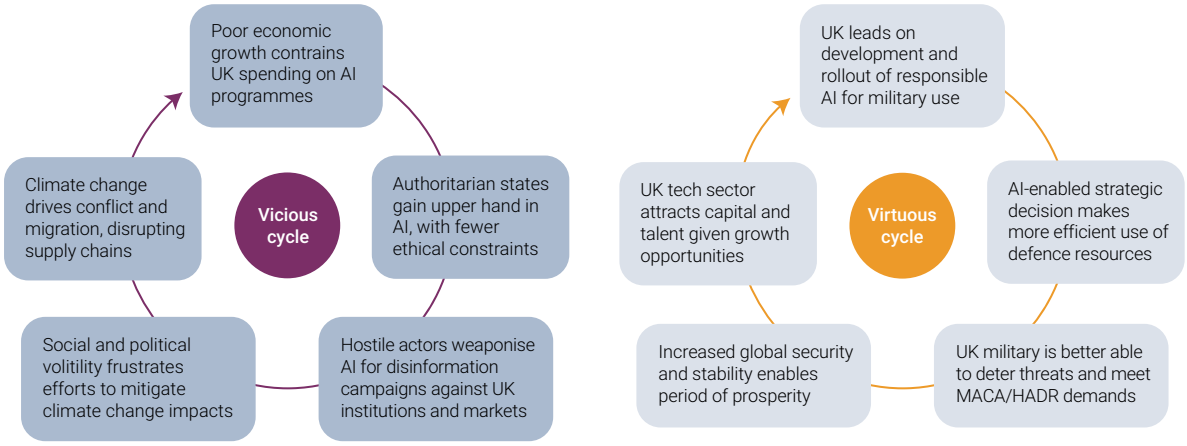
## 3.4. Summary

As will be explored further in Chapter 4, the interplay of these different AI-related impacts at the national level contributes to developments at the global level, and vice versa. This creates feedback loops that can either accentuate risks (i.e. vicious cycles) or further enhance strategic opportunities (i.e. virtuous cycles). Figure 3.3 below provides illustrative (and therefore simplified) examples of such potential feedback loops.

63 Anonymous, interview by authors, 22 March 2024.

64 Kenneth Payne, interview with the authors, 15 March 2024.

**Figure 3.3 Example of virtuous or vicious cycles emerging from strategic impacts of military AI**




*Source: RAND Europe analysis (2024).*

The evidence collected for this short exploratory study suggests that there are significant AI-related opportunities to be had from bolstering governance systems – to the benefit of wider society, the economy and public services, not only the military instrument – and from use of AI decision support and productivity tools to enhance the efficiency of Defence in delivering its tasks with the finite resources available.[65] Conversely, there are significant risks associated with Defence going too fast and making poor choices about the rollout of AI that then create unintended consequences and backlash. Or, alternatively, going too slow and being left behind by more agile competitors. Responsible development of military AI thus means balancing the need to go fast to secure an advantage over hostile actors who would threaten international peace and security, with the need to take time building governance and safeguards.

Crucially, literature and interviews emphasise that, while there are differences between military use of AI and applications in other sectors, it is hard to disentangle the question of AI's impacts on Defence from broader questions over its effects on innovation, skills, policy, ethics, law, regulation and the transformation of governance systems, economies and their underlying societies.[66] Defence will be shaped profoundly not only by military-specific AI systems – be they friendly or hostile – but also by the ways and success with which AI is applied to dealing with other pressing global challenges. These include (re) building social cohesion, economic growth, trust in institutions, public services and a collective response to climate change.[67] With this in mind, the next chapter considers how these impacts from AI at the national level might aggregate at the international level.

65 Representatives of Adarga, interview with the authors, 3 April 2024.

66 Johnson (2021b).

67 Anonymous, interview with the authors, 15 March 2024.

# Chapter 4. **Impact at the international level**

This chapter considers the second category of the framework, namely those risks and opportunities arising from the impacts of military AI at a global level. It focuses on ways in which advances in AI could affect the international system, as well as the intensity and dynamics of cooperation or competition within that system, thereby influencing strategic (in)stability and the escalation ladder. Given its global focus, this is inherently the most abstract of the levels of analysis and discussion in the framework, but it provides a foundation in relevant academic theory for the more granular and practical implications of military AI explored in Chapter 5 (By Competition Type) and Chapter 6 (By Actor Type).

**Box 4.1 Summary of findings: Chapter 4**

This chapter explores how the advent of military AI holds the potential to alter the sources and balance of power between actors, both as they compete for advantage *in* AI or compete for advantage *through* AI.

Similarly, it shows how AI could exacerbate existing pressures on the rules-based international order, including international institutions, law and norms. It also explores how military AI and other emerging disruptive technologies could thereby drive increased insecurity – and thus new demands for Defence – if not proactively addressed through novel governance arrangements.

Finally, this level of the conceptual framework considers the possible impacts of military AI on the intensity and dynamics of strategic competition within the international system; affecting the ways and means through which states seek to achieve strategic advantage for themselves, as well as creating complex feedback loops that could bring about unintended consequences for overall strategic (in)stability.

While the advent of AI in theory creates new reasons for competing powers (e.g. the US, Europe, China or Russia) to work together to reduce AI-related risks for mutual benefit, there are fears among AI experts that it will be difficult to separate the question of managing AI from the wider disfunctions of contemporary geopolitics.

*Source: RAND Europe analysis.*

## 4.1. Understanding the international system

Literature and interviews consulted for this study propose a wide range of lenses through which to understand the impacts of AI on the international system, drawing on international relations theory. For example, realist and neorealists might emphasise the potential influence of military AI on balances of power, affecting competing states' respective capabilities in an inherently anarchic international system.[68] Proponents of liberal theories might consider how AI instead affects state preferences, the functioning of domestic politics, and the scope for soft power or cooperation abroad, including through international institutions.[69] Constructivist readings, meanwhile, might focus on how machine intelligence might affect perceptions among actors and the spread of norms that shape behaviours at the strategic level.[70]

To enable a categorisation of these many different types of impact, it is possible to identify three main pathways through which military AI can affect change in the international system[71]:

- **Systems change**: a change in the nature of the actors within a system (e.g. states, firms, etc.).
- **Systemic change**: a change in the governance of the system.
- **Interaction change**: a change in the interactions between actors (e.g. competition dynamics) and other features of the strategic environment (e.g. climate, technology, etc.).

The following section summarises the potential impacts of military AI on each of these pathways to change in turn, along with associated strategic risks and opportunities. Outcomes at the global level are the net result (systemic change) of games (interactions change) between actors (systems change) competing within the international system, all of which are in turn affected by AI. Table 4.1 therefore makes cross-references to subsequent chapters of the report where there is more detailed discussion of military AI-related impacts, risks and opportunities at lower levels of the conceptual framework.

## 4.2. AI impact: actors, goals, and power

### 4.2.1. Competition over, and through, military AI will change the balance of power

At the level of system change, AI is expected to profoundly affect both the capabilities and preferences of strategic actors. Put differently, the advent of military AI influences not only the ways and means available to a given actor, but also the ends that they pursue. Literature and interviews also differentiate between two forms of strategic competition: that for advantage in AI, and that for advantage through AI, including its applications in a military context. As outlined in Table 4.1, intensifying competition for advantage in development of military AI (or AI more generally) makes resources such as compute, data and talent into critical determinants of a state or firm's

68 Horowitz et al. (2022).
69 Hunter Christie (2022).
70 Fournier-Tombs (2021).
71 Black et al. (2024)

capabilities and thus their influence at the global level.[72]

In turn, deployment of military AI could reshape the balance of power both between different states, and between states and non-state actors of various types. Tech firms responsible for developing AI systems will not only benefit financially from their lucrative products but also exert influence over how AI is governed, regulated and employed (including by the military). Violent extremist organisations, criminal groups, hacktavists or proxy actors will also seek to exploit AI to challenge the capabilities of state militaries. Non-governmental organisations (NGOs) and civil society are meanwhile likely to push for more participatory approaches to AI governance that challenge the state-centricity of the current international system.[73]

For further analysis, see Chapters 3 (Impact on national level) and 6 (Implications by actor type).

**Table 4.1 System change: Impacts, risks and opportunities from AI**

| | Impacts | Risks and opportunities |
|---|---|---|
|  | • AI affects both actor interests and values<br>• Military AI enables increase in power of states who best develop and exploit it<br>• Military AI makes compute, data or talent into critical resources (e.g. 'the new oil')<br>• Growing influence of global AI tech firms challenges sovereignty and the state-centricity of the international system<br>• Some non-state actors, e.g. terrorists, exploit AI to challenge state military power<br>• Other non-state actors, e.g. NGOs, call for more participation in AI governance<br>• In longer term, more fundamental questions about AGI/ASI as an actor in its own right | • Opportunity to take leading role on AI and on military AI specifically<br>• Opportunity for new partnerships on military AI that increase influence internationality<br>• Opportunity to exploit compute, data, offsetting lack of natural resources<br>• Risk of being slower than other, more agile governments to exploit benefits of military AI<br>• Risk that authoritarian states prove able to use AI to enhance repression at home, export it abroad and undermine open democracies<br>• Risk of non-state challenges to sovereignty at expense of national interests (e.g. regulatory capture of certain states by powerful AI firms) |

*Source: RAND Europe analysis (2024).*

72 Johnson (2021a); Waltzman et al. (2020).

73 Anonymous stakeholder, interview by the authors, 15 March 2024; Tallberg et al. (2023).

## 4.3. AI impact: global governance

### 4.3.1. AI exacerbates the pressure on the rules-based international system to adapt current institutions, law and norms to keep pace with new technologies

At the level of systemic change, the literature and interviews emphasise the impact of military AI on several features of governance of the international system[74]:

- Polarity and power transitions
- Institutionalisation
- Norms, including how new norms (e.g. behaviours, ethics) propagate globally and their more explicit codification into mechanisms such as international law, regulation or standards.

Here, the long-term strategic outcomes arising from the advent of military AI are uncertain, as it remains unclear which states will prove best able to adapt to and exploit the possibilities of AI, or how robust and adaptable institutions such as the United Nations will prove. The global governance architecture for AI is currently highly immature. This reflects the novelty of the issues faced, the lack of consensus on what to do about them, and the existence of multiple overlapping and in some cases competing initiatives and forums. Examples from the last year including the Summit on Responsible AI in the Military Domain (REAIM), the US-led Political Declaration, the UK's Bletchley Summit, and more.[75] In this context of fragmented and nascent governance arrangements, the balance of risk and opportunity remains unclear.

**Table 4.2 Systemic change: Impacts, risks and opportunities from AI**

| | Impacts | Risks and opportunities |
|---|---|---|
|  | • Polarity: Uncertain outcomes from impact of AI on multipolarity, either concentrating power for few actors or diffusing to many<br>• Power transitions: AI either accelerates or reverses transition from US dominance to rising powers e.g. China, India<br>• Institutionalisation: AI challenges existing international institutions (e.g. UN) (though these are playing an activist role in trying to shape AI); raises questions about both governance of AI and governance by AI | • Opportunity to benefit from shifts in the balance of power, if able to exploit military AI for strategic advantage<br>• Opportunity to revitalise international institutions to govern emerging technologies<br>• Opportunity to apply AI to international governance challenges e.g. to inform better peace treaties, arms control, negotiations, etc.<br>• Opportunity to be a norm entrepreneur, proactively shaping emergence of new norms |

74 Liu & Maas (2021); Radu (2021); Tarraf et al. (2019); Joe Wang, interview by the authors, 21 March 2024.

75 Bode et al. (2023); Schmitt (2022).

| | Impacts | Risks and opportunities |
|---|---|---|
| | • Law: Software-focused and dual-use nature of AI as GPT challenges model of treaty-based arms controls (but, conversely, AI could support verification for arms control); wider legal questions about military AI and autonomy through lens of IHL/LOAC<br>• Norms: Military AI could affect global norms of behaviour (e.g. levels of autonomy accepted for certain functions), and underlying cultural identities or ethics<br>• Norm propagation: AI provides new tools to shape norms e.g. information operations | • Risk that medium powers are squeezed out in intensifying superpower competition over AI, or forced to choose between incompatible systems from opposing competitors<br>• Risk that the legitimacy of the existing rules-based system is undermined by failure to adapt to an age of AI<br>• Risk that different actors adopt divergent legal views of military AI, undermining IHL/LOAC<br>• Risk that AI enables revisionist actors to promote norms contrary to national interests and values (e.g. undermining human rights) |

*Source: RAND Europe analysis (2024).*

For further analysis, see Chapter 7 (Toolkit of Mechanisms to Shape Risks and Opportunities).

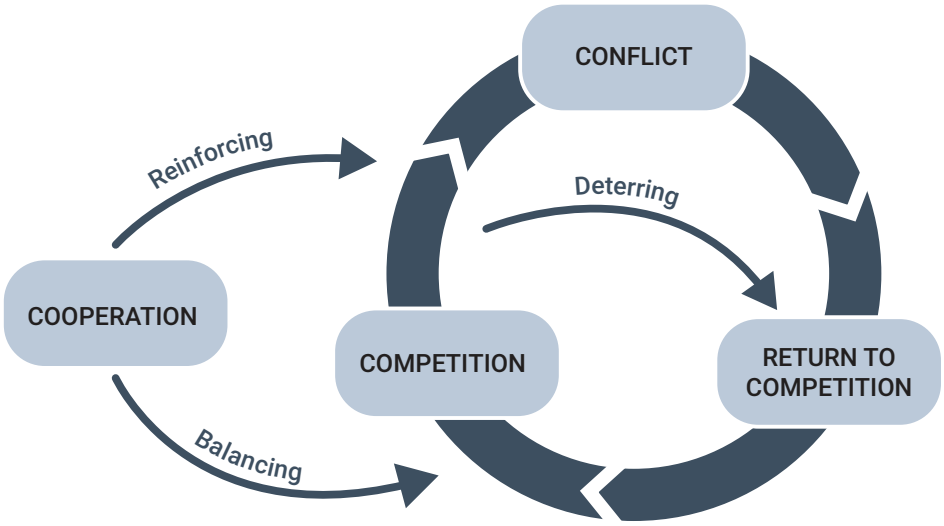
## 4.4. AI impact: strategic competition dynamics

### 4.4.1. Military AI, as a tool of deterrence or warfare, affects the dynamics of cooperation, competition and conflict – with uncertain results

In terms of changes to interactions between different global actors, AI is expected to affect not only military confrontations between states (or other types of actors), but also the full continuum from cooperation through to competition and outright war, including possibly even nuclear escalation.

At the strategic level, interactions between actors are marked by high levels of complexity. Two actors can be cooperating in one area of mutual interest (e.g. countering climate change, or a mutual threat from terrorists seeking to acquire weapons of mass destruction), even as they are competing in others (e.g. to achieve an upper hand in AI) or in conflict in others (whether directly or via proxies). To complicate things yet further, states might seek to compete with one adversary (e.g. the US with Russia) by cooperating with allies and partners (e.g. NATO), influencing non-aligned countries (e.g. India), and deterring other third parties (e.g. Iran) from exploiting their current preoccupation with that rivalry. Furthermore, there are complex feedback loops between these different elements, as shown in Figure 4.1.

**Figure 4.1 Feedback loops across the continuum of cooperation, competition and conflict**




*Source: RAND Europe, adapted from McCoy (2018).*

Literature and interviews suggest that AI, and military AI specifically, will have profound impacts on all elements of this continuum of possible interactions – affecting who cooperates, competes or fights over what, how, why, where, when and with what success.[76]

Here it is useful to distinguish between finite and infinite games as different forms of strategic competition. The former are time-, geography- or issue-bounded competitions between two or more actors, where there are clear rules of the game and ways of determining who has been the victor. Examples of finite games might include a negotiation, a trade dispute or a military confrontation. In these games, the goal is to win, and the advent of military AI may provide new ways and means of doing so.

Infinite games, by contrast, have no clear rules, no agreement between the actors involved over what constitutes success, and no endpoint at which to determine a victor; rather, they just go on, and the goal is simply to keep playing the game.[77] The most obvious example of an infinite game is the continuous strategic competition among states, where the fortunes of different powers ebb and flow, but it is highly uncertain how near-term choices will play out in the long term. Here, military AI is expected to influence the intensity and dynamics of competition, as well as overall strategic (in)stability within the international system as AI-enabled actors may develop more surprising strategies.

### 4.4.2. Experts fear that the advent of AI could intensify strategic competition, even as it provides new imperatives for nations to cooperate for mutual gain

Previous RAND research has identified factors that intensify or moderate levels of competition within the international system. The potential impacts of AI on each of these is summarised in Table 4.3.

76 Gill (2019).

77 Black et al (2023).

**Table 4.3 Impact of AI on intensity of strategic competition**

| Factor | Possible impacts of military AI |
|---|---|
| Intensifying factors | |
| Polarity | See Table 4.2 – Uncertain outcomes from impact of AI on multipolarity, either concentrating power for few actors or diffusing to many |
| Power transition | See Table 4.2 – AI either accelerates or reverses transition from US dominance to rising powers e.g. China, India |
| Uncertainty | Military AI provides new ways of understanding the strategic environment (e.g. intelligence analysis, decision support, predictive analytics) but also introduces new sources of uncertainty and unpredictability, e.g. around escalation dynamics |
| Moderating factors | |
| Interdependence | Currently, even competing national AI sectors are highly interdependent – with cross-border flows of capital, AI talent, algorithms, compute, data, etc. – but efforts are underway to de-couple (especially from China) |
| Institutions | Military AI exacerbates existing political tensions and challenges to international institutions (e.g. UN) and arms control mechanisms |
| Consultative mechanisms | Governments have expressed need for new consultative mechanisms to manage unintended escalation in crises involving military AI, but these remain nascent |
| Democracy | See Chapter 3 – Significant risks associated with use of AI to boost authoritarian regimes (e.g. via surveillance) or undermine democracies (e.g. via disinformation) |
| Nuclear deterrence | See Chapter 5 – Significant potential to affect nuclear deterrence and escalation |
| Multilateral problem solving | Military AI creates imperatives to cooperate (e.g. to avoid escalation, develop governance arrangements) but risks exacerbating the tensions blocking this |
| Context-dependent factors | |
| State identity | See Chapter 3 – AI has profound implications on cultural identities and norms |
| Domestic interests | See Chapter 3 – AI has profound implications on societies and economies |
| Leader goals and character | AI complicates thinking about decision makers, as it provides decision support and becomes a part of the strategic culture of different actors in different ways |
| Type of military forces | See Chapter 3 – AI transforms the Defence enterprise, the military capability development cycle and the types of capabilities delivered (quality, mass, etc.) |

*Source: Mazarr, Blake et al. (2018).*

As is clear from this table, the net result of AI's future impacts on intensifying and moderating factors is uncertain. While much of the literature and many of the experts interviewed for this study argue that military use of AI will lead to intensification of competition and conflict among states, others are more cautious. The latter group expresses hope that the advent of this disruptive technology may catalyse renewed cooperation among strategic rivals on issues of mutual interest (e.g. developing new governance arrangements for AI and other tech to manage the risk of accidental escalation).[78] Ultimately, however, much of the evidence base emphasises that it is difficult to detach the question of AI's impacts on geopolitics from broader, non-AI-related sources of tension (e.g. Russia's war in Ukraine, Taiwan, etc.).[79]

### 4.4.3. Military AI is also likely to affect the dynamics of strategic competition, including how states seek to gain and maintain advantage

As well as affecting the intensity of strategic competition, literature and interviews suggest that AI could have profound impacts on the dynamics of that competition. Previous RAND research has emphasised that the dominant competitive paradigm changes over time to reflect the conditions of the strategic environment in any given period. This reflects the influence of external factors (e.g. changes in climate, technology, etc.) and evolving views among actors as to the perceived rules of the game (i.e. what is permissible in competition or conflict) and the competing players' respective theories of success (i.e. what they each believe provides advantage and leads to beneficial strategic outcomes).[80]

AI is expected to affect this dominant competitive paradigm in several complex but impactful ways, as is the central underlying theme for the remainder of this report:

- Much of the literature and interviews argue that AI may prove one of the **most decisive influences** on strategic competition and its outcomes in the coming decades. This reflects AI's sweeping societal, economic and military applications as set out in Chapter 3, and the features of machine intelligence – competition ultimately being a contest between opposing wills or intelligences.[81]
- There is similarly significant debate and discussion over the possible consequences of military AI for overall **levels of strategic stability**. Military AI adds another layer of complexity and uncertainty to the escalation ladder, exacerbating issues such as multipolarity and cross-domain effects. Here, there is a growing body of research into the effects of AI on both crisis and arms race stability, as the major components of strategic stability.[82] To this end, Chapter 5 explores the theme of deterrence and escalation management in more detail.
- There is particular concern among experts about the possible **intersection of AI with nuclear** weapons, whether through direct integration into nuclear command and control, or through the indirect and likely

78 Johnson (2020c); Luo (2022); Sharma (2023); Ayse Ceyhan, interview by authors, 26 March 2024.

79 Guenduez & Mettler (2023).

80 Black et al. (2023); Mazarr et al. (2024); Mazarr, Frederick et al. (2022).

81 Johnson (2021b); David Galbreath, interview by the authors, 19 March 2024.

82 Anonymous, interview by the authors, 20 March 2024; Onderco & Zutt (2021).

unintended consequences of other military AI systems on escalation dynamics that might lead to nuclear weapons use.[83] Similar concerns are expressed about the potential for AI to enable state and non-state actors to more readily acquire and use **biological weapons** (or, of somewhat lesser concern, chemical ones).[84] To this end, Chapter 5 also examines the issue of weapons of mass destruction and their intersection with AI in more detail.

- There is related concern about AI undermining the **rules-based international system**, intensifying **superpower rivalries** and simultaneously empowering non-state actors, ranging from tech firms to terrorist groups, to challenge the nation state. To this end, Chapter 6 examines the impacts of military AI on different types of actors at the strategic level.

**Table 4.4 Interactions change: Impacts, risks and opportunities from AI**

| | Impacts | Risks and opportunities |
|---|---|---|
|  | Interactions change:<br>• Military AI affects full continuum of possible international interactions, i.e. from cooperation to competition to conflict (see Chapter 5)<br>• Complex and unpredictable effects of interaction between AI with other technological breakthroughs (e.g. biotech, quantum, robotics, etc.)<br>• Complex and unpredictable effects of interaction between AI with other trends and challenges (e.g. climate change)<br>• Enduring role of fog, friction and uncertainty on strategic outcomes, even with application of AI to decision making | • Opportunity to shift dominant competition paradigm towards areas of asymmetric strength (e.g. S&T, alliances and partnerships, soft power) if AI is harnessed and governed in ways benefitting the nation<br>• Conversely, risk that revisionist actors exploit military AI faster and/or shape the global governance (law, norms, etc.) for such technology in ways that restrict freedom of action while giving leeway to less ethical or rule-following actors<br>• Uncertain outcomes from interaction of AI with other major changes underway in the strategic environment – potential 'wicked problem' |

*Source: RAND Europe analysis (2024).*

83 Horowitz (2019); Johnson (2020b); Maurice Chiodo, interview by the authors, 28 March 2024.

84 Anonymous, interview by the authors, 22 March 2024.

## 4.5. Summary

This chapter has outlined the mechanisms by which the rollout of military AI may reshape the strategic environment, focusing on system change, systemic change and interactions change at the international level and the overall patterns of cooperation, competition or conflict. The following chapters delve into more detail and practical examples, unpacking how these impacts might manifest either in terms of affecting defence operations across the continuum of cooperation, competition and conflict (Chapter 5) or in terms of posing different dilemmas for different types of strategic actor (Chapter 6).

# Chapter 5. **Implications by competition type**

This chapter outlines and illustrates ways in which military use of AI may generate strategic risks and opportunities across the full continuum of cooperation, competition and conflict.

This builds on the analysis in Chapter 3 of ways in which AI might transform the Defence enterprise, including the industrial base, and consequently the types of military capabilities that are available to any given actor. Here, the discussion turns to the implications of military AI for the types of operations for which those capabilities might be employed, and the resulting outcomes at the strategic level.[85]

**Box 5.1 Summary of findings: Chapter 5**

The conceptual framework introduced in this report emphasises the need to move away from the narrow focus of much of the academic literature and contemporary policy debates on a few use cases or scenarios (e.g. AI in targeting of airstrikes or AI in nuclear warfighting). Instead, it is important to build a more comprehensive understanding of the myriad ways in which AI might affect the full continuum of competition and conflict.

Prominent practical examples in this regard include AI's impact on:

- Alliances, partnerships and capacity building, for example by creating new technical means of enabling interoperability but also introducing new risks of divergence on ethical and policy guidance on military AI.
- National and societal resilience, including early warning systems and response to various hazards or threats.
- Detection, attribution, countering or conduct of sub-threshold operations below open warfare.
- Escalation dynamics and the effectiveness of deterrence in an increasingly multipolar, AI-influenced world.
- Changes to the military offence–defence balance, increasing the lethality and tempo of military operations and offering new means of delivering multi-domain integration, C2, lethality, logistics and overall mass.
- Nuclear warfighting, as AI becomes integrated into nuclear command, control and communication systems – or threatens those of others – and/or drives new options in terms of arms control or counterforce strikes.

The sections below begin with the benign (i.e. more cooperative) end of that continuum and move up towards outright warfighting and eventual de-escalation back to post-war negotiations and reconstruction.

*Source: RAND Europe analysis.*

85 As a reminder, tactical and operational consequences of military AI fall outside the scope for this short exploratory study, and thus are not examined in this section except where they could, in aggregate, have strategic implications.

## 5.1. Implications for alliances and partnerships

### 5.1.1. AI and digital transformation could give blocs such as NATO an edge, but differing rates of adoption or ethical stances risk divergence among allies

AI has the potential to significantly impact alliances and partnerships in both positive and negative ways. On the positive side, AI holds the potential to enhance interoperability, communication and shared situational awareness among allied forces.[86] This is especially the case if AI is deployed alongside other technologies (e.g. cloud and edge computing, improved cybersecurity, next-generation connectivity, mesh networks, etc.) in support of rollout of a wider digital architecture and transformation of the alliance (e.g. realising the vision of NATO's Multi-Domain Operations backed by Federated Mission Networking).[87]

Additionally, AI-driven decision making tools can help allies and their commanders develop more coherent and coordinated strategies, policies and plans – e.g. through use of AI to enhance intelligence analysis, decision support, wargames, joint training and exercises, or other alliance variations on the use cases already discussed. Ultimately, this can lead to greater unity of effort.[88] LLMs similarly offer practical benefits, such as speeding up translation of written and oral communications between allied personnel speaking different languages or automating other tasks to enhance the productivity of multinational HQs, reducing the inefficiency inherent in any large alliance bureaucracy.[89] This could all lead to more efficient use of resources, more effective joint operations, and thus more credible collective deterrence and defence.

On the negative side, disparities in AI capabilities or policy, ethical and legal stances within an alliance may create tensions, increase friction, or exacerbate power imbalances and concerns over information sharing, security, sovereignty, trust and burden sharing. For example, alliances such as NATO may encounter interoperability challenges if members have varying standards for levels of human control in AI-enabled military systems.[90] Achieving a common approach, whether on ethical issues or technical standards, is necessarily more time-consuming and complex in a consensus-based organisation of 32 separate countries, such as NATO, than doing so unilaterally at a national level. Even if a common approach can be agreed, not all nations will be able to implement it at the same pace. This reflects the stark differences that exist between allies in terms of resources, military forces, industrial bases and technological expertise. As such, 'early adopters' may move ahead with AI-enabled capabilities and new concepts of operations while leaving behind their more change-averse or resource-constrained allies, exacerbating existing inequalities between nations' forces and undermining alliance interoperability.[91]

Technology sharing could help to address these differences, as could more collective development of AI models (e.g. establishing

86 NATO C2COE, interview by the authors, 3 April 2024.

87 Hunter Christie (2022).

88 Burton & Soare (2019).

89 Anonymous, interview by the authors, 3 April 2024.

90 Anonymous, interview by the authors, 25 March 2024.

91 Wilner & Babb (2021).

MLOps pipelines) and related digital technologies (e.g. via software-as-a-service or other procurement mechanisms) on behalf of all or some of an alliance's members – for example through the NATO Communications & Information Agency. Some states with the lead in specific military AI technologies may, however, be unwilling to share these, even with their allies. This caution is reflected, for example, in the US choosing to deepen ties on military AI with the UK and Australia through AUKUS Pillar Two, or with a larger subset of international partners through more informal forums such as the AI Partnership for Defence (AIPfD), while historically holding back its best technologies from many of its other NATO Allies, largely due to fears of leakage to adversaries.

Furthermore, not all allies will be willing to make themselves dependent on other nations' AI offerings and suppliers.[92] The technical challenges associated with AI integration and the risk of dependence on other states' AI systems may further encourage some nations to seek sovereign alternatives (even if less capable and/or more costly), exacerbating these divergences. This could lead to further shifts in the internal dynamics of alliances, give rise to new areas of disagreement over collective priorities relating to AI, and cause tensions amongst allies over industrial policy, procurement and export decisions.[93]

## 5.2. Implications for Defence engagement and capacity building

### 5.2.1. Military use of AI offers opportunities for more effective defence engagement and capacity building, projecting influence and boosting security

AI-driven training and simulation systems can help partner nations develop their military capabilities more rapidly and efficiently, while AI-enabled decision making tools can assist in the allocation of resources and the prioritisation of capacity-building efforts.[94] However, adopting AI in a defence context also raises several challenges. Ensuring that partner nations have the necessary infrastructure, expertise, data, and regulatory, legal and ethical frameworks in place is a major undertaking – and carries reputational risk to the partnership if not managed carefully. This may require significant investments in education, training and technology transfer (a challenge already discussed in relation to formal treaty-based allies but even harder with other partners), as well as the development of new cross-border partnerships between defence organisations, private firms and universities working on AI.

In the near-term, capacity-building efforts could focus on developing AI literacy among foreign military and civilian personnel and fostering responsible AI use.[95] This includes promoting an understanding of the potential benefits and risks associated with AI-driven military capabilities, as well as the development of ethical guidelines and legal frameworks for their use. Additionally, efforts should be made to ensure that AI systems are accessible and affordable for partner nations, to prevent the emergence of

92 Anonymous, interview by the authors, 25 March 2024.

93 Horowitz et al. (2022).

94 Meerveld et al (2023).

95 Kenneth Payne, interview by the authors, 15 March 2024.

new digital divides in the defence domain.[96] In such ways, there is potential for states to emerge as a thought leader in defence AI and a reference partner for other nations seeking to build up their own military AI capabilities and associated policies, structures, training and education courses, infrastructure and more. There are possible lessons here from previous efforts by governments to build partner nations' capacity in the cyber domain, or to educate new or non-space-faring nations on the need for responsible behaviours in outer space. In turn, leadership in this area could support wider security, influence and prosperity goals.

Conversely, of course, there is a risk that adversaries seek to fulfil this role, especially if they market an approach to AI and a toolkit of capabilities that are less ethically constrained. Experts warn, for example, of the risk of authoritarian regimes exporting AI-enabled tools of surveillance and repression, building on models such as China's Digital Silk Road and the Belt and Road Initiative of which it is a part.[97]

## 5.3. Implications for resilience and emergency preparedness

### 5.3.1. AI could improve early warning systems, planning, crisis response and resource allocation, but also create new dependencies and risks to resilience

Many governments, including that of the UK, have placed increasing emphasis on resilience in recent years. This reflects growing military and cyber threats to the homeland, including critical national infrastructure. There are also increasing demands on the armed forces to provide military aid to civilian authorities or to conduct humanitarian assistance and disaster relief operations, e.g. in response to climate change, natural disasters and pandemics.[98]

Here, AI offers potential benefits to Defence, both directly to the military and indirectly via bolstering the resilience of civilian agencies, infrastructure and populations. For example, AI-driven data analysis can help identify threats and vulnerabilities more rapidly and accurately, enabling decision makers to take preventive action before crises arise or escalate.[99] Similarly, AI can be used to optimise the allocation of resources during emergencies, ensuring that critical assets are deployed where they are needed most. This applies not only to military and state agencies but also NGOs, who are beginning to integrate AI and tools such as geospatial mapping using commercial satellite imagery into aid operations (see Section 5.9).

However, the growing reliance on AI also introduces new vulnerabilities in the context of resilience and emergency preparedness. Cyberattacks targeting AI systems could undermine their effectiveness, while algorithmic biases, brittleness, or errors may lead to suboptimal decision making during crises or exacerbate issues around a lack of popular trust in the government response.[100] Ensuring the robustness, security, and transparency of AI models and training data is therefore essential for maintaining resilience.

In turn, AI could itself be a direct threat to resilience. This includes via weaponisation of AI tools by hostile state or non-state

96 Giacomo Persi Paoli, interview by the authors, 2 April 2024.

97 Scharre (2023).

98 Caves et al. (2021).

99 Futter (2022).

100 Liu & Maas (2021); Nilza Amaral, interview by the authors, 18 March 2024.

actors to conduct attacks on critical national infrastructure, supply chains or civilian populations. Rapid progress in the capabilities of commercial LLMs in the past year has, for example, fuelled concerns that these models could be used to support terror attacks, including development of bioweapons to create new pandemics (though, in practice, initial evidence suggests that these fears are overstated with the current generation of LLMs).[101] There are of course also the wider risks around AI safety, including the possible advent of AGI.[102] As such, experts emphasise the need for strategies that boost resilience of critical societal functions to deal with a wide range of possible malicious or accidental negative impacts from AI.[103]

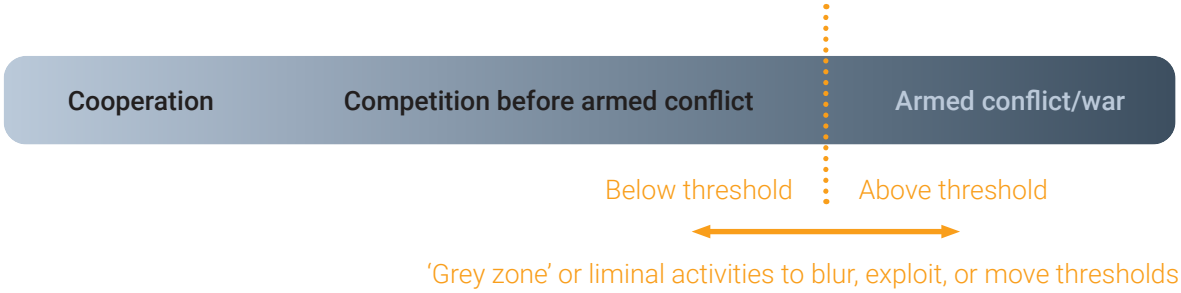
## 5.4. Implications for sub-threshold operations

### 5.4.1. AI may assist with detection and attribution of hostile sub-threshold operations, but also opens new ways of enhancing grey zone tactics

AI has the potential to significantly expand the range of sub-threshold operations, enabling more effective information operations, cyber warfare, and other forms of covert, ambiguous or deniable activity intended to blur thresholds between peaceful competition and overt conflict (see Figure 5.1).

For example, AI-driven disinformation campaigns can be used to manipulate public opinion and undermine the credibility of adversaries, while AI-enabled cyber tools can

**Figure 5.1 Sub-threshold operations in the grey zone of competition and conflict**




*Source: RAND Europe, adapted from US Joint Chiefs of Staff (2019).*

101 Mouton et al. (2023); Mouton et al. (2024); Nelson & Rose (2023).

102 Kenneth Payne, interview by the authors, 15 March 2024.

103 Janjeva et al. (2023).

be used to conduct intelligence gathering, espionage, sabotage and other forms of covert action.[104] Combined with other technologies (e.g. advances in quantum computing), this could for example encourage extraction of vast troves of encrypted data for future decryption once technology allows, potentially using that information to then coerce, embarrass or politically damage and isolate a target individual or government.

However, the growing use of AI in sub-threshold operations also raises the risk of escalation, as AI-driven operations may be misinterpreted or provoke unintended consequences.[105] For instance, the use of AI-driven disinformation campaigns may lead to retaliation and the escalation of conflicts into more overt forms of warfare. Equally, hostile actors could use AI as an excuse to gain plausible deniability for offensive actions against another state: blaming machine malfunction, algorithmic bias or cyber-sabotage by a third party for an attack (e.g. a drone attack on a critical pipeline) and complicating decisions about how to respond, especially in an infosphere polluted with propaganda narratives or a divided alliance.[106] AI deepfakes could similarly be used as part of lawfare,[107] creating fake evidence for spurious political and legal claims.[108]

To manage risks such as these, it is essential for states to develop tacit or explicit rules of engagement and communication channels for sub-threshold operations involving AI.[109] This could include establishing norms for the responsible use of AI in information operations and cyber warfare or targeting critical infrastructure. Also relevant, as explored in Chapter 7, would be AI-related and -enabled TCBMs to reduce the risk of misinterpretation and escalation.

Furthermore, states could invest in the development of AI tools for detection, identification, attribution and verification of hostile sub-threshold operations of various kinds, serving to reduce the effectiveness of such activities and thus deter them in the first place.[110] Here, the combination of AI with open-source intelligence may be especially powerful (e.g. using military AI tools to analyse commercial datasets, satellite imagery, and sensors embedded in infrastructure, drones, etc.). This would enable attribution without giving away covert means of intelligence collection, as well as third-party verification.[111]

## 5.5. Implications for deterrence

### 5.5.1. Both the theory and practice of deterrence will need to adapt to AI to avoid accidental escalation based on misperception or information manipulation

The potential impact of military AI on deterrence is one of the most extensively covered and hotly debated topics within the

104 Anonymous, interview by the authors, 19 March 2024.

105 Scharre & Lamberth (2022).

106 Reinhold & Reuter (2022).

107 In the context of this report, lawfare is characterised as the utilisation of legal systems and institutions to undermine or discredit an adversary, or to dissuade an individual from exercising their legal entitlements.

108 Wilner & Babb (2021).

109 Dortmans et al. (2021).

110 Reinhold & Reuter (2022).

111 For example, as when the UK MOD and U.S. DoD successfully used a mix of declassified and open-source intelligence to call out planned Russian 'false flag' attacks in Ukraine in 2022.

literature and interviews covered in this study. In terms of opportunities, the integration of AI into military systems has the potential to enhance deterrence by increasing the efficiency, accuracy, variety and speed of responses to potential threats.[112] For example, AI-driven early warning systems could help decision makers detect and assess emerging threats more rapidly and accurately, enabling them to take preventive action before conflicts escalate.[113] Similarly, they may assist with data gathering, modelling and thus understanding of adversaries' decision making (e.g. providing better insight into leader psychology or networks of power and influence that affect a given country's strategic culture). AI-enabled military capabilities, whether advanced decision support tools or autonomous systems, could in turn enhance the credibility of deterrence by bolstering a nation's ability to respond rapidly and effectively to aggression. Some advocates of AI argue it is not subject to emotion, fatigue or some of the other limits of human decision making, meaning that human–machine teams may be able to combine the strengths of both types of intelligence to make better decisions about when and how to move up or down the escalation ladder.

However, the growing use of AI in military contexts also creates new challenges for deterrence. Some experts fear that AI has the potential to fundamentally change the cost–benefit analysis of warfare by reducing the fog of war, imposing a superficial rationality on decision making processes, and lessening the perceived human cost of conflict. There are fears that this could lead to an increased willingness to employ force in the first place as a means of resolving disputes and to uncertain escalation dynamics thereafter.

In a crisis, for example, the ability of machines to make certain decisions much quicker than a human or group (e.g. a cabinet) may reduce the window for making important choices. Or, at the very least, one actor may feel compelled to make a hasty decision because they are uncertain to what extent AI and autonomy has been built into, and thus accelerated, their adversary's decision making processes, potentially giving that opponent a decisive advantage unless immediate action is taken.[114]

This temptation to 'shoot first' would only be exacerbated if the rollout of military AI (e.g. for targeting) proved a significant boon to the accuracy and lethality of offensive strikes – whether kinetic or non-kinetic, conventional, or nuclear. Crucially, the offence–defence balance need not necessarily have actually shifted decisively in favour of the offence. It could be enough for one actor to believe that it had, and thus to fear they had to 'use it or lose it' with their own offensive arsenal before a possible decapitating and disarming strike by an adversary.[115] Actors may also overestimate the effectiveness of undertested but overhyped AI or autonomous systems, causing them to make poor judgements about the probability of success or costs of different actions.

Indeed, this issue of potential misperception is a recurring theme: the black box, software-based nature of many AI models and of the control systems for autonomous systems mean that it is difficult for one country to look at another country's C2 or uncrewed assets and make accurate inferences and judgements

112 Onderco & Zutt (2021).

113 Anonymous, interview by the authors, 27 March 2024.

114 Johnson (2019b).

115 Mazarr, Rhoades et al. (2022).

about the extent to which a human remains in, on or out of the loop. Even if they could assess an AI's capabilities or a C2 system's levels of autonomy from afar, these could be changed with a simple software update or change of parameters feeding the algorithms.[116] Furthermore, AI-generated deepfakes could make it harder to discern truth from falsehood, sowing added confusion and uncertainty. Such fakes could therefore manufacture confrontations based on disinformation, or sow distrust among decision makers about the information feeds they are receiving in a crisis, making them fall back on gut instinct.

Together, these sorts of alterations in the decision calculus might be seen to favour pre-emptive actions and weaken the stability of both conventional and nuclear deterrence.[117] The risk of accidental escalation, due to misinterpretation or miscalculation, may increase because of the growing reliance on AI-driven systems and uncertainty about how they have been integrated into the adversary's own command and control.[118] Moreover, the perceived value of traditional deterrence mechanisms, such as the balance of power and mutual assured destruction in a nuclear context (see Section 5.8 below), may be eroded as AI-driven capabilities alter the strategic landscape and introduce new forms of competition and conflict.

Ensuring strategic stability will therefore require new tools and untested approaches.[119] Strategic stability can be broken into two key elements: arms race stability and crisis stability. Addressing the former entails the development of updated doctrines, confidence-building measures and arms control agreements to reflect the impact of military AI. Addressing the latter implies new communication channels and transparency mechanisms to reduce the risk of misinterpretation. AI might also tilt the offence–defence balance towards deterrence by punishment over deterrence by denial by enhancing the certainty and severity of retaliation (e.g. the idea of the 'dead hand' that can launch a second strike), while simultaneously exposing new vulnerabilities.[120] States will also face dilemmas over whether to reveal that they have new military AI capabilities – and how to demonstrate their credibility – since disclosing them is essential if they are to have a deterrent effect, but also lets other actors try to develop countermeasures.

Crucially, game theory and classical deterrence theory – much of it developed at RAND in the 1950s and 1960s – will need updating to reflect new assumptions.[121] Examples are provided in Figure 5.2.

116 Scharre (2018); Scharre & Lamberth (2022).

117 Taeihagh (2021).

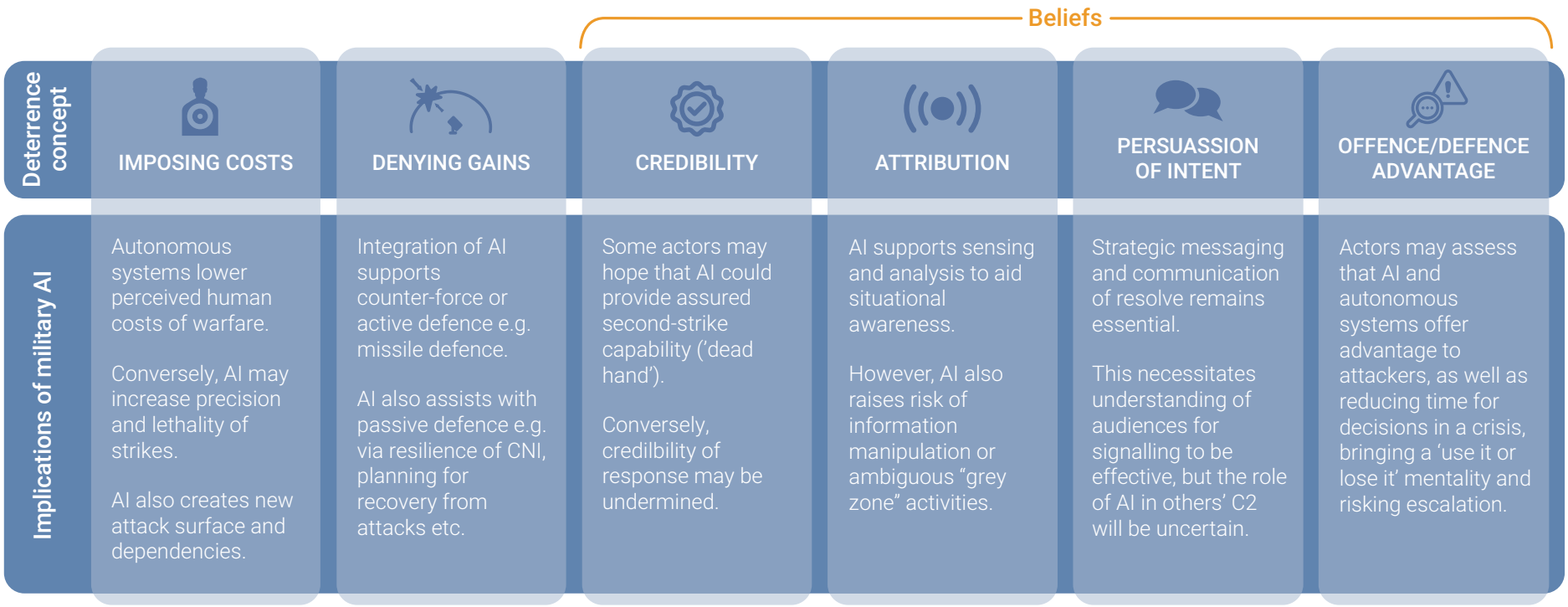
118 Wright (2019).

119 Johnson (2019b).

120 Wilner & Babb (2021); Yu (2023).

121 Johnson (2021a); Payne (2021).

**Figure 5.2 Examples of impact from AI on classical deterrence theory**




*Source: RAND Europe analysis (2024).*

This requires an advancement of deterrence theory that incorporates both views *of* AI, as well as analysis *by* AI, and which remains rooted in modelling the psychology of adversarial behaviour but through the new lens of human–machine teams rather than current human-centric approaches. In this context, 'AI could become an actor, not just a factor', and the second- and third-order effects of this development will need to be understood in detail – potentially especially tricky when facing non-deterministic AI.[122] For example, in past wargames, AI agents have shown levels of risk tolerance and aggressiveness beyond human capacity, but this may not be the case with future systems. Furthermore, new game-theoretic approaches to deterrence will need to also address the complexities of a more multipolar world, increased entanglement between military and civilian systems (including from multiple countries), and an increased attack surface and variety of threat vectors to worry about (e.g. incorporating cyber, electromagnetic, space and cross-domain effects). All these complexities are exacerbated by the added layer of uncertainty presented by AI, though AI may itself provide new tools for studying, modelling and better understanding the dynamics of this changing escalation ladder.

## 5.6. Implications for crisis management

### 5.6.1. AI can support more effective crisis management by facilitating rapid information gathering, decision making and resource allocation

If deterrence has failed, then AI-driven data analysis can help decision makers assess the causes, consequences and potential responses to an unfolding crisis.[123] This may enable them to make more informed choices about how to respond. Similarly, AI could also be used to optimise the allocation of resources during multiple concurrent crises, ensuring that critical assets and capabilities are deployed where they are needed most, reducing the risk that an actor's strategic bandwidth is overwhelmed.

However, the growing use of AI in crisis management also introduces new risks such as AI-induced escalations or accidents caused by malfunctions or the fragility, brittleness, immaturity or insecurity of current AI systems. Such developments require ongoing investments in R&D, with a particular focus on developing secure, robust and transparent AI systems that can withstand manipulation and other forms of disruption. In addition, AI-driven military operations could potentially conflict with common practices of crisis management, such as decelerating military movements to allow time for situation evaluation, cooler decision making and diplomacy towards a political resolution.[124]

## 5.7. Implications for conventional warfighting

### 5.7.1. With other technologies, AI has the potential to transform warfighting by enhancing multi-domain integration, C2, lethality, logistics and overall mass

AI-enabled military capabilities, such as autonomous weapons systems and decision support tools, could enhance the speed, precision and lethality of military operations, leading to more effective and efficient use of force, as well as greater combat mass.[125]

122 Pavel et al. (2023).

123 Onderco & Zutt (2021).

124 Luo (2022).

125 Hou et al. (2023).

For example, AI-driven data analysis could help commanders develop a more accurate and comprehensive understanding of the battlespace, enabling them to make decisions about how to deploy their forces. AI could then enhance targeting of long-range precision fires, supported by use of autonomous systems to help suppress enemy air and missile defences and then conduct subsequent battle damage assessment, all aided by AI-enhanced electronic, cyber and information warfare. Such ideas are the basis of heavy investments by NATO, the US and the UK in new battlespace management systems, resilient networks, secure clouds, edge computing, and sensing grids. The ultimate (and potentially overly ambitious) goal of concepts such as the US's Joint All Domain Command and Control is to create faster 'kill chains' based on an 'any sensor, any shooter' model, whereby friendly assets are networked across all domains and make use of AI to help pass data, targets or taskings from one node to another, contributing to a faster OODA loop than adversaries.[126] Some of this may be hype, doomed to fail for technical, cost or bureaucratic reasons. But some of it is already reality. Uncrewed systems have proven their worth in the battlefields of Ukraine and in the Black Sea, and have been deployed in vast numbers in both the close and the deep battle, initially with a focus on remotely piloted commercial systems but with growing levels of AI and autonomy now being built in.[127]

It is unclear how these trends will affect the offence–defence balance long term, not least given a continuous race to update tactics or countermeasures e.g. counter-unmanned aerial systems (UAS) capabilities to seek advantage. Similarly, it is unclear whether the net result will be to support manoeuvre (the approach that has underpinned UK and NATO doctrine for decades) or make it ever harder to conceal, concentrate and move forces, thus reducing warfare into a positional and attritional grind antithetical to the Western way of warfare.[128] Besides influencing the future of combat, though, AI and autonomous systems will also be vital to transforming logistics and other support to dispersed forces operating in this increasingly contested and transparent battlespace, reducing the force protection burden of using crewed assets for 'last mile' resupply of troops in combat or for casualty evacuation missions.[129]

Predictably, the prospect of growing use of AI in conventional warfighting raises several ethical and operational concerns. Some fear that the increased reliance on autonomous systems, for instance, may lead to an erosion of human control over the use of force. This raises questions about moral and legal responsibility for AI-driven actions, as well as the liability of private firms or the legitimacy of targeting civilian contractors supporting military AI systems.[130] In addition, as discussed in Section 5.5, AI could exert new pressures on strategic stability by accelerating the tempo and lethality of warfare, leading to unforeseen consequences that tip over into use of weapons of mass destruction against either tactical or strategic targets as one side seeks to regain the advantage.[131]

Balancing the benefits and risks of AI in conventional warfighting will require careful

126 Brose (2020).

127 Joe Wang, interview by the authors, 21 March 2024.

128 Rossiter (2021).

129 Black et al. (2024).

130 Futter (2022).

131 Horowitz et al. (2020).

consideration of these different factors. This includes the development of clear guidelines and frameworks for the responsible use of AI in operational contexts, continuing testing and assessments of new AI-enhanced systems that ensure compliance with IHL and LOAC, as well as the establishment of mechanisms for ensuring appropriate levels of human control and accountability over AI-driven military actions.[132] Crucially, too, actors with more restrictive policies and ethical stances of the battlefield use of AI and autonomous systems will require asymmetric means of countering the AI-enhanced capabilities of hostile actors operating with fewer qualms. This may entail a need for increased electronic warfare, counter-UAS and, eventually, counter-swarm capabilities, for example using directed energy or area-effect weapons to be most cost efficient.

In practice, it is likely that levels of autonomy will vary from system to system or even mission to mission depending on a range of risk factors. Examples include: the criticality and complexity of the tasks to be performed; how cluttered the physical environment will be (e.g. it is simpler to train an AI to operate in the relatively uncluttered maritime environment than in a city); how contested access to the electromagnetic spectrum – and thus reach back to a human – will be; levels of confidence in the AI and its technical performance; and the military, ethical, legal and reputational consequences of having a human in, on or out of the loop.[133]

## 5.8. Implications for nuclear warfighting

### 5.8.1. AI could affect nuclear warfighting through improved counterforce capabilities or new threats to command, control and communication systems

While Section 5.5 covered the potential impact of AI on nuclear deterrence, there is also a growing body of literature on what AI-enabled nuclear warfighting might look like should deterrence fail. Some experts argue that the integration of AI into nuclear command, control and communications (NC3) systems has the potential to improve aspects of decision making and reduce the risk of human error, providing decision makers with more accurate and timely information.[134] AI-augmented early warning systems can help decision makers detect and assess emerging nuclear threats more rapidly and accurately, enabling preventive action before conflicts escalate. AI tools also have the potential to augment the verification of arms control agreements, bolster the identification of nuclear testing and aid in the disassembly of nuclear weapons.[135] These potential advantages, however, depend on the readiness of nuclear powers to collaborate and exchange sensitive information, a difficult task in a context of mutual suspicion or rivalry.

The use of AI in nuclear warfighting also creates new risks. AI-enabled intelligence and targeting systems (e.g. to detect nuclear submarines) might heighten the risk of nuclear escalation by jeopardising second-strike capabilities and compromising mutual vulnerability.[136] Any AI models integrated into NC3 could be brittle or

132 FCDO official, interview by the authors, 25 March 2024.

133 Scharre (2018).

134 Meerveld et al. (2023).

135 Onderco and Zutt (2021).

136 Johnson (2021a).

subject to errors or adversarial attacks. Even without direct integration of AI systems into NC3, the utilisation of AI in peripheral systems such as early warning, target identification and battle damage assessment, or conventional counterforce capabilities, could diminish leaders' overall confidence in the survivability of their nuclear forces. This could all encourage pre-emption during a crisis.[137] In addition, countries perceiving themselves to have less secure second-strike capabilities may be more willing to adopt risky forms of autonomy in their nuclear forces, potentially leading to increased crisis instability and escalation risks.[138]

Actions to reduce unintentional escalation risks linked to AI and nuclear systems might include: use of AI for improving arms control verification; setting norms for transparency and accountability; demonstrating unilateral restraint by not deploying destabilising AI capabilities; and bilateral and multilateral stability discussions.[139] Risk mitigations might also involve forbidding the creation of AI systems intended to target adversaries' NC3 and prohibiting AI from having exclusive authority to initiate nuclear weapon launches.

## 5.9. Implications for de-escalation, peacebuilding and reconstruction

### 5.9.1. AI tools could help better design and enforcement of peace settlements, and assist with reconstruction efforts and provision of support to civilian populations

Any war eventually comes to an end, whether through a formal peace treaty or more implicit de-escalation. AI tools hold the potential to assist with all stages of this process, as well as the rebuilding efforts that follow. Already, the United Nations is deploying AI tools in conflict zones such as Libya to engage populations in large-scale digital dialogues, to help identify accommodations through which opposing groups might find some common ground.[140] Similarly, NGOs and researchers are making use of AI tools to analyse what has worked – or failed – when it comes to previous peace initiatives, as well as to analyse the rhetoric, arguments and sentiments presented by different warring sides, to support conflict mediators. AI tools could help to monitor and identify online hate speech, propaganda or changes in public sentiment in real time that might undermine any peace talks or ceasefire. AI systems could similarly be used to support peacekeeping and peace enforcement operations. For example, AI could enhance situational awareness for UN or other (e.g. NATO, African Union) forces, improve their understanding of the local context, or help to detect illicit flows of weapons that could spark violence.

Peace organisations are also looking to how AI might help create the conditions for a more robust and sustainable peace over the long term. This includes the use of AI tools in support of aid distribution, mine detection and clearing, reconstruction of destroyed infrastructure, rebuilding of local economies, provision of healthcare to deal with the mental and physical after-effects of a conflict within affected communities, and investigations to bring war criminals to justice. In Ukraine, for example, NGOs have been using a mix of AI

137 Scharre & Lamberth (2022).

138 Horowitz (2019).

139 Johnson (2022).

140 Fournier-Tombs (2021).

tools, commercial satellite imagery, social media feeds and other open-source data to build up geospatial intelligence on damage to infrastructure, as well as evidence on alleged Russian war crimes.

Conversely, as discussed in preceding chapters, AI also holds significant potential for deepfakes, bots and information-manipulation campaigns that could undermine peace or reconciliation efforts.[141] It is also unclear how well received AI and autonomous systems would be in any peacekeeping setting, or whether their deployment would resolve any of the wider limitations on such initiatives (e.g. restrictions on rules of engagement for peacekeepers, a lack of political will to resolve conflicts more generally, etc.). As such, AI may provide useful new tools, but these may be insufficient if the broader politics of a war are intractable.

## 5.10 Summary

Military use of AI presents a complex landscape of both risks and opportunities across the continuum of competition. On the one hand, AI has the potential to enhance situational awareness, act as a force multiplier, support predictive analysis, improve cyber capabilities, increase precision and lethality, and optimise logistics and supply chain management. AI could lead to better decision making, more effective operations, and increased efficiency and resilience. On the other hand, integration of AI into military systems raises concerns about escalation and instability, human control, and unintended consequences.

There is also concern in the literature and interviews that development of AI technologies may also lead to an 'AI arms race' (though other experts reject this analogy), increasing the tensions between nations and potentially destabilising the geopolitical balance. Technology proliferation is another concern, as the diffusion of AI technologies may enable non-state actors or rogue states to acquire advanced capabilities, leading to increased threats. With such issues in mind, the next chapter turns to the ways in which military AI could impact different types of actor, both state and non-state.

141 Joe Wang, interview by the authors, 21 March 2024.

# Chapter 6. **Implications by actor type**

Having covered the implications of military AI across the continuum of competition in Chapter 5, this chapter examines how AI may affect some kinds of actor differently from others. As discussed before, AI is best understood as a complex socio-technical system, with a vital human element. Risks and opportunities therefore differ depending on the organisational and cultural context in which military AI is adopted – just as the level of ambition, resources or geography of a given nation affects the role it plays within the international system and how it seeks to exploit military AI.

To address these heterogeneous impacts, this final level of the conceptual framework focuses on the differing risks, opportunities and dilemmas that AI may present to different categories of actor, based on:

- **Size or relative power:** superpowers, middle powers, small states
- **Governance type:** democracies, authoritarian regimes
- **Non-state actors:** private firms, violent extremist organisations, organised crime, NGOs, others.

Crucially, the direct impacts of military AI on these different types of actor then have second- and third-order consequences, for example affecting the relative standing of different powers or their relationships.

**Box 6.1 Summary of findings: Chapter 6**

AI risks and opportunities differ by actor. The US and China are the top global players in terms of investment, data, compute, and access to talent, but others are catching up fast and any given country's current position cannot be taken for granted. Crucially, too, other states (including the UK) have an ambition to influence global defence AI developments, but only limited leverage over key actors (whether AI firms or hostile states).

Against this backdrop, key issues that emerge include:

- AI is likely to intensify the strategic, technological and military competition among superpowers (the US and China), to the detriment of international stability. Multilateral initiatives to de-risk military AI are unlikely to work or last without buy-in from both superpowers, meaning other states will want to influence them both.
- AI similarly holds the potential to further erode the credibility of the governance and associated institutions (e.g. the United Nations, Bretton Woods) that have long underpinned global peace, prosperity and stability.
- The rollout of AI also threatens to shift the balance of power between nations, undermine the democratic system of government around the world by promoting AI-enabled systems of surveillance and repression, and empower non-state actors in ways that challenge state control and sovereignty.

Equally, open democracies such as the UK hold potential advantages in terms of attracting AI talent and nurturing private sector innovation. But they face growing pressure to work together through technology partnerships and alliances (e.g. AIPfD, AUKUS) to offset the differing strengths of authoritarian regimes.

*Source: RAND Europe analysis.*

## 6.1. Implications for different types of state

### 6.1.1. It is hard to compare military AI programmes using open-source information, but various indices seek to compare wider national performance on AI

Given the potential benefits discussed in previous chapters, countries around the world are competing to develop and deploy military AI systems – unveiling strategies, creating new or reformed organisational structures, and investing in R&D, experimentation or procurement programmes. Given the classified nature of military AI, as well as the hype found in many governments or firms' public statements on their AI-related ambitions and capabilities, it is difficult to compare countries' progress from open-source information. There is, however, a growing number of indices that seek to assess different nations' capacity, readiness and performance on AI more generally, outside of the military setting. These draw on a mix of qualitative assessments (e.g. of how conducive the regulatory environment is to uptake of AI) and quantitative data (e.g. indicators such as capital investments in AI firms, or size of AI-related workforces). While these comparisons are necessarily reductive, partial and contested, they do illustrate perceptions of the types of nations best placed to take the upper hand in the intensifying competition over AI.

The United States and China routinely top most global indices, with the UK often occupying a spot in the top three or five depending on the indicators considered. Crucially, different nations approaches to AI can look highly asymmetric, with strengths in one area offsetting weaknesses elsewhere. In the case of the UK, for example, it has relative strengths in terms of talent, basic research, certain AI applications, and fields such as AI safety, law or ethics, and the largest number of AI companies of any country in Europe.[142] By contrast, it is seen as less competitive in terms of its regulatory environment, infrastructure (e.g. compute), energy costs or access to finance for companies hoping to scale up without being bought out by foreign investment funds or tech giants (e.g. as in the case of Google acquiring DeepMind).[143]

Similar logic applies to the readiness of different countries to develop, acquire and field military AI, with some national defence establishments being more conducive than others to AI-related innovation, and with differing levels of ambition, urgency or resource. Here, substantial differences exist between the approaches of most states, who remain in a state of relative peace, and the stances taken by countries currently fighting active wars, especially where these are for national survival. Ukraine, for example, has since February 2022 accelerated the integration and deployment of military AI

142 UK Government (2021).

143 Jouan et al. (2024); Keith Dear, interview by the authors, 18 March 2024.

**Figure 6.1 Global AI Index 2024**

| Country | Overall | Talent | Infrastructure | Operating environment | Research | Development | Government strategy | Commercial | Intensity |
|---|---|---|---|---|---|---|---|---|---|
| United States | 1 | 1 | 1 | 28 | 1 | 1 | 8 | 1 | 5 |
| China | 2 | 20 | 2 | 3 | 2 | 2 | 3 | 2 | 21 |
| Singapore | 3 | 4 | 3 | 22 | 3 | 5 | 16 | 4 | 1 |
| United Kingdom | 4 | 5 | 24 | 40 | 5 | 8 | 10 | 5 | 10 |
| Canada | 5 | 6 | 23 | 8 | 7 | 11 | 5 | 7 | 7 |
| South Korea | 6 | 12 | 7 | 11 | 12 | 3 | 6 | 18 | 6 |
| Israel | 7 | 7 | 28 | 23 | 11 | 7 | 47 | 3 | 2 |
| Germany | 8 | 3 | 12 | 13 | 8 | 9 | 2 | 11 | 15 |
| Switzerland | 9 | 9 | 13 | 30 | 4 | 4 | 56 | 9 | 3 |
| Finland | 10 | 13 | 8 | 4 | 9 | 14 | 15 | 12 | 4 |
| Netherlands | 11 | 8 | 16 | 15 | 10 | 13 | 28 | 20 | 8 |
| Japan | 12 | 11 | 5 | 10 | 20 | 6 | 18 | 23 | 25 |
| France | 13 | 10 | 11 | 25 | 15 | 18 | 13 | 10 | 20 |
| India | 14 | 2 | 59 | 12 | 30 | 21 | 38 | 13 | 51 |
| Australia | 15 | 14 | 44 | 62 | 6 | 16 | 14 | 22 | 14 |

*Source: Global AI Index (2024). Note: 'Intensity' refers to a given country's AI capacity relative to its size in terms of population or GDP.*

and autonomous systems at a pace and scale unmatched by almost any other nation, starting from a low base but innovating rapidly due to the existential threat it faces from Russia – even going so far as to establish an entirely new branch of the military focused on uncrewed systems.[144] Israel, similarly, has incorporated battlefield lessons from past rounds of fighting with Hezbollah and the current conflict with Hamas into its accelerated development of military AI and robotic systems.[145] Whether other nations can replicate similar speeds of innovation in a 'pre-war' setting remains to be seen, given their more cautious approach to risk, the pressures

144 Bendett (2023); Tokariuk (2023).

145 Mhajne (2023); Sylvia (2024).

on budgets and their less expedited processes (e.g. for capability development).[146]

The following sections do not seek to comprehensively assess which nations are best placed to overcome such barriers and achieve strategic advantage in and through military AI. Such a comparative analysis is beyond the scope of this short exploratory study but should be a pressing subject for further research. The below discussion instead explores how the risks and opportunities associated with military AI differ for superpowers, middle powers, and small states.

### 6.1.2 Competition over military AI could destabilise the superpower rivalry between the US and China if new mechanisms are not introduced

The advent of military AI poses especially acute opportunities and risks for the United States and China, as the only current superpowers. On the one hand, they each possess technological capabilities and a sheer scale of resources – whether in terms of talent, data, compute or other infrastructure – that other nations cannot hope to match, positioning them to achieve an edge in military AI development and deployment. On the other hand, size can mean less agility, and both the US and China possess vulnerabilities for others to exploit. Though very different in their strategic goals, culture and levers, both countries have not only a lot to gain but also a lot to lose from competition over military AI, if that competition is not carefully managed. Risks include their strategic rival achieving the upper hand, their rivalry spiralling out of control to mutual disadvantage, or other more agile players exploiting AI in asymmetric ways to close the gap with the superpowers' military might, reducing the benefits of size.[147]

The literature and interviews consulted for this study emphasise the intensification of the wider US–China rivalry in recent years as the driving factor in the heavy investments made by both countries in AI in recent years, including by the US military and the Chinese PLA. Fear of being outstripped by the other power is a common trope in the official documents and wider political rhetoric of both sides when it comes to AI.[148] The securitisation of trade and technology policy in recent years (e.g. as with the banning of Huawei from 5G network infrastructure or moves by the US Congress to block TikTok) has similarly affected AI, with the US and Chinese AI sectors currently undergoing a process of painful de-coupling, even if they remain more heavily interlinked than the worsening relations between the two governments might suggest.

For all its concern about a rising China, and the rapidly advancing capabilities of the PLA, the US and its military retain a substantial lead in AI. Areas of particular strength include the attraction of global AI talent and capital to Silicon Valley, a conducive environment for private sector innovation, access to compute, access to R&D funding, a strong university sector, and the presence not only of many of the world's tech giants (e.g. Apple, Google, IBM, Meta, Microsoft, Nvidia) and leading AI firms (e.g. OpenAI, Anthropic), but also defence-specific AI companies (e.g. Anduril, Epirus, Palantir, Shield AI, etc.).[149] The US has significantly increased spending on military AI in recent years, launching a Defense AI Strategy in 2018, establishing the Joint Artificial Intelligence Center (JAIC) within the DoD, and

146 Scharre (2023).

147 Anonymous, interview by the authors, 25 March 2024.

148 Johnson (2021a).

149 Hunter et al. (2023).

most recently signalling an intent to transform its approach to the related field of military autonomy and robotics through its Replicator initiative. Against the backdrop of increased military and economic security threats from China, the US has also taken steps in recent years to limit Chinese access to key enabling technologies for AI development. These includes semiconductors, most notably through the CHIPS and Science Act introduced in 2022. It has similarly imposed new import tariffs.

Equally, many of the trend lines favour China, which already performs highly in terms of quantity of AI-related outputs (e.g. scientific publications, PhD students, etc.) and is working hard to generate more consistency in quality. The Chinese Communist Party (CCP) has identified AI as a strategic priority, both for the maintenance of Party control and for promotion of China's economic prosperity and military strength.[150] In 2017, China published its Next Generation AI Development Plan, designating AI as a 'strategic technology' crucial for international competition. The plan set targets for China to build a domestic AI industry and lead global AI investments by 2030.[151] China's strategy involves a military-civil fusion (MCF) approach, seeking to draw on investments in AI across the government, military, state-owned enterprises and private firms, the latter of which are legally compelled to support the state's (and by extension the Party's) goals.[152]

In practice, MCF is far from the seamless integration that many people outside of China fear, or which the CCP hopes to achieve. Chinese commentators have criticised slow progress towards true MCF.[153] Still, this strategy reflects the Chinese state's greater ability to compel rather than merely incentivise AI industry to support its goals, including military modernisation. It also fits a pattern of using coercive and illicit means to gain a competitive advantage. Such tactics include intellectual property theft on an industrial scale, including from foreign AI firms and universities, as well as targeted use of foreign direct investment (FDI) to gain control over technology supply chains, including raw materials such as rare earth elements.

This strategic focus on AI has been yielding results. China's AI sector overtook that of the US in terms of total AI-related publications as long ago as 2006 (and surpassed the collective output of European scientists in 2017), though papers from AI experts at American institutions tend to be perceived as higher quality, being cited on average 70 per cent more than Chinese equivalents (and 30 per cent more than European research).[154] Still, quantity has a quality all of its own: the sheer number of Chinese scientific publications on AI means that, even if of lower average quality, they collectively surpassed the US in total citations in 2020 and are expected to overtake them as a share of the top 1 per cent of most-cited papers by 2025.[155] China also far outstrips the US in terms of numbers of AI-related PhDs and masters students generated, though many emigrate to study or work (e.g. in Silicon Valley), with only a portion returning.

Crucially, this intensifying competition in and through military AI could affect the wider strategic rivalry between the United States and China in several possible ways. The greatest

150 Kania (2019).

151 Hoadley & Lucas (2018).

152 Qi (2021) .

153 Xue et al. (2021).

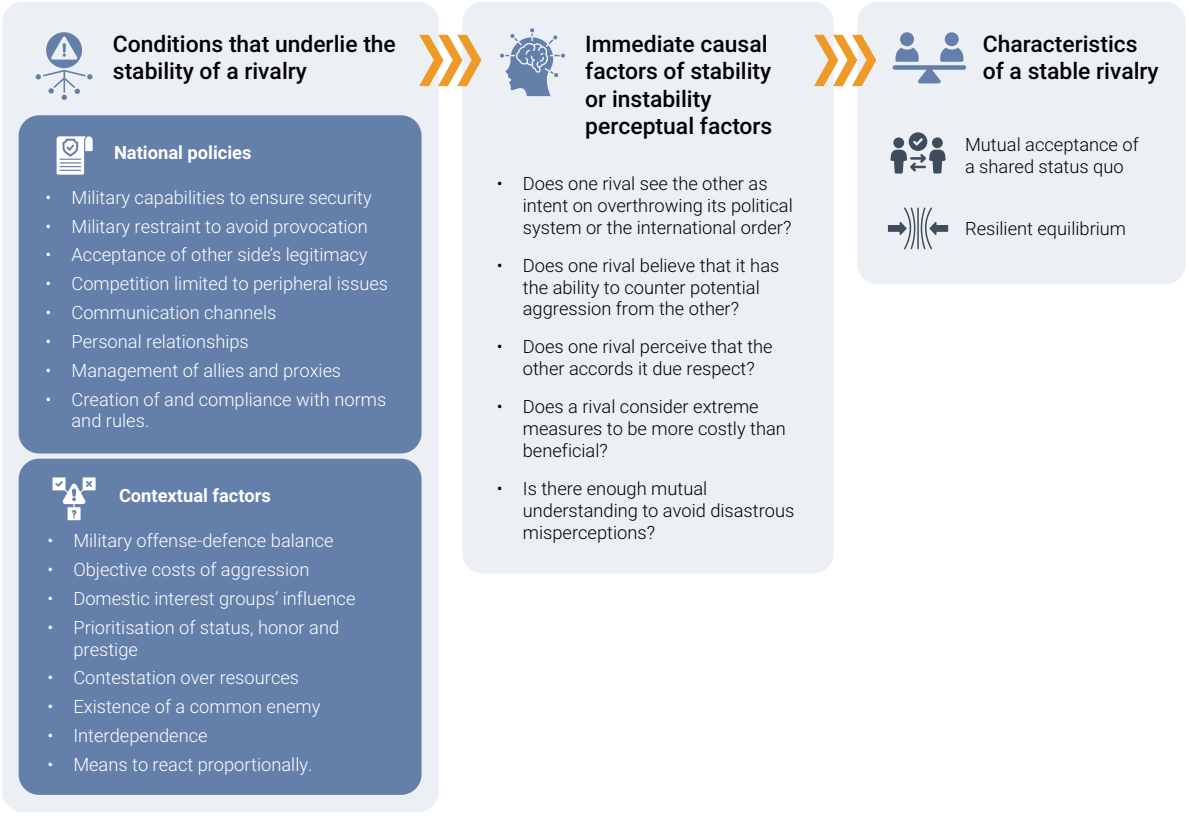
154 Scharre (2023, 30).

155 Scharre (2023, 30).

concern from either side would be that military AI provides a decisive advantage to their rival, affecting conflict outcomes e.g. in any future war over Taiwan.[156] Besides these direct battlefield impacts, military AI systems could also have broader implications for the stability – or otherwise – of their superpower rivalry.

Previous RAND research and historical case studies have identified factors that contribute to the stability of such strategic rivalries. AI could affect almost all of those detailed in Figure 6.2. As examined in Chapters 4 and 5, the rollout of military AI could influence each rival's strategic goals, decision making calculus, access to and trust in information, perceptions and misperceptions, and domestic politics and external relationships (e.g. with allies, partners and proxies). So too could it affect the military offence–defence balance, and the escalation ladder, leading to an arms race or crisis instability.

**Figure 6.2 Factors that contribute to the stability or instability of a superpower rivalry**




*Source: Mazarr, Charap et al. (2021).*

156 As mentioned in Chapter 3, China's PLA is modernising its capabilities and adopting a set of concepts and doctrine intended to prepare it for 'intelligentised warfare' incorporating AI as part of 'systems destruction warfare' and a 'systems confrontation' against an AI-enhanced US military. The US military, in turn, is looking to advanced technologies, including AI and autonomous systems, as means of offsetting PLA advantages in terms of geography (given the proximity of Taiwan to the Chinese mainland), sheer mass, and deployment of anti-access, area denial (A2AD) and long-range strike capabilities that would make it costly for US and allied forces or bases to operate within the first or second island chains of the Indo-Pacific in the opening days and weeks of any conflict.

In recognition of the potentially destabilising effects of AI on their already fraught relations, the leaders of the US and China agreed in November 2023 to establish a dialogue on AI safety and related issues. Yet communication channels between the two militaries remain limited and wider Track 1, 1.5 or 2 efforts to establish a common approach to managing AI-related escalation risks are in the early stages at best.

These uncertain effects of AI on a changing US–China relationship pose risks and opportunities for other nations as they seeks to navigate their own role in shaping competition among the two rivals, as summarised in Table 6.1.

**Table 6.1 Superpowers: Impacts from military AI**

| Opportunities | Risks |
|---|---|
| • Opportunity to deepen cooperation with the US DoD on AI, e.g. via AIPfD and AUKUS Pillar Two<br>• Opportunity to use investments in military AI and autonomous systems to deepen interoperability with US and enhance capacity to project power into the Indo-Pacific region<br>• Opportunity to seek to exert a moderating influence on both rivals, if with comparatively limited DIME levers<br>• Opportunity to develop other partnerships on AI and/or exert more leadership on European defence (e.g. via NATO or the Joint Expeditionary Force) as a hedge against a US preoccupation with China | • Risk that US and China agree and impose a common approach to global governance of AI that does not fully align with national interests<br>• Risk that AI sector is undermined by intensifying US–China trade war, intellectual property thefts and disruptions to technology supply chains (e.g. semiconductors, rare earth elements)<br>• Risk that AI-intensified competition with China further pulls US focus towards Indo-Pacific, to the detriment of European/NATO security<br>• Risk that US prioritises military AI and uncrewed vehicle (UxV) capabilities optimised for countering China (e.g. in an air- and maritime-centric theatre) but not for Russia (in a land-centric theatre)<br>• Risk that US is pulled into conflict with China by AI-induced escalation, with cascading effects on global economy and security<br>• Risk that the rivalry between the US and China could force countries, including security partners, to make choices, with some opting to align with China |

*Source: RAND Europe analysis (2024).*

### 6.1.3. Middle powers face tough choices over how to focus resources, carve out areas of asymmetric strength and influence global governance for military AI

Middle powers are also racing to invest in AI. As outlined in Figure 6.1, this includes countries such as Australia, Canada, France, Germany, India, Japan, South Korea, Turkey and the UK. Such nations of course face all the issues covered in Chapters 3 to 5, but also added dilemmas: too small to have the full breadth and depth of AI capabilities, or degree of influence over the evolution of global governance, that are available to the US and China, but also large enough powers to have agency and to harbour ambitions to exert meaningful influence over the development of military AI.[157]

**Table 6.2 Middle powers: Impacts from military AI**

| Impacts | Risks and opportunities |
|---|---|
| • AI could help middle powers to remain economically and military competitive in age of intensifying US–China rivalry<br>• Middle powers could seek to increase their global influence as early adopters or exporters of military AI systems<br>• Middle powers could seek to play an 'honest broker' or convening role between strategic rivals (e.g. US, China) or in shaping nascent global norms and governance for military AI<br>• Military AI may enhance the capacity of middle powers to address defence and security threats, either alone or in coalitions (reducing reliance on US ally), offsetting deficiencies in personnel, mass<br>• Military AI may produce new minilateral groupings or expand scope of extant forums for defence cooperation (e.g. the Quad)<br>• Additional pressures on nuclear powers and the subset of those who are permanent members of the UN Security Council (P5), e.g. AI-related nuclear escalation concerns for India and Pakistan | • Opportunity to be leading player relative to a nation's size when it comes to military AI<br>• Opportunity to use military AI systems to enhance capability and interoperability in minilateral groups (e.g. the Joint Expeditionary Force), as a framework nation<br>• Opportunity to use cooperation on military AI development to build new political, military and industrial ties with other regional powers (e.g. India, Turkey, Japan)<br>• Opportunity to play role of norm entrepreneur and influence emerging global governance architecture for military AI<br>• Risk that other nations prove more agile in developing and adopting military AI<br>• Risk that EU plays increasingly influential role on AI governance and/or makes more ambitious investments in AI industrial growth and military AI that compete with other (e.g. UK) interests<br>• Risk that divergent policy, legal and ethical stances on military AI strain relations with certain middle powers, or bring added domestic backlash against certain partnerships |

*Source: RAND Europe analysis (2024).*

157 Anderson & Feldgoise (2022).

Of course, approaches to military AI are not only shaped at the national level but also through partnerships. Of particular note here is the European Union, a bloc which has expressed its own ambition to be both a regulatory superpower for new technologies (e.g. via the new EU AI Act) and a more meaningful defence and security actor, including through a new European Defence Industrial Strategy. Having left the EU in 2020, the UK is no longer able to directly influence such developments but is affected by them given the size of the single market and the EU's influence on global regulatory norms (e.g. as with the General Data Protection Regulation).

### 6.1.4. Small states risk being left behind by bigger players, but could have outsized impact at the strategic level if they prove more agile in embracing military AI

Military AI also poses some unique challenges and opportunities for small states. While these nations lack the resources of larger countries, the nature of AI as a set of primarily software-based technologies and a 'force multiplier' offers the potential to further de-couple a given country's military capability or global influence from the size of its population or economy.[158]

As outlined in Figure 6.1, many of the countries ranked most highly in the Global AI Index are small nations seemingly punching well above their weight by embracing innovation, developing niche areas of strength in AI value chains, cultivating close ties between government, industry and academia (e.g. a 'Triple Helix'), and exploiting their relative agility to produce outsized impact: examples such as Singapore, Israel, Switzerland, Finland or the Netherlands.[159] Small states can accrue disproportionate military, industrial and soft power benefits from developing asymmetric strengths in certain niches of military AI, while also seeking to influence the evolution of global governance arrangements (e.g. by trying to influence alliances of which they are a part, or to play a bridging role between larger rival powers). Conversely, because smaller states lack the resources of larger nations, they necessarily have less strategic bandwidth, influence or diversification of their financial and industrial portfolios, leaving them potentially exposed should external events develop contrary to their interest.

158 Johnson (2021b).

159 Williams (2023a).

**Table 6.3 Small states: Impacts from military AI**

| Impacts | Risks and opportunities |
|---|---|
| • Military AI and autonomous systems offset lack of mass or strategic depth (e.g. in territorial terms) for small states<br>• Some small states may be able to move more quickly than large, more cumbersome defence establishments, but lack resources and can less afford to make mistakes<br>• AI can support and reinforce existing Total Defence or societal resilience models in small states with clear existential threats<br>• AI can reinforce niche military and industrial strengths (e.g. built around territorial defence) and in turn drive exports and regional influence<br>• Small states could have outsized benefits from military AI, but are also exposed to uncertain outcomes of how larger players adopt these technologies and seek to establish global governance arrangements | • Opportunity to engage with small states in niche areas of mutual interest on military AI<br>• Opportunity to exert influence through Defence engagement using AI tools<br>• Opportunity to use convening power to build coalitions of like-minded small states, including both AI leaders and developing economies, to influence emerging global governance arrangements for military AI<br>• Risk that some small states engage in a regulatory race to the bottom on AI safety and standards, or military AI specifically, that undermines competitiveness of UK AI sector<br>• Risk that certain small states become critical nodes in global AI value chains that introduce military or economic security threats (as with dependence of Taiwan for semiconductors as key enabler of AI tech) |

*Source: RAND Europe analysis (2024).*

In the field of military AI, smaller states often benefit from the added impetus that comes from heightened popular and private sector awareness of existential security threats (with the first four of the five countries listed above having military conscription and a 'Total Defence' model of citizen and industry participation in national defence). Similarly, these threats are more tightly bounded (i.e. confined to homeland defence, rather than power projection). This means that AI capabilities can be developed to optimise for these specific scenarios, supporting development of military and industrial niches (e.g. Israel's integration of AI into air defences, where it is a global leader) rather than trying to spread R&D or procurement funding or the skills base too thinly.[160] Small states such as those named above also have strong traditions of resilience, emergency preparedness and combating disinformation, creating opportunities to apply AI to reinforce these further.

## 6.2. Implications for different systems of government

### 6.2.1. Open democracies hold advantages in attracting AI talent and innovation, but AI offers new tools for cementing or exporting authoritarianism

The literature and interviews consulted for this study emphasised that AI, whether general-purpose or military-specific, present different risks and opportunities for democracies and authoritarian regimes.

160 Newman (2023).

On the positive side, open democracies tend to be better at attracting and retaining global AI talent, as well as encouraging vibrant private sectors and start-up cultures.[161] They can also draw upon international allies and partners, providing opportunities for cooperation and exchange of ideas and technology across the capability lifecycle for military AI.[162] Conversely, democracies may, as examined in Chapter 4, be more exposed to information manipulation, electoral interference and other acts of political subversion using AI. Concerns about privacy, civil liberties and algorithmic bias may also make it less palatable to utilise certain datasets for training of AI systems, and obviously influence policy, legal and ethical restrictions on LAWS.[163]

Conversely, there are also some specific advantages that AI presents to authoritarian leaders. AI tools may assist such regimes in reinforcing systems of mass repression and surveillance, with Russia investing in AI systems that exploit massive volumes of data on their populations, from video surveillance and Internet traffic to facial or gait recognition and even DNA databases.[164] China, for example, has been accused of developing AI tools that specifically help it with monitoring its Uighur minority, as well as extending the influence of its digital 'social credit' system. Harnessing AI to existing systems of repression could open the door to ever-more repressive regimes that are able to use the speed, automation and pattern recognition of AI to further enhance their control of the information space and crack down on dissenters.

Crucially, too, countries can seek to export this model of AI-enhanced authoritarianism abroad, increasing their own influence and challenging the promotion of democratic values by countries such as the UK.[165] China's desire to spread authoritarian norms is well-documented – China has exported non-AI surveillance platforms to cities in over 80 countries – and the addition of AI-enabled systems is a concerning elaboration on this trend. This also supports China's wider Digital Silk Road initiative, serving both to bolster authoritarianism across regions such as the Middle East, Africa, Latin America or Southeast Asia, and to create new critical dependencies upon China, which can extract data and increase its own local influence, sharpening bifurcation between US and Chinese systems.

Similarly, countries such as Russia and China have historically sought to use defence exports to build relations abroad, arm dictatorships and proxy actors (see Section 6.3.3) and generate funds, often via corruption, with dire consequences both for human rights and regional security. Military AI systems and autonomous systems add to the potential product list, potentially with lower ethical or safety standards in place. Such systems may be of particular interest to authoritarian regimes with concerns about the quality or loyalty of their military personnel. Conversely it remains to be seen how willing highly centralised, top-down systems of rule will be to integrate AI into their decision making. Dictatorships are typically more averse to challenge or initiative by officials at lower echelons, discourage presentation of facts

161 Castro et al. (2019).

162 NATO C2COE, interview by the authors, 3 April 2024.

163 Altmann & Sauer (2017).

164 Robles & Mallinson (2023).

165 Haner & Garcia (2019).

that disagree with the agreed party line, and often foment factionalism or duplicate security institutions (e.g. encouraging competition between an army and a national guard) to keep possible political rivals weak.

Even within democratic countries, commercial and dual-use AI technologies exported from suppliers in authoritarian countries (e.g. drones or CCTV cameras with built-in edge AI) could create new vulnerabilities and potential backdoors within critical national infrastructure. AI tools could also support online monitoring, intimidation or extortion of diaspora communities or foreign dissidents as well as assisting with social engineering, honeytraps and use of deepfakes to exert influence over elected politicians. Such threats pose significant challenges as well as direct threats to the AI sector.

## 6.3. Implications for non-state actors

### 6.3.1. The private sector is an indispensable partner on military AI, but also complicates efforts to ensure sovereignty and legitimacy of AI governance

Private firms are playing an increasing role in not only developing military AI but also in contributing to discussions on national or global governance arrangements for this new technology.

Rolling out military AI necessitates an even closer relationship between industry and the MOD, given the nature of working in MLOps pipelines and of through-life support to software-based capabilities.[166] It also entails a need for mechanisms to share data in a secure manner, improve industry understanding of military use cases, and create feedback loops from end users to enable refinement of AI algorithms based on live operations, all backed by more agile commercial approaches. At the same time, here there are risks such as vendor lock-in, or reliance on foreign suppliers of AI systems or related services (e.g. compute or secure clouds) that constrain a country's freedom of action and security of supply.[167]

The latter role for industry involves drawing in technical expertise to intra- and intergovernmental discussions on addressing AI safety, bias and risk, including responsible development of defence AI and possible solutions to the issues discussed in previous chapters (e.g. unintended escalation). However, here there is a risk that private firms may exert outsized influence over the governance and use of military AI – either globally, or through 'regulatory capture' of certain jurisdictions which embark on a race to the bottom on standards in pursuit of a competitive advantage.[168] Civil society organisations have expressed concern that commercial interests (i.e. profit and shareholder returns) may be prioritised over the public good whenever tech giants have an opportunity to influence the policy process, whether via public consultations or behind-the-scenes lobbying.[169] This raises wider questions about accountability, transparency, legitimacy and anti-trust.

166 Ryseff et al. (2022).

167 Andrew van der Lem, interview by the authors, 22 March 2024.

168 Rupert Barrett-Taylor, interview by the authors, 19 March 2024.

169 Baum et al. (2022).

There are other important cultural differences between private tech companies and the military.[170] These can be a positive, but some tech workers express reservations about working with the armed forces. Survey results highlight a civil–military divide over AI development, with a substantial proportion of Silicon Valley employees and alumni from top computer science programmes feeling uneasy about certain military applications of AI, particularly those involving lethal force.[171] This may limit the ability of some democracies to leverage the full potential of their industries to support military goals, even as other nations are able to compel support from industry (e.g. China's MCF).

There is also a longer-term risk around the erosion of the state's monopoly over the use of force. The increasing involvement of the private sector in military AI development may result in non-state actors, including private security companies, gaining access to advanced military AI technologies and capabilities.[172] Already, social media companies play an increasing political role in regulating the infosphere and tackling issues, such as online extremism, discussed in Section 6.3.2.

### 6.3.2. Proliferation of AI could support extremist organisations with recruitment, planning and the conduct of increasingly sophisticated attacks

AI also has potential applications across the spectrum of terrorist activities. At one end, violent extremist organisations may use Generative AI to support fundraising and recruitment. For instance, Generative AI could be used to create persuasive propaganda materials or to target individuals susceptible to radicalisation online. AI could also be used to acquire knowledge to plan and execute attacks, making them more lethal and precise, for example using ML models to predict the responses of security forces, or to analyse large amounts of data to inform planning.[173] In more extreme cases, terrorists could use AI and autonomous systems as part of attacks directly, for example to conduct strikes on military forces, critical infrastructure or soft targets (e.g. crowds, civil aviation, etc.). Finally, AI could allow the best-resourced terrorist organisations to deploy hybrid forces on the battlefield (akin to ISIS at its peak, or to Hezbollah today), to coordinate the activities of both conventional and irregular forces, or to integrate cyber and physical attacks, likely with few ethical constraints.

Conversely, there are opportunities for states to incorporate AI into counter-terrorism operations both domestically (police and security services) and abroad (military). For example, AI could be used to identify patterns in terrorist activities, or to predict and prevent potential attacks. However, these uses of AI will need to be balanced against concerns about privacy and civil liberties, for example if AI profiling tools were used to support early identification of individuals at risk of radicalisation.

### 6.3.3. Proxy actors are already making extensive use of uncrewed systems, with increasing rollout of military AI and autonomy likely to exacerbate this threat

As alluded to in Section 6.2, hostile states such as Russia, China and Iran may export military AI systems and autonomous systems to proxy

170 Hunter et al. (2023).

171 Ryseff et al. (2022).

172 Anonymous, interview by the authors, 22 March 2024.

173 Haner and Garcia (2019).

actors, allowing them to project power and influence. By controlling the algorithms, these states can enhance their control over proxies while simultaneously boosting the latter's military capabilities. The real-world examples of Houthi use of Iranian drones against Saudi Arabia or in the Red Sea, or of Hezbollah and Hamas attacks on Israel using such systems, illustrate this trend.

Arming of proxies puts increasing pressure on state militaries to find novel ways of dealing with the cost asymmetry of hostile actors using cheap, massed autonomous systems against high-value, low-density traditional military platforms (e.g. ships) and bases. The deployment of AI-enabled systems by proxy actors could also lead to more intense and prolonged conflicts, making it harder for fragile societies to escape a cycle of violence. Enhanced military capabilities may allow proxy actors to resist conventional forces more effectively, prolonging the fighting and increasing the potential for escalation or spillover to neighbouring regions. Similarly, state sponsors of proxy actors could use AI-enabled systems to conduct operations with greater plausible deniability.[174] By relying on proxies to deploy AI, hostile states can distance themselves from direct involvement in conflicts and avoid repercussions, e.g. from AI decisions.

### 6.3.4. Serious and organised crime groups could similarly acquire increasingly sophisticated AI capabilities that pose new threats to international security

Domestically, AI could have sweeping impacts on crime (e.g. AI for fraud, deepfakes for blackmail and extortion, etc.), but this is largely an issue for the police and therefore outside the scope of this report. However, AI and autonomous systems could undermine the security and stability of fragile states, creating the conditions for criminal groups not only to conduct their business but also to directly challenge the local government. This would exacerbate worrying trends seen recently with the increasingly pseudo-military capabilities of drug cartels in Mexico, who have used submarines to smuggle drugs into the US and engaged in direct confrontations with the Mexican military, or the takeover of Haiti by criminal gangs. Transnational crime networks may increasingly be able to acquire capabilities that used to be the purview of sophisticated state militaries. This could include AI-enabled surveillance systems, weapons and offensive cyber tools, potentially leading to a more dangerous security environment. In turn, such networks may be major players in proliferating AI tools and military systems in the first place, as with the global illicit trade in small arms, explosives and technical know-how or materials associated with weapons of mass destruction.

### 6.3.5. NGOs using AI do not pose a direct military threat, but could nonetheless prompt unintended consequences if not handled carefully

As briefly discussed in Chapter 5, NGOs could incorporate AI into planning for aid provision and disaster relief. AI and autonomous systems could be used to predict the impact of disasters, optimise resource allocation, and improve the efficiency and effectiveness of aid delivery. This presents benefits to states from increased civil society capacity to deal with issues such as civilian harms.[175] Conversely, there are some risks associated with using dual-use capabilities. NGOs must

174 Reinhold & Reuter (2022).

175 Janjeva et al. (2023).

be cautious to ensure that their use of AI does not inadvertently contribute to militarisation or conflict. If not managed carefully, the use of AI by NGOs could lead to reputational risks. For example, if an AI used by an NGO makes a mistake or is subject to a hack that leads to harm, this could damage the NGO's reputation and credibility, leading to a wider backlash against aid workers and institutions – with knock-on effects on human security.

## 6.4. Summary

Just as Chapter 5 covered the diversity of ways in which strategic risks and opportunities associated with military AI might manifest across the continuum of competition and conflict, so this chapter has explored the differing ways in which certain types of actor might exploit this technology. These have cascading implications for the international system and for strategic interests: with nations needing to navigate multiple simultaneous challenges to the rules-based order from intensifying superpower competition and the growing influence of non-state actors, all exacerbated by proliferation of military AI.

The next chapter examines ways of mitigating such risks and positioning to exploit opportunities, emphasising the need for a clear proposition from the UK (or any other actor seeking to exert a constructive influence), and coordinated use of all DIME levers to shape global defence AI developments.

# Chapter 7. **Priority issues**

This technical report has sought to lay out a conceptual framework that enables a structured analysis and discussion of the full range of strategic-level risks and opportunities associated with the rollout of military AI. To this end, the preceding chapters have offered a detailed overview of the – admittedly still immature and fast-evolving – evidence base that exists on this important topic. The question then arises: which of these myriad risks and opportunities should nations worry about most, and prioritise action against, given finite time, resources and possible levers of influence?

The remainder of this report aims to inform ongoing work within government to answer this question and develop a strategy for mitigating the strategic risks associated with military AI and positioning states to create or exploit opportunities for advantage in terms of improved security, influence or prosperity:

- The following short chapter draws together the key findings emerging from Chapters 1 to 6, offering an initial (and tentative) assessment by the RAND team of priority areas for action.
- Chapter 8 considers what makes AI-related risks and opportunities different – or similar – to those encountered with other technologies (e.g. nuclear) or domains (e.g. cyber), and possible insights from how these have been tackled (e.g. via arms control, TCBMs, norms).
- Chapter 9 incorporates these insights and maps the major risks and opportunities identified below in Chapter 7 against a toolkit of potential measures through which states might influence the future direction of global defence AI developments, using all DIME levers.

## 7.1. Towards a prioritisation of strategic risks and opportunities

### 7.1.1. One of the most important findings of this study is deep uncertainty around AI impacts; an initial prioritisation is possible, but this should be iterated as evidence improves

If seeking to prioritise a subset of AI-related issues from the many identified through this report and the conceptual framework it has outlined, a clear rationale is required. In broad terms, there are two types of method available for prioritising measures intended to mitigate risks or maximise opportunities:

- **Risk-based methods** are useful where there are high levels of confidence in projections of both the probability and impact of certain AI-related trends (risk being calculated as a function of probability multiplied by impact), leading to strategies that optimise against a narrow set of scenarios.
- **Uncertainty-based methods** are useful where there are low levels of confidence in projections of either probability or impact, leading to a focus on strategies that minimise regret across the widest possible range of plausible scenarios (e.g. employing methods such as robust decision making, or assumptions-based planning).

This latter group of methods may hold most initial promise, with the ambition being to drill into more detailed risk modelling as understanding of what AI can do on a technical level, and how it is being used in the real-world, improves.

### 7.1.2. The RAND team identified priority issues demanding urgent action

It is beyond the scope of this short study – given tight constraints on time, data, and resources – to undertake such a robust analysis of the relative impact and importance of different AI-related risks and opportunities, and thus a definitive prioritisation of which issues governments should focus on, in which order. Conducting such an assessment, and updating it on a rolling basis, should be an urgent task for government, e.g. incorporating greater focus on AI risks into national risk registers or defence planning.

In the interim, the RAND team undertook an initial assessment, based on their analysis of the prominent themes emerging from the interviews and literature consulted for this study. Given the acute uncertainty already mentioned, they sought not to attempt to attach their own estimate of likelihood to individual risks and opportunities, but rather to capture those that AI and defence experts see as most concerning due to having the potential for either a) significant or severe impacts on national interests in the short to medium term, and/or b) potentially catastrophic impacts that may not accrue until the long term, but nonetheless require pressing action now given the high stakes and the long lead times for crafting an effective global response.

On this basis, Table 7.1 outlines ten priority issues that the RAND team identified as emerging from each level of the framework detailed in Chapters 1–6. While not exclusively related to military AI, the team also included one macro-trend that is highly prominent – if contentious – within the contemporary debate on AI-related risks and opportunities, namely the possibility of humanity achieving AGI.

**Table 7.1 Prioritising risks and opportunities for action**

| SELECTED PRIORITY RISKS AND OPPORTUNITIES | | | INITIAL ASSESSMENT OF IMPACT | | |
|---|---|---|---|---|---|
| **Framework category** | | **Issue** | **Significant** i.e. potential to disadvantage in sub-threshold | **Severe** i.e. potential to disadvantage in conventional war | **Catastrophic** i.e. potential for catastrophe or existential threat |
| **National** | | Economic disruption and warfare | ● | | |
| | | Information-manipulation (e.g. deepfakes) | ● | ● | ● |
| | | Changes to defence productivity, mass and lethality | ● | ● | |
| **International** | By actor type | Erosion of RBIO and governance institutions | ● | ● | |
| | | AI-enabled repression (and export thereof) | ● | | |
| | | Empowerment of non-state actors (e.g. bioweapons) | ● | ● | ● |
| | By conflict type | Changes to military offence-defence balance | ● | ● | |
| | | Impact on escalation dynamics | ● | ● | ● |
| | | Impact on nuclear | ● | ● | ● |
| **Macro-trends** | | Prospects for AGI and non-alignment | ● | ● | ● |

*Source: RAND Europe analysis.*

### 7.1.3. Whether these manifest as risks or opportunities will depend on how quickly and effectively states adapt to intensifying competition over and through AI

Of the issues above, literature and interviews suggest the most potentially impactful are those relating to:

- **Information manipulation**, such as AI deepfakes, which could not only drive political, economic and social problems but also skew military decision making in times of crisis.
- **Empowerment of non-state actors** with asymmetric capabilities that challenge the dominance of state militaries or, in the worst-case scenario, new tools of mass destruction (e.g. bioweapons).
- The interlinked impacts of AI on the **offence–defence balance** between

adversaries, on **escalation dynamics** toward warfighting, and on the **stability of nuclear deterrence**. These issues are especially concerning amidst intensifying superpower rivalries and in a world already grappling with other drivers of insecurity (e.g. Ukraine, Israel–Iran, Taiwan, migration, climate change).

- The potential catastrophic safety and security risks associated with any future **advent of AGI**.

Below this level, there are still major potential issues to worry about in terms of the disruptive impacts of weaponised AI both on the domestic politics and economy that sets the direction and resources for Defence, and on the rules-based international order that underpins global security, stability and prosperity. There is also significant concern about the extent to which AI could tip the balance in favour of repressive and authoritarian modes of governance in many parts of the world, while threatening to subvert democratic politics, pollute the information environment and undermine will-to-fight at home.

However, many of these potential risks could also manifest as opportunities – with the balance of pros and cons from the rollout of AI hinging on how quickly and effectively states are able to adapt institutions such as national MODs/DoDs and the Armed Forces to exploit the benefits of AI, and how well they exert influence internationally to shape global behaviours around AI in a direction that suits their national interests and values. To address these challenges, governments such as the UK's must develop a comprehensive strategy that considers the complex interplay of technological advances, geopolitical competition, and evolving norms and rules in the international system.

To this end, the next chapter turns to possible insights from how governments and their militaries have addressed analogous issues arising in the past from other novel and disruptive technologies or domains.

# Chapter 8. **Lessons from other domains**

When grappling with the challenge of military AI, governments can draw insights and potential lessons from the successes and failures of initiatives to govern disruptive technologies and bolster strategic stability in other domains. This chapter provides a brief overview of models employed in such areas, drawing on the literature review and interviews, before moving to consider what makes today's AI-related risks and opportunities similar or different to those encountered in the past. On this basis, it then considers possible transferrable lessons that could inform ongoing development of strategies for shaping global defence AI developments, using the toolkit of measures outlined in Chapter 9.

**Box 8.1 Summary of findings: Chapter 8**

AI has some fundamental characteristics that differentiate it from other disruptive technologies (e.g. nuclear) and which demand a bespoke approach to developing risk management and governance strategies. These include its status as a set of dual-use GPTs, its primarily software-driven, if hardware-enabled nature, and its high levels of proliferation and democratisation across borders due to the central role of the private sector in AI innovation.

Nonetheless, there are possible lessons – both success stories and cautionary tales – from how governments and militaries have worked with allies, partners and even sworn enemies to manage the risks arising from analogous technologies or domains in the past.

Recurring themes include the need for patience; building a common understanding and levels of mutual trust over time through a multi-track series of dialogues; identifying 'quick wins' on certain issues and promoting behavioural norms, and transparency and confidence-building measures (TCBMs) as a prelude to more formal agreements; bringing in verification to ensure compliance with those agreements; and then engaging in inclusive, participatory discussions with a range of stakeholders to scale any emergent norms from minilateral to global level.

The ongoing development of a cross-governmental strategy for addressing the risks and opportunities associated with military AI at the strategic level should build upon such insights, as well as other design principles identified through the literature review and interviews conducted for this short study. These suggest that governments should employ a toolkit of different mechanisms combining: i) efforts to boost uptake of AI and maximise its benefits to Defence; ii) efforts to restrict or slow non-state and terrorist actors' own uptake of military AI and impose costs upon them; and iii) efforts to shape global, minilateral and bilateral governance arrangements for AI.

*Source: RAND Europe analysis.*

## 8.1. Existing models of risk management

### 8.1.1. Governments should incorporate insights from other domains into their approach to managing the strategic risks of military AI, including the need for patience

When grappling with the challenge of military AI, states can draw insights and potential lessons from the successes and failures of initiative to govern disruptive technologies and bolster strategic stability in other domains. Examples of governance models attempted elsewhere are provided in Table 8.1, based on a list of potentially analogous issue areas identified through literature review and interviews.

**Table 8.1 Potential models from other domains and sectors**

| Domain/Sectors | Approach |
|---|---|
|  Nuclear | • Initial attempts at US–Soviet dialogue in 1950s largely unsuccessful; took the Cuban Missile Crisis of 1962 to catalyse the Partial Test Ban Treaty of 1963.<br>• Emphasis on multi-track approach (Track 1, 1.5, 2).<br>• Informed by development of game theory to guide prioritisation of risks.<br>• Emphasis on managing both arms race stability (i.e. via arms control and non-proliferation agreements) and crisis stability (i.e. via communication channels).<br>• Shift over time from initial belief in possibility of winning a nuclear exchange towards promulgation of idea of mutually assured destruction (MAD).<br>• Series of arms control treaties between the biggest superpowers of the Cold War, including limitations on numbers, ranges and positioning of certain weapons to address perceived drivers of greatest escalation risk.<br>• Backed by verification mechanisms and mutual acceptance of need to not interfere with national technical means of verification or early warning and NC3 systems.<br>• Influential mediating role for small states (e.g. Sweden) with niche expertise.<br>• Promotion of wider global limitations on spread of nuclear weapons through, above all, the Non-Proliferation Treaty (NPT), and the promotion of a political ambition to move eventually towards a nuclear weapons-free world. (In practice, the NPT has had mixed impact, with some states abandoning nascent nuclear programmes but others developing them regardless. Some criticism that the NPT represented many of the world's most powerful states entrenching a hierarchical arrangement that favoured them and then not delivering on disarmament pledges. Some other states [e.g. Japan] have opted to retain latent capacity to 'rush for the bomb' in lieu of acquiring nuclear weapons.)<br>• Increasing challenges adapting to the 21st century. (Arrangements designed for a bipolar world have not proven resilient in the face of the twin challenge of deteriorating US–Russia relations since the invasions of Ukraine in 2014 and 2022, and increased multipolarity given the rise of China, a nuclear-armed North Korea and concerns about Iranian nuclear programme.) |

| Domain/Sectors | Approach |
|---|---|
|  <br>Missiles | • Intersection with nuclear: addressing both payload and delivery technologies.<br>• In Cold War, formal treaty agreements to restrict certain intermediate range nuclear-capable missiles to de-risk possibility of nuclear confrontation in Europe.<br>• Non-treaty arrangements (e.g. Missile Technology Control Regime) to prevent proliferation of missile systems capable of carrying weapons of mass destruction, including export controls on weapons, components, materials and technology. |
|  <br>Conventional forces | • Flagship Treaty on Conventional Armed Forces in Europe, intended to build trust among signatories, reduce escalation risk, and avoid a costly security dilemma.<br>• Backed by verification mechanisms (e.g. with Armed Forces in Europe (CFE) Treaty's basic provisions on aerial overflight expanded upon via the Treaty on Open Skies).<br>• Again, as with nuclear arms control, recent fraying of both formal mechanisms and informal norms in the face of wider deterioration of the strategic environment. |
|  <br>Biological and chemical weapons | • Establishment of strong norms against use of biological or chemical weapons after the end of the First World War, with neither side using them in the European theatre of the Second World War despite large stockpiles and a total war scenario.<br>• Formalisation of Geneva Protocol and later the Biological Weapons Convention (BWC) and Chemical Weapons Convention (CWC) to cement these norms.<br>• For BWC, use of confidence-building measures but failed negotiation of a formal verification regime to monitor compliance.<br>• For CWC, greater progress towards destruction of weapons stockpiles (if with controversies along the way, especially in Iraq and Syria).<br>• Challenges with dual-use nature of some underlying technologies and know-how.<br>• Increasing pressures on BWC from democratisation of technologies associated with producing bioweapons, including new risks from the intersections with AI. |
|  <br>Landmines and cluster munitions | • Subset of nations sign up to agreements in effort to promote a global norm and exert political pressure on non-signatories.<br>• In practice, biggest military powers (including US, Russia, China) and most prolific uses of landmines opt out, limiting the overall global impact.<br>• Full-scale Russian invasion of Ukraine and drawdown of US military aid in 2023–2024 reopened debates over military benefits of fielding cluster munitions. |

| Domain/Sectors | Approach |
|---|---|
|  Maritime | • Centuries-long evolution of norms and law.<br>• Formalisation of UN Convention on the Law of the Sea.<br>• To reduce risk of accidental escalation of confrontations at sea, US and Soviet Union established Incidents at Sea Agreement in 1972 to provide mechanisms for reporting, de-escalating and learning from any incidents, and establishing rules of acceptable behaviour and ways of sharing information.<br>• Incidents at Sea Agreement seen as successful model and potential inspiration for other domains (e.g. outer space). |
|  Civil aviation | • Empowered International Civil Aviation Organisation (ICAO) that sets standards and drives cooperation on commercial aviation, safety, crash investigations, etc.<br>• Strong economic incentives for states to comply with ICAO. |
|  The Internet | • Globally distributed computer network, without any single owner or overseer.<br>• Reliance on decentralised, multi-stakeholder and voluntary system of interconnected actors drawn from civil society, industry, academia, governments and international institutions – with major players including the Internet Governance Forum (IGF), Internet Corporation for Assigned Names and Numbers (ICANN), Internet Architecture Board (IAB), Internet Engineering Task Force (IETF), Internet Research Task Force (IRTF) and World Wide Web Consortium (W3C).<br>• Cooperation on shared policies and standards to ensure interoperability.<br>• Principles of decentralisation and fair, equal, open access for all.<br>• In more recent years, concerns over the potential for fragmentation of the Internet (the Splinternet) as competing nations seek to control their national networks or isolate them from the wider world (e.g. in times of crisis, or for censorship).<br>• In parallel, growing concerns about potential centralisation of excessive power in the hands of a small number of tech giants (e.g. to privilege their own online services), prompting a backlash (e.g. the Web 3.0 movement). |
|  Social media | • Significant impacts on society, as well as security and safety (e.g. spreading disinformation, propaganda and terrorist recruitment campaigns).<br>• Need to balance any regulation with free speech, data and privacy concerns.<br>• Dominance by very large US and Chinese social media companies.<br>• Lack of global governance arrangements; some national legislative and regulatory initiatives (e.g. UK's Online Safety Bill), but fragmented approach and overall reliance on social media companies to police content on their own platforms.<br>• Prioritisation of most extreme content and impacts (e.g. safety risks to children). |

| Domain/Sectors | Approach |
|---|---|
|  <br>Cyber | • Emphasis on inclusive, multi-stakeholder approaches and technical dialogue.<br>• Limited progress towards norms of agreed behaviour in terms of cyber warfare, but more towards regulations, guidelines, standards on less contentious issues.<br>• Challenge of separating from wider geopolitical tensions. |
|  <br>Space | • Dual-use technology that was initially controlled by state actors but increasingly with commercial firms driving innovation, with some parallels to AI.<br>• Cold War competition over space, but utopian ambition to preserve space as 'province of all [hu]mankind' and arena for scientific exploration and discovery.<br>• Policy debates over whether space represents global commons, or not.<br>• Some progress towards legal regime for space, most notably the Outer Space Treaty, and targeted bans to deal with certain escalatory scenarios (e.g. nuclear weapons placement in outer space), but sizeable gaps in space law, including lack of clarity over issues such as weaponisation or militarisation of space.<br>• Formalisation of UN role through the International Telecommunications Union (ITU) and the parallel if at times disjointed efforts of the Committee on the Peaceful Uses of Outer Space (COPUOS) and work on Preventing an Arms Race in Outer Space (PAROS).<br>• Increasing challenges from democratisation and proliferation of space technology, leading to more players, debris and threatening behaviours.<br>• UK-led initiative to promote normative approach to promoting responsible space behaviours through UN Open-Ended Working Group (ultimately vetoed by Russia and others but yielded some benefits and showed utility of participatory approaches that brought in industry and civil society organisations' perspectives).<br>• By contrast, proposal of formal treaty instruments by Russia, China, but concerns that these are means to entrap democratic nations.<br>• In face of limited progress towards a global solution, declarations of self-restraint from some nations (e.g. bans on testing anti-satellite missiles) and signing of minilateral agreements, e.g. Artemis Accords, to seek to influence global norms.<br>• Increasing calls for more robust intergovernmental organisation to address issues such as space traffic management or astronaut rescue. |

*Source: RAND Europe analysis (2024).*

Of the different technologies or domains covered in the table above, those most frequently compared to AI include nuclear weapons (given their potentially catastrophic, even species-ending implications), bioweapons (given the outsized, non-linear impacts and the challenges around countering proliferation, including to non-state actors) and the Internet, cyberspace and space (given their dual-use nature and the centrality both of digital technologies and private firms in shaping global developments).

Still, AI has some important characteristics that make it different to other disruptive developments that governments and militaries have sought to regulate, govern and risk manage in the past. To this end, the following section considers what possible lessons might be transferrable to the unique context of AI.

## 8.2. Transferrable learning

### 8.2.1. AI risks – and opportunities – differ from those in other areas, but that does not mean that certain insights from past initiatives cannot be carried across

There are some important differences between the strategic risks and opportunities arising from military AI and the governance challenges encountered in other domains. Important considerations include:

- The fact that AI is a set of **dual-use GPTs** with both military and civil applications means that there are acute trade-offs for governments to make between, say, pursuing the economic or social benefits of rapid AI adoption and managing the potential national security risks. Even more than most disruptive technologies, AI demands a joined-up approach across government to balance the competing policy imperatives of different departments.
- The fact that AI focuses on **machine intelligence** means that it is 'an actor not just a factor' in decision making, with uniquely direct implications for issues such as military command and control, or strategic stability.[176] Other issues, such as nuclear, missile or cyber threats, change what strategists think about. AI changes how they think and who – or what – does the thinking.
- The fact that AI is **primarily software-based** brings added practical difficulties to any attempt to counter the cross-border proliferation of algorithms; discern from afar the capabilities and levels of autonomy or human oversight of other actors' military AI systems; or enforce compliance of said systems with binding instruments such as legal treaties, or softer guidance and norms.
- These difficulties are exacerbated by the fact that innovation in AI is **driven by the private sector**, especially big US or multinational firms, with more limited government control. Crucially, as explored in Chapter 6, commercial incentives (i.e. maximising profit and returns to shareholders) and the culture of tech firms can sometimes clash with government policy goals, with the private sector reaping the financial rewards of technology disruption even as the public sector must deal with the negative externalities of rolling out AI.
- At the same time, the fact that AI is **hardware-enabled**, relying on access to compute, data, power and supporting infrastructure, as well as a highly skilled workforce, means that there are still opportunities to target the spread of certain physical and human enablers of military AI.
- The heightened **uncertainty** emphasised in Chapter 7 means that many risks are still poorly understood, in terms of both probability and impact, when compared to those in other domains where such issues have been studied for decades and where real-world experience has yielded empirical data on how these risks play out in practice.

176 Pavel et al. (2024).

Conscious of such characteristics, governments should consider what possible lessons – either success stories or cautionary tales – they can derive from historical experience in other areas. There are parallels, for example, between initiatives such as REAIM or the Political Declaration with efforts to test and establish norms of responsible behaviour in other fast-changing domains (e.g. space, cyber); between calls for communication channels over incidents involving military AI or autonomous systems and those already established in the nuclear or maritime contexts; between the desire to build new verification mechanisms and some of the successes or challenges encountered elsewhere (e.g. bioweapons); in the emphasis on TCBMs as a basis for enabling deeper dialogue and more ambitious agreements down the line; and in the wider difficulty of divorcing technical questions about governance from the broader geopolitics, and insulating any treaties from deteriorating relations between key players. Table 8.2 offers an overview.

**Table 8.2 Potential transferrable lessons from other domains and sectors**

| Lesson | Description |
|---|---|
| Create forums for dialogue, even with adversaries | As seen in nuclear negotiations, dialogue is crucial in building consensus and managing risks. Governments should engage in Track 1, 1.5, and 2 diplomacies to foster cooperation and understanding among stakeholders. |
| Develop new models and theory to aid understanding and prioritisation of risks | As with nuclear or bio risks, governments should invest in the development of game theory and robust modelling activities to improve understanding and guide the prioritisation of AI risks and inform strategic decision making. |
| Promote responsible behaviour and self-restraint | Like in the space domain, governments should encourage responsible behaviour and self-restraint among AI stakeholders, both domestically and internationally. |
| Consider the balance between minilateral versus global agreements | In the absence of global consensus, governments should explore minilateral agreements and initiatives to address priority issues and influence the subsequent evolution of global norms. This could even include making unilateral declarations. |
| Exercise patience but proactively build momentum towards more ambitious agreements using TCBMs | As seen with the conventional forces, governments should work towards agreements that reduce the risk of 'arms races' in military AI, while recognising that these may take years – or even decades – to achieve given the need to build mutual trust as a prerequisite for more ambitious formal agreements (e.g. arms control).<br>In the meantime, it is nonetheless possible to establish communication channels to manage crises and prevent escalation, contributing to that build-up of trust. |
| Develop and implement verification mechanisms | As seen in the nuclear, conventional forces and chemical weapons domains, verification mechanisms are crucial in ensuring compliance with agreements. Governments should support the development of such mechanisms for military AI. |

| Lesson | Description |
|---|---|
| Recognise the challenges of dual-use tech proliferation and democratisation, including via participatory approaches that engage emerging economies and non-state actors | As has been done in the biological, chemical and space domains, governments should address the challenges posed by dual-use technologies and the democratisation of AI, including the potential for unintended consequences and misuse.<br>As seen in the cyber and space domains, for example, governments could promote inclusive, multi-stakeholder approaches that involve government, industry and civil society to develop norms and regulations for AI. |
| Adapt to the changing strategic environment | Governments should recognise that the strategic environment is constantly evolving, as seen in domains such as cyber or space. States should be prepared to adapt their approach to managing AI risks in response to these changes, to ensure that any governance arrangements do not rapidly become obsolete as tech advances. |

*Source: RAND Europe analysis.*

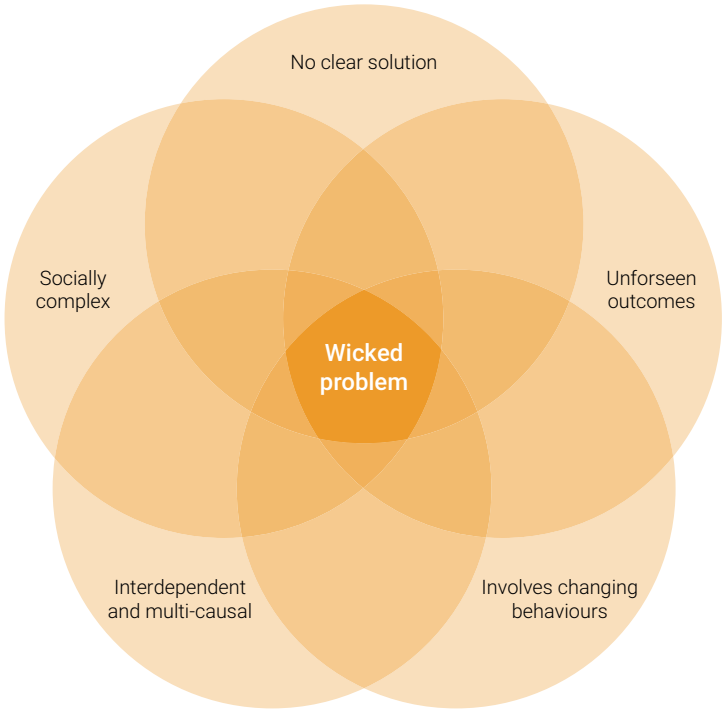
### 8.2.2. This study is not intended to craft a comprehensive strategy for dealing with military AI, but its research suggests possible building blocks for such a plan

Chapter 9 examines the toolkit of possible measures available to governments for shaping global defence AI developments in a favourable direction, beginning with actions at the national level and then moving to how states might influence developments internationally. Before that, though, the literature and interviews consulted for this study identified several design principles that should guide the formulation of a more detailed strategy for managing risks and opportunities arising from military AI:

- **The starting point should be humility and recognition that collective understanding of AI risk is still very immature – in many cases, 'uncertainty' would be a more appropriate term:** Efforts to properly understand, let alone quantify, the strategic-level risks are in their infancy when it comes to many aspects of military AI, or of AI in general. There is a lot of hype around this subject, but comparatively little hard evidence on which to base potentially highly consequential decisions in the face of deep uncertainty. Furthermore, what evidence and analysis does exist is contested, meaning that different actors (e.g. the UK versus the US versus Russia versus China) hold differing interpretations (e.g. of levels of risk), making it harder to agree on how to prioritise certain issues or begin to design solutions.
- **In this way, discussions with other global stakeholders about governance of military AI should focus on 'problem finding', not just 'problem solving':**[177] Concrete solutions can already be found to some of the technical and policy issues raised by military AI. But the myriad strategic risks and opportunities outlined in this report reflect a broader 'wicked problem' of how to manage a set of technologies as cross-cutting, fast-paced, and disruptive as AI.

177 Liu & Maas (2021).

**Figure 8.1 'Wicked problem' of global governance of military AI**




*Source: RAND Europe analysis (2024).*

A 'wicked problem' is one that cuts across different policy areas, institutional remits, and international borders; one that is complex and multi-causal, with uncertain consequences and cascading second- and third-order effects that are hard to predict; one that is interdependent with other issues (e.g. a deterioration of US–China relations, or a breakdown of Western engagement with Russia following the latter's full-scale invasion of Ukraine in 2022), making it hard to address in isolation; one that is understood and interpreted differently by different stakeholders, with no consensus on what the biggest or most urgent problems are, let alone how to solve them; one that has no single solution, but rather requires a wide range of interventions by different actors to change behaviours, with no single actor possessing all the levers needed; and one that is ultimately a political, social and cultural challenge as much as a military-technical one, requiring willingness to compromise. This means that global governance of military AI is not simply something that can be 'solved'; rather, it is a continuous 'sensemaking process regarding socio-technical change', with partial solutions to some of the issues raised by military AI opening new, potentially unforeseen issues that will need to be managed.[178]

- **Faced with such uncertainty and a lack of consensus on problem-framing, an iterative approach is needed, built around a learning process:** Improving shared understanding of the problems and dilemmas thrown up by the advent

178 Minkkinen & Mäntymäki (2023).

of military AI is thus an urgent priority, as it sets the baseline for then negotiating possible solutions at the international level. This should focus on research and learning; sharing the fruits of that learning to build common understanding of AI-related risks among relevant experts and stakeholders at the global level (e.g. via larger forums such as the UN or REAIM, or more selective groupings such as AIPfD); moving over time from an initial patchwork of different minilateral initiatives, partial solutions and 'quick wins' to build a broader consensus around norms of behaviour and risk mitigations for military AI; and eventually consolidating towards a more comprehensive and robust governance architecture, ideally one where any institutional arrangements or treaties are designed to be future-proof and not be rapidly rendered obsolete by the pace of technological change in AI.

Based on the literature and interviews, priorities for building a better collective understanding include deeper research and dialogue on:

- The reality of risks and opportunities at the strategic level, behind the hype and rhetoric.
- Cascading second- and third-order effects.
- The intersections with other emerging and disruptive technologies (e.g. bio, quantum).
- Different potential future scenarios for military AI and/or governance, and the factors and path dependencies that might funnel the world towards outcomes that would generate the biggest regret (e.g. unintended escalation to nuclear warfighting, or bioweapons use).
- 'What actually works?' in mitigating risk – or maximising opportunity – both in terms of technical and policy solutions, considering the full lifecycle of military AI (e.g. restricting certain actors' access to talent, compute, and data needed to build military AI, shaping the design features of military AI systems to build in certain safeguards, or agreeing norms of behaviour for where, when and how military AI systems are or are not used).
- The perspectives of other stakeholders (e.g. adversaries) on military AI, the underlying decision making logic and assumptions, and a state's leverage to shape these in a direction favourable to its strategic objectives when it comes to global defence AI developments.[179]

- **This should include efforts to bridge the divide between those experts and stakeholders focused on near-term risks (or opportunities) and those focused on existential ones:** This is a false dichotomy. Governments, tech giants and the wider AI community should have sufficient bandwidth to deal with both at the same time. Indeed, finding local solutions to smaller technical or policy challenges associated with military AI (e.g. reducing the risk of AI targeting NC3 systems, or developing new tools or agreements that prevent forms of information manipulation

179 These can be fed by various other analytical activities (many of which can be enabled by AI) including futures and foresight methods (to anticipate possible futures challenges); scenario analysis and risk modelling (to consider the possible impacts of different developments); net assessment; and wargaming, modelling and simulation.

using AI) can create more favourable conditions for dealing with macro-level problems such as GCRs, e.g. by enhancing resilience, or building trust and confidence among competing powers.

- **Ultimately, AI is a socio-technical system, and therefore the political dimension is as important as the technical one – governments need clear goals, a theory of success for influencing the international system, and an integrated approach to maximising levers at the national level:** To this end, Figure 8.2 overleaf summarises the toolkit of mechanisms through which governments can exert influence over global developments in defence AI, mapped against the conceptual framework discussed in the preceding chapters of this report. This toolkit combines:
  - » Efforts to boost the uptake of AI and maximise its benefits to Defence
  - » Efforts to limit the adoption of military AI by non-state and terrorist actors, or hostile / rogue states, while also imposing costs on them to influence their actions
  - » Efforts to shape global, minilateral and bilateral governance arrangements for AI.

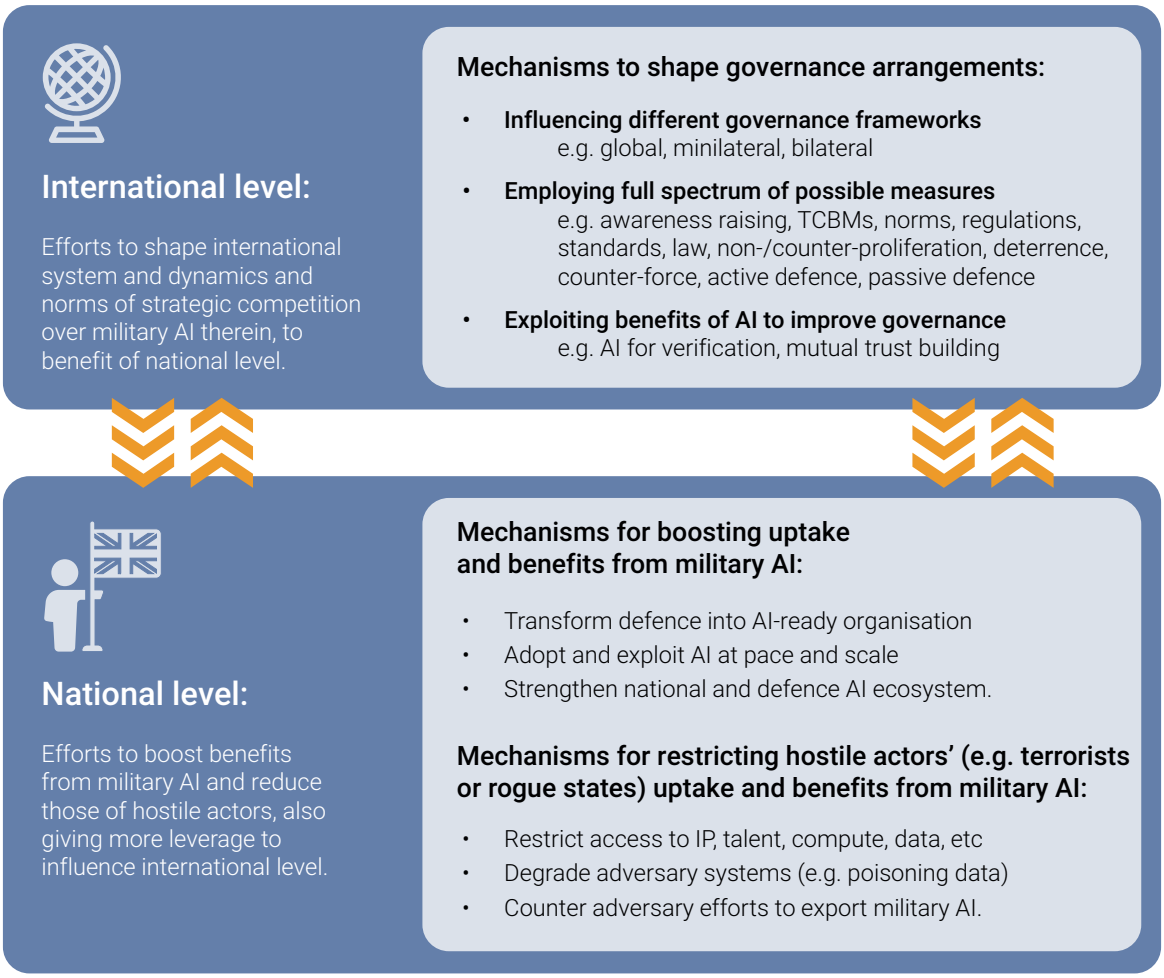
  Collectively, such measures should be mutually reinforcing, increasing the potential and propensity for strategic advantage through military AI at the national level, as well as leverage over emerging governance arrangements and norms at the international level. This reflects the different categories of the framework of risks and opportunities discussed in this report.

  Crucially, many levers sit outside government, necessitating an integrated approach across departments and with allies, partners, industry, academia and civil society. Though the global governance architecture for military AI – or AI in general – is nascent at best, states can learn from how they have tackled analogous challenges posed by disruptive technologies in the past. As shown in this chapter, there are transferrable lessons from other domains (e.g. nuclear, bio, space), mechanisms (e.g. arms control) and frameworks (e.g. the UK Missile Defence Centre's Five Pillars[180]), though AI has distinctive characteristics that require a bespoke approach.

  In essence, then, governments must balance competing to build domestic ecosystems for AI, and military AI specifically, that maximise their opportunity for advantage, while simultaneously collaborating to shape global governance arrangements in such a way as to best protect shared interests and values. This combines elements of zero- and positive-sum relations with other actors, e.g. competing to secure a military advantage but also working to reduce the risk of unwanted escalation. Getting this balance right will be a delicate dance, since negotiating with adversaries from a position of strength is important to exert influence over their behaviours, but dialogue, trust and compromise are also required to avoid a breakdown in cooperation on the biggest issues, including GCRs associated with AI.

180 The Five Pillars framework covers non-proliferation, deterrence, counterforce, active defence, and passive defence (e.g. resilience) measures. Given the breadth of the challenges posed by the advent of military AI, this study proposes expanding upon the Five Pillars to include measures such as both 'hard' and 'soft' law (e.g. norms) or TCBMs.

**Figure 8.2 Toolkit: mechanisms to shape risks and opportunities**




*Source: RAND Europe analysis (2024).*

The following chapter examines each element of this toolkit in turn, beginning with measures through which states can pursue strategic advantage at the national level, before turning to ways in which they can seek to influence the emerging architecture of global governance arrangements for military AI.

# Chapter 9. Toolkit of measures to exert influence

Building on the insights and principles outlined in Chapter 8, this chapter turns to the toolkit of different mechanisms that Defence, and governments more broadly, might employ to hedge against or mitigate the emerging risks identified so far and to maximise the opportunities associated with military use of AI.

**Box 9.1 Summary of findings: Chapter 9**

This chapter expands upon the toolkit of practical measures through which to shape global defence AI developments in directions that accord with UK and allies' interests and values. These fall into three categories.

Mechanisms to boost AI-related benefits:

1. Accelerate investment in and adoption of AI across Defence (and government and society more widely), while increasing national resilience and preparedness for hostile use or accidental misuse of AI.

Mechanisms to restrict AI adoption and benefits for adversaries:

2. Adopt a campaigning approach to restrict, slow or increase costs to non-state actors, terrorist actors and hostile states of deploying military AI.

Mechanisms to shape global governance arrangements for military AI:

3. Play a leading role in awareness raising, problem finding and sharing learning about military AI risks.
4. Develop TCBMs in conjunction with international partners – and competitors.
5. Promote an inclusive, participatory approach to build an emerging global consensus on norms of responsible behaviour around military AI, as a prelude to future more robust agreements.
6. Promote parallel development of minilateral mechanisms for reducing nuclear- and bio-related AI risks in smaller but potentially more agile forums with more bounded remits and more practical outputs.
7. Investigate ways to incorporate AI into verification and compliance mechanisms, and vice versa.
8. Over time, consolidate the current fragmented landscape of AI governance initiatives into a more concrete architecture, e.g. either through existing institutions (e.g. the UN) or new ones.

*Source: RAND Europe analysis.*

## 9.1. Mapping risks and opportunities against the toolkit

Recalling the prioritisation of risks and opportunities in Chapter 7, it is possible to map these against the elements of the toolkit outlined at a high level in Chapter 8 and expanded upon in the sections below. Crucially, different activities are intended to be mutually supporting, e.g. efforts to bolster national adoption of AI increase credibility as a thought leader in discussions on global governance, and provide leverage over the behaviours of adversaries, encouraging them to engage in risk mitigations around AI for mutual benefit.

The following sections expand upon each of these different mechanisms in turn, beginning with those to address the impacts of military AI at the national level (recalling Chapters 3 and 6) before moving to address those at the international level (recalling Chapters 4 and 5).

## 9.2. Mechanisms to boost AI adoption and benefits for Defence

### 9.2.1. The UK and allies should accelerate investment in and adoption of AI across Defence, while also seeking to increase resilience against hostile usage of AI

This category of the toolkit shown in Figure 8.2 and Table 9.1 focuses on measures to bolster uptake of military AI within Defence and to maximise the benefits these bring to the UK and like-minded nations. To these ends, the UK Government has released a raft of policy documents and plans in recent years:

- National AI Strategy (2021)
- AI Action Plan (2022)
- Defence AI Strategy (2022)
- Establishing a Pro-Innovation Approach to Regulating AI (2022)
- Royal Navy's Artificial Intelligence Adoption Roadmap, British Army's Approach to Artificial Intelligence, and Royal Air Force's Autonomous Collaborative Platforms Roadmap (2023–2024)
- Defence AI Playbook (2024).

It is not the intention of this report to recapitulate the many policy changes, structural and procedural reforms, or investments in R&D, capability development or workforce skills that are detailed within such official documents – nor to formally evaluate or compare different nations' progress towards them, which would require a much larger study. But several themes emerge from the literature and interviews:

- **The importance of maintaining momentum:** The UK and its allies (e.g. the US) must build on promising initial progress while remaining conscious of the potential disruption induced by elections, new strategic defence reviews and spending review cycles.
- **The need for access to AI talent, data and compute:** Governments are right to identify access to talent and compute, and a modernised approach to data management and sharing, as key enablers of AI. Still, all three areas represent potential bottlenecks that could hinder realisation of ambitious goals for leadership in military AI and influence over global AI development.
- **The importance of realising a more agile approach to capability development:** The UK and other allies have taken welcome steps to accelerate not only experimentation but also fielding of AI systems. Governments should seek to embed MLOps pipelines and spiral development practices across Defence more generally: aiming to rapidly acquire the 60–80 per cent solution and then iterate from there using end user feedback,

**Table 9.1 Mapping of priority issues for governments against the toolkit**

| | | Impact at national level | | | Impact at international level | | | | | | Macro trends |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Implications by actor type | | | Implications by competition type | | | |
| | | Economic disruption and warfare | Information -manipulation (e.g. deep-fakes) | Defence productivity, mass, lethality | Erotion of RBIO and governance institutions | AI-enabled repression (and export thereof) | Empower-ment of non-state actors (e.g. bioweap-ons) | Changes to military offence -defence balance | Impact on escalation dynamics | Impact on nuclear deterrence | Prospects for AGI and non-alignment |
| Mechanisms to boost AI adoption and benefits for UK defence | Accelerate investment in and adoption of AI across defence, while increasing reslilience against hostile use of AI | Improved counter-measures or economic security | Improved means to detect, attribute and remove manipulated data or narratives | Ensuring UK outperforms adversaries in and through military AI | | | | Developing both AI and counter-AI (e.g. counter-C4ISTAR) capabilities | Ensuring UK has a range of deterrent options to influence adversary behaviour | Ensuring UK has a range of deterrent options to influence adversary behaviour | |
| Mechanisms to restrict AI adoption and benefits for adversaries | Adopt a campaigning approach to restrict, slow or increase the costs to adversaries of deploying military AI | | | | Countering revolutionist actors' efforts to exploit AI to undermine RBIO | Countering authoritarian efforts to deploy tools of repression globally | Countering proliferation and misuse of AI and using AI for CT/CVE | Undermining adversaries' efforts to achieve military advantage over the UK | | | |
| Mechanisms to shape emerging governance arrange-ments for military AI | Play a leading role in awareness raising, problem finding and sharing learning about military AI risks | Building understanding of economic security risks from AI | Building understanding of technical solutions to AI deepfakes | Building understanding of barriers and enablers of transforma-tion | Rejuvenating global diplomacy via consensus building on AI risks | Articulating risks of depenfencies on Chinese and Russian AI tools | Establishing common cause with all states in stopping terrorist misuse of AI | Sharing learning about AI risks with others as basis for dialogue | Sharing learning about AI risks with others as basis for dialogue | Sharing learning about AI risks with others as basis for dialogue | Cutting through hype and debate to better understand the real risks of AGI |

| | | Impact at national level | | | Impact at international level | | | | | | Macro trends |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Implications by actor type | | | Implications by competition type | | | |
| | | Economic disruption and warfare | Information -manipulation (e.g. deep-fakes) | Defence productivity, mass, lethality | Erotion of RBIO and governance institutions | AI-enabled repression (and export thereof) | Empower-ment of non-state actors (e.g. bioweap-ons) | Changes to military offence -defence balance | Impact on escalation dynamics | Impact on nuclear deterrence | Prospects for AGI and non-alignment |
| Mechanisms to shape emerging governance arrange-ments for military AI | Develop transparency and confidence building with key allies (e.g. US) and competitors (e.g. China) | Reducing economic instability by estabilishing norms for use of AI | Reducing hostile interference by estabilishing norms for use of AI | | Seeking to reduce overall tensions e.g. in superpower rivalries | | | Building trust to avoid 'use it or lose it' mentality and premptive strikes | Building trust to avoid 'use it or lose it' mentality and premptive strikes | Building trust to avoid 'use it or lose it' mentality and premptive strikes | Seeking to establish common rules for sake, responsible research into AGI |
| | Promote an inclusive, participatory approach to build norms of behavious on military AI, focusing on 'quick wins' | Addressing inherent interconnec-tivity of global economy | Addressing inherent interconnec-tivity of global infosphere | | Boosting legitimacy via engagement of diverse state and non-state voices | | | Recognising role of private firms in shaping military AI applications | Promoting soft norms as prelude to more ambitious formal agreements | | Making progress on practical aspects of AI safety e.g. Red Teaming models |
| | Promote parallet development of minilateral mechanisms for reducing nuclear- and bio- related AI risks | | | | Seek to reduce overall tensions e.g. in superpower rivalries | | Establishing common cause with all states in stopping terrorist misuse of AI | | De-risking the most pressing potential flashpoints and conflict scenarios | Working with nuclear powers to agree on issues such as AI in NC3 | |

| | | Impact at national level | | | Impact at international level | | | | | | Macro trends |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Implications by actor type | | | Implications by competition type | | | |
| | | Economic disruption and warfare | Information -manipulation (e.g. deep-fakes) | Defence productivity, mass, lethality | Erotion of RBIO and governance institutions | AI-enabled repression (and export thereof) | Empower-ment of non-state actors (e.g. bioweap-ons) | Changes to military offence -defence balance | Impact on escalation dynamics | Impact on nuclear deterrence | Prospects for AGI and non-alignment |
| Mechanisms to shape emerging governance arrange-ments for military AI | Investigate ways to incorporate AI into verification and compliance machanisms, and vice versa | | | | Reinforcing mutual trust with hard empirical data ('trust but verify') | | Countering proliferation of military AI to non-state actors | | De-risking the most pressing potential flashpoints and conflict scenarios | Exploiting AI to assist with verification for nuclear arms control | |
| | Over time, consolidate the current fragmented landscape of AI govenance initiatives into a more concrete architecture | | | | Supporting more robust institutions and streamlining competing initiatives or fora | | | | | | Making progress on addressing AI safety at global rather than national level |

*Source: RAND Europe analysis. Note: the coloured bar beneath each priority risk or opportunity reflects the assessment of its impact from Figure 7.1 (with darker being higher impact).*

rather than the traditional waterfall approach to procurement.[181] Notably, though, this is far from the first time governments have attempted to embrace agile acquisition, and previous rounds of reform have faltered due to factors such as short-termism, lack of resource, internal opposition and turf wars, a lack of senior ownership, a culture misaligned with reform goals, and bureaucratic inertia. Governments must learn lessons from past failures and ensure this attempt at transformation sticks.

- **The need to harness the full benefits of collaboration:** Governments such as the UK's are already engaged in a range of defence AI collaborations, including through AIPfD, AUKUS and Five Eyes (FVEY). They will have new such opportunities with, for example, the extension of AUKUS Pillar Two projects to other countries besides the core partners (e.g. Japan), the maturation of NATO's Defence Innovation Accelerator for the North Atlantic (DIANA), or various bilateral cooperation initiatives. Crucially, though, governments must ensure that they extract maximum value from this burgeoning panoply of collaboration: this entails ensuring a strong holistic approach working in close partnership with industry and academia to maximise influence over programme goals, workshare, etc.; addressing issues such as intellectual property rights, data and standards early; and ensuring sovereignty concerns are addressed. Similarly, governments should continue to explore how they can best exert a constructive influence over the defence AI industry, and the AI sector more broadly, which is unlike traditional defence industry.
- **The importance of resource:** Realising the ambitions outlined above takes sizeable resource. While the UK and allies have invested in various AI (and related, e.g. autonomy) programmes, despite other budgetary pressures, there remains a disconnect between political rhetoric and financial realities – as is true of levels of defence spending more widely, given the increased threats facing the world since February 2022. The scale of the strategic opportunities and risks associated with military AI, as explored in this report, suggests that more ambition is needed if the UK and like-minded nations are to outcompete their rivals. In turn, embracing AI across the Defence enterprise could bring efficiencies that drive financial savings.
- **The importance of being willing to slay 'sacred cows' to exploit the full benefits of AI:** Relatedly, some of the literature and interviews consulted for this study argue that most militaries not engaged in an active war (i.e. unlike Ukraine) remain too risk-averse when it comes to disinvesting from legacy structures, capabilities and programmes to make space for more innovative, AI-enabled solutions. They suggest that governments need to embrace a different attitude to risk commensurate with changing threats, not to mention the pace of change in AI.
- **The need to increase understanding and buy-in across Defence:** Implementing controversial reforms, defunding certain programmes and potentially increasing defence AI budgets all require buy-in from stakeholders at different levels: inside Defence itself, across government, at the political level and in the public. This

181 Retter & Dee (2024).

entails a need to affect cultural change within Defence, including through strong leadership but also through continuing efforts to raise awareness of the applications, benefits and limitations of AI (e.g. educating personnel on AI bias), to highlight success cases arising from deployment of military AI, and to show the link to strategic outcomes.

- **The need to demonstrate continued focus on responsible development of military AI:** Finally, the literature and interviews underscored the importance of not only pursuing a responsible and ethical approach to the development of military AI, but of being seen to do so. This speaks to the need for continuing public dialogue and proactive engagement with legislators, NGOs, civil society organisations, academics, the media and others to explain governmental approaches.

Furthermore, the evidence gathered for this study also emphasised another dimension – one that has arguably been overlooked, as it falls outside of the remit of any single government department or agency, reflecting the dual-purpose nature of AI technologies. Specifically, the literature and interviews emphasised the need for urgent action to enhance the resilience of governments, economies, the infosphere and society to withstand possible shocks associated with AI safety (e.g. malfunctions, biases or GCRs such as AGI) or threats associated with hostile use of AI (e.g. AI to influence elections, disrupt financial markets, promote propaganda, etc.).

Many of these actions would fall outside the direct control of Defence, but national MOD/DoDs and militaries would nonetheless have an important supporting role to play (not least given their growing expertise in AI) as part of a cross-governmental effort to bolster resilience in preparation for the new demands of an age of AI.[182]

## 9.3. Mechanisms to restrict AI adoption and benefits for non-state and terrorist actors, and hostile and rogue states

### 9.3.1. Governments should adopt a campaigning approach to restrict, slow or increase the costs to adversaries of deploying military AI, to help shape their behaviours

As discussed in Chapters 3 and 4, strategic advantage is not only about maximising one's own potential and propensity for advantage, but also taking steps to reduce that of hostile actors and/or to influence them away from hostile courses of action in the first place. This reflects the dialectical nature of strategy.

Here, this category of the toolkit shown in Figure 8.2 focuses on actions that governments could take across the full capability lifecycle for military AI. Crucially, many of the levers required to undertake such action are owned by different parts of Defence, or by other departments, allies, partners, industry or academia. This entails the need for a campaigning approach to exert influence, deter and outcompete hostile and rogue actors. If AI is truly as significant a determinant of future strategic advantage as many government policy documents – or this report – suggest, then it follows that a similarly proactive and joined-up approach is needed to directly and indirectly degrade adversaries' capacity to develop and deploy military AI. Furthermore, successful actions of this kind will also buy time for governments to overcome

182 Janjeva et al. (2023).

barriers to adoption or improvement of their own military AI.

Literature and interviews suggested a range of possible actions that could form part of such a campaign:

- **Clear understanding of different adversaries' perceptions, goals, strengths and limitations when it comes to military AI, and AI in general, informed by net assessment:** As noted in Chapter 4, there is some evidence that competing nations tend to overstate their adversaries' progress towards the rollout of military AI, or assume a greater willingness to take certain actions (e.g. lowering levels of human control) than may actually be the case, all while being all too painfully aware of the policy, financial, bureaucratic and technical barriers on their own side. More generally, there is a need for a revised understanding of how the strategic culture and decision making processes and logic of other nations or non-state actors are evolving because of the incremental integration of military AI alongside traditional human-centric C2. Both imperatives speak to the need for enhanced dialogue among competing nations as part of TCBMs, as well as intelligence gathering and analysis to build a better understanding of how other actors, including non-state actors, are approaching the integration of AI and how they stack up in relative terms.
- **Measures to restrict, or at least increase the cost to adversaries of, access to military AI-related talent, IP, data, compute and other infrastructure:** Especially when targeting hostile non-state actors or rogue states, this entails non-proliferation measures adapted to the specific context of AI. While it may be harder to stem the flow of AI technologies compared to traditional military capabilities, given their software-based nature, some of the literature and interviews consulted for this study noted that there were nonetheless ways to make life harder for nefarious actors. For example, this could mean putting in place cybersecurity measures to restrict access to training data or export controls on essential hardware (much as the US and Netherlands are currently seeking to limit China's capacity to produce the most advanced semiconductors), or even monitoring energy usage to identify hidden computing facilities, given their need for power and cooling. And while it may be difficult to restrict the proliferation of dual-use AI technologies, additional controls can be placed on bespoke military AI systems. There is also a need for action to boost investment and supply chain security, prevent IP theft, and secure AI talent. This builds on new legal powers (e.g. the UK's National Security and Investment Act, which classifies AI as a 'high risk' sector), but likely requires further awareness raising across the AI sector, including both private firms and universities, about possible threat vectors.
- **Measures to actively degrade hostile actors' military AI capabilities:** While the previous point addresses efforts to restrict or slow hostile actors from developing military AI systems, governments should also ensure they have the tools needed to degrade such systems once they have been deployed – whether covertly below the threshold of open conflict (e.g. via cyber-espionage and sabotage, or poisoning of training data for algorithms) or more overtly in any potential warfighting scenario (e.g. via battlefield use of kinetic, electronic and cyber means of taking down hostile autonomous systems). This fits within a broader requirement for counter-C4ISTAR capabilities, as alluded

to in Chapter 4, given the intensifying competition for tactical, operational and strategic advantage through superior sensing, access to information and decision making.[183] Understanding how adversary AI systems operate, their role within military decision making processes, and their biases, assumptions or other cognitive limitations could help states devise ways of exploiting such vulnerabilities. Crucially, too, the human in or on the loop within an adversary's C4ISTAR complex is also a target: it may be enough to erode the enemy commander's trust in the information and analysis being presented to them by an AI system, for example through information and psychological operations, rather than actually degrading that AI system directly. Longer term, nations could also seek to impose additional costs on hostile actors by developing a diversified portfolio of different military AI and related capabilities of their own, imposing additional intelligence and R&D costs on adversaries needing to spread their resources thinly to develop countermeasures.

- **Measures to counter adversaries' attempts to export military AI:** As outlined in Chapter 6, governments do not only face the threat of adversaries deploying military AI to support their own armed forces. They should also investigate further how best they can work with allies and partners to prevent hostile or rogue states from exporting military AI systems to third parties, thereby expanding their influence with proxies, spreading a model of AI-enabled authoritarianism, and undermining regional security. China and Russia have long proven effective in exporting surveillance technologies, private security companies and traditional defence equipment, especially in parts of the Middle East, Africa, Latin America or the Indo-Pacific. While governments are unlikely to want to export their most sensitive technologies, they could nonetheless work with industry, allies and partners to develop a competing proposition to export markets – stressing the risks of dependence on countries such as China and extolling, by contrast, the benefits of responsible defence AI development. This could include providing capacity building for other nations seeking to develop their military and industrial capabilities when it comes to military AI, or to stand up systems of democratic oversight for AI more generally, as well as political and economic inducements to promote such behaviours.
- **A mix of deterrence, coercion, persuasion and inducements to shape adversaries' behaviour:** This theme of influencing behaviour applies to adversaries themselves. Given its scope, and the focus on the strategic risks and opportunities that arise from military AI's use in an increasingly competitive geopolitical environment, this report has necessarily focused on implications through the lens and language of strategic advantage. Yet it must be re-emphasised that governments can derive advantage through dialogue and cooperation even with their adversaries, wherever this enables positive-sum collaboration on issues of mutual interest (e.g. avoiding accidental nuclear escalation due to AI).[184] The first bullet point in this section emphasised the need for improved

183 Black et al. (2024); Lucas et al. (2024).

184 Cave & Ó hÉigeartaigh (2019).

understanding of how adversaries are thinking about and approaching military AI, and subsequent bullets have outlined robust measures through which governments might enhance their ability to undermine, coerce or deter adversaries seeking to deploy military AI in ways that run contrary to their interests or values. Yet states must also develop ways of persuading or inducing otherwise hostile actors into compromise over the most problematic military AI risks, and of being clearer about their own stance to avoid misperceptions. This requires careful calibration, employing a mix of 'carrot' and 'stick' as needed to exert influence over other actors' thinking and actions.

It is to this final question of shaping and influencing the emerging norms of behaviour around military AI, including the impacts of this set of disruptive technologies on more formalised governance arrangements at the global level, that the rest of this chapter now turns.

## 9.4. Mechanisms to shape and influence governance arrangements

### 9.4.1. Governments should initially focus on 'quick wins' (e.g. easier issues, non-binding mechanisms or smaller groups), building trust and momentum towards eventual consolidation of a more robust governance architecture

This category of the toolkit shown in Figure 8.2 focuses on ways in which states can seek to exert a constructive influence on the norms and dynamics of military AI governance at an international level. Recent years have seen a flurry of relevant forums and initiatives launched at either the minilateral or global level, some focused on exploratory dialogue, some on scoping voluntary principles, and others on developing more ambitious proposals for new regulations or policy guidelines.

A mapping of global AI initiatives by Global Partners Digital identified more than 50 that are currently active, with prominent examples including[185]:

- **Initiatives under the auspices of the United Nations** or its various committees and agencies, including the UN Secretary General's High-Level Advisory Body on AI (HLAB-AI), negotiations over the Global Digital Compact (GDC), the activities of the International Telecommunications Union (ITU), Internet Governance Forum (IGF) or Human Rights Council (HRC), and efforts relating to the adjacent fields of military autonomy and robotics, such as the open-ended Group of Government Experts (GGE) on lethal autonomous weapons systems.
- **Initiatives under the auspices of established regional blocs,** such as the Council of Europe's AI Treaty (CAI), the EU AI Act, the G7 Hiroshima Process, the Organisation for Economic Cooperation and Development's AI Policy Observatory and AI Principles, the work of the World Health Organization (WHO) on AI risks to global health security, or the examination of military applications and risks of AI through NATO ACT and the NATO Centres of Excellence. This also includes collaboration between the EU and US through the EU-US Trade and Technology Council (TTC), with its remit on, inter alia, coordinating the approach to digital technologies, including through a TTC Joint Roadmap on Trustworthy AI and risk management. Similarly, it includes consultations among

185 Kaspar et al. (2023).

Brazil, Russia, India, China, and South Africa through the BRICS.

- **Initiatives through new minilateral forums** or events created, at least in part, for this express purpose, such as the Responsible AI in the Military Domain (REAIM) conference first held in the Netherlands in February 2023 (and scheduled for a second iteration in South Korea in 2024), or the UK's AI Safety Summit at Bletchley Park in November 2023.
- **Vision statements** promoted by different, often overlapping coalitions seeking to build new global norms. Examples include the Call to Action endorsed by 57 nations at the end of REAIM in 2023, or the US-initiated Political Declaration launched in parallel, which has been endorsed by 52 nations as of March 2024 and shares many similar principles but is implicitly framed more around tackling (perceived) irresponsible behaviours, most notably from the US's rival, China.[186]
- **Initiatives driven by industry**, such as the Frontier Model Forum, voluntary commitments on safe, secure and transparent development of AI agreed between Amazon, Anthropic, Google, Inflection, Meta, Microsoft and OpenAI in July 2023, or the second round of commitments made by eight further US companies (including defence focused Palantir) in August of that year.
- **Networks, conferences, and informal dialogues** (including Track 2) among academic, think tank and civil society representatives from different nations, such as forums bringing together US, Chinese and European AI experts to exchange ideas on possible 'red lines' of common agreement.

### 9.4.2. Navigating this increasingly complex web of overlapping, often duplicative initiatives is a burden on the diplomatic bandwidth of any nation

This is a dizzying array of initiatives. Indeed, it has been argued that the 'central problem among states who may wish to identify the implicit rules of the road for using military AI resembles less of a prisoner's dilemma and more of a coordination problem'.[187] Faced with this challenge, states need not have a definitive plan at this stage for which of the above forums – or more – they intend to prioritise in the long term for addressing a specific military AI-related risk or opportunity, as they will invariably need to remain flexible given the immaturity of so many of the relevant groupings and initiatives. Still, states should continue to interrogate where and how they can best exert a constructive influence on the development of specific initiatives, whether leading from the front or more quietly from the back, and make sure that forums yield concrete outputs and serve higher objectives, rather than just being yet another talking shop.

Here, the literature and interviews consulted for this study emphasise the need to employ a diverse assortment of measures to shape the emerging governance architecture and ultimately enhance strategic stability in the face of military AI risks, as outlined below:

#### Given gaps in understanding, there is an urgent need for awareness raising, problem finding and information sharing about military AI risks

186 Javadi & Onderco (2024).

187 Horowitz et al. (2020).

As shown in Chapter 7, even the most AI-savvy governments and militaries have a relatively nascent understanding of the full risks (or opportunities) associated with different AI technologies, or their application in a military context. These shortfalls in nuanced understanding are only more acute in the case of nations with less sophisticated public or private sector expertise in AI.

This entails a need for awareness raising, education and outreach activities to establish a baseline of common understanding as the foundation for subsequent dialogue, to dispel myths and hype, and to enhance mutual comprehension of different stakeholders' emerging perspectives on topics such as the most pressing risks, the possible 'quick wins' for collective action and areas of disagreement.

Here, the UK can draw upon its asymmetric strengths in terms of private sector and academic expertise on AI safety, risk and ethics, as well as the intersections of AI with different sectors (including defence) and disciplines (including law, social sciences, and economics). It can also leverage its convening power as an influential military and diplomatic player, including building on the success of the Bletchley AI Summit, the UK's participation in the AIPfD, AUKUS, Five Eyes, NATO, UN Security Council and other forums, and the lessons from other domains, such as space, where the UK has played an 'honest broker' role to bring together diverse actors to build a common understanding of risks and definitions.

### Governments should develop transparency and confidence-building measures with both allies and competitors to build trust and reduce escalation risks

Before TCBMs can be implemented, it is essential to put in place lines of communication with potential adversaries to address military AI-related risks, both before and during a crisis. The experience of other domains, such as the dialogues that have existed around nuclear, chemical or biological weapons, or missiles and other conventional forces, suggests a need for a multi-track approach.

This should combine high-level political exchanges, more technical consultations on specific issues of mutual interest, the potential for joint research or exchange of information on AI-related risk modelling, and a variety of dialogue formats (e.g. Track 1, 1.5 or 2). Importantly, it should also include military-to-military relations and direct communication channels to help avoid misunderstandings and accidental de-escalation in a crisis, though in practice certain states may be reluctant to make much use of such channels amidst wider tensions.[188]

Beyond the above-mentioned dialogues, there are a range of other potential TCBMs through which governments might seek to bolster strategic stability when it comes to the rollout of military AI.[189] These could include unilateral declarations of restraint (such as a moratorium on incorporation of AI into NC3 systems) intended to reassure adversaries and encourage reciprocation, de-risking certain mission areas – building on examples from other domains, such as cyber or space.[190] It could also mean publication of relevant doctrine, policy guidelines, information on operations involving new AI-enabled military capabilities, or a standard approach to incident reporting, as in the maritime domain or civil aviation.

188 Pavel et al. (2024); Geist (2024).

189 Horowitz et al. (2020).

190 Horowitz & Scharre (2021).

More creatively, it could even mean proactive measures to signal the levels of autonomy or human oversight under which uncrewed military systems are operating, especially in times of heightened tension; or, relatedly, to avoid jamming other nations' uncrewed assets to ensure they have at least the option of reach back to a human decision maker, reducing the risk of accidental escalation due to a machine decision, brittle model, or error.

Other initiatives could focus on reducing AI biases and vulnerabilities that could lead to unintended consequences, for example by working to bolster the cybersecurity of AI systems. While TCBMs are insufficient to grapple with the risks of military AI in isolation, they may serve to help avoid worst-case scenarios, and to build trust as a basis for more ambitious governance discussions.[191]

### Governments should promote an inclusive, participatory approach to build global norms of responsible behaviour around military AI, as a prelude to formal agreements

Building on all the above, states could seek to promote the development of new norms of responsible behaviour at the international level, shaping these norms as a basis for more ambitious and formalised agreements in future. Such norms could either be:

- **Positive norms**, i.e. committing to a certain level of human control over AI systems for given military tasks, and ways of handling errors or other incidents, as in the air and maritime domains.
- **Negative norms,** i.e. committing to restrict development or deployment of certain capabilities or avoid certain uses of military AI systems in the field, as with restrictions on landmines, cluster munitions, or weapons of mass destruction.

Based on the priority issues identified in this study, examples of areas for norm development include:

- Codes of conduct and guidance for responsible R&D on military AI, such as on how to consider safety and proliferation risks from the outset, scrutinise foundation models, ensure cybersecurity, or ensure a test and certification regime that reduces risks of error or non-compliance with IHL and LOAC.
- Agreeing common approaches and standards on issues such as incident reporting, data or privacy.
- Countering the malicious generation and spread of AI deepfakes and other propaganda tools, and especially the manipulation of information likely to result in significant economic disruption, societal harms or military escalation risks (e.g. targeting financial markets, critical national infrastructure or NC3 systems).
- Pursuing international consensus on de-risking contentious issues such as LAWS, building on the existing work of bodies such as the open-ended GGE under the Convention on Certain Conventional Weapons (CCW).
- Pursuing international consensus on de-risking the potential future development of AGI.

Closely related to normative approaches are 'soft law' instruments that seek to encourage and reinforce certain behaviours – either from states or industry – but which are not binding. While these are necessarily less concrete than

191 Puscas (2022).

binding mechanisms, they do have certain advantages:

- On the one hand, they can serve as a testing ground for new ideas to see what works and how actors respond, teasing out the dividing lines on bigger issues.
- On the other, they are much easier to agree or implement than binding mechanisms, and there is a reduced concern than certain states (normally, though not exclusively, authoritarian countries such as Russia, China or Iran) will either defect from these binding agreements or use them as the basis for lawfare against those parties who comply.
- They also provide an opportunity to engage with a much wider range of perspectives, including a more diverse range of nation states beyond the most powerful militaries or biggest economies (e.g. countries in the so-called 'Global South'), the private sector, academia and civil society organisations. Here, the multi-stakeholder models employed to shape norms around other digital technologies, e.g. governance of the Internet or cyberspace, could provide useful lessons, as could the opening of the recent UK-initiated UN Open-Ended Working Group (OEWG) on space to give voice to non-state perspectives.

By comparison, 'hard law' instruments have long been seen as the gold standard for regulating interstate competition using dangerous or contentious technologies (e.g. arms control for nuclear weapons). However, many of the landmark Cold War treaties have unravelled in recent years amidst a sharp deterioration of US–Russia relations and the shift to a more multipolar world (e.g. with the rise of China, India and others). Furthermore, in other domains (e.g. space), authoritarian actors have often proposed legally binding treaties as a means of constraining the freedom of action of those (i.e. democratic) states who are bound to comply with international law, even as the authors plan to bend or break their own rules.

### Governments should, in parallel, develop minilateral mechanisms for reducing nuclear- and bio-related AI risks as an urgent priority that cannot await global consensus

In the near term, there appears to be limited scope for formal global treaties to restrict the proliferation or deployment of military AI – not least given the aforementioned characteristics of dual-use, software-driven AI technologies which only make them harder to regulate in this manner compared to other capabilities (e.g. nuclear warheads, missiles). Nonetheless, such agreements could remain a longer-term ambition for smaller groupings of states or for specific issue areas. Literature review and interviews suggested a number of priority areas:

- Engaging with nuclear-armed powers (e.g. the P5, or more ambitiously other powers such as India and Pakistan, or even North Korea) to promote a tightly bounded dialogue to de-risk the potential intersection of AI with nuclear weapons. Practical examples could include agreements of common definitions around AI and its potential applications and use cases in the nuclear domain, and/or to retain a human decision maker in the loop for any decision on employment of nuclear weapons, and/or to not target other countries' NC3 systems using AI or uncrewed systems. (Conversely, use of AI-based targeting systems could reduce the need for nations to hold as many nuclear weapons and provide a bargaining chip with other nuclear powers, setting the conditions for dialogue around future

arms control.[192] Such trade-offs reflect the complexity of addressing nuclear issues in a multipolar world marked by heightened tensions and the deterioration of Cold War agreements.)

- Engaging with like-minded nations to agree practical actions for reducing the risk of hostile non-state actors such as terrorist organisations acquiring access to basic enablers of military AI (e.g. compute, data or AI talent) or misusing AI for their own malicious purposes (e.g. to assist with planning of attacks or, most alarmingly, acquisition of biological weapons and other capabilities).

### Governments should investigate ways to incorporate AI tools into verification and compliance mechanisms, and vice versa

As noted in Chapters 5 and 7, the software-driven nature of military AI systems poses challenges from a verification perspective. But that does not mean it is not worth technical research into means of building an initial, incomplete verification system and then improving it over time. Equally, states could choose to champion research into the use of AI as a tool to support verification and compliance for other agreements, e.g. conventional or nuclear arms controls. This could help to defuse geopolitical tensions more generally and thereby create more favourable conditions for finding accommodations on the global governance of military AI.

### States should, over time, aim to consolidate the current fragmented landscape of AI governance initiatives into a more concrete architecture

Ultimately, the ambition should be to shift focus from an initial flurry of 'quick wins' towards building more concrete, formalised and binding measures at the global level. This speaks to the need for a mix of bilateral dialogues with key allies or adversaries alongside minilateral engagements with coalitions of the willing (e.g. AIPfD, FVEY), in parallel with more democratic, inclusive and multi-stakeholder forums (e.g. REAIM). The latter can thereby focus on awareness raising, technical dialogue and norm shaping with a broader audience that also incorporates the Global South, as well as industry, academia and civil society perspectives. Some of the more realpolitik conversations that need to take place on topics such as the impact of AI on nuclear escalation or cyber warfare can take place in more closed settings.

In time, the landscape of governance mechanisms for military AI may thus evolve from its current fragmented, polycentric and largely exploratory model towards a set of more widely agreed core definitions and ideally norms of behaviour, accompanied by a consolidation around a smaller number of more concrete international agreements or forums.

A recurring lesson from other domains or sectors is the fact that building consensus or compromises can take years or decades: initial discussions may not seem to yield much in the way of concrete results, but they nonetheless set the foundation for later negotiations on concrete agreements. Conversely, there is the serious risk that such maturation of governance arrangements for AI does not occur at sufficient pace to forestall some of the potential strategic shocks that could arise, intentionally or otherwise, from certain military uses of AI – with meaningful progress towards global governance instead coming only after some disaster or near-disaster as a catalytic event, as was the case with nuclear risk

192 Puwal (2024).

mitigations and arms controls in the Cold War after the Cuban Missile Crisis of 1962.[193]

## 9.5. Summary

This chapter has examined the potential building blocks of a strategy or plan for mitigating the risks and opportunities posed by increasing military use of AI globally. It has proposed a toolkit of measures that seek to bolster the UK's and like-minded nations' own AI capabilities and resilience to withstand AI-related shocks; to increase their ability to impose costs upon hostile or rogue actors, including as a means of deterring or influencing them to the negotiating table; and to shape the evolution of military AI governance at the international level using all DIME levers.

The final chapter provides overall conclusions from the study, as well as identifying possible next steps.

193 Geist (2024).

# Chapter 10. Conclusion and next steps

The development and adoption of AI by state militaries and armed non-state groups is ushering in significant changes to the character of competition and conflict. Literature and interviews suggest that the advent of AI could be one of the defining features of the 21st century, with profound ramifications across all areas of government policy, the economy, wider society, and the military.

Against this backdrop, the MOD and FCDO commissioned RAND to build a conceptual framework to bring structure and nuance to thinking about the emerging strategic risks or opportunities from this growing military use of AI, as presented in **Chapter 2** of this report.

The framework aims to address current deficiencies in the research and policy debates over military AI, including a narrow focus on certain risks (e.g. LAWS or nuclear-related GCRs) at the expense of others, and the substantial effects of hype, rhetoric and uncertainty. Building on literature review and interviews, this exploratory study has examined the strategic risks and opportunities that arise for UK Defence across the different sub-categories of the conceptual framework, exploring how military use of AI could impact:

- **Chapter 3:** The international system, and the intensity and dynamics of strategic competition or collaboration within that system.
- **Chapter 4:** The potential and propensity of individual actors (e.g. the UK) to achieve strategic advantage within that persistent global competition.
- **Chapter 5:** The full continuum of cooperation, competition and conflict: from alliance-building through to deterrence, crisis management, conventional warfighting or even nuclear exchanges.
- **Chapter 6:** Actor type: differing between superpowers (e.g. US and China), medium powers (e.g. UK) and small states; democracies and authoritarian regimes; or state and non-state actors.

This study has also considered the toolkit through which states might seek to influence global military AI trends in their favour. This has involved:

- **Chapter 7:** Identifying priority issues for action, while recognising that high levels of uncertainty around the pace, direction and likely impacts of military AI will necessitate further research and discussion to refine this initial assessment.
- **Chapter 8:** Examining possible transferrable lessons and insights from risk mitigations employed in other domains and technology areas, as well as what makes military AI-related impacts different.
- **Chapter 9:** Outlining a mix of practical measures that governments could employ to bolster their own benefits from military AI, restrict those of hostile actors, and exert a constructive influence over emerging global governance arrangements for defence AI.

Importantly, the tentative conceptual framework and associated findings or high-level recommendations as presented in this report are the product of a quick-turnaround study conducted in one month. They are intended not as definitive answers but rather as the basis for further, more detailed

research and analysis, as well as additional consultations. To this end, the RAND team identified several gaps in the literature and evidence base that could merit further investigation:

- **Iterative refinement of the conceptual framework,** including through expert workshops.
- **Systems mapping** to better understand the intersections and potential feedback loops that might exist between different strategic risks or opportunities that have hitherto been considered largely in isolation from each other, and by different communities of AI and defence experts.
- **Scenario analysis** (e.g. using techniques such as hierarchical cluster analysis or backcasting) to explore the potential ramifications of alternative future worlds for military AI and its governance (or lack thereof), as well as any path dependencies and their implications.
- **Wargaming** to explore the potential implications of different scenarios, as well as possible courses of action for leading governments and responses from other actors, both allies and adversaries.
- **Deep dives** into specific areas of strategic risk or opportunity (e.g. AI deepfakes in information operations, or requirements for C4ISTAR capabilities to degrade hostile military AI systems), technical capabilities of AI systems, elements of the toolkit for exerting influence over military AI developments (e.g. TCBMs), lessons from other domains and sectors (e.g. nuclear, space), or lessons from the evolving perspectives of other actors, both state and non-state.
- **Net assessment** to assess the relative potential and propensity of different states to achieve a strategic – rather than merely tactical or operational – advantage in and through military use of AI, and to improve upon existing global AI indices to enable a comparative analysis that is much more tailored to the defence setting.
- **Pre-mortems or Red Teams** to stress-test possible government interventions on military AI.
- **Assumptions-based planning** exercises to identify the load bearing assumptions that underpin any strategy or plan for shaping military AI risks, opportunities or governance; assess which of those might be vulnerable should external conditions shift; generate an early warning system of signals that would indicate such a change in conditions; and identify shaping actions to reduce the likelihood of such strategic shocks and hedging actions to reduce their impact if they do.

# References


Akhtar, Rabia. 2017. 'Making of the Seventh NWS: Historiography of the Beginning of the Nuclear Disorder in South Asia.' *The International History Review* 40(5): 1115–33. doi: 10.1080/07075332.2017.1404482

Altmann, Jürgen & Frank Sauer. 2017. 'Autonomous Weapon Systems and Strategic Stability.' *Global Politics and Strategy* 59(5): 117–42. doi: 10.1080/00396338.2017.1375263

Ashby, Heather. 2024. 'Pitting Existential Risks Against Near-Term AI Risks is a False Dichotomy.' Instick, 25 January. As of 28 March 2024: https://inkstickmedia.com/pitting-existential-risks-against-near-term-ai-risks-is-a-false-dichotomy/

Bachman, Sascha Dominik Dov & Richard V. Grant. 2023. 'The Need for An Australian Regulatory Code for the Use of Artificial Intelligence (AI) in Military Application.' *American University National Security Law Brief* 13(2). As of 29 August 2024: https://digitalcommons.wcl.american.edu/nslb/vol13/iss2/2

Baker, Mauricio. 2023. 'Nuclear Arms Control Verification and Lessons for AI Treaties.' *arXiv*. doi: 10.48550/arXiv.2304.04123

Barbé, Esther & Diego Badell. 2019. 'The European Union and Lethal Autonomous Weapons Systems: United in Diversity?' In *European Union Contested*, edited by Elisabeth Johansson-Nogués, Martin C. Vlaskamp & Esther Barbé, 133–52. Cham: Springer.

Barreiros, Daniel & Ítalo Barreto Poty. 2021. 'The US Strategy for Short-Term Military Artificial Intelligence Development (2020–2030).' *AUSTRAL: Brazilian Journal of Strategy & International Relations* 10(19): 199 –216. As of 28 March 2024: https://seer.ufrgs.br/index.php/austral/article/view/97219/63506

Bassett Cross, Rob, Harry Halem & Gabriel Elefteriu. 2024. 'AI Power and British Strategy.' Council on Geostrategy, Policy Paper, 22 March. As of 28 March 2024: https://www.geostrategy.org.uk/research/ai-power-and-british-strategy/

Baum, Seth D., Robert de Neufville, Anthony M. Barrett & Gary Ackermann. 2022. 'Lessons for Artificial Intelligence from Other Global Risks.' In *The Global Politics of Artificial Intelligence*, edited by Maurizio Tinnirello, 103–32. New York: Chapman and Hall/CRC.

Beauchamp-Mustafaga, Nathan. 2024. *Exploring the Implications of Generative AI for Chinese Military Cyber-Enabled Influence Operations: Chinese Military Strategies, Capabilities, and Intent*. Santa Monica, Calif.: RAND Corporation. CT-A3191-1. As of 28 March 2024: https://www.rand.org/pubs/testimonies/CTA3191-1.html

Bellas, Alexandria, Stefawn Perrin, Brandon Malone, Kaytlin Rogers, Gale Lucas, Elizabeth Phillips, Chad Tossell & Ewart de Visser. 2020. 'Rapport Building with Social Robots as a Method for Improving Mission Debriefing in Human-Robot Teams.' In *2020 Systems and Information Engineering Design Symposium (SIEDS): University of Virginia (Virtual Conference), Charlottesville, Virginia, USA, 24 April 2020*. doi: 10.1109/SIEDS49339.2020.9106643

Bendett, Samuel. 2023. 'Roles and Implications of AI in the Russian-Ukrainian Conflict.' Center for a New American Security (CNAS), 20 July. As of 28 March 2024: https://www.cnas.org/publications/commentary/roles-and-implications-of-ai-in-the-russian-ukrainian-conflict

Bitzinger, Richard A. & Michael Raska. 2022. 'Chinese and Russian Military Modernization and the Fourth Industrial Revolution.' In *Russia-China Relations: Emerging Alliance or Eternal Rivals?*, edited by S. Kirchberger, S. Sinjen & N. Wörmer, 121–40. Cham: Springer.

Black, James, Alice Lynch, Kristian Gustafson, David Blagden, Pauline Paille & Fiona Quimbre. 2022. *Multi-Domain Integration in Defence: Conceptual Approaches and Lessons from Russia, China, Iran and North Korea*. Santa Monica, Calif.: RAND Corporation. RR-A528-1. As of 28 March 2024: https://www.rand.org/pubs/research_reports/RRA528-1.html

Black, James, Diana Dascalu, Megan Hughes, Benedict Wilkinson, Maeve Ryan, Ahron Bregman, Peter Carlyon, Jennifer Cheung, Lawrence Freedman, Rebecca Lucas, Alessio Patalano, Patrick Porter, Fiona Quimbre, Sam Stockwell & Mann Virdee. 2023. *Strategic Advantage in a Competitive Age: Definitions, Dynamics and Implications*. Santa Monica, Calif.: RAND Corporation. RR-A1959-1. As of 28 March 2024: https://www.rand.org/pubs/research_reports/RRA1959-1.html

Black, James, Rebecca Lucas, John Kennedy, Megan Hughes & Harper Fine. 2024. *Command and Control in the Future: Concept Paper 1: Grappling with Complexity*. Santa Monica, Calif.: RAND Corporation. RR-A2476-1. As of 28 March 2024: https://www.rand.org/pubs/research_reports/RRA2476-1.html

Bode, Ingvild & Hendrik Huelss. 2018. 'Autonomous Weapons Systems and Changing Norms in International Relations.' *Review of International Studies* 44(3): 393–413. doi: 10.1017/s0260210517000614

———. 2023. 'Constructing expertise: the front- and back-door regulation of AI's military applications in the European Union.' *Journal of European Public Policy* 30(7): 1230–54. doi: 10.1080/13501763.2023.2174169

Bode, Ingvild, Hendrik Huelss, Anna Nadibaidze, Guangyu Qiao-Franco & Tom F. A. Watts. 2023. 'Prospects for the Global Governance of Autonomous Weapons: Comparing Chinese, Russian, and US practices.' *Ethics and Information Technology* 25(5). doi: 10.1007/s10676-023-09678-x

Boulanin, Vincent, S. M. Amadae, Shahar Avin, John Borrie, Justin Bronk, Martin Hagström, Michael C. Horowitz, Anja Kaspersen, Chris King, Jean-Marc Rickli, Frank Sauer, Dimitri Scheftelowitsch, Page O. Stoutland & Petr Topychkanov. 2019. *The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk, Volume I, Euro-Atlantic Perspectives*. Stockholm: SIPRI.

Brose, Christian. 2020. *The Kill Chain: Defending America in the Future of High-Tech Warfare*. New York: Hachette.

Burton, J. & S. R. Soare. 2019. 'Understanding the Strategic Implications of the Weaponization of Artificial Intelligence.' In *2019 11th International Conference on Cyber Conflict*, 1–17. doi: 10.23919/CYCON.2019.8756866

Butcher, James & Irakli Beridze. 2019. 'What is the State of Artificial Intelligence Governance Globally?' *The RUSI Journal* 164(5–6): 88–96. doi: 10.1080/03071847.2019.1694260

Büthe, Tim, Christian Djeffal, Christopher Lütge, Sabine Maasen & Nora von Ingersleben-Seip. 2022. 'Governing AI – attempting to herd cats? Introduction to the special issue on the Governance of Artificial Intelligence.' *Journal of European Public Policy* 29(11): 1721–52. doi: 10.1080/13501763.2022.2126515

Cabinet Office. 2023. 'National Risk Register 2023.' As of 28 March 2024: https://www.gov.uk/government/publications/national-risk-register-2023

Calderaro, Andrea & Stella Blumfelde. 2022. 'Artificial intelligence and EU security: the false promise of digital sovereignty.' *European Security* 31(3): 415–34. doi: 10.1080/09662839.2022.2101885

Castro, Daniel, Michael McLaughlin & Eline Chivot. 2019. *Who is Winning the AI Race: China, the EU or the United States?* Center for Data Innovation. As of 18 June 2024: https://datainnovation.org/2019/08/who-is-winning-the-ai-race-china-the-eu-or-the-united-states/

Castro, Daniel & Michael McLaughlin. 2021. *Who is Winning the AI Race: China, the EU or the United States?* Center for Data Innovation. As of 28 March 2024: https://www2.datainnovation.org/2021-china-eu-us-ai.pdf

Cave, Stephen & Sean Ó hÉigeartaigh. 2019. 'An AI Race for Strategic Advantage: Rhetoric and Risks.' *Ethics & Society*. As of 28 March 2024: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3446708

Caves, Ben, Rebecca Lucas, Livia Dewaele, Julia Muravska, Chris Wragg, Tom Spence, Zudik Hernandez, Anna Knack & James Black. 2021. *Enhancing Defence's Contribution to Societal Resilience: Lessons from International Approaches*. Santa Monica, Calif.: RAND Corporation. RR-A1113-1. As of 28 March 2024: https://www.rand.org/pubs/research_reports/RRA1113-1.html

Cho, Sungrim, Woochang Shin, Neunghoe Kim, Jongwook Jeong & Hoh Peter In. 2020. 'Priority Determination to Apply Artificial Intelligence Technology in Military Intelligence Areas.' *Electronics* 9(12): 2187. doi: 10.3390/electronics9122187

Cirincione, Greg, Tien Pham, Andrew Ladas, Brian Stanton & Gregory Fischer. 2019. 'Design and implementation of the U.S. Army Artificial Intelligence Innovation Institute.' Proc. SPIE 11006, Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications, 110060B, 10 May. As of 28 March 2024: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/11006/2524026/Design-and-implementation-of-the-US-Army-Artificial-Intelligence-Innovation/10.1117/12.2524026.full#_=_

Cox, Jessica & Heather Williams. 2021. 'The Unavoidable Technology: How Artificial Intelligence Can Strengthen Nuclear Stability.' *The Washington Quarterly* 44(1): 69–85. doi: 10.1080/0163660X.2021.1893019

Cozad, Mark, Jeffrey Engstrom, Scott W. Harold, Timothy R. Heath, Sale Lilly, Edmund J. Burke, Julia Brackup & Derek Grossman. 2023. *Gaining Victory in Systems Warfare: China's Perspective on the U.S.-China Military Balance*. Santa Monica, Calif.: RAND Corporation. RR-A1535-1. As of 31 March 2024: https://www.rand.org/pubs/research_reports/RRA1535-1.html

Criddle, Cristina & Eleanor Olcott. 2024. 'Chinese and Western Scientists Identify "Red Lines" on AI Risks.' *Financial Times*, 18 March.

Davis, Stephen I. 2022. 'Artificial Intelligence at the Operational Level of War.' *Defense and Security Analysis* 38(1): 74–90. doi: 10.1080/14751798.2022.2031692

Dear, Keith. 2019. 'Will Russia Rule the World Through AI? Assessing Putin's Rhetoric Against Russia's Reality.' *The RUSI Journal* 164(5–6): 36–60. doi: 10.1080/03071847.2019.1694227

Dexe, Jacob & Ulrik Franke. 2020. 'Nordic Lights? National AI Policies for Doing Well by Doing Good.' *Journal of Cyber Policy* 5(3): 332–49. doi: 10.1080/23728871.2020.1856160

Ding, Jeffrey. 2018. 'Deciphering China's AI Dream The context, components, capabilities, and consequences of China's strategy to lead the world in AI.' Future of Humanity Institute, Oxford University. As of 28 March 2024: https://www.fhi.ox.ac.uk/wp-content/uploads/Deciphering_Chinas_AI-Dream.pdf

Defence Innovation Board. 2019. 'AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense.' As of 28 March 2024: https://media.defense.gov/2019/Oct/31/2002204458/-1/-1/0/DIB_AI_PRINCIPLES_PRIMARY_DOCUMENT.PDF

Djeffal, Christian, Markus B. Siewert & Stefan Wurster. 2022. 'Role of the State and Responsibility in Governing Artificial Intelligence: A Comparative Analysis of AI Strategies.' *Journal of European Public Policy* 29(11): 1799–21. doi: 10.1080/13501763.2022.2094987

Dortmans, Peter, Joanne Nicholson, James Black, Marigold Black, Carl Rhodes, Scott Savitz, Linda Slapakova & Victoria M. Smith. 2021. *Supporting the Royal Australian Navy's Strategy for Robotics and Autonomous Systems: Building an Evidence Base*. RAND Corporation, RR-A929-1. As of June 13, 2024: https://www.rand.org/pubs/research_reports/RRA929-1.html

Dov Bachmann, Sascha-Dominik & Richard V. Grant. 2023. 'The Need for an Australian Regulatory Code for the Use of Artificial Intelligence (AI) in Military Application.' *National Security Law Brief* 13(2). As of 2 April 2024: https://digitalcommons.wcl.american.edu/nslb/vol13/iss2/2/

Egel, Daniel, Ryan Andrew Brown, Linda Robinson, Mary Kate Adgie, Jasmin Léveillé & Luke J. Matthews. 2022. *Leveraging Machine Learning for Operation Assessment*. Santa Monica, Calif.: RAND Corporation. RR-4196-A. As of 2 April 2024: https://www.rand.org/pubs/research_reports/RR4196.html

Enemark, Christian. 2005. 'United States Biodefense, International Law, and the Problem of Intent.' *Politics and the Life Sciences* 24(1–2): 32–42. doi:10.2990/1471-5457(2005)24[32:USBILA]2.0.CO;2

Engstrom, Jeffrey. 2018. *Systems Confrontation and System Destruction Warfare: How the Chinese People's Liberation Army Seeks to Wage Modern Warfare*. Santa Monica, Calif.: RAND Corporation. RR-1708-OSD. As of 28 March 2024: https://www.rand.org/pubs/research_reports/RR1708.html

Fischer, Sophie-Charlotte. 2022. 'Military AI Applications: A Cross-Country Comparison of Emerging Capabilities.' In *Armament, Arms Control and Artificial Intelligence. Studies in Peace and Security*, edited by T. Reinhold & N. Schörnig, 39–55. Springer, Cham. doi: 10.1007/978-3-031-11043-6_4

Fitzpatrick, Mark. 2019. 'Artificial Intelligence and Nuclear Command and Control.' *Global Politics and Strategy* 61(3): 81–92. doi: 10.1080/00396338.2019.1614782

Fossaceca, John M. & Stuart H. Young. 2018. 'Artificial Intelligence and Machine Learning for Future Army Applications.' In *Proceedings SPIE 10635, Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR IX, 4 May 2018*. doi: 10.1117/12.2307753

Fournier-Tombs, Eleonore. 2021. 'Towards a United Nations Internal Regulation for Artificial Intelligence.' *Big Data & Society* 8(2). doi: 10.1177/20539517211039493

Futter, Andrew. 2022. 'Disruptive Technologies and Nuclear Risks: What's New and What Matters.' *Global Politics and Strategy* 64(1): 99–120. doi: 10.1080/00396338.2022.2032979

Garcia, Denise. 2018. 'Lethal Artificial Intelligence and Change: The Future of International Peace and Security.' *International Studies Review* 20(2): 334–41. doi: 10.1093/isr/viy029

Geist, Edward. 2024. 'Building a Foundation for Strategic Stability with China on AI.' The RAND Blog, 2 April 2024. As of 2 April 2024: https://www.rand.org/pubs/commentary/2024/04/building-a-foundation-for-strategic-stability-with.html

Gerstein, Daniel M. & Erin N. Leidy. 2024. *Emerging Technology and Risk Analysis: Unmanned Aerial Systems Intelligent Swarm Technology*. Santa Monica, Calif.: RAND Corporation. RR-A2380-1. As of 2 April 2024: https://www.rand.org/pubs/research_reports/RRA2380-1.html

Gill, Amandeep Singh. 2019. 'Artificial Intelligence and International Security: The Long View.' *Ethics & International Affairs* 33(2): 169–79. doi: 10.1017/s0892679419000145

Goritiyal, Chandravadan & Laxmi Goritiyal. 2020. 'An Entrepreneurial Opportunity in Civil Aviation & Defence Aerospace Sector in India.' *Pacific Business Review International* 12(12): 126–32. As of 2 April 2024: http://www.pbr.co.in/2020/2020_month/June/13.pdf

Gray, Maggie & Amy Ertan. 2021. 'Artificial Intelligence and Autonomy in the Military: An Overview of NATO Member States' Strategies and Deployment.' CCDCOE, January. As of 2 April 2024: https://ccdcoe.org/library/publications/artificial-intelligence-and-autonomy-in-the-military-an-overview-of-nato-member-states-strategies-and-deployment/

Guenduez, Ali Asker & Tobias Mettler. 2023. 'Strategically constructed narratives on artificial intelligence: What stories are told in governmental artificial intelligence policies?' *Government Information Quarterly* 40(6): 1–13. doi: 10.1016/j.giq.2022.101719

Haner, Justin & Denise Garcia. 2019. 'The Artificial Intelligence Arms Race: Trends and World Leaders in Autonomous Weapons Development.' *Global Policy* 10: 331–37. doi: 10.1111/1758-5899.12713

Heath, Timothy R., Clint Reach & Michael J. Mazarr. 2024. *The Societal Basis for National Competitiveness: Chinese and Russian Perspectives*. Santa Monica, Calif.: RAND Corporation. RR-A2611-1. As of 28 March 2024: https://www.rand.org/pubs/research_reports/RRA2611-1.html

Hicks, Kathleen. 2021. 'Implementing Responsible Artificial Intelligence in the Department of Defense.' US Department of Defence. As of 28 March 2024: https://media.defense.gov/2021/May/27/2002730593/-1/-1/0/IMPLEMENTING-RESPONSIBLE-ARTIFICIAL-INTELLIGENCE-IN-THE-DEPARTMENT-OF-DEFENSE.PDF

Hoadley, Daniel S. & Nathan J. Lucas. 2018. *Artificial Intelligence and National Security*. Congressional Research Service. As of 28 March 2024: https://digital.library.unt.edu/ark:/67531/metadc1157028/

Hornung, Jeffrey W., Scott Savitz, Jonathan Balk, Samantha McBirney, Liam McLane, and Victoria M. Smith. 2021. *Preparing Japan's Multi-Domain Defense Force for the Future Battlespace Using Emerging Technologies*. RAND Corporation. PE-A1157-1. As of 22 March 2024: https://www.rand.org/pubs/perspectives/PEA1157-1.html

Horowitz, Michael C. 2018. 'Artificial Intelligence, International Competition, and the Balance of Power.' *Texas National Security Review* 1(3): 36–57. doi: 10.15781/T2639KP49

———. 2019. 'When Speed Kills: Lethal Autonomous Weapons Systems, Deterrence and Stability.' *Journal of Strategic Studies* 42(6): 764–8. doi: 10.1080/01402390.2019.1621174

Horowitz, Michael C. & Lauren Kahn. 2023. 'Bending the Automation Bias Curve: A Study of Human and AI-based Decision Making in National Security Contexts.' *arXiv*. doi: 10.48550/arXiv.2306.16507

Horowitz, Michael C., Lauren Kahn & Casey Mahoney. 2020. 'The Future of Military Applications of Artificial Intelligence: A Role for Confidence-Building Measures?' *Orbis* 64(4): 528–43. doi: 10.1016/j.orbis.2020.08.003

Horowitz, Michael C., Shira Pindyck & Casey Mahoney. 2022. 'AI, the International Balance of Power, and National Security Strategy.' In *The Oxford Handbook of AI Governance*, edited by Justin B. Bullock, Yu-Che Chen, Johannes Himmelreich, Valeri M. Hudson, Anton Korinek, Matthew M. Young & Baobao Zhang, 914–36. New York: Oxford University Press.

Horowitz, Michael C. & Paul Scharre. 2021. 'AI and International Stability: Risks and Confidence-Building Measures.' Center for a New American Security, 12 January. As of 28 March 2024: https://www.cnas.org/publications/reports/ai-and-international-stability-risks-and-confidence-building-measures

Horowitz, Michael C., Paul Scharre & Alexander Velez-Green. 2019. 'A Stable Nuclear Future? The Impact of Autonomous Systems and Artificial Intelligence.' *arXiv*. doi: 10.48550/arXiv.1912.05291

Hou, Yong, Zhiqiang Wang & Erning Zhai. 2023. 'Artificial Intelligence Technology Pushes Forward the Modernization of Firepower Weapon Equipment.' Proc. SPIE 12720, 2022 Workshop on Electronics Communication Engineering, 127200F, 28 June. doi: 10.1117/12.2668167

Hughes, Megan, Richard Carter, Amy Harland & Alexander Babuta. 2024. 'AI and Strategic Decision Making: Communicating Trust and Uncertainty in AI-Enriched Intelligence.' Centre for Emerging Technology and Security, The Alan Turing Institute. As of 26 April 2024: https://cetas.turing.ac.uk/publications/ai-and-strategic-decision-making

Hummel, Kristina & Paul Cruickshank. 2022. 'Special Issue: The Biological Threat – Part One.' *CTC Sentinel* 15(4). As of 28 March 2024: https://ctc.westpoint.edu/wp-content/uploads/2022/04/CTC-SENTINEL-042022.pdf

Hunter, Lance Y., Craig D. Albert, Christopher Henningan & Josh Rutland. 2023. 'The Military Application of Artificial Intelligence Technology in the United States, China, and Russia and the Implications for Global Security.' *Defense & Security Analysis* 39(2): 207–32. doi: 10.1080/14751798.2023.2210367

Hunter Christie, Edward. 2022. 'Defence Cooperation in Artificial Intelligence: Bridging the Transatlantic Gap for a Stronger Europe.' *European View* 21(1): 13-21. doi: 10.1177/17816858221089372

Ish, Daniel, Jared Ettinger & Christopher Ferris. 2021. *Evaluating the Effectiveness of Artificial Intelligence Systems in Intelligence Analysis*. Santa Monica, Calif.: RAND Corporation. RR-A464-1. As of 28 March 2024: https://www.rand.org/pubs/research_reports/RRA464-1.html

Jacobsen, Jeppe T. & Katrine Nørgaard. 2024. 'Reading Security Imaginaries as Fantasies – Loss, Desire, and Enjoyment in the Military Quest for Explainable AI.' *Millennium: Journal of International Studies*: 1–25. doi: 10.1177/03058298231225753

Janjeva, Ardi, Nikhil Mulani, Rosamund Powell, Jess Whittlestone & Shahar Avin. 2023. 'Strengthening Resilience to AI Risk: A Guide for UK policymakers.' CETaS Briefing Paper. As of 28 March 2024: https://cetas.turing.ac.uk/publications/strengthening-resilience-ai-risk

Javadi, Mahmoud & Michal Ondero. 2024. 'What Does Global Military AI Governance Need?' European Leadership Network, 2 February. As of 28 March 2024: https://www.europeanleadershipnetwork.org/commentary/what-does-global-military-ai-governance-need/

Jensen, Benjamin M., Christopher Whyte & Scott Cuomo. 'Algorithms at War: The Promise, Peril, and Limits of Artificial Intelligence.' *International Studies Review* 22(3): 526–50. doi: 10.1093/isr/viz025

Johnson, James. 2019a. 'Artificial Intelligence & Future Warfare: Implications for International Security.' *Defense & Security Analysis* 35(2): 147–69. doi: 10.1080/14751798.2019.1600800

———. 2019b. 'The AI-Cyber Nexus: Implications for Military Escalation, Deterrence and Strategic Stability.' *Journal of Cyber Policy* 4(3): 442–60. doi: 10.1080/23738871.2019.1701693

———. 2019c. 'The End of Military-Techno *Pax Americana*? Washington's Strategic Responses to Chinese AI-Enabled Military Technology.' *The Pacific Review* 34(3): 351–78. doi: 10.1080/09512748.2019.1676299

———. 2020a. 'Artificial Intelligence: A Threat to Strategic Stability.' *Strategic Studies Quarterly* 14(1): 16–39. As of 28 March 2024: https://www.jstor.org/stable/26891882

———. 2020b. 'Artificial Intelligence in Nuclear Warfare: A Perfect Storm of Instability?' *The Washington Quarterly* 43(2): 197–211. doi: 10.1080/0163660X.2020.1770968

———. 2020c. 'Delegating Strategic Decision-Making to Machines: Dr. Strangelove Redux?' *Journal of Strategic Studies* 45(3): 439–77. doi: 10.1080/01402390.2020.1759038

———. 2021a. '"Catalytic Nuclear War" in the Age of Artificial Intelligence & Autonomy: Emerging Military Technology and Escalation Risk Between Nuclear-Armed States.' *Journal of Strategic Studies*. doi: 10.1080/01402390.2020.1867541

———. 2021b. 'Deterrence in the Age of Artificial Intelligence & Autonomy: A Paradigm Shift in Nuclear Deterrence Theory and Practice?' *Defense & Security Analysis* 36(4): 422–48. doi: 10.1080/14751798.2020.1857911

———. 2022. 'Inadvertent Escalation in the Age of Intelligence Machines: A New Model for Nuclear Risk in the Digital Age.' *European Journal of International Security* 7(3): 337–59. doi: 10.1017/eis.2021.23

Jouan, Nicholas, Ondrej Palicka & James Black. 2024. 'Developing AI Capacity and Expertise in UK Defence.' As of 18 June 2024: https://committees.parliament.uk/writtenevidence/127769/html

Jyothi, A. P., Anirudh Shankar, Ashwath J. R. Narayan, Kanyadara Bhavya, Svv Sai Madhu Sudhan Reddy & A. Yashwanth. 2022. 'AI Methodologies in Upcoming Modern Warfare Systems.' *2022 IEEE International Conference on Current Development in Engineering and Technology (CCET)*. doi: 10.1109/CCET56606.2022.10080014

Kania, Elsa B. 2019. 'Chinese Military Innovation in the AI Revolution.' *The RUSI Journal* 164(5–6): 26–34. doi: 10.1080/03071847.2019.1693803

Kaspar, Lea, Maria Paz Canales & Michaela Nakayama Shapiro. 2023. 'Navigating the Global AI Governance Landscape.' Global Partners Digital, 31 October. As of 28 March 2024: https://www.gp-digital.org/navigating-the-global-ai-governance-landscape/

Kim, Hyunsoo. 2022. 'Suggestions for the Role of AI in the Arms Control and Non-Proliferation of WMD.' *Robotics & AI Ethics* 7(2): 57–67. doi 10.22471/ai.2022.7.2.57

Kim, Jongheon. 2023. 'Traveling AI-Essentialism and National AI Strategies: A Comparison Between South Korea and France.' *Review of Policy Research* 40(5): 705–28. doi: 10.1111/ropr.12552

Konaev, Margarita, Husanjot Chahal, Ryan Fedasiuk, Tina Huang & Ilya Rahkovsky. 2020. *U.S. Military Investments in Autonomy and AI*. CSET Policy Brief. doi: 10.51593/20190044

Levesques, Antoine. 2024. 'Early Steps in India's Use of AI for Defence.' IISS, 18 January. As of 2 April 2024: https://www.iiss.org/online-analysis/online-analysis/2024/01/early-steps-in-indias-use-of-ai-for-defence/

Lee, Chung Min, Jung Ku-hyun, Lee Jung-eun, Hana Anderson, Jacob Feldgoise & Juhern Kim. 2022. 'How South Korea Is Honing a Competitive Edge.' Carnegie Endowment for International Peace, 22 November. As of 18 June 2024: https://carnegieendowment.org/research/2022/11/how-south-korea-is-honing-a-competitive-edge?lang=en&center=europe

Lewis, Larry. 2017. 'Insights for the Third Offset: Addressing Challenges of Autonomy and Artificial Intelligence in Military Operations.' CNA, September. As of 2 April 2024: https://www.cna.org/reports/2017/insights-for-the-third-offset

Liebig, Laura, Licinia Güttel, Anna Jobin & Christian Katzenbach. 2022. 'Subnational AI policy: shaping AI in a multi-level governance system.' *AI & Society*. doi: 10.1007/s00146-022-01561-5

Liu, Hin-Yan & Matthijs M. Maas. 2021. '"Solving for X?" Towards a Problem-Finding Framework to Ground Long-Term Governance Strategies for Artificial Intelligence.' *Futures* 126(22). doi: 10.1016/j.futures.2020.102672

Longpre, Shayne, Marcus Storm & Rishi Shah. 2022. 'Lethal Autonomous Weapons Systems & Artificial Intelligence: Trends, Challenges, and Policies.' MIT Science Policy Review, 29 August. As of 2 April 2024: https://sciencepolicyreview.org/2022/07/mitspr-191618003019/

Lowther, Adam & Curtis McGiffin. 2024. 'America Needs a Dead Hand More Than Ever.' War on the Rocks, 28 March. As of 28 March 2024: https://warontherocks.com/2024/03/america-needs-a-dead-hand-more-than-ever/

Lucas, Rebecca, Conlan Ellis, James Black, Peter Carlyon, Paul Kendall, John Kendall, Stephen Coulson, and Louis Jeffries. 2024. *Command and Control in the Future: Concept Paper 2: The Defence C2 Enterprise.* Santa Monica, Calif.: RAND Corporation. RR-A2476-2. As of June 18, 2024: https://www.rand.org/pubs/research_reports/RRA2476-2.html

Luo, Shuxian. 2022. 'Addressing military AI risks in U.S.-China crisis management mechanisms.' *China International Strategy Review* 4: 233–47. doi: 10.1007/s42533-022-00110-5

Maas, Matthijs M. 2019. 'How viable is international arms control for military artificial intelligence? Three lessons from nuclear weapons.' *Contemporary Security Policy* 40(3): 285–311. doi: 10.1080/13523260.2019.1576464

Maas, Matthijs M., Kayla Matteucci & Di Cooke. 2023. 'Military Artificial Intelligence as Contributor to Global Catastrophic Risk.' In *The Era of Global Risk*, edited by SJ Beard, Martin Rees, Catherine Richards & Clarissa Rios-Rojas, 237–84. Cambridge, UK: Open Book Publishers. doi: 10.2139/ssrn.4115010

Marcinek, Krystyna & Eugeniu Han. 2023. *Russia's Asymmetric Response to 21st Century Strategic Competition: Robotization of the Armed Forces*. Santa Monica, Calif.: RAND Corporation. RR-A1233-5. As of 28 March 2024: https://www.rand.org/pubs/research_reports/RRA1233-5.html

Mazarr, Michael J. 2022. *The Societal Foundations of National Competitiveness*. Santa Monica, Calif.: RAND Corporation. RR-A499-1. As of 28 March 2024: https://www.rand.org/pubs/research_reports/RRA499-1.html

Mazarr, Michael J., Jonathan S. Blake, Abigail Casey, Tim McDonald, Stephanie Pezard, and Michael Spirtas. 2018. *Understanding the Emerging Era of International Competition: Theoretical and Historical Perspectives.* Santa Monica, Calif.: RAND Corporation. RR-2726-AF. As of June 18, 2024: https://www.rand.org/pubs/research_reports/RR2726.html

Mazarr, Michael J., Samuel Charap, Abigail Casey, Irina A. Chindea, Christian Curriden, Alyssa Demus, Bryan Frederick, Arthur Chan, John P. Godges, Eugeniu Han et al. 2021. *Stabilizing Great-Power Rivalries*. Santa Monica, Calif.: RAND Corporation. RR-A456-1. As of 28 March 2024: https://www.rand.org/pubs/research_reports/RRA456-1.html

Mazarr, Michael J., Alexis Dale-Huang & Matthew Sargent. 2024. *The Emerging Competitive Paradigm: A Contest of Effective Governance*. Santa Monica, Calif.: RAND Corporation. PE-A2611-1. As of 28 March 2024: https://www.rand.org/pubs/perspectives/PEA2611-1.html

Mazarr, Michael J., Bryan Frederick & Yvonne K. Crane. 2022. *Understanding a New Era of Strategic Competition*. Santa Monica, Calif.: RAND Corporation. RR-A290-4. As of 28 March 2024: https://www.rand.org/pubs/research_reports/RRA290-4.html

Mazarr, Michael & Tim McDonald. 2022. *Competing for the System: The Essence of Emerging Strategic Rivalries*. Santa Monica, Calif.: RAND Corporation. PE-A1404-2. As of 28 March 2024: https://www.rand.org/pubs/perspectives/PEA1404-2.html

Mazarr, Michael J., Ashley L. Rhoades, Nathan Beauchamp-Mustafaga, Alexis A. Blanc, Derek Eaton, Katie Feistel, Edward Geist, Timothy R. Heath, Christian Johnson, Krista Langeland et al. 2022. *Disrupting Deterrence: Examining the Effects of Technologies on Strategic Deterrence in the 21st Century*. Santa Monica, Calif.: RAND Corporation. RR-A595-1. As of 28 March 2024: https://www.rand.org/pubs/research_reports/RRA595-1.html

McCoy, Kelly. 2018. 'In the Beginning, There Was Competition: The Old Idea Behind the New American Way of War.' *Modern War Institute,* 11 April. As of 22 December 2023: https://mwi.usma.edu/beginning-competition-old-idea-behind-new-american-way-war/

McFarland, Tim. 2022. 'Reconciling Trust and Control in the Military Use of Artificial Intelligence.' *International Journal of Law and Information Technology* 30(4): 472–83. doi: 10.1093/ijlit/eaad008

Meerveld, H.W., R.H.A. Lindelauf, E.O. Postma & M. Postma. 2023. 'The Irresponsibility of Not Using AI in the Military.' *Ethics and Information Technology* 25(14). doi: 10.1007/s10676-023-09683-0

Menthe, Lance, Dahlia Anne Goldfeld, Annie Tingstad, Sherrill Lingel, Edward Geist, Donlad Brunk, Amanda Wicker, Sarah Lovell, Balys Gintautas, Anne Stickells et al. 2021. *Technology Innovation and the Future of Air Force Intelligence Analysis: Volume 2, Technical Analysis and Supporting Material*. Santa Monica, Calif.: RAND Corporation. RR-A341-2. As of 2 April 2024: https://www.rand.org/pubs/research_reports/RRA341-2.html

Metcalf, Mark. 2022. 'The PRC Considers Military AI Ethics: Can Autonomy be Trusted?' *Frontier in Big Data* 5(991392). doi: 10.3389/fdata.2022.991392

Mhajne, Anwar. 2023. 'Israel's AI Revolution: From Innovation to Occupation.' *Sada*, Carnegie Endowment for International Peace, 2 November. As of 28 March 2024: https://carnegieendowment.org/sada/90892

Miller, Rodney, John Mehrman & Mike Marlow. 2010. 'Risk Management Challenges of Multi-Payload Launch Missions Executed by the DoD Space Test Program.' *2020 IEEE Aerospace Conference Proceedings*. doi: 10.1109/AERO.2010.5446864

Ministry of Innovation, Science & Technology. 2023. 'Israel's Policy on Artificial Intelligence – Regulations and Ethics.' As of 28 March 2024: https://www.gov.il/BlobFolder/policy/ai_2023/en/Israels%20AI%20Policy%202023.pdf

Minkkinen, Matti & Matti Mäntymäki. 2023. 'Discerning Between the "Easy" and "Hard" Problems of AI Governance.' *IEEE Transactions on Technology and Society* 4(2): 188–94. doi: 10.1109/TTS.2023.3267382

Mori, Satoru. 2018. 'US Defense Innovation and Artificial Intelligence.' *Asia-Pacific Review* 25(2): 16–44. doi: 10.1080/13439006.2018.1545488

Mouton, Christopher, Caleb Lucs & Ella Guest. 2023. *The Operational Risks of AI in Large-Scale Biological Attacks: A Red Team Approach.* Santa Monica, Calif.: RAND Corporation. RR-A2977-1. As of 28 March 2024: https://www.rand.org/pubs/research_reports/RRA2977-1.html

———. 2024. *The Operational Risks of AI in Large-Scale Biological Attacks: The Results of a Red Team Study.* Santa Monica, Calif.: RAND Corporation. RR-A2977-2. As of 28 March 2024: https://www.rand.org/pubs/research_reports/RRA2977-2.html

Nadibaidze, Anna. 2022. 'Great Power Identity in Russia's Position on Autonomous Weapons Systems.' *Contemporary Security Policy* 43(3): 407–35. doi: 10.1080/13523260.2022.2075665

———. 2024. ‘Technology in the Quest for Status: The Russian Leadership’s Artificial Intelligence Narrative.’ *Journal of International Relations and Development*. doi: 10.1057/s41268-023-00322-1

Nadibaidze, Anna & Nicolò Miotto. 2023. ‘The Impact of AI on Strategic Stability is What States Make of It: Comparing US and Russian Discourses.’ *Journal for Peace and Nuclear Disarmament* 6(1): 47–67. doi: 10.1080/25751654.2023.2205552

Nelson, Cassidy & Sophie Rose. 2023. *AI-Facilitated Biological Weapon Development*. Centre for Long-Term Resilience. As of 28 March 2024: https://www.longtermresilience.org/post/report-launch-examining-risks-at-the-intersection-of-ai-and-bio

New, Joshua. 2018. ‘Why the United States Needs a National Artificial Intelligence Strategy and What It Should Look Like.’ Center for Data Innovation, 4 December. As of 2 April 2024: https://datainnovation.org/2018/12/why-the-united-states-needs-a-national-artificial-intelligence-strategy-and-what-it-should-look-like/

Newman, Marissa. 2023. ‘Israel Quietly Embeds AI Systems in Deadly Military Operations.’ *Bloomberg*, 16 July. As of 2 April 2024: https://www.bloomberg.com/news/articles/2023-07-16/israel-using-ai-systems-to-plan-deadly-military-operations

———. 2024. ‘Thiel’s Palantir, Israel Agree Strategic Partnership for Battle Tech.’ *Bloomberg*, 12 January. As of 2 April 2024: https://www.bloomberg.com/news/articles/2024-01-12/palantir-israel-agree-to-strategic-partnership-for-battle-tech

Nurkin, Tate & Ryo Hinata-Yamaguchi. 2020. ‘Emerging Technologies and the Future of US-Japan Defense Collaboration.’ Atlantic Council, 17 April. As of 2 April 2024: https://www.atlanticcouncil.org/uncategorized/emerging-defense-technologies-and-the-future-of-us-japan-defense-collaboration/

Onderco, Michal & Madeline Zutt. 2021. ‘Emerging Technology and Nuclear Security: What Does the Wisdom of the Crowd Tell Us?’ *Contemporary Security Policy* 42(3): 286–311. doi: 10.1080/13523260.2021.1928963

Otani, Yasuo & Naohiko Kohtake. 2019. ‘Applicability of Civil and Defense Dual Use to Space Situational Awareness System in Japan.’ *Space Policy* 47: 140–47. doi: 10.1016/j.spacepol.2018.11.001

Parker, Edward. 2024. *The Chinese Industrial Base and Military Deployment of Quantum Technology*. Santa Monica, Calif.: RAND Corporation. CT-A3189-1. As of 2 April 2024: https://www.rand.org/pubs/testimonies/CTA3189-1.html

Pavel, Barry, Ivana Ke, Michael Spirtas, James Ryseff, Lea Sabbag, Gregory Smith, Keller Scholl & Domenique Lumpkin. 2023. ‘AI and Geopolitics: How Might AI Affect the Rise and Fall of Nations?’ RAND Expert Insights, 3 November. As of 28 March 2024: https://www.rand.org/pubs/perspectives/PEA3034-1.html

Payne, Kenneth. 2018. ‘Artificial Intelligence: A Revolution in Strategic Affairs?’ *Global Politics and Strategy* 60(5): 7–32. doi: 10.1080/00396338.2018.1518374

———. 2021. *I, Warbot*. London, UK: C Hurst & Co.

———. 2024. ‘Warbot 2.0: Reflections on the Fast-Changing World of AI in National Security.’ Binding Hook, 28 March. As of 28 March 2024: https://bindinghook.com/articles-book-binder/warbot-2-0-reflections-on-the-fast-changing-world-of-ai-in-national-security/

Puscas, Ioana. 2022. 'Confidence-Building Measures for Artificial Intelligence.' United Nations Institute for Disarmament Research. As of 28 March 2024: https://unidir.org/wp-content/uploads/2023/05/Confidence-Building_Final.pdf

Puwal, Steffan. 2024. 'Should Artificial Intelligence be Banned from Nuclear Weapons Systems?' *NATO Review*. As of 12 April 2024: https://www.nato.int/docu/review/articles/2024/04/12/should-artificial-intelligence-be-banned-from-nuclear-weapons-systems/index.html

Qi, Haotian. 2021. '"Smart" Warfare and China-U.S. Stability: Strengths, Myths, and Risks.' *China International Strategy Review* 3(2): 278–99. doi: 10.1007/s42533-021-00094-8

Radu, Roxana. 2021. 'Steering the Governance of Artificial Intelligence: National Strategies in Perspective.' *Policy & Society* 40(2): 178–93. doi: 10.1080/14494035.2021.1929728

Ray, Trisha. 2018. *Beyond the 'Lethal' in Lethal Autonomous Weapons: Applications of LAWS in Theatres of Conflict for Middle Powers*. Observer Research Foundation, Occasional Paper No. 180. As of 7 June 2024: https://www.orfonline.org/research/beyond-the-lethal-in-lethal-autonomous-weapons-applications-of-laws-in-theatres-of-conflict-for-middle-powers

Reinhold, Thomas & Christian Reuter. 2022. 'Cyber Weapons and Artificial Intelligence: Impact, Influence and the Challenges for Arms Control.' In *Armament, Arms Control and Artificial Intelligence: Studies in Peace and Security*, edited by Thomas Reinhold & N. Schörnig, 145–58. Cham: Springer. doi: 10.1007/978-3-031-11043-6_11

Retter, Lucia & Stuart Dee. 2024. 'Pace Through Integration? UK Defence Attempts Procurement Reform, Again.' The RAND Blog, 20 March. As of 28 March 2024: https://www.rand.org/pubs/commentary/2024/03/pace-through-integration-uk-defence-attempts-procurement.html

Retter, Lucia, Julia Muravska, Ben Williams & James Black. 2021. *Persistent Challenges in UK Defence Equipment Acquisition*. Santa Monica, Calif.: RAND Corporation. RR-A1174-1. As of 28 March 2024: https://www.rand.org/pubs/research_reports/RRA1174-1.html

Roberts, Guy B. 2003. *Arms Control without Arms Control: The Failure of the Biological Weapons Convention Protocol and a New Paradigm for Fighting the Threat of Biological Weapons*. Colorado: USAF Institute for National Security Studies. As of 28 March 2024: https://archive.org/details/DTIC_ADA435071/mode/1up

Roberts, Huw, Alexander Babuta, Jessica Morley, Christopher Thomas, Mariarosaria Taddeo & Luciano Floridi. 2023. 'Artificial Intelligence Regulation in the United Kingdom: A Path to Good Governance and Global Leadership?' *Internet Policy Review* 12(2). doi: 10.14763/2023.2.1709

Robles, Pedro & Daniel J. Mallinson. 2023. 'Catching Up With AI: Pushing Toward a Cohesive Governance Framework.' *Politics & Policy* 51(3): 355–72. doi: 10.1111/polp.12529

Roff, Heather M. 2019. 'The Frame Problem: The AI "Arms Race" isn't one.' *Bulletin of the Atomic Scientists* 75(3): 95–98. doi: 10.1080/00963402.2019.1604836

Rossiter, Ash. 2021. 'AI-Enabled Remote Warfare: Sustaining the Western Warfare Paradigm?' *International Politics* 60: 818–33. doi: 10.1057/s41311-021-00337-w

Ryseff, James, Eric Landree, Noah Johnson, Bonnie Ghosh-Dastidar, Max Izenberg, Sydne J. Newberry, Christopher Ferris & Melissa A. Bradley. 2022. *Exploring the Civil-Military Divide over Artificial Intelligence*. Santa Monica, Calif.: RAND Corporation. RR-A1498-1. As of 28 March 2024: https://www.rand.org/pubs/research_reports/RRA1498-1.html

Sajduk, Błażej. 2019. 'Theoretical Premises of the Impact of Artificial Intelligence on International Relations and Security.' *The Copernicus Journal of Political Studies* 2: 157–80. doi: 10.12775/CJPS.2019.017

Schaefer K.E., Perelman B., Rexwinkle J., Canady J., Neubauer C., Waytowich N., Larkin G., Cox K., Geuss M., Gremillion G., Metcalfe J.S., DeCostanza A. & Marathe A. 2021. 'Human-Autonomy Teaming for the Tactical Edge: The Importance of Humans in Artificial Intelligence Research and Development.' In *Systems Engineering and Artificial Intelligence*, edited by William F. Flawless, Ranjeev Mittu, Donald A. Sofge, Thomas Shortell & Thomas A. McDormott, 115–48. doi: 10.1007/978-3-030-77283-3_7

Scharre, Paul. 2018. *Army of None: Autonomous Weapons and the Future of War*. W.W. Norton & Co.

———. 2023. *Four Battlegrounds: Power in the Age of Artificial Intelligence*. W.W. Norton & Co.

———. 2024. 'The Perilous Coming Age of AI Warfare.' Center for a New American Security. 29 February. As of 28 March 2024: https://www.cnas.org/publications/commentary/the-perilous-coming-age-of-ai-warfare

Scharre, Paul & Megan Lamberth. 2022. 'Artificial Intelligence and Arms Control.' *arXiv*, 22 October. doi: 10.48550/arXiv.2211.00065

Schmidt, Eric. 2022. 'AI, Great Power Competition & National Security.' *Daedalus* 151(2): 288–98. doi: 10.1162/daed_a_01916

Schmitt, Lewin. 2022. 'Mapping Global AI Governance: A Nascent Regime in a Fragmented Landscape.' *AI & Ethics* 2: 303–14. doi: 10.1007/s43681-021-00083-y

Secretariat of Science, Technology and Innovation Policy (Cabinet Office, Government of Japan). 2022. *AI Strategy 2022 (tentative translation of its overview)*. As of 2 April 2024: https://www8.cao.go.jp/cstp/ai/aistratagy2022en_ov.pdf

Secretary of the Air Force Public Affairs. 2023. 'Japan MoD, US DoD Sign Joint Agreement for AI, UAS Research.' 22 December. As of 2 April 2024: https://www.af.mil/News/Article-Display/Article/3624158/japan-mod-us-dod-sign-joint-agreement-for-ai-uas-research/

Shahi, Megan & Adam Conner. 2023. 'Priorities for a National AI Strategy.' Center for American Progress, 10 August 2023. As of 28 March 2024: https://www.americanprogress.org/wp-content/uploads/sites/2/2023/08/AIpriorities-PB.pdf

Sharma, Sanur. 2023. 'Trustworthy Artificial Intelligence: Design of AI Governance Framework.' *Strategic Analysis* 47(5): 443–64. doi: 10.1080/09700161.2023.2288994

Shasha, Yu & Fiona Carroll. 2021. 'Implications of AI in National Security: Understanding the Security Issues and Ethical Challenges.' In *Artificial Intelligence in Cyber Security: Impact and Implications. Security Challenges, Technical and Ethical Issues, Forensic Investigative Challenges*, edited by Reza Montasari & Hamid Jahankhani, 157 75. Cham: Springer. doi: 10.1007/978-3-030-88040-8_6

Shen, Yang, Xiang Ye & Di Zhai. 2023. 'Research and Application of Intelligent Technology for Preventing Human Error in Nuclear Power Plant.' In *Engineering Psychology and Cognitive Ergonomics, HCII 2023, Lecture Notes in Computer Science 14018*, edited by D. Harris & WC. Li, 345–59. doi: 10.1007/978-3-031-35389-5_24

Sigfrids, Anton, Jaan Leikas, Henrikki Salo-Pöntinen & Emmi Koskimies. 2023. 'Human-Centricity in AI Governance: A Systemic Approach.' *Frontiers in Artificial Intelligence* 6. doi: 10.3389/frai.2023.976887

Slapakova, Linda, Abby Fraser, Megan Hughes, Maria Chiara Aquilino & Kristin Thue. 2024. *Cultural and Technological Change in the Future Information Environment*. Santa Monica: RAND Corporation. RR-2662-1. As of 28 March 2024: https://www.rand.org/pubs/research_reports/RRA2662-1.html

Slapakova, Linda, Paola Fusaro, James Black & Peter Dortmans. 2022. *Supporting the Royal Australian Navy's Campaign Plan for Robotics and Autonomous Systems: Emerging Missions and Technology Trends*. Santa Monica, Calif.: RAND Corporation. RR-A1377-1. As of 28 March 2024: https://www.rand.org/pubs/research_reports/RRA1377-1.html

Stanley-Lockman, Zoe. 2021a. *Military AI Cooperation Toolbox: Modernizing Defense Science and Technology Partnerships for the Digital Age*. CSET Issue Brief. As of 28 March 2024: https://indianstrategicknowledgeonline.com/web/Military%20AI%20Cooperation.pdf

———. 2021b. *Responsible and Ethical Military AI Allies and Allied Perspectives*. CSET Issue Brief. As of 28 March 2024: https://cset.georgetown.edu/wp-content/uploads/CSET-Responsible-and-Ethical-Military-AI.pdf

Steel, J. Jordan, Katherine L. Bates & Michael D. Barnhart. 2019. 'Investing in our Nation's Future Military Leaders' Synthetic Biology Knowledge to Understand and Recognize Threats and Applications.' *Synthetic Biology* 4(1). doi: 10.1093/synbio/ysz024

Sylvia, Noah. 2024. 'Israel's Targeting AI: How Capable is it?' RUSI, 8 February. As of 2 April 2024: https://www.rusi.org/explore-our-research/publications/commentary/israels-targeting-ai-how-capable-it

Taddeo, Mariarosaria and Luciano Floridi. 2018. 'Regulate Artificial Intelligence to Avert Cyber Arms Race.' *Nature* 556: 296–98. doi: 10.1038/d41586-018-04602-6

Taeihagh, Araz. 2021. 'Governance of Artificial Intelligence.' *Policy & Society* 40(2): 137–57. doi: 10.1080/14494035.2021.1928377

Tallberg, Jonas, Eva Erman, Markus Furendal, Johannes Geith, Mark Klamberg & Magnus Lundgren. 2023. 'The Global Governance of Artificial Intelligence: Next Steps for Empirical and Normative Research.' *International Studies Review* 25(3). doi: 10.1093/isr/viad040

Tarraf, Danielle C., William Shelton, Edward Parker, Brien Alkire, Diana Gehlhaus, Justin Grana, Alexis Levedahl, Jasmin Léveillé, Jared Mondschein, James Ryseff et al. 2019. *The Department of Defense Posture for Artificial Intelligence: Assessment and Recommendations*. Santa Monica, Calif.: RAND Corporation. RR-4229-OSD. As of 28 March 2024: https://www.rand.org/pubs/research_reports/RR4229.html

Tegmark, Max. 2017. *Life 3.0: Being Human in the Age of Artificial Intelligence*. London, UK: Allen Lane.

Thornton, Rod & Marina Miron. 2020. 'Towards the "Third Revolution in Military Affairs".' *The RUSI Journal*, 165(3): 12-21. doi: 10.1080/03071847.2020.1765514

Tokariuk, Olga. 2023. 'Ukraine's Secret Weapon – Artificial Intelligence.' Center for European Policy Analysis, 20 November. As of 28 March 2024: https://cepa.org/article/ukraines-secret-weapon-artificial-intelligence/

UK Government. 2021. 'National AI Strategy.' 22 September. As of 28 March 2024: https://www.gov.uk/government/publications/national-ai-strategy

UK Ministry of Defence. 2020. 'Integrated Operating Concept.' 30 September. As of 28 March 2024: https://www.gov.uk/government/publications/the-integrated-operating-concept-2025

———. 2022. 'Defence Artificial Intelligence Strategy'. 15 June. As of 2 April 2024: https://www.gov.uk/government/publications/defence-artificial-intelligence-strategy

U.S. Department of Defense (US DoD). 2018. 'Summary of the 2018 Department of Defence Artificial Intelligence Strategy – Harnessing AI to Advance Our Security and Prosperity.' As of 28 March 2024: https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF

———. 2022b. *U.S. DoD Responsible Artificial Intelligence Strategy and Implementation Pathway*. As of 28 March 2024: https://media.defense.gov/2022/Jun/22/2003022604/-1/-1/0/Department-of-Defense-Responsible-Artificial-Intelligence-Strategy-and-Implementation-Pathway.PDF

Vedula, Padmaja, Abbie Tingstad, Lance Menthe, Karishma R. Mehta, Jonathan Roberts, Robert A. Guffey, Natalie W. Crawford, Brad A. Bemish, Richard Payne & Erik Schuh. 2023. *Outsmarting Agile Adversaries in the Electromagnetic Spectrum*. Santa Monica, Calif.: RAND Corporation. RR-A981-1. As of 2 April 2024: https://www.rand.org/pubs/research_reports/RRA981-1.html

Vestner, Tobias. 2024. 'From Strategy to Orders: Preparing and Conducting Military Operations with Artificial Intelligence'. In *Research Handbook on Warfare and Artificial Intelligence*, edited by Robin Geiß and Henning Lahmann. As of 3 April 2024: https://www.gcsp.ch/publications/strategy-orders-preparing-and-conducting-military-operations-artificial-intelligence

Waltzman, Rand, Lillian Ablon, Christian Curriden, Gavin S. Hartnett, Maynard A. Holliday, Logan Ma, Brian Nichiporuk, Andrew Scobell & Danielle C. Tarraf. 2020. *Maintaining the Competitive Advantage in Artificial Intelligence and Machine Learning*. Santa Monica, Calif.: RAND Corporation. RR-A200-1. As of 28 March 2024: https://www.rand.org/pubs/research_reports/RRA200-1.html

Whyte, Christopher. 2020. 'Problems of Poison: New Paradigms and "Agreed" Competition in the Era of AI-Enabled Cyber Operations.' *2020 12th International Conference on Cyber Conflict (CyCon)*: 215–32. doi: 10.23919/CyCon49761.2020.9131717

Williams, Dan. 2023a. 'Israel Aims to be "AI Superpower", Advance Autonomous Warfare.' *Reuters*, 22 May. As of 2 April 2024: https://www.reuters.com/world/middle-east/israel-aims-be-ai-superpower-advance-autonomous-warfare-2023-05-22/

———. 2023b. 'From Rockets to Recruitment, Israel's Military Refocuses on AI.' *Reuters*, 13 June. As of 2 April 2024: https://www.reuters.com/business/aerospace-defense/rockets-recruitment-israels-military-refocuses-ai-2023-06-13/

Wilner, Alex & Casey Babb. 2021. 'New Technologies and Deterrence: Artificial Intelligence and Adversarial Behaviour.' In *NL ARMS Netherlands Annual Review of Military Studies 2020*, edited by F. Osinga & T. Sweijs, 401–17. doi: 10.1007/978-94-6265-419-8_21

Wong, Yuna Huh, John Yurchak, Robert W. Button, Aaron B. Frank, Burgess Laird, Osonde A. Osoba, Randall Steeb, Benjamin N. Harris & Sebastian Joon Bae. 2020. *Deterrence in the Age of Thinking Machines*. Santa Monica, Calif.: RAND Corporation. RR-2797-RC. As of 28 March 2024: https://www.rand.org/pubs/research_reports/RR2797.html

Wright, Nicholas D., ed. 2019. *Artificial Intelligence, China Russia, and the Global Order*. Alabama: Air University Press. As of 3 April 2024: https://www.jstor.org/stable/resrep19585

Wyatt, Austin, James Ryseff, Elisa Yoshiara, Benjamin Boudreaux, Marigold Black, and James Black. 2024. *Towards AUKUS Collaboration on Responsible Military Artificial Intelligence: Co-Design and Co-Development of AI Among the United States, the UK and Australia*. Santa Monica, Calif.: RAND Corporation. RR-A3079-1. As of 27 March 2024: https://www.rand.org/pubs/research_reports/RRA3079-1.html

Xue, Ying, Chai Fang & Ying Dong. 2021. 'The Impact of New Relationship Learning on Artificial Intelligence Technology Innovation.' *International Journal of Innovation Studies* 5(1): 2–8. doi: 10.1016/j.ijis.2020.11.001

Yu, Yifan. 2023. 'Military AI's Impacts on International Strategic Stability.' *Applied and Computational Engineering* 4(1): 20–25. doi: 10.54254/2755-2721/4/20230339

Yusuf, Shahir Mohd, Samuel Cutler & Nong Gao. 2019. 'Review: The Impact of Metal Additive Manufacturing on the Aerospace Industry.' *Metals* 9(12): 1286. doi: 10.3390/met9121286

Zhang, Li Ang, Jia Xu, Dara Gold, Jeff Hagen, Ajay K. Kochhar, Andrew J. Lohn & Osonde A. Osoba. 2020. *Air Dominance Through Machine Learning: A Preliminary Exploration of Artificial Intelligence-Assisted Mission Planning*. Santa Monica, Calif.: RAND Corporation. RR-4311-RC. As of 2 April 2024: https://www.rand.org/pubs/research_reports/RR4311.html

Zhu, Rongsheng, Ziwen Feng & Qi Chen. 2022. 'Evolution of International Security Specifications for Artificial Intelligence.' *2022 International Conference on Information System, Computing and Educational Technology (ICISCET)*: 217–22. doi: 10.1109/ICISCET56785.2022.00060

# Annex A. Methodology

This annex provides additional information on the research approach and methodology, expanding upon the short summary provided in Chapter 1.

## A.1. Research approach

To accomplish the research objectives, RAND conducted a multi-method study that was divided into six work packages (WPs). Five of these (WP1–5) focused on technical aspects, including data collection and analysis, while WP0 involved project management. The approach to each WP is detailed below:

- **WP0 Project management:** WP0 included regular communication between RAND, the UK MOD and FCDO, as well as quality assurance, adherence to research ethics principles and practices, appropriate data security protocols, and continuous risk management. This was achieved through a regular cycle of weekly meetings, as well as internal project meetings.
- **WP1 Scoping:** WP1 aimed to establish the scope and desired outcomes of the study, as well as identifying the primary stakeholders.
- **WP2 Data collection:** WP2 aimed to gather data from various documentary sources. To achieve this, RAND conducted a two-pronged literature review, which occurred concurrently with stakeholder and expert engagement activities in WP3. For more, see Section A.3.
- **WP3 Stakeholder and expert engagement:** WP3 aimed to collect insights on advancements in defence AI, along with their strategic implications, by engaging with government stakeholders and subject-matter experts through semi-structured interviews and feeding in insights from both webinars and parliamentary inquiries underway alongside the study. For more, see Section A.3.
- **WP4 Analysis:** WP4 concentrated on the evaluation of the AI strategic risks and opportunities identified in WP2 and WP3. The aim was to generate a conceptual framework for strategic military AI risks and opportunities, along with broader findings and implications for the MOD and FCDO.
- **WP5 Reporting:** WP5 focused on drafting a final report and slides that incorporated the findings of all previous WPs, coupled with quality assurance and a review of the draft by the MOD and FCDO.

## A.2. Data collection methods

To maximise the breadth and rigour of the evidence base and analysis within the tight time constraints for this four-week study, the RAND research team employed a multi-method research approach principally involving conducting desk research and stakeholder interviews.

### A.2.1. Literature review

RAND conducted a two-pronged literature review as part of WP2, running in parallel with stakeholder and expert engagement activities of WP3.

The first aspect of the literature review focused on the strategic implications of AI for defence. The review included relevant academic studies, grey literature and past RAND reports to identify a broad range of implications of AI technologies that were expected to affect the military and defence and security more generally. Due to time and resource constraints, the literature review was conducted as a narrative literature review. Relevant literature was identified through a structured set of Boolean string searches in relevant open-source databases and subscription services. The team used RAND Knowledge Services, RAND's in-house library, to identify, access and generate a longlist of literature of 1,500 sources. The research team then scanned and shortlisted these sources to identify the most relevant literature (~200 sources), which was then reviewed using a structured data extraction approach. This gathered key information regarding four key research areas:

1. Strategic implications of military use of AI
2. Frameworks for regulating the impacts of military use of AI
3. National approaches to regulating the impacts of military use of AI
4. Risk mitigation strategies for military use of AI

The review of national approaches included analysing data in relation to Australia, China, Germany, India, Israel, Japan, Russia, South Korea and the United States, representing a cross-section of UK allies, partners and competitors. The extraction also gathered implications of emerging technologies more broadly, where considered relevant alongside the AI-specific data. This approach ensured a consistent review of individual sources and facilitated the understanding of AI key risks and challenges identified by the existing literature.

The second aspect of the two-pronged literature review examined conceptual frameworks and risk management or regulatory/governance approaches from other sectors that could be applicable to mitigating the strategic risks of AI. This aspect of the literature reviewed mechanisms aimed at addressing specific strategic military challenges, such as arms control agreements on nuclear, chemical and biological weapons, or transnational governance issues such as climate change, outer space or other forms of tech regulation.

### A.2.2. Stakeholder and expert engagement

The objective of the stakeholder and expert engagement carried out in WP3 was to gather and consolidate insights on advancements in military use of AI, as well as the strategic implications of such developments. The semi-structured interviews were conducted virtually, and each lasted up to 60 minutes, guided by a set protocol but with scope to ask follow-ups based on responses. Annex B contains a list of interviewees.

In addition to the interviews, the RAND team fed in insights from a series of timely workshops or webinars held during the scoping or delivery period of this short study, as outlined in Table A1.1.

**Table A1.1 Workshops or webinars incorporated into RAND study**

| Date | Organisation(s) | Topic | Speakers/Panellists |
|---|---|---|---|
| 30 January 2024 | Sciences Po Paris | Military Applications of AI | James Black (RAND), Sarah Grand-Clement (United Nations Institute for Disarmament Research) |
| 22 February 2024 | Royal Navy's Strategic Studies Centre (RNSSC) | Strategic Implications of Emerging Technology | James Black (RAND), RNSSC and University of Cambridge representatives, anonymous |
| 12 March 2024 | RAND Europe, RAND US, RAND Australia and the Swedish Defence Research Agency (FOI) | Impact of AI on Future of Defence and Deterrence | RAND and FOI experts, anonymous |
| 13 March 2024 | Defence Nuclear Organisation | Impact of AI on Nuclear and Strategic Stability | Marina Favaro (Anthropic) |
| 22 March 2024 | British Army's Centre for Historical Analysis and Conflict Research | Impact of Autonomy on Warfare | Paddy Walker (RUSI, Imperial War Museum's Institute for Public Understanding of War and Conflict) |
| 26 March 2024 | Vienna Centre for Disarmament and Non-Proliferation | AI in the Military Domain: Technical, Legal and Ethical Perspectives | Thomas Reinhold, Elisabeth Hoffberger-Pippan (both Peace Research Institute Frankfurt), Alexander Blanchard (Stockholm International Peace Research Institute) |
| 26 March 2024 | Center for a New American Security (CNAS) | Autonomy and International Security: Confidence-Building for the Indo-Pacific | Thomas Shugart, Paul Scharre (both CNAS) |

*Source: RAND Europe (2024).*

In addition, the RAND team also incorporated insights emerging from two ongoing parliamentary inquiries running alongside the study, namely the Commons Defence Select Committee's ongoing inquiry into implementation of the Defence AI Strategy, as well as the Lords International Relations and Defence Committee's inquiry into the lessons emerging from the war in Ukraine, including around AI and uncrewed systems – for both of which the RAND study lead, James Black, provided oral evidence.[194]

194 UK Parliament (2024a); (2024b).

# Annex B. List of interviews

This annex provides a breakdown of institutions and individuals engaged in semi-structured interviews as part of WP3. RAND approached officials in government departments such as the MOD, the FCDO, the Department for Science, Innovation and Technology (DSIT), and the Cabinet Office, as well as officials at NATO and the United Nations, and various external experts in the field of defence AI.

This latter category included representatives from a mix of academic institutions, think tanks and experts directly involved in AI research, defence and security research, or pertinent working groups such as the Global Commission on Responsible Use of AI in the Military Domain (GC-REAIM). The RAND team similarly engaged with experts in defence and the defence AI industries. While the primary focus was on UK-based organisations, several interviews engaged European or US-based AI experts. In some cases, multiple representatives from one organisation were involved in the same interview.

A full breakdown of interviews is provided in Table A2.1.

**Table A2.1 List of interviews**

| Organisation | Name | Position |
|---|---|---|
| Adarga | Rob Bassett Cross | Founder and CEO |
| | Charlie Maconochie | SVP Public Sector |
| | Ollie Carmichael | Position not provided |
| | David Green | Position not provided |
| | Dylan Thomas | Position not provided |
| | Seb Matthews | Position not provided |
| Alan Turing Institute | Rupert Barrett-Taylor | Research Fellow |
| | - | Anonymous |
| Arondite | Will Blyth | CEO |
| Atlantic Council | Tate Nurkin | Non-resident Senior Fellow |
| BAE Systems | - | Anonymous |
| BASIC | Dr Chris Spedding | Policy Fellow |
| | - | Anonymous |

| Organisation | Name | Position |
|---|---|---|
| British Army | Major Patrick Hinton | Air Defence HQ |
| Centre for Historical Analysis and Conflict Research (CHACR) | Major General (Rtd.) Andrew Sharpe | Director |
| Center for Naval Analyses | Dr Heather Roff | Senior Research Scientist |
| Centre for the Study of Existential Risk | Dr Maurice Chiodo | Research Associate |
| Chatham House | Nilza Amaral | Programme Manager – International Security |
| | - | Anonymous |
| Department for Science, Innovation and Technology | - | International Policy, Strategy and Multilaterals |
| | - | International AI Policy |
| | - | Anonymous |
| | - | Anonymous |
| Defence Science and Technology Laboratory | Andy Corcoran | Head of International Relationships, AI Policy Directorate |
| European Leadership Network | Alice Saltini | Research Coordinator |
| | Dr Rishi Paul | Senior Policy Fellow |
| | - | Anonymous |
| Faculty AI | Andrew van der Lem | Head of Defence Team |
| Foreign, Commonwealth, and Development Office | - | Euro-Atlantic Security Policy |
| | - | Emerging and Disruptive Technologies, Cyber, Space and Intelligence |
| | - | Anonymous |
| Fujitsu | Dr Keith Dear | Managing Director – Centre for Cognitive and Advanced Technologies |
| Google DeepMind | Dr Lucy Lim | Research Scientist |
| Government Communications Headquarters (GCHQ) | - | Anonymous |
| | - | Anonymous |

| Organisation | Name | Position |
|---|---|---|
| | - | Anonymous |
| | - | Anonymous |
| Hague Centre for Strategic Studies | Dr Tim Sweijs | Director of Research |
| King's College London | Professor Kenneth Payne | Professor of Strategy, member of GC-REAIM |
| Ministry of Defence | - | Counter-Proliferation and Arms Control Centre |
| | - | Anonymous |
| | - | Anonymous |
| | - | Anonymous |
| | - | Anonymous |
| NATO Command and Control Centre of Excellence | Major Marcel Schrrenburg | Staff Officer |
| Palantir | - | Anonymous |
| | - | Anonymous |
| RAND Europe | Peter Watkins CBE | Associate, former Director-General Strategy and International at MOD and Director of Defence Academy |
| Royal United Services Institute | Dr Pia Huesch | Research Analyst |
| RUSI | Noah Sylvia | Research Analyst |
| Sciences Po Paris | Professor Ayse Ceyhan | Political Scientist |
| Special Competitive Studies Project | Dr Joe Wang | Senior Advisor |
| United Nations Institute for Disarmament Research | Dr Giacomo Persi Paoli | Head of Programme – Security and Technology, member of GC-REAIM |
| University of Bath | Professor David Galbreath | Professor of War and Technology |

*Source: RAND Europe (2024).*