# Exploring the Security Implications of Artificial Intelligence in Military Contexts

Amy Ertan

Information Security Group

Centre for Doctoral Training in Cybersecurity

Royal Holloway, University of London

Supervisors:

Dr. Rikke Bjerg Jensen

Prof. Keith Martin

May 2022


# Abstract

Artificial Intelligence (AI) has been described as "revolutionary" for modern warfare. This thesis focuses on the complex and evolving military AI innovation landscape underexplored to date. Existing research on the implications of AI in military contexts is emerging yet nascent despite the apparent pivot to drive the adoption and use of AI-enabled systems in militaries across the globe. This thesis aims to identify the security implications of military AI innovation and explore U.K., U.S., and NATO approaches to adopting AI in a military context. This research employs a multiple methods approach to engage with this rapidly emerging field, analysing available non-classified literature and policy documentation, observational methods, and expert interviews. This thesis finds that defence-focused communities are increasingly interested in military AI innovation and associated security implications, recognising that AI will have a far-reaching impact beyond creating new capabilities. This research highlights a growing awareness of AI among security-focused defence practitioners and policy experts that states must adopt AI to secure or maintain perceived military advantages against adversaries. There are significant challenges relating to AI in military contexts, which include: (1) technical aspects such as cyber security and data protection, (2) organisational aspects including barriers to procurement and the culture of military innovation (alongside wider knowledge and awareness), and (3) strategic aspects including the increased speed of warfare in which humans may not have time to understand or interfere with AI-enabled processes, and potential widening in capabilities in light of "AI arms race" activity. Finally, this research highlights the role of international military organisations as a productive mechanism for developing norms and common approaches to the use of military organisations and explores NATO's opportunities and challenges to contribute to this domain. These discussions form methodological and policy-related contributions to scholarship and aim to inform academic and policy-focused communities. This research will be of particular interest to defence professionals dedicating their efforts to mitigate the negative security implications of AI in military contexts.

# Declaration of Authorship

I, Amy Ertan, hereby declare that this thesis and the work presented in it is entirely my own. Where I have consulted the work of others, this is always clearly stated.

Signed:  Date: 4 May 2022

# Acknowledgements

I have been fortunate to have received fantastic support throughout this thesis. I would first like to thank my supervisors, Dr Rikke Bjerg Jensen and Professor Keith Martin. Rikke, thank you for trusting me as I set out to explore a very messy landscape, and for your detailed feedback, consistent encouragement, and fantastic methodological guidance. Your support was unwavering even when it looked like I was running in circles, and I will never forget it. Keith, your feedback has vastly improved the quality of this thesis, and I appreciate your gentle challenges that frequently made me rethink my approaches and refine my writing for the better. I am grateful to the staff and vibrant research community of the Centre for Doctoral Training (CDT) at Royal Holloway. Special thanks to Dr Jorge Alis Blasco and Professor Lizzie Coles-Kemp, who helped nurture my interest in this research space. The thoughtful structure of the CDT and the staff's openness to discussing ideas with students helped equip me with the skills to carry out this research, and the community of staff and students have also made this thesis journey much less intimidating and lonely than it might have been. The many brilliant characters in the office made the department such a positive place; special shoutouts to Rob and Liam for their infectious humour. I would like to express my thanks for the financial support from EPSRC and the CDT, without which none of this research would have been possible. Special appreciation to Claire Hudson for helping me sort out most of the logistics for this research. Thanks to Rob, Sanja, Georgia, Angela and Lydia for giving me the incredible experience of competing in Cyber 9/12.

Next, I am grateful to the countless researchers, practitioners, and public-sector staff who offered me advice and direction throughout the course of this research. This thesis would have little to show without the thirty-eight interviewees who volunteered their time and energy to deliver incredibly valuable insights into this field of research. Furthermore, I am grateful to have benefited from fellowship positions and encouragement from the Institute of Technology and Society (Rio) Data Protection Fellowship programme, the Belfer Center for Science and International Affairs Cyber Project Fellowship programme, the Financial Services-Information Sharing and Analysis Centre BCD Scholarship Programme, and the NATO CCDCOE Visiting Scholars programme. Several staff members in and beyond these organisations went out of their way to support my development as a researcher, with special thanks to Anne Benishek, Berta Pappenheim, Dr Christian Perrone, Laura Brent, Lauren Zabierek, and Dr Martin Hawley. I want to express my gratitude to the academics who reminded me what it's all about: Nick Robinson for welcoming me to the CDT and imparting elder wisdom throughout, Tarah Wheeler and Bruce Schnier, who reminded me always to stay curious and

speak up for what is right; and all those who collaborated with me to explore our research passions, in particular Dr Andrew Dwyer, Agnes Venema, Edward Hunter Christie, Maggie Gray, and Dr Tim Stevens. Laura Shipp, Georgia Crossland, and Keele are inspiring researchers and friends in equal measure, and I'm excited to see them take the post-PhD world by storm. I am grateful to those who supported me here in Tallinn throughout the final stages of this thesis, including Helena, Ingrid, Jake, Jan, Martin, Mike, Piret, Henrik, Lisa, Phil, Krissu and Rainer, Wes, the dream (cyber) team with Pilleriin, Bobby, Gry-Mona and Aurimas, and the lovely winter swimming community.

I am grateful to have always been surrounded by loved ones. I am indebted to Dr Matthew Bodle, who gave me the courage to first apply for the CDT programme and always believed in me. My family, particularly Oguz 'the Bulldozer' Ertan, who always checked I was getting enough rest, and Aras, who offered an excellent stream of cat pictures over the years. Thank you to my mother, who always supported my endeavours and warmly welcomed me home unexpectedly halfway through the PhD. I appreciate all my friends, who do not understand what I do but fiercely support me anyway: Abbie, Cassie, Cate, Claire, Colette, Emily, Frankie, Kavita, Louise, Luke, Nick, Peadar, Sophie, Sun, and Tom. Thank you to Helen and Rachel, whose support and guidance helped protect my sanity in the latter stages of writing. And, of course, heartfelt appreciation for my husband, Anton Veeremets, who has offered unfaltering support and is extremely excited for me to take up non-thesis related (!) hobbies.

# Publications and presentations

Elements of this research have been published or presented to audiences. This thesis draws on and references co-authored material:

- Gray, Maggie and Amy Ertan, "Artificial Intelligence and Autonomy in the Military: An Overview of NATO Member States' Strategies and Deployment. CCDCOE. Main Report and Appendices. Retrieved from https://ccdcoe.org/library/publications/artificial-intelligence-and-autonomy-in-the-military-an-overview-of-nato-member-states-strategies-and-deployment/.
- Christie, Edward H., and Amy Ertan. (Accepted/In Press). NATO and Artificial Intelligence. Routledge Companion to Artificial Intelligence and National Security Policy
- Robinson, Nicholas, Alex Hardy and Amy Ertan. (Accepted/ In Press). Estonia: a curious and cautious approach to artificial intelligence and national security, Routledge Companion to Artificial Intelligence and National Security Policy
- Christie, Edward H., Amy Ertan, Matthias Klaus and Laurynas Adomaitis. (Submitted). Regulating Autonomous Weapon Systems: Are Existing Principles Sufficient?
- Ertan, Amy. (Accepted/ In Press). The challenges associated with AI military innovation: examining UK practitioner perspectives. *In Hybridity, Conflict, and the Global Politics of Cybersecurity*. The Hague Program on International Cyber Security. Expected Summer 2022.

Aspects of this research were presented through academic posters and talks at the annual HP/HPE (Virtual) Colloquium on Information Security hosted at Royal Holloway, the Doctoral Symposium for Defence and Security, and the Harvard Kennedy School. I have also acted as a non-blind reviewer for the following:

- Babuta, Alexander, Marion Oswald, and Ardi Janjeva. "Artificial intelligence and UK national security: policy considerations." 2020.
- Schneier, Bruce. "Invited Talk: The Coming AI Hackers." In International Symposium on Cyber Security Cryptography and Machine Learning, pp. 336-360. Springer, Cham, 2021.

When citing relevant literature, this thesis relies exclusively on published documents that are either open access (from academics, think-tanks, official government documentation or statements, and reputable media outlets) or published by academic journals. There are two exceptions to this:

- An academic book chapter that was shared with me by the author before the book's formal release: Dwyer, Andrew. "A Foundry of Artificial Intelligence? The case of UK national security." In *Routledge Companion to Artificial Intelligence and National Security Policy*. (Accepted/ In Print). 2022.
- On a laptop, I was shown on-screen the final draft versions of the U.K. AI Defence Policy and the Executive Summary of the U.K. AI Defence Strategy on Monday, 28th March 2022, before the public release of each document. I was not allowed to take the documents with me and had about fifteen minutes to read the content. I was allowed to take limited written notes; however, I do not quote from this within the thesis as I realise the documents may be edited further before public release: U.K. Ministry of Defence. "Defence AI Strategy". Expected Spring/Summer 2022; U.K. Ministry of Defence. "Defence AI Policy". Expected Spring/Summer 2022.

# Affiliations

During this PhD, I held positions that contributed to this research by helping develop my network to approach interviewees. I became a Predoctoral Cybersecurity Fellow with the Belfer Center for Science and International Affairs in July 2020 (with the position ending in July 2022), where my research proposal aligned significantly with the U.S.-focused portion of this thesis. In August 2020, I joined the NATO Cooperative Cyber Defence Centre of Excellence as a Visiting Scholar, through which I conducted the NATO-focused part of this thesis. As the Visiting Scholars position initially concluded in October 2020, my affiliation with the Centre continued through three temporary contracts with the Centre, the last of which ends on 26 August 2022.[1]

[11] These affiliations are detailed and reflected on in more detail in Chapter 1 section 1.3.

# Table of Contents

# List of Tables and Images

# Chapter One: Introduction

## 1.1. Background

My interest in the dynamics of, and discourse relating to, emerging technologies and security was piqued early on during the Centre for Doctoral Training (CDT) programme at Royal Holloway, University of London. During the first year on the programme, I visited "Security & Counter Terror Expo 2018", one of the U.K.'s largest security-focused expositions. The massive warehouse in which the trade show took place revealed a broad selection of physical and cyber security systems, from armoured vehicles to threat intelligence. I visited an intelligent drone demo and was unpleasantly surprised at the salesperson's tone and language as they showed off their product. After explaining how the drone could capture visuals from two miles away, the salesperson excitedly described the drone as "all the intelligence you could ever need…. there are so many commercial opportunities... we're keen to speak to anyone that finds this interesting".[2] There was an apparent dismissal of privacy, ethical, and other broader concerns of these deployed systems. It was a wake-up call that not every company places a value on responsible, conscientious development. This moment of indignation ("How can they get away with saying that? Don't we *all* have to think about privacy?") was the beginning of many rabbit holes, delving into national security strategy and "intelligent" or "algorithmic" warfare, amongst other topics.

There are broad concerns relating to security, responsible use, and stability that deserve careful consideration. This thesis explores these issues while many of them are still unfolding, exploring how states hope to integrate AI effectively and protect themselves against adversaries' malicious use of such systems. AI is argued to be a force multiplier of associated technical phenomena, including big data and autonomy, contributing to faster-paced conflict and a compressed decision cycle. The formulation of AI policies for the area of defence is a recent but fast-moving area of work, and the attention and activity relating to AI in military contexts has increased significantly since I first approached this research in 2018. At that time, no state had released a military AI strategy, and very few academic papers reflected on post-2016 military applications of AI-enabled technology. The subsequent four years have seen growing discussions in state defence departments, with militaries increasingly realising how critical AI would become across sectors - including national security and

[2] Researcher's written notes with quotes taken verbatim, Security & Counter Terror Expo 2018.

military contexts. At the time of writing, the United States (U.S.) and France were the only NATO members with published national AI defence strategies (US Department of Defense, 2018; French Ministry of the Armed Forces, 2019). The United Kingdom (U.K.) announced their intention to release a defence-focused AI strategy in 2021 (see UK Ministry of Defence, 2021, 42). At the time of submission, this AI strategy has not been published. In October 2021, NATO members also agreed on the first NATO AI strategy to help coordinate military innovation and responsible use of AI across the Alliance (Sprenger, 2021).

The military interest in AI-enabled technology has been well-noted (Scharre and Horowitz, 2018; Hoadley and Lucas, 2019; NATO Science and Technology Organization, 2020). There has been a growing consensus amongst researchers and policymakers that states cannot afford to ignore the implications of AI in warfare without falling behind strategic competitors and must therefore invest in, develop, and adopt AI where they have the resources to do so. Recent years have seen a rapid proliferation of national AI strategies (Hill, 2020). Several states are investing significantly in military AI capabilities, most prominently the U.S., Russia (Hoadley and Lucas, 2018), and China (Kania, 2017). The U.S., China and Russia are actively seeking to advance their capabilities to employ AI for military applications (Kania, 2017b). Some states have released publicly accessible documents outlining military approaches to AI, including France (Ministere Des Armées, 2019) and Germany (German Army Concepts and Capabilities Development Center, 2020).[3]

This thesis draws on the findings of a literature review, semi-structured interviews, document analysis, and observant practice-based techniques to present several findings. First, there has been a significant increase in awareness amongst defence-focused communities, including policy-focused staff, on how AI may impact military contexts. There is evidence that the attention towards military AI innovation has increased in terms of greater discussions and increased financial investments at a national and NATO level, with this increased interest showing few signs of deacceleration. Second, this research strongly suggests that AI will have a far-reaching impact beyond creating new capabilities. AI complements other emerging trends and is being integrated into applications including but not limited to intelligence processing, situational awareness and decision-assistance, cyber-physical systems including automated weapons systems, as well as a swathe of enterprise-level logistics and maintenance. Third, this research highlights a growing awareness of AI among security-

[3] For a detailed overview on how the U.K and U.S. have approached AI in the military context see Chapter 5 and Chapter 7 respectively. For an overview on how NATO has approached AI see Chapter 6.

focused defence practitioners and policy experts during the course of this research, as well as a sentiment demonstrated both at the national and NATO level that the U.S., U.K and like-minded states must adopt AI to secure or maintain perceived military advantages against adversaries. Fourth, significant challenges relating to AI in military contexts include (1) technical aspects such as cyber security and data protection, as well as mitigating attacks both *on* AI systems such as data poisoning or adversarial AI, and *from* AI-enabled tools used to target a system maliciously; (2) organisational aspects including barriers to procurement, public-private partnerships, and the culture of military innovation (including knowledge and awareness); and (3) strategic aspects including the malicious use of AI by adversaries, the increased speed of warfare in which humans may not have time to understand or interfere with AI-enabled processes, and potential widening in capabilities in light of "AI arms race" activity. Fifth, and finally, this research highlights the role of international military organisations as an effective mechanism for developing norms and common approaches to the use of military organisations and explores NATO's opportunities and challenges to contribute to this domain.

## 1.2. Research Scope and Research Questions

This thesis addresses two overarching research questions:

(1) What are the implications of AI innovation in military contexts?

This research question explores the phenomena of military AI innovation, understanding how militaries and government defence departments perceive and approach AI in military contexts through the period of this research. This research finds that AI is expected to hold profound impacts across all aspects of warfare, from increasing efficiency in supporting office-based and logistics efforts to enhancing speed and accuracy in battlefield environments. The deployment of AI has a broad range of security implications that are increasingly the focus of attention for military communities. These include but are not limited to: technical challenges to ensure the integrity and reliability of AI systems; operational challenges ensuring adopted systems are used responsibly and safely; and strategic implications relating to strategic stability, AI competition amounting to "arms race" dynamics, and broader geopolitical contexts.

(2) How are actors attempting to mitigate challenges identified in relation to the development and use of AI in military contexts?

This research aims to develop a holistic overview of existing military AI innovation dynamics across the U.K., U.S., and NATO members, identify knowledge gaps and security challenges, and outline the emerging mitigation strategies undertaken by various actors. The thesis draws on the data collected in this research to understand how AI technology is being approached, invested in, deployed, and secured (against) in military contexts. Becoming familiar with this research area has meant drawing literature from various disciplines, including computer science, geopolitics, and security studies, to explore themes such as military innovation and strategic stability. This research finds a number of emerging strategies undertaken by various states, and NATO, to mitigate the perceived challenges of military AI. Such actions and activities include: significantly increased funding and resource dedicated to relevant AI R&D, for example, within the U.S. DoD; the creation of dedicated bodies for AI in national defence within the U.S and U.K.; the development of doctrine and formal adoption of strategies including the U.S. DoD AI Strategy, the U.K. Defence AI Strategy and the NATO AI Strategy; and broader efforts to develop international consensus, including on norms and responsible use, via discussions at NATO and through state-coordinated activities like the U.S. AI Partnership for Defense. In such a rapidly evolving innovation landscape, emerging mitigation techniques are likely to continue to evolve as the field matures.

## 1.3. Researcher affiliations

During the course of this research, I have held two affiliations that have been central to facilitating aspects of this research: a Cybersecurity Fellow position at the Belfer Centre for Science and International Affairs and a Visiting Scholars position at the NATO Cooperative Cyber Defence Centre of Excellence (CCDCOE).

The Belfer Center for Science and International Affairs (often referred to as the Belfer Center) is a research institute situated within the Harvard Kennedy School of Government at Harvard University. Research undertaken by fellows, students, and staff at the Center focuses on international security, diplomacy, environmental challenges, and science and technology policy themes.[4] The Belfer Center runs a “Cyber Project”, an initiative focusing on interdisciplinary policy-focused research on cyber

[4] For more information on the Belfer Center see: https://www.belfercenter.org/about

security themes, within which they welcome academic fellows.[5] I applied for the Cyber Project fellowship in December 2019 and was accepted to start in July 2020.[6] This research application proposed focusing on how the U.S. approaches AI in military contexts through document analysis and interviews with U.S. Department of Defence (DoD) experts. The research undertaken forms the research findings for Chapter 7. This affiliation was almost entirely non-resident due to the pandemic. The Belfer Center affiliation was valuable in connecting me with other Fellows, allowing for the periodic exchange of ideas, and helping me develop initial connections to reach out to potential interviewees (described further in Chapter 2).

The NATO CCDCOE (also called "the Centre") is a NATO-accredited interdisciplinary research centre focusing on cyber defence. Based in Estonia, the Centre comprises staff from 35 sponsoring and contributing nations and provides research on the focus areas of strategy, law, technology and operations.[7] My time with the NATO CCDCOE commenced with a 6-week Visiting Scholars position with the Strategy Branch in August 2020, which took place physically in Tallinn. After the formal conclusion of the position, I remained in touch with the Centre and (separately to this research) have engaged with temporary part-time employment contracts as a researcher, the most recent of which spans January 2022 to May 2022. The Visiting Scholars affiliation was valuable in connecting me to relevant staff at the Centre itself, who introduced me to appropriate staff to interview or were interviewed themselves, and in providing credibility to invite relevant NATO employees to be interviewed for this research. I reflect on this approach, and the impact on my researcher positionality, in Chapter 2.

## 1.4. Thesis outline

Following this introduction, Chapter 2 presents a literature review on the expansive range of academic, think-tank, military, public, and occasional industry-created research material pertaining to the area of military AI. This is a broad area in which literature is continually being updated in line with technological breakthroughs and the changing international geopolitical landscape. Hence, this chapter is necessarily framed within the scope set out by the research questions. Examining available military definitions of AI before exploring material on where AI is currently in use and material exploring near-future feasible applications in military contexts provides the groundwork to address

[5] For more information on the Cyber Project see: https://www.belfercenter.org/project/cyber-project
[6] The Fellowship with the Belfer Center concludes in July 2022.
[7] For more information on the NATO CCDCOE see: https://ccdcoe.org/

the first research question on the impact of AI technologies on warfare. From this, literature is drawn from various disciplines to explore the security implications of AI in this context. These implications range from technical challenges and themes of trust, human-machine teaming in how users interact with and oversee AI systems, to the strategic implications for AI at the level of national policy and international stability. In addressing the research question on how states are approaching the challenges associated with AI, the literature review examines suggestions across the literature on how the negative implications of military AI may be prevented or mitigated.

Chapter 3 presents an overview of the methodological approach to this thesis and justifies the use of the selected methods. This chapter sets out the rationale for scoping this research and literature review and describes the subsequent data collection focusing on the U.K., U.S, and NATO approaches to military AI. The chapter explains how the data collection process was structured, detailing the chosen methods of observant practice as a form of ethnographic fieldwork, interviews with relevant experts in the field of defence and AI technology, and document analysis. Next, the chapter lays out the timeline and thesis journey, summarising challenges faced in access, network-building, and the disruption caused by COVID-19-related factors. Finally, this chapter discusses my positionality as a civilian academic researcher and thus an "outsider" in military-focused/militarised spaces to a degree, acknowledging my part-time contribution as a researcher with the NATO CCDCOE throughout the later stages of my research.

Chapter 4 presents the findings of my observant practice-based research. I attended six physical events (three in the U.K., with others in the U.S., Belgium, and Estonia). Analysing fieldnotes taken during events reveals divided perspectives among conference and trade show participants with varying degrees of caution and pro-AI enthusiasm. The chapter reflects on how event spaces form an environment in which discussions occur and evolve, analysing the content of discussions at the event to understand how attendees understand and approach military AI themes. This chapter reflects on themes such as private sector innovation and military procurement, human-machine teaming and trust in AI technologies, and the value of meeting in semi-closed spaces. The chapter presents how various defence communities were beginning to grapple with military AI topics and determine effective yet secure military approaches to AI adoption during 2019-2020.

Chapters 5, 6 and 7 focus on the U.K., U.S., and NATO approaches to military AI themes. The chapters introduce the context of each respective innovation landscape, an analysis of data collected on each theme, and a brief discussion placing the findings into the context of the literature and broader

research questions. The U.K. chapter, Chapter 5, examines the context of the U.K. military AI innovation landscape and presents the findings of interviews with relevant experts across the U.K.'s defence sector to explore the underlying drivers for AI development and adoption in the military domain. Through the NATO-focused interviews in Chapter 6, this research explored possible roles for NATO in mitigating the potential challenges created or amplified by AI technologies. Looking at the U.S. approach in Chapter 7, this research included interviewing employees with experience in AI policy at the U.S Department of Defense, capturing their perspectives on the U.S approach to AI and maintaining stability in the international landscape. Acknowledging that there have been significant developments in the policy landscape over the course of this thesis, these research findings are presented in the order they were conducted. Chapter 8 presents the main discussion section, where the research findings are brought into the conversation to address the research questions above. Including examples of how the findings build on existing literature and contribute to currently underdeveloped research areas, this chapter provides the main summary view of this research. A conclusion presents a holistic view of the research's limitations and opportunities for further exploration.

# Chapter Two: Literature Review

## 2.1. Introduction

The literature relating to this research area is multidisciplinary, complex, and overwhelmingly dynamic. Literature focusing on the strategic implications of AI is "sparse" (Payne, 2018, 7) and remains underexplored despite an increased focus in recent years in line with recent developments, including national strategies and policies (Johnson, 2018; Hill, 2020). Understanding the dynamics and implications of AI innovation in military contexts requires consideration of literature across disciplines, including computer security, international relations and security theory, and strategic studies. This literature review offers an overview of current AI innovation in defence and security contexts, first illustrating the working definitions of AI and existing applications of AI in defence contexts before focusing on the literature on the security implications of AI-enabled emerging technologies. Drawing together different bodies of work to explain the evolving approach to military AI by state and non-state innovators, this literature review provides a narrative on how AI is approached and is used in military contexts. In so doing, sections 2.2, 2.4 and 2.5 of this chapter in particular address the first research question, highlighting the emerging and predicted implications of military AI. The discussion of the emerging innovation landscape informs an approach to the second research question on how states are approaching security challenges associated with military AI innovation and deployment. Particularly in sections 2.3-2.6, the literature review highlights how various states, and groups of states including NATO, have taken distinct approaches to military AI terminology, innovation plans and strategic doctrine, and emerging governance commitments relating to the "responsible" use of AI in military contexts.

This research has had to carefully define its scope, acknowledging that AI is a technological enabler that does not exist in a vacuum. The dynamic nature of the field and the way in which AI innovation may complement or impact other technical, operational, and strategic processes, means that there are insights to be gained from the analysis of broad trends in military innovation. Reviewing the literature on even subsets of emerging technologies reveals an "overwhelming array" of material that can challenge attempts to make sense of a topic (Siemens and Tittenberger, 2009, 41). This literature review draws on relevant literature that contributes most directly to the discussion on AI innovation

in a military context and thus has chosen to focus on some themes rather than others.[8] For example, this review draws heavily on scholarship including computer science and information security literature that highlights the technical shortcomings of currently feasible AI techniques, a range of primary state reports alongside think-tank scholarship examining current military AI innovation, and strategic and policy-focused literature examining the strategic implications and emerging responses to AI innovation across the international landscape. This thesis does not refer in depth to historical literature examining the foundations of militaries over time, though acknowledges that fields including defence economics and the history of military industrialisation (including scholarship on historical relationships between the private and public sector for defence purposes) would be useful in informing the context of modern trends. More generally this literature review does not draw heavily on scholarship from military theory or related scholarship focusing on the epistemology of war. To some extent, this focus on recent literature means accepting claims within such scholarship on the novelty and uniqueness of AI as an emerging disruptive technology in military contexts. For example, while this chapter does acknowledge an ongoing debate within the literature on how far military AI represents a revolution in military affairs (see footnote 27), this reference is used as supporting evidence from broader literature that supplements modern policy-focused scholarship. Finally, this chapter includes literature which relates most directly to the theme of military innovation which has often meant strict scoping of literature that may provide helpful context. For example, while the organisational sciences and psychology fields contain ample scholarship on how organisational cultures affect decision-making generally (see Schneider et al., 2017; Nonaka and von Krogh, 2009) this chapter primarily draws on scholarship that focuses on military-specific cultural contexts (see section 2.4).

### 2.1.1 A rapidly evolving field

Intense recent interest in military AI means constant evolution in the international landscape, both in policy and research terms.[9] In approaching this literature review, I drew inferences and connections between relevant literature and data collected through the course of this thesis.[10] By developing the literature review around the concepts identified through the course of this research, I was able to adapt the scope of the review. This approach is consistent with this research's broader grounded theory

[8] See Chapter 3, section 3.1 for a methodological discussion on research scope.
[9] This literature review considers research published up to December 2021.
[10] While this thesis will generally use neutral third person language to refer to this research, occasionally first-person language to discussing researcher positionality and the way in which I, as the researcher, engaged with the research process.

approach (as described in Chapter 3, section 3.2), which focuses on inductively drawing out concepts based on research findings rather than applying pre-selected assumptions or theoretical lenses. This approach allowed me to explore the literature on themes raised through the research that may not have been identified initially. Following grounded practice and the reflective nature of drafting a literature review in a rapidly evolving field, some concepts were emphasised as I became more familiar with the research terrain and analysed collected data, in line with the reflective approaches described in Chapter 3, section 3.3.2. As discussed in section 2.1.1, in prioritising the examination of key themes that emerged through this research, this research acknowledges there will be literature that is not explored in-depth in this literature review that is nonetheless relevant.

Throughout this research it was apparent that academic literature, such as articles published in academic peer-reviewed journals, can often lag behind other forms of literature. This thesis also cites original research and analysis pieces produced through research institutes and think tanks or specialist media outlets like WarontheRocks or Lawfare Blog, which have a much faster review process than academic journals and conference submissions. This scope allowed for a broader range of perspectives, including where researchers were not publishing as academics but as independent or commissioned researchers, which allowed the capture of views from former and current defence staff as cited through this thesis. This approach also allowed this research to draw on research published rapidly in response to real-world activity. Moreover, several research institutions have produced numerous insightful reports valuable for this research, including, for example, the Center for Security and Emerging Challenges (CSET).[11] Particularly for policy-focused literature, it was assessed that this research would benefit from drawing on relevant papers from such a wide range of sources. In selecting sources, I recognised that specific publications, or publication venues, will often have an agenda that should be considered when reviewing the literature, particularly when the research has not been peer-reviewed.[12] While reviewing the literature, sources were chosen based on the reputation of those involved in the publishing process. My growing familiarity with the research field through this research enabled me to select appropriate references with confidence. This literature review draws on various materials, including peer-reviewed academic literature, governmental and

[11] CSET is a policy research institute based in Georgetown University (Washington DC) and regularly publishes research papers regarding artificial intelligence in a military context, with a primary focus on U.S. national security. For more information see https://www.cnas.org/artificial-intelligence-and-global-security https://cset.georgetown.edu/publications/

[12] Acknowledging an institutes research agenda does not necessarily mean dismissing the research. For example, this thesis considers multiple publications from CSET and the Center for New American Security (CNAS) with the awareness that both research organisations hold a U.S.-centric approach to their analysis. This awareness helped me put their publications into context when drawing together literature from different sources.

intergovernmental organisations' publications or statements, and research and reflections from a range of policy literature, including reports from research institutes. Where relevant, I have specified the nature of the publishers to acknowledge their scope and agenda.

### 2.1.2 Chapter Structure

This literature review is set out into five parts. The first, "What is AI", considers efforts to define AI and AI-related concepts and is presented in section 2.2. This section sets out the scope of the technologies considered in this thesis contained within the umbrella term "AI". Drawing together literature on the nature of AI technology, this review characterises AI as a broad set of technologies that do not exist in a vacuum, but which complement and acts as a force "enabler" for a range of other capabilities and technological systems (Cox and Williams, 2021; Horowitz 2018; Johnson, 2020). The second part, presented in section 2.3, focuses on AI applications in military contexts and describes where AI is deployed in military environments. Next, section 2.4 reviews the literature on military AI innovation, noting the increasingly prominent role of the private sector in the context of military technologies. Section 2.5 maps out a range of security implications of AI-enabled technology in military contexts, divided into three subsections. The first offers an overview of strategic security considerations, including discussions within the literature on arms race dynamics, growing capability gaps as some militaries adopt AI faster than others, and Russia and China's approach to military AI innovation. The second subsection details major technical and operational challenges relating to current development and adoption, including but not limited to technical vulnerabilities of an AI algorithm and related infrastructure, including training data, adversarial AI, and the challenges associated with human-machine teaming and the compressed decision-making cycle due to faster AI-enabled warfare. The final section sets out how the implications mapped out often apply to, and might be addressed via, multilateral organisations, and sets out the literature on how alliance structures may play a role in adapting to emerging technology in military contexts. The final part of the literature review is set out in section 2.6 and focuses on the responsible use of AI in military contexts. This section draws on extensive literature relating to lethal autonomous weapons systems (LAWS) to discuss current relevant attempts to set out legal frameworks, international norms, and ethical principles.

## 2.2. What is AI?

### 2.2.1 Disagreements on definitions

While this literature review focuses specifically on AI technology in a military context, it is helpful to understand key concepts and terminology for AI overall. An overview of the diverse range of definitions for AI provides a clear demonstration as to why approaches to AI innovation often seem fragmented, as communities use "AI" to discuss different concepts. Similarly, understanding the technical fundamentals has contributed to shaping this thesis will discuss AI as an umbrella term for modern AI techniques.

In popular discourse, AI terminology is "a surprisingly fuzzy concept" (Kaplan and Haenlein, 2019, 15), a vague phrase that is "widely used and loosely defined" (Brennan, Howard and Neilsen, 2018, 1). Experts disagree on what AI *is*, never mind what it can do or how it can be regulated (Brennan, Howard and Neilsen, 2018), and the academic literature has a range of definitions for AI and AI techniques which have changed over time (Dufour, 2018; Russell and Norvig, 2016). The challenge to find a consensus on what AI means has been taking place throughout the field's history (see Sweeney, 2003). In the military planning context, the confusion about terminology has hampered attempts to explore AI's implications, given the constant grappling over what AI is and how it may be applied (Burton and Soare, 2019).

At a fundamental level, AI should be considered a "collection of ideas, technologies, and techniques" (Brennan, Howard, and Nielsen, 2018). This depiction aligns with Kaplan and Haenlein's (2019, 15) description of AI as an umbrella category of technologies that may be categorised by evolutionary stages (see below for a discussion on first and second wave systems) or techniques.

Several definitions refer to human intelligence as a reference point, as demonstrated in Table 2.1:

*Table 2.1: A selected cross-section of AI definitions in academic literature*

| Authors | AI Definitions |
|---|---|
| McCarthy et al., 1955 | AI is: "making a machine behave in ways that would be called intelligent if a human were so behaving". |

| | |
|---|---|
| Minsky, 1968, v | AI is: "the science of making machines do things that would require intelligence if done by men". |
| Gartner, 1999, 33-34 | AI as the: "biopsychological potential to process information…to solve problems or create products that are of value in a culture". |
| Kaplan and Heinlein, 2019, 5 | AI is: "a system's ability to correctly interpret external data, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation." |

As Table 2.1 highlights, various definitions exist across the literature, with some but not all including comparisons with human intelligence. The literature highlights a fragmented approach to understanding terms as different researchers define AI differently, introducing terminology ambiguity that poses a risk to the field and hampers straightforward discussions (Sweeney, 2003; Kraaft et al., 2020). Comparing survey responses by AI researchers and reviewing AI policy documentation, Kraaft et al. (2020) highlight that different communities favoured different definitions, with AI researchers preferring definitions that emphasized technical aspects and policymakers drawing on definitions that relate systems to human intelligence. Kraaft et al.'s (2020) research highlight that the technical definitions favoured by AI researchers tend to refer to currently feasible techniques. In contrast, policy communities were found to use more forward-looking definitions, and policy-focused efforts may "may overemphasize concern about future technologies at the expense of pressing issues with existing deployed technologies" (Kraaft et al., 2020, 1).

Military and defence-focused definitions for AI do not differ inherently from non-defence literature. Table 2.2 highlights definitions offered in defence-focused U.K., U.S., and NATO policy documentation.

*Table 2.2: Examples of U.K., U.S., AND NATO AI definitions in published reports*

| | |
|---|---|
| U.K. Ministry of Defence. Joint Concept Note 1/18 Human-Machine Teaming (2018, 60). | "The performance by computer systems of tasks normally requiring human intelligence, such as translations between languages."<br><br>The report cites this definition as the Concise Oxford English Dictionary, 12th Edition. This definition is adopted by this thesis for AI, with further clarifications on techniques outlined in this section. |

| | |
|---|---|
| The U.S. Department of Defense AI Strategy Executive Summary (2019, 5) | "AI refers to the ability of machines to perform tasks that normally require human intelligence – for example, recognizing patterns, learning from experience, drawing conclusions, making predictions, or taking action – whether digitally or as the smart software behind autonomous physical systems." |
| NATO Science & Technology Organization's (STO's) 2020-2040 S&T Report (2020, 50). | *As above, draws on the U.S. definition.* |

As Table 2.2 highlights, these definitions correspond with Kraaft et al.'s (2020) assessment that policy literature links AI to human intelligence. The U.S. Department of Defence Strategy Executive Summary, and the NATO STO definitions, also align strongly with the Oxford Reference definition, which describes AI as: "The theory and development of computer systems *able to perform tasks normally requiring human intelligence*, such as visual perception, speech recognition, decision-making, and translation between languages" (emphasis added).[13]

"What many AI researchers do when they say they are doing AI contradicts what some AI researchers say is AI" (Sweeney, 2003, 3)

Definitions matter. A lack of consensus on AI concepts means it can be challenging to measure progress in the field (Sweeney, 2003). Moreover, Kraaft et al. (2020, 1) describe how the lack of consensus on agreed definitions "hampers the possibility of conversation" around AI, particularly regarding policy-facing discussions on regulatory or legal matters relying on defined terms. This section now introduces distinctions and debates on terminology and how this thesis further defines and approaches AI.

[13]https://web.archive.org/web/20220420225104/https://www.oxfordreference.com/view/10.1093/oi/authority.20110803095426960

*Note A: AI over time and the distinction between 'First Wave' and 'Second Wave' AI*[14]

AI definitions have also changed over time in line with the evolutionary stages of AI (Kaplain and Haenlein, 2019), and what was considered AI in 1980 is not given the same term today. The NATO Science and Technology Organization's "Science & Technology Trends 2020-2040: Exploring the S&T Edge" report (2020, 50)[15] highlights a distinction between categories of AI, including "first wave" or "knowledge-based" AI and "second wave" or "data-based" AI that is reflected elsewhere including but not limited to institutional policy-level research reports published via the European Union (EU) (Boucher, 2020), and the U.S. (Homeland Security Science and Technology Advisory Committee, 2017). First wave, or knowledge-based AI systems, have existed for decades and are typically not referred to as AI today, relying on rules-based decision-making using "if-then" logic to determine an action. This category of systems cannot learn from their experiences or apply their reasoning beyond this "if-then" process (O'Leary, 2021). In contrast, "second wave" or "data-based" AI systems solve specific problems using statistical algorithms to find patterns in data (Christie, 2020). Data-based AI is trained on large data sets and includes machine learning (ML) techniques – which involve "programming computers to optimize a performance criterion using example data or past experience" (Alpaydin, 2009, xxxi). There are various subsets to ML, including Deep Learning (DL), which uses complex layers of adjustable computing elements (Russell & Norvig, 2021).

First wave AI has been employed in a military context for decades, with examples of systems with fully autonomous modes, including the American Patriot air defence missile system since the 1990s (Hawley, 2019) and the Aegis defence system since the 1980s (Scharre, 2018). Papers in the 1980s explored how first wave AI systems could support the military through applications including decision-making assistance and situational awareness (Gilmore, 1985), models and simulations (Erickson, 1985; Shinar, Siegel and Gold, 1988), and autonomous vehicles (Din, 1987).

The data-based learning techniques described in Note A have seen significant innovation in recent years due to research breakthroughs relating to artificial neural networks, increased computing power,

[14] This literature review will use sectioned off boxes to introduce and acknowledge concepts that are important to understand the broader context of this research, but which will not be explored in depth due to limitations on scope and word count.

[15] The report offered a categorisation of critical and emerging technologies that persisted into the Emerging and Disruptive Technologies Coherent Implementation Strategy (2021) and set out NATO's approach to emerging technologies. See Chapter 6 for more detail on key NATO documents regarding AI in military contexts..

and the greater availability of data through corresponding technology innovation in big data.[16] It is important to note that both first and second wave AI systems described in Note A are forms of "narrow AI", or weak AI. Narrow AI refers to AI "systems designed to do deliberately constrained tasks" in a particular environment (Horowitz, 2018, 38). For example, a voice recognition AI algorithm is designed to process and would not help interpret unrelated data. Today's AI is considered "brittle", working best in the system in which it was designed (Payne, 2018, para 8). Modern technology is still some way from achieving "third wave AI" capabilities, also known as "artificial general intelligence" or "strong AI". These terms refer to the AI systems that approach super-intelligence, going beyond narrow tasks to approach other challenges in a way that approaches human or super-intelligence (Babuta, Oswald, and Janjeva, 2020). Current AI capabilities still fall firmly within the narrow, second wave AI category as AGI does not currently exist (Ramamoorthy and Yampolskiy, 2018).

Third wave AI is considered out of scope for this thesis. This scoping decision aligns with Horowitz's (2018) argument that third wave AI capabilities are unlikely to emerge within the next generation. It is the implications of narrow AI capabilities, Horowitz writes, that "are most likely to affect militaries — and with them the balance of power — over the next two decades" (2018, 42). This view is also reflected in U.K. government material, with an MoD Joint Concept on Human-Machine Teaming taking care to confirm that the scope of any reflections, assumptions, and statements referred only to narrow AI capabilities (Ministry of Defence, 2018). A further statement caveat that the report's evaluation would be invalid and outdated once AI approached the realm of general intelligence (Ministry of Defence, 2018). This literature review follows this logic and uses "AI" to refer to second wave, narrow (and largely ML-based), techniques.

[16] "Big Data" consists of vast datasets that require non-traditional methods of storage, analysis, and visualisation due to the massive amounts of varied, complex, and high velocity data being considered. Analysing big data may reveal new insight and information including previously hidden patterns and correlations (Sagiroglu and Sinanc, 2013).

*Note B: What can ML do, and how?*

Having outlined the scope of AI for this thesis so far, it is important to be aware that the working term is *still* not a monolith but represents an umbrella category for a range of computational techniques. For example, while Note A highlighted the AI subset of ML techniques, discussions on AI can refer to narrow sub-sets referring to different computational techniques and methods. Unsupervised ML allows an algorithm to draw inferences from unlabelled data and can be particularly adept at identifying anomalies. In contrast, supervised learning allows for an algorithm to be trained on correctly labelled data, which is then used to infer predictions with opportunities in data classification (Tamir, 2020). As a sub-category of AI, ML can use large amounts of data as a basis for pattern recognition, classification of items, and future predictive ability. These models rely on the data in which they are trained. There are already many cases where ML algorithms outperform humans in an "increasing range of narrow pattern recognition and prediction tasks" (Christie, 2020, para 6).

Sub-families of ML offer advances for different technical breakthroughs. As one example, a class of techniques based on artificial neural networks, DL, has attracted researchers interested in problems relating to multi-dimensional data. DL algorithms have beaten previous records across spaces, including DNA analysis and predicting gene expression, image recognition, and natural language processing (Lecun, Bengio & Hinton, 2015). Since the turn of the millennium, ML has emerged from "laboratory curiosity*"* to widespread application and commercial use (Jordan and Mitchell, 2015, 255). Improving through experience, AI shows an aptitude for practical data-heavy tasks from computer vision to natural language processing. The applications to date have been highly diverse and are omnipresent across sectors; from finance (Bahrammirzaee, 2010) to healthcare (Yu, Beam and Kohane, 2018); from use in digital assistants such as Google Home or Amazon Alexa (Maedche et al., 2019) to image (and facial) recognition used in various technologies across industry and government applications. (Raji et al., 2020). This thesis will explore the relevance of such ML AI applications, now referred to as AI applications for the purposes of this thesis, in military contexts.

Current AI algorithms carry out tasks in ways that are distinct from human reason processes, and modern AI techniques can perform actions unpredicted by its creators or observers. Prominent non-military examples have highlighted this case: in 2016, the "Go" board game competition between professional Lee-Sedol vs Deepmind's AI system "AlphaGo" is often cited as a breakthrough in AI

and strategic computational victory (Yudkowsky, 2016). At one point, observing professionals assumed AlphaGo was making mistakes, only to be taken aback when it was revealed the reasoning system had succeeded, in part benefiting from approaching the game without human assumptions. Instead, armed with instrumental efficiency and probabilistic calculations, the AI system won three of the five games against the human expert. This perceived creativity will have implications beyond the benign and is likely to display in future instances of AI, through which an offensive tool may operate in ways that lie outside activity planned by the human attacker.

The following section examines how AI is currently being applied in military contexts, discussing available research outputs, investments, and predictions on the future of military AI. It is a swiftly moving landscape. A U.S. Congress-commissioned report by the National Security Commission on Artificial Intelligence (NSCAI)[17] describes how a "few months" of fierce activity around AI great power competition as an "AI revolution" that has "shaken the strategic terrain" in the U.S (2020, 6).

[17] The NSCAI was established as an independent commission through the 2019 National Defense Authorization Act to advise on how the U.S could advance the development and adoption of AI techniques in national security and defence contexts. The Commission submitted a series of quarterly supports to Congress, culminating in its final report in March 2021 and dissolvement in October 2021. For further detail see Chapter 7 section 2.2.

*Note C: AI and Autonomy*

Autonomy does not necessarily mean the same as artificial intelligence. Agreement on what autonomy means varies within the literature. However, the term is frequently used to refer to systems with ML technology (i.e., when discussing autonomous weapons systems with ML-enabled capability). The U.S. military recognises autonomy as a goal, where autonomy is defined as a capability (or a set of capabilities) that enables a particular action of a system to be automatic or, within programmed boundaries, "self-governing" (DoD Defense Science Board, 2012, 1). This definition does not necessitate AI technology, as defined within this thesis. As intelligent systems become more capable of decision-making in complex environments, they may be granted greater autonomy and become capable of achieving their goals without human direction (Vignard, 2014) in ways that may or may not utilise AI.

Franklin and Graesser (1996, 25) provide the following definition of an autonomous agent: "An autonomous agent is a system situated within and part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and as to effect what it senses in the future". Scharre and Horowitz (2015, 5) have a more straightforward definition when writing about autonomous weapons, with autonomy simply referring to "the ability of a machine to perform a task without human input" and state that this does not necessarily include AI technology. In contrast, while U.K.'s Defence Science and Technology Laboratory (Dstl)[18] defines AI with reference to human intelligence and acknowledges that while there are numerous definitions, their report sets out definitions of autonomy and autonomous systems that assume AI components (Dstl, 2021).[19] The attempt at a non-scientific but practical stance reflects a similar perspective laid out by the U.S. in the Defense Science Board report, which describes attempts to define autonomy as "a waste of both time and money" and advocates for a pragmatic approach (DoD Defense Science Board, 2012, 24).

This thesis will refer to autonomous weapons where the systems are paired with AI technology.

[18] Dstl is an executive agency sponsored by, and part of, the U.K. Ministry of Defence (MoD). Dstl focuses on science and technology for the defence and security field, coordinating the funding mechanisms Defence and Security Accelerator (DASA) and incorporating an 'AI Lab' at its site in Porton Down, U.K. This thesis denotes the agency as 'Dstl', as opposed to 'DSTL', in line with the agency's website and agency-produced materials.

[19] Dstl's definition for autonomous systems: 'A system containing AI-based components that allow it to exhibit autonomy' and for autonomy: 'The characteristic of a system using AI to determine its own course of action by making its own decisions' (Defence Science and Technology Laboratory, 2021, 4).

### 2.2.2 AI as an enabler and force multiplier[20]

In acknowledging uncertainty about an emerging technology's future, Molas-Gallart (1997) emphasises that technology is dual-use when it has current, or potential, military applications. Often, a technology developed by either sector may be adopted – in some cases, unexpectedly – by the other. Many AI applications can be used for military and non-military applications and are considered dual-use (Gilli, 2019; Nouwens and Legarda, 2018; Stowsky, 2004).[21] Examples of AI use cases across civilian and military contexts may include, for example, information processing or autonomous vehicles, a factor enhanced by the connection between defence communities and AI research and development efforts (Geist, 2016). In this way, AI can have significant consequences beyond the military domain to impact wider society (Carlo, 2020). Stowsky (2004) argues that policymakers must consider the dual-use nature of capabilities when considering the implications of technological development.

Furthermore, AI operates as a force multiplier in the context of broader technological developments, including in communications, data processing, and cyber capabilities (Johnson, 2020; Payne, 2018). For example, an AI-enabled intelligence processing tool can directly enable faster, more precise analysis and have the extended impact of freeing up personnel to use this information to coordinate military units or determine the most optimal next steps (Johnson, 2019). Similarly, while the concept of evolving malware is already recognised (Meng et al., 2016), the future may involve malware that can hold increased autonomy. AI-enabled malware may help launch more effective attacks against a target's infrastructure in ways that human operators may not predict in advance (UNIDIR, 2018). A NATO Science and Technology Organization report identified how AI represents a "fulcrum around which big data will be turned into actionable knowledge, and, ultimately, a NATO decision advantage" (2020, 14). Johnson (2020) points out that the overarching trends of data processing, cyber capabilities, and communication "revolutions" (18) would have occurred without AI - but are magnified significantly by AI's manifestation.

[20] The U.S. Department of Defense defines a force multiplier as "A capability that, when added to and employed by a combat force, significantly increases the combat potential of that force and thus enhances the probability of successful mission accomplishment." (JP 3-05.1) (US DoD). This thesis generalises this definition to refer to the multiplication or leveraging of force in ways that may include, or refer solely to, non-kinetic effects.

[21] This thesis adopts Cowan and Foray's (1995, 851) definition of dual-use technologies, as technologies "developed and used both by the military and space sectors on the one hand and by the civilian sector on the other".

Overall, Finlan (2020) summarises the fundamental advantages of AI in managing information flow and, therefore, coordination in battle. Particularly as the amount of data received by military leaders is expected to increase exponentially, it appears "inconceivable" that future warfare will not benefit from AI-enabled systems (Finlan, 2020, 6). Further exploration of the implications of AI in military contexts will be considered in section 2.5.

## 2.3 Military AI Applications

This thesis adopts RAND's (2020) simplified categorisation of applications in the defence space: enterprise AI, mission support AI, and operational AI. Such categorisation assists in distinguishing between the implications of AI in different safety-critical (or less extreme) environments. Enterprise AI incorporates all enterprise environment uses of AI, including financial and human resources management systems. These are relatively low-risk applications with few safety issues should there be a malfunction (RAND, 2020). Within the defence and military context, AI can be utilised across many office and supporting applications, from logistics and software optimisation tools (Konaev et al., 2021), to human resources and tasks, including soldiers' wellbeing (Poulin et al., 2014; Haner and Garcia, 2019). Enterprise applications of military AI can deliver significant performance enhancements, and policy-focused literature has recommended greater investment in military enterprise AI as a relatively non-controversial and low-risk environment in which AI holds considerable promise (Konaev and Chahal, 2020). Reflecting on the second research question, such an approach represents one potential innovation path through which states might adopt military AI while avoiding the more safety-critical risks relating to AI deployment in less predictable battlefield environments. As Chapters 4-7 will detail, it appears it is too early to tell how far states are prioritising certain AI applications over others as part of broader risk mitigation initiatives for military AI.

The second category covers mission support AI, which covers the intermediate use of tools where the environments and implications fall between operational and enterprise applications (RAND, 2020). This category includes tools for logistics and maintenance and intelligence, surveillance and reconnaissance applications. With the capability to rapidly process vast amounts of information, AI can significantly assist in information management and thus enhance situation awareness (Finlan, 2020). AI can be used for the automation of intelligence (Sharkey, 2011; Selyanin, 2021), surveillance and reconnaissance in the military context (Davis, 2019), including via unmanned surveillance

systems (Department of Defence, 2014). A study by Street et al. (2018) highlighted that combining NATO data with ML and analytics capabilities could provide improved decision-making support at the Commander level. Researchers at Marine Corps University have started experimenting to determine how adaptive learning may be revolutionised via AI-enhanced wargaming platforms (Jensen, Cuomo, and Whyte, 2018). AI can also be applied in command, control, communications, and intelligence (C3I) systems to assist in complex and adversarial environments (Johnson, 2020, 17). AI also facilitates enhanced cyber capabilities to defend and attack computer networks (Johnson, 2020) and enables accelerated cyber-attack and defence activity (NSCAI, 2019).

The third category, operational AI, represents AI capabilities that can be deployed in uncertain, dynamic environments with a higher cost of failure (Tarraf et al., 2019). For example, AI can be used in enhanced missile defence, including automatic target recognition technologies and munitions (Johnson, 2020). The European Defence Agency has an active research project (as of spring 2022) exploring using AI-enabled radar communications systems to increase resilience when faced with electronic warfare (European Defence Agency, 2020). In addition, AI may be implemented in physical infrastructure, including unmanned systems such as search and rescue robotics (Nouwen and Legarda, 2018, Christie, 2020; Sharkey, 2011; Stodola, Drozd and Nohel, 2020) or across a broad range of autonomous weapons systems (Haner and Garcia, 2019). Boulanin and Verbruggen (2019) describe five categories of existing autonomous military weapons systems: active protection systems, air defence systems, robotic sentry weapons; guided munitions; and loitering munitions.[22] The implications of autonomous unmanned systems could include swarms for strike missions (Johnson, 2020) and intelligence, surveillance, and reconnaissance (Allen and Chan, 2017). Johnson (2020) further highlights how drone swarms have additional use cases, including enhanced weapons delivery systems and electronic or cyber warfare capabilities, and may allow adversaries to use low-cost swarm systems to overwhelm sophisticated defence capabilities. A U.S.-focused interview-based study of 39 military-focused academic and industry experts revealed the view that enterprise AI is seen as closest to deployment than mission-support AI (Tarraf et al., 2019). Mission-support AI is also likely to be widely deployed sooner than any broader deployment of operational AI (Tarraf et al., 2019). These interviews revealed that all three categories of AI are believed to deliver significant

[22] Loitering munitions are weapons which can be maintained in the air for some time before engaging in a targeted attack. The IAI Harop is an example of a loitering munitions system with fully autonomous capabilities. See Scharre, Paul. "Autonomous weapons and stability." PhD diss., King's College London, 2020.

future advantages to the U.S. military, a view supported by existing scholarship (Tarraf et al., 2019). In a divergence from the current categorisation of conflict, Finlan (2020) predicts that by 2045, combat specialists may not be split by domain (air, sea, and land, for example). Instead, warfare will occur within an integrated landscape where AI technologies and conventional forces are entirely interoperable and harmonised (Finlan, 2020). This suggestion implies a significant shift in the fundamental nature of warfare. At the same time, the moving away from analysis of military operations across silo-d domains suggests a different form of how military conflict is understood.

Militaries are likely to deploy AI-enabled capabilities in the cyber domain increasingly in the near future (Taddeo and Floridi, 2018; Kania, 2017a). AI-enabled autonomy could be used in cyberspace operations, including offensive capabilities (Kania, 2017b). The U.S. DoD Directive on Autonomy in Weapons systems specifically excluded cyberspace operations from the condition of "appropriate levels of human judgement" (Department of Defense, 2012, 7).

Turning to information warfare, AI falsified materials, also known as "deepfakes", can be weaponised to challenge national security by undermining democratic discourse, trust in institutions, and international diplomatic efforts (Chesney and Citron, 2018; NCSAI, 2019). In this way, military AI holds significant implications ranging from operational disruption to challenging strategic narratives as well as international political stability. With a significant range of applications and associated harms, Citron and Chesney (2018) highlight how deepfakes can harm military or intelligence operations or capabilities, most obviously through disinformation campaigns at a strategic or operational level. They offer the example of how a falsified video of an American soldier in Afghanistan killing local civilians could potentially cause violent local protests in the short term and undermine strategic narratives of the conflict more broadly (Citron and Chesney, 2018). Similarly, sophisticated forgeries represent another security threat. An enemy may compromise sensitive documents through cyber espionage and then release a combination of genuine (compromised) and falsified material (Allen and Chan, 2017; Citron and Chesney, 2018). Deepfakes may then be released in which the compromised actor "verifies" the forgeries, confusing verification and causing challenges for the targeted actor's foreign policy (Allen and Chan, 2017). Johnson (2021b) presents a feasible scenario in which deepfake videos could be used to feed misleading evidence to foreign military intelligence to provoke conflict escalation. For example, Johnson highlights that "State A" could facilitate the release of a deepfake video in which "State B" senior leadership appears poised to launch a pre-emptive strike on "State C". If "C" does not recognise the footage as a deepfake and escalates in response, "B" will respond in kind, resulting in what Johnson (2000, 2021, 2021b) has

termed “inadvertent escalation”. Using AI to facilitate military deception can complicate military coordination on a national and multinational level, for example, by creating distrust between allies (Lin-Greenberg, 2020).

## 2.4. The evolving military AI innovation landscape[23]

The drive to adopt AI in the military context represents an opportunity to improve and enhance capabilities in ways that draw on non-defence-specific commercial innovation while attempting to mirror developments in the civilian economy (Fischer, 2020; Sayler, 2019). Payne (2016) describes the AI research agenda as disparate and eclectic, commenting that while militaries are increasingly interested in the strategic security implications and AI, much of the research conducted by private sector and academic researchers is not orientated towards the military. With military doctrine heavily shaped by external factors including battle terrain or geography, budget constraints, the balance of power between nations and technological change (Bickel, 2018), militaries have had to adapt to a complex landscape that includes the proposes and challenges relating to military AI innovation. In addition, militaries wishing to adopt AI must face the broader challenges of military innovation, not only by offering adequate resource dedication but also in terms of possessing the organisational capacity to adapt accordingly (Horowitz, 2010). For AI and related emerging technologies, militaries are facing an intensifying innovation landscape which has prompted the creation, and restructuring, of new institutions.[24]

Warfare has, of course, always been shaped by emerging technologies of the time (Sechser, Narang, and Talmadge, 2019; Burmaoglu and Sarıtas, 2016). War is a catalyst for technological innovation (Singer, 2009, 45), with the military funding technical research spanning applications from the first mechanical calculator to Babbage's 1822 "difference engine", the latter considered the groundwork for modern computers (2009, 45). Bellais and Fiott (2017) highlight the 21st-century trend of disruptive innovation from the commercial sector, innovating heavily in AI. They argue for the benefits of greater public-private cooperation to make the most of opportunities in innovation to best enable innovation in defence capabilities in an adapting transformation landscape (Bellais and Fiott, 2017). Indeed, militaries have focused on the advantages of strengthening public-private cooperation. States have shown an awareness to streamline procurement of military AI technologies, with commitments to improve engagement with academia and industry contained within the DoD AI Strategy Executive Summary (2019) and the NATO AI Strategy (2021). While the U.S. has been

[23] This literature review offers an overview of the military landscape but does not cover the U.K., NATO or U.S. approaches in detail. A review of relevant literature and policy documentation is contained within the context sections of Chapter 5 for the U.K., Chapter 6 for NATO, and Chapter 7 for the U.S.

[24] Examples of these institutions include the creation and restructuring of the JAIC in the U.S. context, and the creation of the Defence AI Centre in the U.K. context. An overview of relevant institutional structures will be described further as relevant in Chapters 6-8.

making this transition through the Third Offset Strategy and creating DARPA's Defense Innovation UnitX to drive "civ-mil" innovation (Bellais and Fiott, 2017), the U.K. MoD and the U.S. DoD have each launched various mechanisms over the period of this research, including innovation coordinators and dedicated procurement platforms for AI respectively, to encourage private-sector engagement with dual-use or military focused technologies. While these initiatives will be discussed in more detail in Chapters 5, 6, and 7 respectively, the recent creation of innovation mechanisms and facilitators represents one way in which states are attempting to circumvent the adoption challenges for military AI. Bellais and Foitt (2017) acknowledge significant challenges that the commercial disruption has brought to the defence innovation landscape, with lowering barriers to entry in what was previously a relatively closed market.[25] Policy recommendations across the literature advise militaries to continue to advance public-private partnerships relating to military AI technology (Kania, 2017, 41; Fischer, 2020). These need not be limited to national-level security collaboration; Gilli (2020, 35) proposes the creation of a NATO "Artificial Intelligence, Integration and Implementation-Enabling Centre (A3IC)", which could support AI adoption across NATO.

Writing in 2009, Singer asserts that the U.S. military set the agenda for AI innovation, funding as much as 80% of all AI research in the U.S. (Singer, 2009). More recent literature provides an alternative interpretation of the emerging landscape. Cummings (2017) reports a shift in R&D development from the military to commercial dominance, finding that defence investments in AI represent a small proportion of the R&D by commercial information technology firms, including Google, Amazon and Facebook. Particularly in autonomous systems, commercial innovation far outstrips military development (Cummings, 2017). Cutting-edge AI technologies are now commercially driven (Whittaker, 2021; Cummings, 2017; FitzGerald and Parziale, 2017), with much of current AI innovation developed as a general purpose for a civilian market (Christie, Buts and Du Bois, 2021; FitzGerald and Parziale, 2017). This dynamic has been recognised in the literature for some time; Stowsky (2004) highlighted how most military technologies used by the U.S. had commercial origins, with dual-use technologies readily available via a competitive marketplace. Whittaker (2021) goes as far as to compare the technology industry's control over modern AI to the

[25] This finding appears to align with Christensen's (1997) influential scholarship on disruptive innovation, which highlights how established incumbent market players can often fail to compete with less established disruptive innovators. The recommendation Christensen (1997) provides is for organisations to be pro-disruption and for large companies to create and support small nimble divisions to engage in, rather than ignore, disruptive innovation themselves. King and Baatartogtokh (2015) critique Christensen's theory as too simplistic to explain entire shifts away from incumbent players but highlight that "the theory of disruptive innovation provides a generally useful warning about managerial myopia".

U.S. military entrenchment in scientific research through the Cold War period.

Whittaker (2021) highlights the concentrated dominance of large technology corporates when it comes to modern AI, with industry actors holding the most data and knowledge on the technology and the role of deploying and profiting from the technology. This scenario creates a challenge for the state, and their defence establishment, who are dependent on civilian AI innovation to a greater degree than with other less mature technologies (Christie, Buts and Du Bois, 2021). This challenge is compounded by the government's relatively modest relative purchasing power and increasing commercial sector funding sources (FitzGerald and Parziale, 2017). Industry dominance over AI innovation creates a challenge for communities, including academics, who want to research the consequences of AI but are faced with a lack of access (Cummings, 2017). This landscape deprives community advocators and academics of the knowledge and ability to challenge the industry actors responsible for designing and deploying AI-enabled systems (Cummings, 2017). These assertions are supported by a review of AI media coverage in 2018, which highlighted that 60% of news articles were indexed to industry products or announcements (Brennen, Howard and Neilsen, 2018). The review concluded that a greater variety of voices were needed to contribute to a nuanced debate on AI, including academics, politicians, civil servants, and activists, to actively engage with and complement private sector perspectives.

A survey of papers from 57 highly regarded computer science conferences revealed that since breakthroughs in DL in 2020, there has been a significant rise in participation from large technology firms and elite universities (Ahmed and Wahed, 2020). The survey also concluded that large technology firms primarily collaborate with elite universities, further contributing to a de-democratisation of AI research. Industry innovators also shape the kinds of innovation, focusing more on narrow AI and short-term benefits (Klinger, Mateos-Gargia, and Stathoulopoulos, 2020; Ahmed and Wahed, 2020). Such approaches may entrench AI technologies that are sub-optimal long-term or neglect longer-term research considerations (Klinger, Mateos-Garcia, and Stathoulopoulos, 2020). The computation resources required to research modern AI are another barrier to innovation (Ahmed and Wahed, 2020).

While commercial innovation may dominate in cutting-edge AI, military innovation in AI may result in a spillover of technological applications beyond the military context. A survey of defence patents revealed how dual-use technologies emerged from the military into civilian applications (Engers, 2013). By the 1990s, the prominent view among policymakers was that civilian technology was more

advanced than defence technology, with military research and development of limited interest for commercial industry (Cohen and Foray, 1995; Council on Competitiveness, 1991). Cowan and Foray (1995, 851) challenged the observed consensus in 1995 that the military had little to offer the civilian sector in terms of research and development as "too simple" an argument. Instead, Cowan and Foray highlight the advantages of military R&D compared with commercial research: militaries are more willing to experiment to achieve a marginal edge that contributes to military advantage (1995), and at the time highlighted the optimism of the DARPA as an "explorer of young technologies" (1995, 863). While DARPA continues to invest in fundamental AI technology research (Horowitz, 2018), private sector actors are achieving dominant gains in cutting-edge technical innovation (FitzGerald and Parziale, 2017). The path forward for military innovation is argued to benefit from a multistakeholder and not a purely state-funded approach (Sisson et al., 2020). Where there are collaborative military-civilian research environments, military research can benefit technological diversity (Cowan and Foray, 1995), improving the commercialisation of the technology (Stowsky, 2004).

### 2.4.1 Non-military appetite for private-public partnerships

There is some question as to how willing non-military innovators, such as academics and private sector organisations, are to work on military projects. Academics may not trust the state not to subvert their work for military perspectives. Singer (2009) describes the historical example of a NASA engineer who was told their research on radars would be used to map Venus, finding out later it was implemented in cruise missiles. Academics may also refuse to engage because of the "slippery slope" where project funding may be conditional on scoping away from the academics' initial research motivations and towards a military setting (Singer, 2009, 172). This "slope" shifts the academic from not initially working on military applications to relying on military leaders for buy-in and funding (Singer, 2009, 172). Singer (2009) further reflects on historical examples of the academic refusal to do military work, with disputes between scientists and the U.S. military on the Manhattan Project,[26] and the U.S. and Soviet-based scientists working on atomic and nuclear bomb technology, who subsequently left and campaigned for nuclear disarmament. However, Singer (2009) acknowledges

[26] The Manhattan Project was a U.S. military programme set up in 1942 to develop the atomic bomb and represented a huge collaborative effort drawing on the scientific and industrial sectors. There are several instances in which Manhattan Project scientists expressed discomfort and argued against the goals of the project. For further detail see Price, Matt. "Roots of dissent: The Chicago Met Lab and the origins of the Franck Report." Isis 86, no. 2 (1995): 222-244l; Gosling, Francis George. *The Manhattan Project: making the atomic bomb*. Diane Publishing, 1999.

that these people are a "tiny minority" and that military funding is seen as necessary for many research departments. The defence, in this case, is that academics feel their research can be utilised for different purposes outside their control; the academics themselves can claim to stay out of politics (Singer, 2009, 173). Another complicating angle is the international aspect of academia (NSCAI, 2019, 13). The Interim Report for Congress presented by the NSCAI (2019) stresses how international collaboration, including with non-U.S. allies, is highly beneficial but also raises the potential threat of U.S expertise and technology being shared with their adversaries. This dynamic among tech workers, some have requested their employers limit or avoid engagement with the DoD – though others have embraced the opportunity to work for state-led security and defence reasons (NSCAI, 2019).

The U.S. wishes to maintain a first-mover advantage in the context of international security (Chadi, 2021). In this context, "first-mover advantage" refers to the concept of those who adopt AI first and thus benefit most from any technology advantages. In contrast, those behind them are nudged into the position of "technical dependence and military vulnerability" (Gilli, 2020, 43). However, while being the first to innovate gives the innovator a head-start in utilising that technology, the innovator has borne the brunt of the development and research costs, committing to developing the technologies even when they are uncertain of its success (Singer, 2009; Horowitz, 2018). Late adopters may benefit from hindsight and the ability to "free-ride" on the first mover's investments, copying promising initiatives and leveraging the first-mover's initial success. Unlike an investment in tanks, or the atomic industry, once initial AI research has delivered value the resulting innovation can be cheap and highly scalable (Singer, 2009). It is, therefore, difficult to maintain one's first mover advantage as capabilities rapidly diffuse, *especially* when they are dual-use capabilities available via the private market (Horowitz, 2018). This dynamic has led to the situation in which adversaries can simply buy similar capabilities from the commercial sector or within the wider arms market (Singer, 2009).

Looking at possible mitigations to the spread of insecure AI applications or capabilities which might be misused by adversarial actors, options for arms controls appear limited. While there are calls to limit some forms of military AI developments, such as lethal autonomous weapons systems (LAWS) (Jeangène-Vilmer, 2021; Schuller, 2017),[27] there are a number of challenges identified in the literature on security and technological innovation. It is almost impossible, and certainly impractical, to prevent

[27] See also state submissions to the UN GGE. For more detailed discussion on LAWS, AI, and responsible deployment see Christie, Ertan, Adomaitlis and Klaus, *forthcoming.*

dual-use technological innovation from being accessed by adversarial actors or competitors (Stowsky, 2004).[28] It is difficult for governments to prevent the transfer of expertise or technology to adversaries, particularly where innovation occurs in the private sector (Kania, 2017, 40). Geist (2016) highlights that arms control agreements hinge on verification of compliance which is very difficult to enforce due to the nature of AI research. Arms control mechanisms may also be undesirable in principle: international collaboration on AI can be beneficial, especially in terms of safety and standards (Kania, 2017, 30). Cummings (2017) argues that with superior technologies often available in the commercial sector, controls amounting to a ban on autonomous military systems would likely be impractical.

Trade shows and exhibition spaces are also another way for communities to transfer knowledge, understand, and imagine defence technologies (McCann, 2011; Jackman, 2016). Jackson's (2016) observations of drone-focused trade shows found that such gatherings were valuable sites for industry manufacturers and service providers, policymakers, regulators, and academic communities to discuss the possible futures of defence. Beyond Jackson's work and, to an extent, Rech's (2015) observations on British military airshows, there appears to be no specific recent literature that employs this form of observation-based ethnography to gather perspectives on the military AI landscape via military-focused physical events.

### 2.4.2 The culture and management of military innovation

The field of military innovation literature has extensively discussed the dynamics of innovation in relation to civil-military relations, interservice politics (the relationship between military organisations within a state, focusing largely on resource scarcity as a model for innovation), intraservice politics (focusing on the distinct nature of, and often competition between, military branches within one state), and organisational culture (Grissom, 2006). Examining these structural models for military innovation, Grisson (2006) illustrates a largely top-down model of military innovation as senior civilian decision-makers and military service leaders determine the military's approach to change. While there is some limited evidence for bottom-up innovation, Grissom highlights the gap in conceptual literature on military innovation that are underexplored in the scholarship. Nonetheless, these models of military innovation highlight the complexity of factors and

[28] Stowsky uses this argument to argue that expert controls are severely limited in assuring national security goals and protecting IP, as limits on dual-use innovation, such as expert knowledge transfer or publication controls are unlikely to ultimately prevent adversaries gaining access to material (2004, 266).

structural norms that shape how states approach military innovation, and how approaches may differ within branches of one military. As one example, different branches within a military may be more willing or capable to adopt technology, whether due to inter-service rivalry (Grissom, 2006) or since as the Navy or Air Forces tend to be more reliant on technology, they tended to view technology more favourably than, for example, the Armed Forces (Mahnken, 2008).

How technology is adopted into military contexts is shaped in part by the organisational culture of the military in question (Kier, 1997; Mahnken, 2008). Reflecting on the U.S. over the sixty years post-World War Two, Mahnken (2008, 2) argues that while the relationship between technology and service culture is complex and often two-way in nature, generally the U.S armed services "have molded technology to suit their purposes". Kier (1997) proposes that the origins of military doctrine are also shaped heavily by military culture, as each military's organisational culture shapes how the military responds to threats. Applying these arguments to the modern context of military innovation and adoption of AI, the choices states have made to invest in AI are influenced by the organisational culture of their militaries. This may be shown through the U.S.' bold signalling through the release of the first defence-focused AI strategy by any state, which represents a continuation of the U.S self-positioning of military superiority as observed by Mahnken (2008). Such militaries institutions have cultures that effectively represent enduring "personalities", distinct between states (Builder, 1989). This difference between cultures can go some way to explaining why states might choose different approaches to military AI innovation.

Alongside the cultural norms across each military, and even *within* military branches (Mahnken, 2008), a series of organisational norms also shape a military's approach to technical innovation. Often the very norms of established practice within institutionalised structures make it extremely difficult to deliver effective product innovation (Dougherty and Heller, 1994). Within military institutions, Bickel (2018) argues that individuals have a significant role to play in shaping change, particularly those staff who create doctrine and are thus able to influence across levels of leadership in a military. However, Jensen's (2018) focus on the U.S. military describes large bureaucratic structures which can suppress individual creativity in relation to doctrine development. This aligns with broader arguments about how larger militaries can be disadvantaged by the organisational change required to accommodate change relating to innovation, with major doctrinal changes a challenge to bureaucratic institutions (Horowitz, 2010; Dougherty and Heller, 1994). Mitigating these organisational barriers to effective military innovation might be achieved through the use of dedicated "safe space" environments for professionals to think about innovation without fear of failure, and through

widespread advocacy of non-traditional ideas that proactively challenge bureaucratic cultures (Jensen, 2018). This need for institutional consideration aligns with broader recommendations to support legitimate innovation in large enterprises, where Dougherty and Heller (1994, 200) argue managers must "weave the activities of product innovation into their institutionalized system of thought and action, not merely change structures or add values".

Military AI innovation also requires a reconceptualisation of various assumptions. Critical literature focusing on the social construction of technology (SCOT) (Bijker et al., 1987) has highlighted that developments around digital technologies have forced a rethinking of how to effectively conceptualise information technologies (Baalen, Fenema and Loebbecke, 2016). Many theoretical approaches in information technologies, including disruptive theory (Christensen, 1997) do not address the organisational or cultural aspects of technical innovation (Baalen, Fenema and Loebbecke, 2016). Instead, Leonardi and Barley (2008, 160) describe and critique the assumption of "technology determinism" across much of digital technology discuss, a phrase suggesting technical forces shape behaviour. Instead, Baalen, Fenema and Loebbecke (2016) argue for an extended SCOT framework to place additional focus on themes including socio-digital contexts and human-machine interaction. Applying this to the context of military AI innovation, research published by NATO (STO, 2021), the U.K MoD (2021) and the U.S. DoD (2021) on themes including human-machine interaction show a growing acknowledgement of human aspects. As discussed in Chapters 6-8, the U.S. and U.K. have increasingly dedicated resources towards organisational restructuring though research efforts relating to "human aspects" of military AI innovation remain nascent (STO, 2021; NSCAI, 2021).[29]

The scholarship describing organisational challenges and the broader innovation environment was reflected throughout the interview findings and in discussions taking place through observant practice. The U.S., U.K., and NATO have acknowledged organisational challenges and suggested that it is a priority to address procurement challenges to best leverage commercial and dual-use solutions from the private sector (STO, 2021; DoD, 2019; NSCAI, 2021; Ministry of Defence, 2021). Several responses have emerged through this thesis as possible attempts to mitigate challenges to adopting AI, in ways that inform the second research question. For example, the U.S. DoD creation of the Tradewind platform intends to "streamline rapid procurement and agile delivery of AI capabilities for the Department of Defense... partnering with commercial, academic and industry

[29] Private conversation with Dstl staff, 2019.

partners" (JAIC Public Affairs, 2021, para 1-2) and is discussed further in Chapter 7. In terms of international activity, the creation of the NATO Defence Accelerator for the North Atlantic and the NATO Innovation Fund (Brussels Summit Communiqué, 2021) demonstrates two major developments of the Alliance's response to an innovation landscape where breakthroughs come from outside the public sector, discussed further in Chapter 6.

## 2.5. Mapping the security implications of AI in military contexts

The Malicious AI Report (Brundage et al., 2018) highlights that AI represents threats to digital, physical and political security via three key aspects. The first challenge lies in how AI facilitates the expansion, or scaling-up, of existing threats. Once AI can automate or improve on human processes, launching an attack will require relatively less expertise, time and intelligence from the adversary. Second, new threats are introduced, as AI systems will create new weapons for the adversary. Additionally, AI systems may be targeted by attackers. Third, and finally, a change to the typical character of threats as the nature of the threat landscape will be changed by AI. Increased precision and information processing also leads to a greater impact of misuse of these systems. It could lead to the exploitation of new flaws and exacerbate challenges relating to cyber attribution. The following sections offer an overview of the literature that describes these challenges and addresses the first research question, distinguishing between literature focusing on high-level strategic security themes and literature focusing on lower-level technical and operational challenges.

### 2.5.1 Strategic security implications

Strategic literature examining AI's impact on military and strategic stability is nascent and relatively recent (Payne, 2018; Johnson, 2021). Existing policy-focused literature focuses primarily on AI's technical, ethical or legal implications (Burton and Soare, 2019). Recent years, however, have seen increased interest in the theme of AI implications from international relations scholars (Johnson, 2019), many of whom have highlighted several strategic implications of AI integration into military contexts. Under the umbrella category of the international security landscape, the first subsection will look at strategic implications of stability and international order, including the conduct of warfare, "arms race" dynamics, and perceived adversary use of AI. The second subsection will look at the implications of technical challenges, human-machine teaming, and operationalising AI systems

reliably. The reflections in the literature are widespread though it is worth remembering the uncertainty of predictions of technology sets as broad as AI. AI will work differently between various applications (Soare and Burton, 2019). Horowitz et al. (2018) assert that AI can be more accurately compared to electricity than a particular type of application, and there is consensus with this view across strategic literature (Johnson, 2021; NSCAI, 2019; Cox and Watts, 2021).

A significant range of opportunities and challenges have been identified with the use of AI in military contexts. In focusing on the security implications of AI, this thesis recognises the significant advantages AI may bring in enabling greater autonomy, speed, and accuracy. The thesis reviews the significant concerns in the literature around the associated risks and potential negative implications of AI, especially when deployed prematurely or inadequately.

*Note D: Strategic terminology and AI concepts*

Strategic literature uses a range of terms to discuss AI concepts in the context of strategic security and the military, from "military technological dominance" (Kania, 2017), military AI (Holland, 2020), AI for defence (Taddeo, 2021; 2019; Boulanin et al., 2019) "AI in the military", AI in warfare (Payne, 2019; Johnson, 2019; Kania, 2017), and AI and national security (Allen and Chan, 2017; Sayler, 2020) to refer to phenomena around international competition and geopolitical tensions. There are different interpretations of these terms. "AI and National Security" may focus on domestic applicability, surveillance and border defences beyond an explicitly military focus. Few authors specify exactly how they draw the line between military or (broader) defence contexts. Researchers have sometimes established a discussion of AI in specific applications such as weapons (Scholtz and Galliott, 2018) and military robotics (Noorman and Johnson, 2014) rather than defining the environment in which the AI is deployed. Likewise, AI in the military is not limited to AI on the battlefield or in military operations, and this research draws on literature relating to broader issues on strategic stability. This thesis, therefore, uses the term "military contexts" to refer to the scope of my research, focusing on the military sphere in a way that extends beyond military operations to consider broader issues of strategic stability and great power competition.

Many of the strategic implications of military AI technology are not unique to this set of technologies (Johnson, 2020, 18). It is the destabilising effects of AI that matter, relating to the speed at which innovation is taking place, the speeding of warfare as enabled by AI, and the intersections of AI in a

context of international competition that pressures states to prematurely deploy unsafe AI capabilities (Johnson, 2020, 18).

While there is significant uncertainty as to the implications of AI applications in the military (Horowitz, 2016; Johnson, 2020), Johnson (2020, 28) highlights increased "fog and friction" as a likely "ubiquitous outcome" due to the complexity of deployed AI technologies in defence. Some academics go as far as to say AI may be "*the* revolution in military affairs"[30] (Finlan, 2020, 5, emphasis unchanged from source). In facilitating autonomous warfare not limited by humans, AI is believed to change the psychological nature of strategic approaches to war (Payne, 2018b). Some academics believe that once AI goes beyond narrow AI capabilities to achieve singularity, general AI (the approximate point at which AI equals or exceeds human intelligence) could happen as early as the 2040s and significantly exceed human cognition by 2070 (Finlan, 2020; Del Monte, 2018). The strategic implications for militaries have been described as radical (Cave and ÓhÉigeartaigh, 2018), profound (Payne, 2018), and revolutionary (Allen and Chan, 2017).

The literature describes how the opportunities afforded by AI offer a competitive advantage in warfare; having more information or more accurate capabilities provides an edge that marks technical superiority in conflict. According to NATO, AI "has the potential for revolutionary impact on NATO operations and capabilities" (NATO STO, 2020, 14). A significant factor here is speed, as AI is poised to accelerate the pace of conflict (Dufour, 2018; Johnson, 2021). However, there are substantial possible downsides should AI speed up warfare to the extent that humans can neither comprehend the environment (Singer, 2009) nor keep up to the extent that they can effectively make human decisions, resulting in a loss of control (Scharre, 2017). Johnson (2020, 16) further highlights the lack of debate on how AI may compound risks relating to accidental or inadvertent escalation in warfare. Errors by AI systems may be unpredictable and difficult to detect, especially as algorithms increase in complexity, particularly in a military environment, where adverse consequences could have safety-critical implications, including for human safety (Kania, 2017b).

[30] The term "revolution in military affairs" (RMA) refers broadly to significant military effects brought around by emerging technologies, concepts or organisational shifts (Raska, 2021), and has been discussed extensively within security studies and defence-focused research communities since the mid-90s (Raska, 2021). There are arguments across policy-focused literature that the implications of AI in military contexts equate to a RMA (Thornton and Miron, 2020; Raska, 2021), as well as alternate suggestions that AI technologies instead represent an *evolution*, not *revolution*, in military contexts (Fiott, 2017). This thesis considers an evaluation of these conceptual debates as being beyond the scope of this thesis but will refer to RMA-focused literature where content articulates *how* the implications for AI in military contexts represents potentially unprecedented change.

Much of the literature on emerging technologies and nuclear deterrence focuses on the destabilising impact of AI technologies (Johnson, 2020; Johnson, 2021, Cox and Williams, 2021; Boulanin et al., 2019; Horowitz et al., 2020). The way AI-augmented conventional capabilities speed up warfare also poses risks for nuclear-decision making (Horowitz, Scharre, and Velez-Green, 2019) in a way that increases the risk of "nuclear confrontation" (Johnson, 2020, 16) with growing reliance contributing to "catastrophic mistakes" (Geist and Lohn, 2018, 22). Prematurely adopting unverified and unreliable AI technology could have "catastrophic implications" in the context of nuclear conflict (Johnson, 2020, 17). While expert workshops revealed a consensus among nuclear policy experts that AI is expected to undermine nuclear stability by 2040 (Geist and Lohn, 2018), Cox and Williams (2021) have highlighted that the impact of AI will depend on its application (Cox and Watts, 2021). AI may contribute to stabilising effects, for example, through integration in arms control systems. In terms of potential mitigations to destabilising effects, international law should codify principles in such a way that requires humans to remain in the loop to be thus able to intervene with applications relating to nuclear deterrence and weaponry and to limit the deployment of autonomous nuclear-armed weapons to avoid unintentional nuclear escalation (Cox and Watts, 2021).

While the implications for asymmetric political warfare are significantly underexplored (Polyakova, 2018), researchers have highlighted how AI creates systemic risks for military and strategic stability while enabling disruptive military activity (Kania, 2017b). While various definitions of strategic stability exist, Bidwell and Macdonald (2018, 10) propose the most precise modern definition in line with Acton (2013, 121): strategic stability references the situation in which "neither side has nor perceives the inventive to use nuclear weapons first out of the fear that the other side is about to do so". The related concept of "arms race stability" is outlined by Bidwell and Macdonald (2018, 9) as the "absence of perceived or actual incentives to augment a nuclear force - qualitatively or quantitatively - out of the fear that in a crisis an opponent would gain a meaningful advantage by using nuclear weapons first". This discouragement reflects the recognition that deployment by one state may trigger deployment by another and cause a "repeating action-chain cycle" that is both destabilising and resource-intensive for all involved (Bidwell and Macdonald, 2018, 9).

AI systems that do not use human-based reasoning to conduct a strategy may act in ways thoroughly unexpected by human observers. This attribute has implications when such systems are deployed in military contexts (Ayoub and Payne, 2016). AI increases the speed of the observation, orientation, decision, and action (OODA) loop decision-making (Johnson, 2020, 17). In a tactical and operational sense, AI might be used to carry out a tactical activity such as storming an enemy - dynamically

coordinating and manoeuvring networked technology to achieve a stated goal. AI can also enhance strategic decision-making, processing vast amounts of data to update assumptions quicker than human counterparts, without fatigue or emotion. Payne and Ayoub (2016) highlight how AI is not susceptible to groupthink or human heuristics.

Payne (2018, 7) compares the U.S. statements on AI as a "third offset" against the description of nuclear power as the first "offset strategy".[31] He argues that while both nuclear and AI technologies provide capabilities to change both strategy and the balance of power significantly, AI is arguably more revolutionary. AI technology, Payne states, might be leveraged across a full spectrum of applications of force with the potential to impact all conflicts, from minor to the most intense warfare. Fiott (2017) highlights the security challenges the U.S. third offset represents to NATO: as the U.S. attempts to harness commercially developed technology to offset the growing military capabilities of Russia, China, and Iran, it risks widening the gap between the U.S. and NATO's European allies unless NATO is willing to engage in a similar strategy. The adoption of the NATO AI Strategy (2021) acknowledges the importance of capability development in a way that appears to address Fiott's (2017) concerns, detailing the Alliance's own interpretation of an offset strategy which aims to maintain technological competitiveness against adversaries. More broadly, the U.K. Integrated Review (2021) and NATO products including the Emerging and Disruptive Technologies Roadmap (2018) NATO 2030 Agenda (2021) and AI Strategy (2021) detail how the U.K. and NATO are engaging in military AI adoption to maintain military superiority. Committing to investment, innovation and increasing discussion on military AI themes, the documents inform an answer to the second research question as actors engage in innovation to risk being left behind, or relatively disadvantaged as adversaries adopt AI capabilities.

Continuing to reflect on the international landscape, Soare and Burton (2019) noted intense military AI-related activity undertaken by the U.S., China and Russia in particular. There is some identification of "near-peer competitors" for those at the forefront of military emerging technologies (Forrest, 2020), with Haner and Garcia (2019) listing the top AI-driven autonomous weapons systems competitors in the United States, China, Russia, South Korea, and the European Union. The most likely challenger to the U.S. military AI advantage is China (Johnson, 2021; Kugler, 2021; Horowitz, 2018). China's desire to innovate in this space, alongside the increasing capabilities of AI, is a cause

[31] An "offset" aims to compensate against disadvantages in conventional weaponry. The first "offset strategy" can be understood as nuclear weaponry, the second to the information technology revolution in the 1970s, and the third offset strategy relating to AI technologies (Payne, 2018). For further detail on the U.S. Third Offset Strategy see Chapter 7.

for concern and offers significant uncertainty on the future of military and strategic stability (Kania, 2017b; Chadi, 2021; Johnson, 2021). However, Kania (2017b) points out that strategic stability is in the interests of the U.S., Russia, and China, all of whom acknowledge the benefits of avoiding (unintended) escalation.

*Note E: Hype*

AI has been noted as having "a bit of a Hollywood problem" (Madhaven, 2016, para 1). Boulanin (2019, 13) highlights a "vast gap" between AI's actual capabilities and the expectations of the public and policymaker communities. Beyond the military sphere, scholars have accused media coverage of tending towards sensationalism both in massively over-hyping the potential benefits and simultaneously speculating with predictions of robot-fuelled nightmares (Craig, 2018). It is a challenge to disentangle actual AI capabilities from "mismanaged expectations and premature hype" in the defence space (Tarraf et al., 2019, 130), and such hype around AI makes it easy to "overstate the opportunities and understand the challenges" of "AI in the military sphere" (Johnson, 2020, 16). Scholars have argued for more research into the implications of AI, especially as it is increasingly deployed in high-consequence environments such as defence and security (West, Whittaker and Crawford, 2019). Chadi (2021) and Johnson (2021) argue that decision-makers must consider nuanced dynamics between the opportunities and vulnerabilities made possible by military AI technologies and that experts across academia and research institutions must investigate how military AI interacts with various security scenarios.

The situation is neither hopeless nor unsalvageable. With researchers keen to promote their breakthroughs, Payne (2016) argues that there is significant information and public debate on AI at a generalist level. This trend results in a greater range of available information for those interested in reading further, particularly compared to the more coordinated and covert nuclear weapons research of the 20$^{th}$ century (Payne, 2016). Brennen, Howard and Neilsen (2018, 5) highlight that while there is definite sensationalism in the media, with articles blurring the distinction between "what is actually possible and what is aspirational", many articles, including on LAWS, remain grounded in real-world events. Instead, their concern is that much public reporting is weighted toward industry products and concerns and, therefore, disproportionately amplifies the self-interested perspectives of industry actors (Brennen, Howard and Neilsen, 2018). The authors argue that these risks undercut critical public concerns and limit the perspectives of others (Brennen, Howard and Neilsen, 2018). Finally, the pressure to innovate is not solely driven by hype; a failure to "adapt to adopt" the opportunities afforded by many AI applications already implemented in commercial environments would be difficult to justify in international security terms (Soare, 2021).

It is important to note that hype as a broader phenomenon in military contexts is nothing new. Elhefnawy (2018, 16) urges readers to look at historical examples in which militaries have

repeatedly held overwhelming expectations for technological adoption, with new military technologies holding a "dismal" track record for delivering on their original promises. Commenting on the renewed excitement around military AI themes, Elhefnawy *(*2018*)* suggests the same exaggeration is happening again across popular discourse.

There is more room for academic exploration of interdisciplinary technology and international security implications (Dremliuga, 2020). For politics and international relations academics, AI represents a "mysterious and largely unknown" area related to great power competition and instability (Cladi, 2021, 1). Johnson's research, among others, are emerging efforts to tie discussions on the military implications of AI to existing (and to international relations academics, familiar) discussions on strategic stability, deterrence, and nuclear risk (Chadi, 2021). Dremliuga (2020) argues that systems of effective and non-controversial rules for military AI-enabled autonomous weapons systems AI are expected to hold profound implications for strategy (Ayoub and Payne, 2016). Reflecting on technological advances in 2009, Singer argued that "man's monopoly of warfare is being broken. We are entering the era of robots at war" (2009, 41).

*5.11 The AI Arms Race*

There are a number of references to an "AI arms race" in policy-focused literature on AI in military contexts (Slijper, Beck, and Kayser, 2018; Venema, 2021). In 2019 Soare and Burton (2019) noted an emerging arms race, particularly between the U.S., China and Russia. The pace of innovation and perceptions of how adversaries may use AI technologies often invokes the term "AI arms race" (Hoadley and Lucas, 2018, 17), though the phrase has been criticised for oversimplifying a complex, competitive landscape (Roff, 2019; Imbrie, Kania, and Laskai., 2020). Murrey (2020) describes international military interest as a "technology adoption race", pointing out that AI is not inherently a weapon and acknowledging that AI competition includes broader economic and civilian aspects. Similarly, Cave and ÓhÉigeartaigh (2018, 36) recognise concerns about a military AI race but frame the race for strategic advantage in AI as "a race for technological superiority". Roff (2019) and Scharre (2021) both argue that the "AI arms race" phrase is unhelpful - with there being no such phenomena summarising trends across the broad range of AI applications. Roff (2019) instead argues that AI is a tool within a vast arsenal and represents a capability always applied to an associated task and should thus never be considered in isolation. Roff's (2019) proposal of an 'AI+' framework centres on the fact that AI never exists in a vacuum, arguing that the technology - and its harms - must be understood in a task-specific manner. The considerations may be in terms of which AI applications

cause greater harm (Roff gives the example of autonomous weaponry) or in which AI applications enable malicious attacks to be generated or carried out on a much wider scale, for example, facilitating the creation of Deepfakes. Widespread understanding of AI technology through this lens, Roff (2019) argues, would assist designers and builders of AI tools to realise they are never engaging in unprejudiced, value-neutral projects that exist in a vacuum but are contributing products that affect individuals and communities.[32] Roff (2019) uses the "A+" framework to state that AI arms race rhetoric is unhelpful. Instead, there is international competition, with variations of technological proliferation and diffusion. Roff (2019) cautions against the use of potentially escalatory arms race language that unnecessarily securitises international AI innovation, arguing against assertions in the literature that claim international AI arms race dynamics already apply in military contexts (Geist, 2016; Thornton and Miron, 2020).

Jankowski (2021, 1) highlights an example of arms race dynamics in practice, highlighting how Russia's recognition of how fundamental emerging technologies will be to future military and defence activities makes joining the arms race "less of choice and more of an existential necessity". Taddeo confirms their view that an AI arms race is happening (interviewed in Venema, 2021, quoting Taddeo). Gilli and Gilli (2016, 83) similarly call for increased investments in unmanned and autonomous systems and associated countermeasures to "preserve its technological lead over friends and foes". Looking toward the U.S., U.S. military researchers have also advised efforts to "promote and protect" national emerging technology advantages "to ensure continued technology, economic, and military edges remain" (Forrest, 2020, 32). Scholars have argued that U.S. government-commissioned reports on military AI convey a clear message to other states that the U.S. intends to lead the AI arms race (Venema, 2021). Due to the dual-use nature of AI, if a state develops broader technical superiority in AI, there is the significant implication that it will also establish military dominance (Cave and ÓhÉigeartaigh, 2018). Fedasiuk, Melot and Murphy's (2021, IV-5) analysis of Chinese military investment suggests that within the next decade, the People's Liberation Army (PLA) will continue to develop AI, in particular, to erode U.S. military advantage, particularly in undersea warfare and through the disruption of military information systems. The analysis concludes that it is not yet clear how AI will impact military power dynamics in the Indo-Pacific region.

[32] "Value neutral" is used in this context to refer to an impartial situation which is free from the influence of beliefs, attitudes, or values.

Slijper, Beck and Kayser (2018, 36) highlight the need to prevent future arms race escalation, calling on the need for a memorandum relating to the development or use of LAWS and confidence-building measures to have the assurance that other states will not develop such systems. As well as calling for states to establish legally binding mechanisms to prevent LAWS from autonomously engaging to attack targets, the report recommends private sector commitments to clear policies and guidelines to dissuade industry engagement with LAWS development (Slijper, Beck and Kayser, 2018).

The debate on whether arms dynamics are present goes far back to the Cold War era (Wohlstetter et al., 1974). There is more debate than consensus in discussions on arms race theory. Glaser (2000) notes that theorists appear to fall into two camps: one side believing that arms races benefit state security when facing an aggressive adversary, versus theorists who believe arms races undermine stability and strain relations. Glaser (2000) critiques the lack of a complete theory to understand arms race dynamics, though the term's utility means it has carried through to modern discussions on AI arms races (Geist, 2016). Scharre (2018, 330) describes the principle in practice where "the main rationale for building fully autonomous weapons seems to be the assumption that others might do so", which risks becoming a "self-fulfilling prophecy". Scholars have argued that the AI arms race contributes to international instability in the context of insufficient international regulation relating to areas where AI would be deployed, for example, in cyberspace, making escalation more likely (Venema, 2021; Taddeo and Floridi, 2018).

While the emerging response so far highlights that states such as the U.S. and Russia are engaging with, rather than attempting to avoid, arms race dynamics (Venema, 2021; Jankowski, 2021), in a wish to remain competitive, the literature contains broad recommendations on mitigating AI arms race escalation in a military context. Taddeo and Floridi (2018) argue for the definition and enforcement of regulations for state behaviour in cyberspace. Taddeo (Venema, 2021, quoting Taddeo) argues that determining whether to use AI should be based on a risk-benefit assessment that assesses the proportionality and necessity of any activity. In the absence of effective arms control regimes, Geist (2016) calls on the international AI research community to come together to monitor technical developments relating to military applications of AI and to develop a security culture that transcends national boundaries. Geist (2016, 320) argues that in contributing to monitoring efforts, researchers can communicate in the form of diplomatic exchange to prevent the development of AI systems that can undermine stability, such as systems designed to challenge strategic nuclear deterrents. In primarily focusing on competitive dynamics between major powers, policy-focused

literature did not concentrate in any detail on how non-state adversaries may incorporate AI-enabled technologies for malicious use.

### *2.5.12 Speculation: widening capability gaps*

Some states have more resources at their disposal to adopt AI into broader military capabilities, and scholars have reflected on the implications of this. As the U.S. seeks to maintain their technological military superiority by investing heavily in emerging technologies,[33] there may be a widening technological capability gap between the U.S. and other NATO allies who cannot afford to innovate in this area (Fiott, 2017; Allen et al., 2017). This gap may introduce or deepen challenges relating to interoperability between alliances such as NATO and thus weaken collective defence opportunities (Fiott, 2017; Gilli, 2020). Dufour (2018) details this argument in a NATO policy brief to highlight that the communication between sensor, command and control, battle management, and broader offensive and defensive AI-enabled systems enable a form of "algorithmic warfare". In this scenario, the decision-action cycle is compressed "to such an extent that countries not connected to the system will be unable to keep up" (Dufour, 2018, 3). This circumstance challenges NATO cohesion if members across the alliance fail to prepare for the impact of emerging technologies such as AI (Dufour, 2018). Furthermore, with states utilising AI to facilitate asymmetric capabilities, AI may promote asymmetric warfare to benefit more powerful militaries who have successfully integrated AI capabilities (Johnson, 2019; Polyakova, 2018).[34]

Potential gaps do not necessarily relate to the financial investment states can dedicate to military or dual-use AI development. In a NATO Defence College policy brief, Dufour (2018, 4) highlights that smaller, more "nimble" states can use AI technology as a force multiplier to *close* capability gaps between NATO members. For smaller militaries, successful integration of AI and autonomous technologies means military forces can leverage their strengths to "punch in the aggregate" (Finlan, 2020, 16). This leverage is particularly crucial as Payne (2018, 9) argues that militaries that can successfully develop and deploy AI will gain a "dramatic increase in fighting power relative to those who cannot". The release of the NATO AI Strategy (2021) demonstrates NATO's response in this regard, highlighting several plans for dynamic adoption which include enhancing interoperability for the alliance and leveraging the adoption efforts of NATO and Allied agencies and is discussed further in Chapter 6.

[33] See Chapter 7 for more detail.
[34] Assymetric warfare is defined as "conflicts between actors with wide disparities in power" (Arreguín-Toft, 2005, 2).

*5.13 Perceived adversaries to Western states*[35]

Innovation in military AI is not occurring in a vacuum (Johnson, 2021) but in the context of perceived "great power competition" (US Air Force, 2019, 2, Hill; 2020; NSCIA, 2019, 6; Imbrie, Kania and Laskai, 2020, 6; Jankowski; 2021, 1; Cox and Williams, 2020, 81). The literature tends to focus on actions by U.S., Russia, and China when discussing strategic stability and power competition. Modern strategic competition between the U.S., Russia, and China has been described as more complex and unpredictable relative to the 20th century (Jankowski, 2021), as economic interdependencies are entrenched alongside strategic disagreements, leading to a complex, layered landscape of "simultaneous competitions" (Jankowski, 2021, 1). Both Russia and China have signalled a recognition that AI will significantly influence future conflict (NSCAI, 2021). There is evidence that both states are actively focusing on the applicability of AI for modern security and defence purposes (NSCAI, 2021; Soare, 2021; Kugler, 2021). Perceived threats from China and Russia are evaluated distinctly across academic literature and policy-focused reporting. This literature review will briefly summarise each state's approach to military innovation and security mitigations in a way that provides context for the second research question.[36]

*China's approach*

In a broader economic and technological sense, China has moved ahead of the U.S. in some areas of innovation (O'Rourke, 2020), and Kania (2017) predicts this power to extend to military and security power, particularly given China's stated intentions to dominate in military AI. Such ambitions have been framed within China's release of the "New Generation AI Development Plan", a national strategic roadmap on how China wishes to "build China's first mover advantage" and "lead the world" by 2030 (Horowitz, 2018, 45; Webster et al., 2017).[37] The plan lays out China's objective to enhance military-civil fusion to develop dual-use capabilities, which may apply in various areas within military and defence applications (Kania, 2017b). A report by the International Institute for Strategic

[35] The research for this thesis was completed prior to Russia's invasion of Ukraine in February 2022. The researcher acknowledges that the defence landscape has shifted dramatically because of this war. At the stage of submission of this thesis, the researcher has assessed that it would be beyond the scope of this research to reflect on possible updated interpretations of Russia's commitments in the field, with any attempt likely to offer only limited value in the context of a continuously unfolding landscape.

[36] It is beyond the scope of this literature review to describe in detail the innovation landscape in both countries. For effective overviews on China's activity to date see Fedasiuk, Merlot and Murphy, 2021 and Kania, 2017. For more detail on Russia's approach to AI in military context see Gokhberg, Sokolov, and Chulok (2017) and Edmonds et al. (2021).

[37] An English transition of the document is available via "New America", a U.S.-focused public policy think-tank, at: https://www.newamerica.org/cybersecurity-initiative/digichina/blog/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/

Studies highlights China's pursuit of dual-use technology and major civil-military integration (Nouwens & Legarda, 2018). The report argues that Chinese promotion of national innovation while simultaneously accessing foreign emerging technology has China well-placed to catch up to and surpass U.S. and European emerging capabilities in and beyond the military arena (Nouwens & Legarda, 2018). Burton and Soare (2019) highlight that states such as China enjoy a relative advantage compared with the U.S. and Europe due to greater control of industry assets, including data.

Chinese policy refers to "intelligentized warfare" to refer to the transformations AI will enable in military contexts and has many active investments and initiatives focusing on this area (Kania, 2017b). Regarding where AI systems may be integrated into military contexts, a CSET report reviewing Chinese approaches to military AI analysed 343 public AI-related contracts (Fedasiuk, Melot and Murphy (2021). The analysis revealed that China appears to be focusing on intelligence analysis, predictive maintenance, information warfare, and navigation and target recognition in autonomous vehicles (Fedasiuk, Melot and Murphy (2021). Nouwens and Legarda (2018) highlight the Chinese PLA's use of AI-enabled UAVs for reconnaissance and surveillance and autonomous vehicles in conflict environments and highlight the Chinese state's interest in automated weapons systems.

Fedasiuk, Melot and Murphy's (2021, IV-5) research shows a rapid intensification in these efforts in recent years, with China's investment potentially matching that of the U.S. military as of 2021. The Chinese military has achieved "extraordinary progress in procuring AI systems for combat and support functions", drawing on private Chinese corporates and the international market (Fedasiuk, Melot and Murphy, 2021, v).

This progress has several strategic implications identified in the literature. Fedasiuk, Melot and Murphy (2021) assess that by 2031, China's investments in dual-use AI are likely to erode U.S. military advantages in areas such as sea warfare, with possible implications for the power balance in the Indo-Pacific region. Their report highlights how China may exploit both international supply chain weaknesses and lapses in due diligence to access technology from the U.S. and other states, including the advanced chip technology market (Fedasiuk, Melot and Murphy, 2021). Their findings are reflected in other literature, which describes how China hopes to leverage AI to change the very nature of warfare (Guoning and Shoulin, 2010).

*Russia's approach*

Russia is engaged in various forms of strategic activity relating to AI innovation, including political-military competition for power (Jankowski, 2021). While few non-Russian language resources detail Russia's approach to military AI, Jankowski (2021) highlights Russia's military advances in AI, particularly maritime innovation. Russia has a national strategy for the development of AI through 2030, developed by state-owned bank Sberbank, and has created the National Defense Management Center, which reportedly uses AI for information processing purposes (Edmonds et al., 2020). The Russian military also utilises and innovates heavily in autonomous systems, with progress in robotics and swarm technology (Jankowski, 2021; Kania, 2017b). In 2018, the Russian Ministry of Defence published a statement on AI with recommendations, including calls for a new defence research campus to develop AI solutions for military users (Ministry of Defense of the Russian Federation, 2018). These proposals did not contain details or timelines on how such efforts may be coordinated (Bendett, 2018), and it is unclear how far these were implemented.

Overall, Russia is described as lagging behind the U.S. and China in AI capabilities (Petrella, Miller and Cooper, 2020; Markotkin and Chernenko, 2020) and broader military and economic terms (Gokhberg, Sokolov and Chulok, 2017). Jankowski (2021, 2-3) highlights how Russia has limitations on its technological innovation, including a preference for domestic suppliers, limiting the available market for technologies and international sanctions restricting Russian access to cutting-edge technology imports. Petrella, Miller and Cooper (2020) highlight the challenges of limited expertise in the military as Russia fails to attract and retain AI experts and highlight how reprioritisation of focus due to COVID-19 may further challenge investment. Nonetheless, Thornton and Miron (2020) highlight the Russian use of AI-enabled information warfare as an activity with significant strategic implications and is tantamount to a revolution in military affairs.

*Note F: Responding to China and Russia: recommendations in policy-focused literature*

The literature offers several broad recommendations to Western states to mitigate the strategic implications of Chinese and Russian state AI innovation in this context, mainly calling for the states to consider the implications of such activity. Nouwens and Legarda (2018) call on European states to start thinking strategically about dual-use AI innovation, bearing in mind China's activities and stated intentions on civil-military fusion, while Thornton and Miron (2020) cautioned against underestimating Russian intentions relating to AI in military contexts. Researchers recommend the U.S. continue robust export controls on, for example, semiconductor chip hardware (Buchanan, 2020; NSCIA, 2019). There is a recognition in the literature that once an AI innovation is realised, it can be difficult, if not almost impossible, to prevent technology transfer to others (Kania, 2017; NSCIA, 2019, 40; Christie, Buts and Du Bois, 2020, 8). Export control regimes can also be amended at will in pursuit of commercial interests, with Gilli and Gilli (2016) highlighting the example of French and British sales of autonomous cruise missiles, Scalp EG/Storm Shadow to Saudi Arabia. For example, France and the United Kingdom have exported their joint air-launched cruise missile, the Scalp EG/Storm Shadow, to Saudi Arabia. While Chapters 5-7 offer an overview of emerging U.K., U.S. and NATO responses to the changing landscape, the broad range of recommendations and gaps highlighted throughout the literature suggest that the landscape is still immature and in flux. States are prioritising adopting at pace to minimise the risk of falling behind adversaries, and the development and frameworks or restrictions for AI technology may or may not emerge as the landscape matures.

### 2.5.2 Technical and operational security implications

Buchanan (2005) comments that the last sixty years have resulted in cumulative technological breakthroughs, partly facilitated by access to more extensive and more detailed data collections (complemented by the "big-data" hype). As discussed in section 2.4, integrated AI capabilities are poised to hold implications across the board to the extent that Finlan (2020) predicts that by 2045, militaries without AI capabilities will be heavily disadvantaged against AI-supported adversaries. However, there are a number of technical and operational challenges relating to AI development and deployment that are unanswered at the time of submission of this thesis. This section offers an overview of current technical barriers and concerns raised in the literature on the operational risks of integrating AI systems into military environments.

*2.5.21 AI and technical challenges: cyber security*

This section examines first how AI may be used as a tool at a technical level - through tools, techniques and procedures. For cyber security, for example, AI is being used to significantly enhance cyber defence protections and develop sophisticated attack techniques and methods (Soare and Burton, 2019). Second, this section examines various challenges associated with current AI technologies, drawing on recent literature reflecting on the technical shortcomings of narrow AI.

There is a wide range of cyber security and computer science-focused literature exploring the technical implications of AI-enabled tools which enable or amplify attacker techniques, set out in Note G.

*NOTE G: AI-enabled cyber attacks*

AI capabilities can feasibly be utilised at each stage of a cyber attack's lifecycle, from initial reconnaissance to data exfiltration after a successful intrusion (Chomiak-Orsa, Rot and Blaicke, 2019). Jeong et al. (2018) highlight the uses of ML for cyber intelligence, surveillance and reconnaissance in closed military networks. Weaponised intelligent agents will be able to "remember" information gathered through reconnaissance to plan, infer and carry out the most effective attack strategy (Guarino, 2013). In automating non-trivial intelligence, ML requires less effort and time-staking research by the attacker, offering the chance to launch attacks on a grander scale (Guarino, 2013*;* Whyte, 2020). Specific AI-enabled tools enhance opportunities for attackers to learn how to evade detection while intelligence-gathering (Guarino, 2013) effectively. In addition to reconnaissance, AI may also be used to weaponise and deliver a malicious payload, enabling the actual carrying out of an attack. Nichols et al. (2017) highlight how fuzzing uses AI to test invalid and unexpected inputs to an application to identify a weakness, simplifying finding vulnerabilities significantly. Email delivery provides one example of AI-facilitated cyber exploits. While machine-learning spam filters are nowadays popular with cyber defenders, filters have resulted in an "evolutionary scenario" in which attackers evolve their tools to minimise detection (Guzella and Caminhas, 2009). In 2016, HoneyPhish (Gallagher, 2016) used natural language processing to create targeted spear phishing, while DeepPhish (Correa Bahnsen, 2018) demonstrated how threat actors might enhance their attacks using AI algorithms to bypass AI-enabled phishing detection. Seymour and Tully (2016) explore the idea of AI used to create an automated end-to-end spear-phishing tool on Twitter, using material taken from a target's timeline and shortened URLs to personalise effective attacks.

Overall, the breakthroughs explored in Note G contribute to an attacker's capabilities in ways that can automate and scale sophisticated-level cyber-attacks without requiring significant sophistication from the attackers (Whyte, 2020; Brundage et al., 2018). These threat scenarios should be of considerable concern to militaries and defence structures likely to be targeted by these attacks, many of which exist or are developing through prototypes (Whyte, 2020).

While researchers were quick to highlight the benefits in time-resource of AI for cyber defence teams, there is limited literature exploring how AI may enable malicious actors to effectively identify targets and weaknesses, or which actors may benefit from such technology. More broadly, outside the cyber domain, McDonald (2018) highlights examples where unmanned aerial vehicles have been re-

purposed by non-state actors for intelligence, surveillance and reconnaissance purposes. Reports for U.S. Congress underscore how adversaries are likely to engage in counter-AI techniques, with new techniques revealing new vulnerabilities, though does not go into technical detail (NSCAI 2019).

While AI both facilitates existing defence capabilities and introduces new possible offensive techniques (Brundage et al., 2018), there is a range of challenges to be considered relating to the security and robustness of AI systems, associated data, and related infrastructure. Difficulties include the lack of explainability, poor quality data or data poisoning efforts, vulnerability to attack or subversion of the algorithm, and adversarial AI techniques. Again, some of these challenges are not *specific* to AI systems; the security of an AI will depend on "traditional" cyber security defences of the algorithm and training data (Schneier, 2021). However, due to how AI systems learn, new vulnerabilities can be introduced, which should be considered by developers and policymakers (Brundage et al., 2018), including data poisoning techniques and adversarial AI.

*Note H: AI systems and training data: vulnerabilities to data poisoning*

AI systems are vulnerable to manipulation through data-poisoning, where an attacker inserts incorrect data into the ML process to undermine or manipulate its operations, for example, resulting in the misclassification of malicious activity and preventing early warning systems (Huang et al., 2011). A small amount of bad data can have a significant effect on an AI's performance, with one study finding that when 3% of the training data to an ML algorithm was poisoned, the algorithm's error rate worsened from 12 to 23% (Steinhardt, Koh and Liang, 2018). Adversaries could poison a data set by replacing data via inside actors, compromising a system to swap data, placing false data in public data sources, or adapting their behaviour to form a false precedent in collected training data (Geist and Lohn, 2018). These techniques represent a challenge, for example, in potentially subverting the AI-enabled cyber defence tools already incorporated into national defence postures (Whyte, 2020).

Critical theorists have also highlighted the challenge of inadequate labelling of training data, questioning how external researchers can trust an AI algorithm's categorisation of a "threat", such as an ISIS pickup truck, when they do not have information on how training data was labelled in the first place (Suchman, 2020).

Adversarial AI techniques refer to the malicious input methods designed to fool an AI algorithm (Tygar, 2011), which might include bypassing spam filters (Barreno et al., 2006), malware classifiers (Laskov and Lippmann, 2010) or anomalous network traffic detecting systems (Gardiner and Nagaraja, 2016). Several machine learning models, including state-of-the-art techniques, are vulnerable to adversarial examples (Szegedy et al., 2014; Goodfellow, Shlens and Gzegedy. 2014, Fink,2019). Particularly where an attacker can query a defender's algorithm, 'zero-knowledge black-box attacks' are enabled, and an attacker can test what inputs might cause data misclassification despite having no prior knowledge of the algorithm design (Biggio and Roli, 2018). Much like data poisoning, a small amount of malicious data can drastically alter an AI algorithm's performance. Adding noise to an image in ways that are not perceivable to the human eye has caused ML image recognition systems to misclassify an image with 99% confidence, highlighting the potential for deception (MITRE, 2017, 30).

Image recognition technology can bring significant advantages by using AI to identify an object. However, researchers have demonstrated how an algorithm's image classifiers can be misled into mis-recognising an item. Athalye et al. (2018) revealed how a neural network misdiagnosed a model turtle as a rifle. Such experiments proved the existence of adversarial 3D models and raise fundamental concerns that weapons may be hidden or falsified in the place of mundane, non-harmful objects, as an adversary desires. This demonstration builds on theoretical threat analysis of a neural network's vulnerability to adversarial examples, as Biggio and Roli (2018) highlighted. Such attacks may be carried out manually, autonomously, or using AI techniques (Tabassi et al., 2019). Adversarial examples remain an open challenge (Horowitz et al., 2018). However, some literature has highlighted examples of 'desirable' attacks in which adversarial ML can be used to defend civil liberties by giving citizens the chance to glean information from surveillance algorithms used by repressive states as they feed input in to determine the algorithm's methods (Albert et al., 2020). To this end, Albert et al. (2020) argue that adversarial ML techniques can be used for democratic ends to empower the subjects of such systems. However, in the military context, this visibility would likely be viewed undesirably by military staff explicitly tasked with maintaining security.

Once deployed, there are still risks in defending the integrity and confidentiality of an AI system and associated data. The lack of large military datasets for AI often leads to a reliance on third-party data sets, which adversaries might be able to access to investigate the defender's AI systems or use to benefit their own AI systems (STCTTS, 2019). AI is "brittle" and narrow, which poses challenges for unpredictable battlefield environments (Payne, 2016) and raises the risk of accidents (NSCAI, 2019).

More generally, AI algorithms are also susceptible to errors due to bias in their development or training data (Osoba and Welsner, 2017; Fry, 2018). Particularly looking at the implications of AI in the public sector, there has been a lively public debate on how these technical limitations of AI can have negative social implications. Several academics are writing popular science books on the negative consequences of big data and AI, particularly where algorithms influence or replace human-led decision-making (O'Neil, 2016; Fry, 2018). Recently, several academics have published books beginning to explore the strategic connotations of AI in military-specific contexts (Payne, 2021; Johnson, 2021). The implications of AI at a technical level are still explored, intuitively, primarily within computer science and cyber security literature (e.g., Athalye et al., 2021; Chomiak-Orsa, Rot and Blaicke, 2019). Within the context of defence sector innovation, Carlo (2020) argues for the development of risk mitigation tools to be implemented in and around an AI algorithm to prevent "any kind of bias" in the system.

AI can also be used to automate and scale techniques used in technical attacks and through information warfare efforts. For example, by learning which techniques are likely to bypass spam filters successfully, ML algorithms hold the possibility to increase the delivery of malicious emails. The use of created content through tools like NLP algorithm GPT-2 would allow for the mass-production of spam and phishing content,[38] leveraging information gained through reconnaissance in ways that can significantly automate the initial stages of a cyberattack against military infrastructure. Broader innovations in convincing fraudulent content generation include AI development of "deepfakes" (Citron & Chesney, 2019; Whyte, 2020), which consists of convincing faked content that may be used creatively to add credibility to social engineering techniques. Particularly if launched across channels, the broad applicability and increasing sophistication of deepfakes make it harder for defenders to detect and monitor malicious activity, enabling attackers to leverage the technology to achieve higher delivery success rates. This technique has implications for national security, particularly when such attacks may aim to compromise sensitive government or military systems, but also via the threat of destabilisation that widespread deepfakes may cause. The attributes of such AI-enabled cyber attacks may also facilitate "greater obfuscation and strategic misdirection" for national defenders (Whyte, 2020, 18), adding layers of confusion via the introduction of a greater frequency of more sophisticated attacks.

[38] GPT-2 is an open-source AI language model that uses deep learning to generate human-like synthetic text, predicting the next word based on previous words in a dataset. See more: https://openai.com/blog/better-language-models/

NSCAI reports (2019, 2020a, 2020b, 2021) have highlighted two vectors of vulnerability relating to AI, the first being what adversaries could do with AI technology and the second acknowledging the wide-ranging consequences if AI is not employed safely. Adversaries will likely develop countermeasures to deliberately compromise each other's AI algorithms and data (Kania, 2017b) Calling on militaries to address these challenges when integrating AI, Kania (2017b) encourages militaries to consider: robust testing for military AI systems; focusing on reliability and safety in uncontrolled environments; and creating redundancies in military AI systems, including systems for intelligence-related activity, to enable verification and evaluation of AI outputs and detect any errors.

*2.5.22 Operational risks*

Taylor (2002) raises several challenges around how intelligent systems can contribute to military decision-support systems, particularly relating to human oversight, which are considered just as relevant today. Such AI systems operate at a pace exceeding "the cognitive and physical ability of human decision-makers to control or even comprehend events" in combat (Johnson, 2020, 29-30). AI systems do not make decisions in a similar way to humans (Payne, 2018), and in complicating cognitive decision-making, AI enables warfare in a way that risks detaching and potentially undermining human agency (Jensen, Whyte and Cuomo, 2020).

Errors committed by AI algorithms in intelligence, surveillance or reconnaissance capabilities or command systems, for example, create new risks of misperception if the algorithm misidentifies a threat or fails to recognise critical intelligence and can contribute to misinformed decision-making (Kania, 2017b). Furthermore, there is the challenge of AI systems as "black box" structures, which describe how the opaque nature of modern ML techniques prevents humans from understanding the algorithm's decision-making process (Hua, 2019; Boulanin, 2019). NATO's STO notes that "integration of AI into combat models and simulation, enterprise systems, decision support systems, cyber defence systems and autonomous vehicles will allow for rapid and more effective human-machine decision making" (2020, 15). In theory, such integration means human analysts can delegate some tasks to AI-enabled systems to dedicate their time to tasks that require human insights. However, there are several associated risks associated with human-machine teaming. Automation bias is a peril of AI decision-making, as are the dangers of human deskilling. Human operators may not remember how to do tasks that have been delegated to AI systems and, therefore, cannot provide redundancy should the AI system turn out to be faulty (Venema, 2021). There are related risks of

overreliance in which operators may defer judgement to flawed computer systems (French, Kiju and Rose, 2019). For example, one study found that pilots who were given imperfect autopilot systems in flight simulations were more likely to both make errors and miss errors that were not explicitly highlighted by the system (Horowitz et al., 2018). In some cases, operators followed the system prompts to make decisions that directly contradicted their training (Horowitz et al., 2018). The combination of AI brittleness and unclear human control has had critical consequences. In 2003, American Patriot missile defence systems were involved in two fratricides, misidentifying allied jets as missiles (Boulanin, 2019). A subsequent investigation found that the operators and Patriot community were "trusting the system without question" (Scharre, 2018, 144). As early as 2002, a Dstl paper highlighted the uncertainties on how humans would interact with intelligent and automated systems in ways that preserve and optimise a human-centric system (Taylor, 2002).

Researchers have warned that maintaining caution and realistic human control in war is critical due to legal, ethical and operational risks (Horowitz, 2016). Human autonomy must be maintained within human-machine team structures (Venema, 2021). Related literature highlights the importance of "meaningful human control", or MHC, of AI systems in military operating environments (Taddeo, McNeish and Blanchard, 2021; Boardman and Butcher, 2019; Kania, 2017b). Broadman and Butcher (2019) highlight how humans must have genuine and sufficient control and influence over-AI enabled systems. However, Maas (2019) highlights that MHC may not be possible with such systems due to AI-enabled technologies speeding up activities and leaving little time for human analysis or any errors.

On the other hand, Payne and Ayoub (2015, 816) argue against the notion that humans would be removed from the battlefield in favour of powerful AI systems, with strategic decision-making retaining an "inevitable human flavour" to account for the calculations that human comprehension and free will make possible. MHC was a central concept during the UN's 2016 meetings on "Certain Conventional Weapons" and the use of LAWS. Moreover, the risks of AI detaching human agency in warfare, as systems increasingly augment or shape decision-making, must be mitigated through informed decision-makers, sufficient accountability mechanisms, and governance to prevent irresponsible integration of AI systems (Jensen, Whyte and Cuomo, 2020). Literature calls for defence establishments to invest in education and talent development for AI expertise (Christie, 2020; Kania, 2017; Cox and Watts, 2021).

AI systems are coded by human engineers and incorporate these engineers' assumptions, which may include biases (Johnson, 2020, 29). Research has highlighted the importance of diversity among AI researchers to mitigate potential bias (Kuhlman et al., 2020; West, Whittaker and Crawford, 2019). In particular, scholarship has emphasised that diversity within industry, particularly within large technology firms, lacks in racial (Kuhlman et al., 2020), gender (West, Whittaker and Crawford, 2020), and multidisciplinary terms. West, Whittaker and Crawford (2020) argue that industry has less diversity than academia.

There is strong support within the literature that maintaining human oversight via human-in-the-loop systems is preferable to fully AI-enabled autonomous systems, and concerns that the speed of AI-enhanced warfare may necessitate the removal of human intervention over time (Boulanin, 2019; Foy, 2013, Horowitz, 2018; Noorman and Johnson, 2014). Language in the literature also focuses on enabling better human decision-making through "AI-augmented" capabilities (Jensen, Whyte and Cuomo, 2021, 529), or "augmented intelligence" (AuI), 'supporting the decision-making rather than determining it' (GCHQ, 2021, 6). While remaining aware of the challenges of overreliance, scholars also highlight the necessary trust that users must have in the technology to rely on it appropriately (Venema; 2021; Taddeo, McCutcheon and Floridi, 2019). Kania (2017b) argues for the inclusion of potential failsafe "circuit breaker" mechanisms in fully autonomous military AI systems to allow for de-escalation in cases where the AI has caused unintended engagement.

There is very little detailed analysis in the literature on how AI may affect the distribution of the human workforce for military personnel. The non-military scholarship also speculates how AI may shift tasks from human analysts to AI systems, with limited consensus on how or when AI will replace human tasks in certain roles. Prominent economist and Nobel Laureate Joseph Stiglitz (2018) highlights disagreements between researchers on which roles might be automated, articulating that the scale of workforce reorganisation will depend on government policy. Frey and Osborne's (2013) analysis of 702 (non-military) occupations predicted that an estimated 42% were at risk of automation, with a strong negative relationship between both an occupation's wages and education attainment requirements and the occupation's risk of automation.

A range of operational challenges must be addressed to facilitate effective AI adoption in military contexts. Christie (2021) argues that the dynamic adoption of military AI capabilities requires modern and iterative approaches to technical development alongside adequate access to both data and skilled expertise. Integration of AI-enabled systems will need to account for operational risks relating to the

possible compromise or misapplication of those technologies. Scharre (2016) outlines how risks associated with AI-enabled autonomous weaponry systems could have disastrous consequences for human life, such as unintended fratricide or civilian targeting, should the system not work as intended. AI-enabled systems are likely to be susceptible to "normal accidents", accidents and close calls stemming from the set-up of AI systems, which is particularly challenging given the speed at which systems operate, limiting human intervention in error cases (Maas, 2019). There are different views across the literature on how aggressively states should pursue innovation versus taking a cautious approach to minimise accidents, instability, or the "immoral weaponization of AI" (NSCAI, 2019, 13). At the same time, there is a need to urgently consider the implications of AI as a technology trend and militaries should invest in risk mitigation or AI-derived errors or escalation (Kania, 2017b). There is significant evidence that the U.S. and U.K. have turned their attention to the reliability of AI; the U.S. has highlighted the need for stronger testing and evaluation frameworks for AI systems (Defense Innovation Board, 2019) while the U.K. Defence AI Strategy (2022) emphasises that AI should be deployed with care and only where its use is justified.

Finally, Gilli and Gilli (2016) also highlight bottlenecks to technical innovation in the form of limited resources. The advanced R&D required to develop intelligent systems requires highly specialised expertise, experience, and facilities. Even the U.S., France, and the U.K have struggled to adopt drones or autonomous weapons systems (Gilli and Gilli, 2016). This point counters arguments that military hardware spreads rapidly, a trend that was expected to be exacerbated by globalisation and technological advancements (Goldman and Eliason, 2003; Goldman, 2004).

### 2.5.3. National vs multilateral approaches

Examining the second research question, the literature highlights a range of activities as states and alliances attempt to adapt to, and mitigate the challenges relating to, military AI. These include the state's use of international mechanisms for collaboration, as the strategic implications highlighted in sections 2.5.1 and 2.4.2 often transcend national boundaries. Scholars have called on national security agencies to coordinate with like-minded nations and international alliances, including Five Eyes,[39] to approach military AI themes (Venema, 2021), relating to the implications of developing AI capabilities and norms for AI use in military contexts. There are several existing international

[39] The term 'Five Eyes', or FVEY, refers to the intelligence-focused alliance structure of five states: Australia, Canada, New Zealand, the U.S., and the U.K)

initiatives, from the U.S.-coordinated AI Partnership for Defense, discussed in Chapter 7, section 7.1, to research and capability-development programmes at NATO, as outlined in Chapter 6, section 6.1. At the level of doctrine and international norms, the NATO AI Strategy was agreed upon by members in Autumn 2021 and is discussed further in Chapter 6, section 6.1. International governments have discussed the use of LAWS since 2014 via the Convention on Certain Conventional Weapons (CCW) Group of Governmental Experts (GGE) (UNIDIR, 2018).

Policy-focused literature highlights how alliances may be able to enhance or leverage state approaches. Literature focused on AI in the military context often referred to NATO as a relevant coordinator of multilateral activity, including within NATO-affiliated reports (Gilli, 2021; Gilli, 2020; Dufour, 2018). Some challenges, such as international standards and military operability with other states, require collective thinking, which may be undertaken by organisations such as NATO (Pepe, 2020; Gilli, 2020). Gilli and Gilli (2016, 83) argue that it is in great powers' interests to share emerging capabilities with their allies to avoid duplication of effort and praise the U.S.' relatively liberal export policies that facilitate such technological transfers. Fiott (2017, 434) highlights the importance of "strategic unity" between European and U.S. allies, arguing that European states benefit from binding their strategic approaches with the U.S.' positioning to contribute to emerging international doctrine.

The literature also highlights that AI challenges multinational military operations in ways that military alliances must address in the near future. Lin-Greenberg (2020) notes that the resource and data-sharing required to integrate and use AI systems in military contexts create additional burden-sharing and operability challenges. Alliance decision-making will also be challenged due to the increased speed of AI-enabled warfare, reducing available time for discussions on military planning and courses of action (Lin-Greenbert, 2020). Lin-Greenberg (2020) argues that allies must establish multilateral guidelines, including standardisation guidelines, that promote data-sharing between partners and introduce technical measures to allow for data-sharing. This argument aligns with Dufour (2018), who stresses potential coordination challenges and argues that members of alliances must remain cogniscient of operability challenges and manage the adoption of AI to minimise a widening capability gap between allies. Furthermore, developing international norms and agreements on AI allows alliances to ensure that shared values and ethical stances can inform national approaches (Gilli, 2020). This literature suggests significant room for improvement in international responses to the military AI innovation landscape; while the NATO AI Strategy (2021) articulates the will to increase

interoperability and capability building in ways that may address calls in the literature, it remains to be seen how such commitments will be operationalised.[40]

There is little analysis in the literature on how states may balance international cooperation against national investment for dual-use and military applications of AI, and emerging national doctrine may offer some clarity over time. The U.S. DoD AI Strategy repeatedly highlights the U.S.' intention to engage with allies and partners, as discussed in Chapter 7. Nonetheless, some capability efforts and projects are unlikely to be shared with allies. The French military AI document explicitly sets out the attention to be self-reliant in military affairs, including military AI capabilities, setting out a commitment to preserve "technological sovereignty" to "ensure the confidentiality and control" of information (Ministere Des Armées. 2019, 10).

## 2.6. Encouraging responsible development and use

There are significant unsolved ethical and legal challenges relating to the use of AI technologies in military contexts (Morgan et al., 2020). In the context of any perceived arms race dynamics, the pressure on states to rapidly develop and integrate AI capabilities alongside the lack of established norms on responsible use of AI in military contexts risks a "race to the bottom" with unsafe or unethical systems being implemented (Morgan et al., 2020*).* This section informs a response to the second research question, outlining actors' attempts to promote responsible practices relating to AI in military contexts, namely through the applicability of international law, ethical frameworks and principles, and norms development.

Within discussions on the application of international law, there is discussion across the literature about how AI in warfare impacts the principles of laws and norms, including how technologies like autonomous weapons systems operate concerning just war principles such as *jus in bello* (Bode and Huelss, 2018; Horowitz, 2016). Recent years have highlighted an emerging field of literature analysing how the military might approach ethical challenges. Ethical risks include the "morality of AI decision-making" as operators defer moral judgement to AI-enabled decision-assistance capabilities (French et al., 2019). Horowitz (2016, 33) highlights the risk of autonomous weapons systems enabling an operator's "moral offloading of responsibility" in ways that may undermine human dignity. Many discussions around the ethical and responsible use of AI in military contexts

[40] As discussed further in Chapter 6.

are linked with implications such as safety, control, and operational risks as set out in section 2.5. This section focuses specifically on initiatives to determine and facilitate the responsible use of AI in military contexts.

### 2.6.1 Drawing on LAWS debates

Discussions about the legal and ethical use of AI in military contexts and warfare often focus on the operational risks and ethical implications of LAWS (Kania, 2017). There has been "extensive and burgeoning" discussion within LAWS-focused literature (Scharre, 2016), much of which focuses on ethical and legal implications (see Anderson and Waxman, 2017; Asaro, 2012; Scharre, 2016; Taddeo et al., 2021; Morgan et al., 2020), and this offers some insight into the implications of AI more generally. This section will draw on the literature on LAWs and emerging literature on AI in the broader military context.

*Note I: The applicability of existing international law to LAWS*

While there is no specific legal framework addressing AI-enabled military activity, extensive literature reflects on how international law, including international humanitarian law (IHL), applies to LAWS (Arkin, 2008; Wagner, 2014). Often such literature draws on philosophical and legal concepts to focus on fundamental principles of international law, including proportionality and necessity, and discussions on how far national Rules of Engagements may sufficiently cover AI-enabled systems, including LAWS (Arkin, 2018; Jackson and Kuenzli, 2018). States party to United Nations Convention on Certain Conventional Weapons-coordinated discussions on LAWS (2015) have stated that international law applies regardless of the means of warfare and therefore applies to the use of LAWS. However, research has highlighted how the difficulties in judging an AI system's intent complicate the application of these principles, as existing laws typically assume human culpability and responsibility (Bathaee, 2018). There are different views within the literature on how to determine responsibility and accountability under IHL when LAWS operate autonomously and where human insight may be critically limited (Wagner, 2014, Müller, 2021). Arkin (2008, 2018) argues that gaps may be mitigated somewhat through active consideration of responsible design and employment. Similarly, debates within the literature hold different perspectives on how IHL relates to the challenges of MHC, as discussed in section 2.5.22. Some researchers argue that AI technologies can adhere to international law, integrate MHC (Hua, 2019, Arkin, 2018), and potentially contribute to more robust IHL compliance due to better performance (Sassòli, 2014). In contrast, others feel less optimistic about the ability of LAWS to undertake the ethical judgements required to comply with the IHL principles of proportionality and distinction (Wagner, 2014). McDougall (2019) argues that LAWS should not be deployed until there is evidence of MHC=. It is also not yet clear how such AI-enabled systems, for example, challenge the tenets of just war theory (Horowitz 2016). Reviewing dominant discussions across the literature reveals no clear consensus on how a legal framework may be operationalised to account for LAWS and AI-enabled military systems in a way comprehensive enough to meet the above challenges but flexible enough to account for future technological developments.

Beyond state-level discussions through the CCWGGE avenues and government international discussions to set a national agenda, the literature does not refer to multistakeholder initiatives to collaborate on approaches to ethical or legal frameworks for military AI technologies. According to some researchers, there has been little publicly accessible debate focusing on autonomous weapons technologies (Haner and Garcia, 2019). While there are limited opportunities for civil contributions,

public support is generally weighted against the use of autonomous weapons. A poll of over 18,000 respondents across 26 countries shows that 61% oppose the use of LAWS (Ipsos, 2019). Reviewing discussions of AI ethics across UK media outlets, Brennen, Howard and Neilsen (2018) found that most coverage only goes as far as to reference the existence of calls for discussion on ethics rather than presenting a discussion with nuance. The researchers argue that such calls for ethical debate rely on the argument that AI's influence will be radically profound., and Brennen, Howard and Neilsen (2018) call for urgent consideration of the ramifications of the technology on everything from global trade to politics.

### 2.6.2 Evaluating ethical implications

The ethical implications have been identified within strategic studies literature and in broader multidisciplinary literature focused on the responsible deployment of AI in military contexts (Horowitz, 2016; Kania, 2017; Morgan et al., 2020), many of which have been outlined in section 2.5. Examples of ethical implications include: how AI may be deliberately used as an offensive tool to cause harm (Brundage et al., 2018); how humans may not have the time or information to understand AI systems, reasoning (Hua, 2019); the risk of "moral deskilling" of the military profession as AI and autonomous systems makes the human assessment of military virtues including integrity and honour redundant or devalued (Vallor, 2013); bias, where means errors in AI systems are not challenged by human operators or those in the chain of command (French, Lee, Kibben and Rose, 2019), and any way AI-systems may challenge compliance with international laws, including humanitarian international law (Arken, 2018; Taddeo et al., 2021). However, literature also highlights the possibility of AI having positive ethical implications if designed and deployed correctly. Horowitz (2016) presents the counterpoints that concerns around unpredictability may be ill-founded, with the military unlikely to deploy unreliable technologies. AI may reduce the risk of civilian casualties, for example, with AI simulation to predict collateral damage and therefore choose less harmful alternatives (Singer, 2009). The use of life image processing could help situational awareness and avoid mishaps, while AI-enabled unmanned systems can enable more accurate data collecting (Singer, 2009). Singer (2009, 398) cites the bombing of civilians in 1999 in Kosovo, where NATO pilots did not risk flying low enough to identify refugee buses correctly and instead bombed what they assumed were tanks. Singer (2009) argues that today's AI-enabled UAVs can avoid this mistake. Horowitz (2016) also points out the theoretical benefit of LAWS being less likely to kill unnecessarily, with humans both more likely to engage incorrectly due to rage, revenge, or error, and

less likely to have perfect accuracy compared with LAWS. Critical studies literature disagrees with several optimistic accounts of military AI-facilitated operations. For example, Suchman (2020, 175) argues that AI will exacerbate discriminatory and indiscriminatory targeting while being politically and legally unaccountable, arguments reflected in the discussions of responsibility and MHC as discussed in section 6.2.1.

### 2.6.3 The private sector's approach: responsible AI innovation

Considering the second research question, the past few years have seen the proliferation of statements on how commercial organisations intend to approach AI innovation ethically, showing how the private sector appears to be developing its own ethical guidelines for AI innovation. In 2018 news of Google's cooperation with U.S. DoD Project Maven[41] prompted backlash from Google's tech workers and across the U.S. technology commercial sector (Whittaker et al., 2018). Following Google's non-renewal of the contract and subsequent announcement of AI Principles' (Pichai, 2018), a range of technology corporations, including Facebook and Microsoft, introduced their own AI ethics boards and frameworks (Whittaker et al.,). The Institute of Electrical and Electronics Engineers (IEEE) adopted its code of ethics to include AI aspects (IEEE, 2018).[42] Researchers at the Association for Computing Machinery (ACM) launched a campaign to reflect on the potential negative impact of conducted research as part of a paper's peer review process (Whittaker et al., 2018; Hecht et al., 2018).

While acknowledging the utility of broader debate and stated commitments by such institutions, Whittaker et al. (2018) argue that self-imposed governance structures have limited utility. Self-governance initiatives allow institutions to pursue "ethics-washing", acknowledging challenges exist without any subjection to regulatory oversight nor any commitment to transform problematic processes (Wagner, 2018, 87). A behavioural study with software engineering students and professional software engineers found that participants were no more likely to consider ethical aspects of their work when explicitly instructed to abide by the ACM code of ethics than a control group (McNamara, Smith and Murphy-Hill, 2018). This finding raises doubts about whether codes of ethics represent a practical pathway to ethical design approaches (McNamara, Smith and Murphy-Hill,

[41] Project Maven was a rapid prototyping-focused programme focused on AI-enabled image processing. For more detail see Chapter 7.
[42] The IEEE is a registered professional association for technical engineers and relating disciplines. For more detail see https://www.ieee.org/.

2018). Whittaker et al. (2018, 31) advise against trusting corporate institutions and instead argue for external oversight mechanisms to be put in place and "a cultivation of ethical norms and values" across engineering and technology communities.

*Note J: Principles shaping the conversation*

Examining the principles of ethical design for AI, Green, Hoffman and Stark (2019, 9) found that "high-profile values statements", such as publicly declared organisational AI principles, are "powerful instruments for constructing and imposing a shared ethical frame on a contentious conversation". The researchers found that those designing the principles often co-opt the language of their critics and frame their AI ethics commitments to imply the inevitability of AI. In so doing, designers provide a deterministic view and deliberately limit the conversation by shaping the "moral background" for AI ethics discussions across communities (Green, Hoffman and Stark, 2019, 8). By setting the language of the debate, principle writers can designate experts to address the issues they have defined, with problems, therefore "shielded from democratic intervention" (Green, Hoffman and Stark, 2019, 8). Green, Hoffmann and Stark's (2019) work draws on design and business ethics disciplines. It is, therefore, most aligned with organisational AI statements and guidelines. However, the point on writers of self-imposed guidelines holding power to shape the conversations of external actors carries across to discussions on responsible AI principles at a national level.

### 2.6.4 International and state perspectives: responsible AI

Actors have also responded to the challenges associated with military AI under the banner of "responsible AI". At NATO, a 2018 Emerging and Disruptive Technologies Roadmap was designed to identify opportunities and threats of technologies at a policy level (NATO, 2018). Hill (2020, 149) argues that the document may have provided the basis for drawing out "some commonly accepted legal and ethical principles" on the military applications of AI technology. Beyond NATO, the European Defence Agency has several projects considering the ethical and legal use of military AI technology (Boulanin, 2020). The EU has increasingly been a forum for discussion on AI ethics, with the European Commission's work on 'Trustworthy AI' providing a strong potential starting point for talks on military AI ethics (Boulanin, 2019). A 2018 EU resolution agreed that 'meaning human control' was a condition of acceptable use of lethal autonomous weapons (European Parliament,

2018). The U.S. is the only state with an official policy on lethal autonomous weapons, with directive 3000.09 forbidding autonomous engagement (Slijper, Beck, and Kayser, 2018; DoD, 2012).

Beyond research initiatives and ongoing discussions, there are existing examples of state commitments to responsible AI principles. The U.S. DoD has formally adopted five ethical principles for AI in defence (2019). The NATO AI Strategy includes six principles,[43] while the French armed forces have also published a document that discusses AI ethics (Ministere Des Armées. 2019). The U.S. more generally has the most state-affiliated reports on military AI that refer to ethical themes, including the NSCAI Congress-commissioned reports, which discuss ethics concerning U.S. values, including freedom and democracy (NSCAI 2019; 2020; 2021). This discussion is, of course, not limited to NATO members or allies; China has published several documents relating to AI governance (not specifically for defence) (Roberts et al., 2019). In July 2019, China's Ministry of Science and Technology published "Governance Principles for a New Generation of Artificial Intelligence: Develop Responsible Artificial Intelligence", which listed eight principles: harmony and friendliness; fairness and justice; inclusiveness and sharing; respect for privacy; security and controllability; shared responsibility; open cooperation, and agile governance (Laskai and Webster, 2019). China has developed national legislation, including the development of several voluntary academia- and industry-focused frameworks and the National Ethics Committee on Science and Technology, which has oversight of AI development (Kania, 2018). In 2019, the UN CCW GGE on LAWS, for which China and Russia are participants, endorsed 11 guiding principles for the use of LAWS (GGE Report, 2019). However, Kania (2018) highlights that China's definition of LAWS is ambiguous in describing technology that is both impossible to terminate and which would kill indiscriminately, while Russia has argued against what it considers "excessive regulation" (Jankowski, 2021, 5). GGE discussions have revealed a diverse and disparate set of opinions on how IHL applies to LAWS and how states believe challenges should be mitigated (GGE Report, 2018). As highlighted through submissions to the GGE, the distinctions between states highlight the complexity of agreeing to appropriate frameworks or principles. Furthermore, Taddeo argues that current approaches and discussions relating to national military AI doctrines are insufficient in not addressing shared values and principles relating to human dignity, human rights, and just war theory (Venema, 2021, *interviewing Taddeo*).

Despite the increasing proliferation in interest and investment in military AI innovation, there are few examples of proposed ethical frameworks for AI in military contexts (Taddeo et al., 2021; Wasilow

[43] See Chapter 6 and Chapter 7 for more detail on the U.S. and NATO respective approaches to responsible AI.

and Thorpe, 2019). One proposal by Taddeo et al. (2021) draws on the DoD's five ethical principles for defence (2019) and identifies five principles that should be implemented in coherence with existing laws and regulations: justified and overridable uses; just and transparent systems and processes; MHC; human moral responsibility, and reliable AI systems. Wasilow and Thorpe (2019, 37-41) also propose an ethical assessment framework designed to aid policymakers, technology developers, and decision-makers identify and consider relevant ethical challenges, including compliance with fundamental just war principles, reliability and trust considerations, and societal effects. No other defence or military-focused ethical frameworks or assessment tools were found in the literature. Christie (2020) highlights the benefits of such frameworks when agreed on a national and alliance level, with principles offering a baseline reference point on national positions and, therefore, a starting point for future consultations with allies.

One way to limit the deployment of undesirable or overly risky AI systems is to codify certain principles, such as human control and oversight, into international law (Cox and Watts, 2021). In this way, laws can help best achieve the benefits of AI as a stable force (Cox and Watts, 2021). At the same time, rapid advancements in technology can often lead to legislative gaps while policymakers attempt to regulate and mitigate risks (Soare and Burton, 2019). The literature highlights several such gaps in debates about how AI may or may not comply with international law beyond the LAWS-focused literature examined above. Focusing on how AI could be used to facilitate unacceptable military deception, Chelioudakis (2017) found the Law of Armed Conflict rules were flexible enough to account for and remain unchallenged by deceptive AI machines available at the time. This finding contrasts with Anderson, Reisner, and Waxman (2014), who propose adapting the Laws of Armed Conflict (LOCA) to account for greater autonomy on the battlefield. They suggest focusing on the ends rather than whether a battlefield actor is a human, AI-machine, or a mix, focusing on whether any committed act is in line with the core obligations of LOCA (Necessity, Distinction, Proportionality and Humanity). A sample study by Canfil (2021) found that specifically worded laws can undermine the law's adaptability to emerging technologies and defends the application of suitably ambiguous, and therefore less brittle, international law as a mechanism to constrain misuse of AI.

Implementing international norms may de-escalate undesirable implications of military AI innovation, encouraging responsible AI capability building and setting out agreements on where AI should and should not be employed in conflict (Gilli, 2020). Taddeo argues that national efforts on AI principles development are insufficient. States must cooperate with others to develop shared values underpinning the use of AI in the defence context and set up mechanisms to uphold such values

accordingly (Taddeo as interviewed in Venema, 2021).

## 2.7. Conclusion

This literature review has reviewed the main themes and discussion points relating to AI innovation in military contexts. First, the literature explores how AI themes are understood and defined. The literature highlights a range of definitions that have yet to be formally agreed upon between state actors, despite an emerging convergence towards definitions that favour comparisons with human intelligence. Differing definitions highlight the widespread ambiguity of "AI" terminology, which may make conversations confusing and unproductive between, or even within, communities. This ambiguity remains an open challenge, particularly as technology continues to evolve and requires those writing about AI technology to define their conception of the term clearly. This chapter also sets out how the term is understood for the purpose of this thesis, drawing on the literature to understand AI as a set of capabilities drawing primarily on ML techniques, which can demonstrate "narrow" intelligence in constrained circumstances. This scoping of AI allowed this research to explore how and in which areas militaries are integrating AI capabilities and the near-term expectations on where AI will contribute to military operations and support functions.

Turning to a review of the military AI innovation landscape, the literature highlights the changing relationship between militaries and the private defence sector over time, particularly as cutting-edge AI research takes place increasingly outside the realm of military auspices. The literature also highlights how far organisations may or may not want their technology utilised for military purposes, with no clear consensus on how this will affect future military procurement of services. Analyses of military innovation tend to focus on the U.S., China, and Russia, with limited academic scholarship focusing specifically on, for example, the U.K. approach to military AI innovation. Primarily, strategic policy analysis focuses on themes including US-China competition (Kania, 2017; Imbrie and Kania, 2020; Horowitz, 2018) and the broader strategic lens through which innovation is perceived as a race for dominance.

The review of the implications of increasing AI integration in military contexts revealed a wide range of security considerations. Broadly categorising these security implications highlights a developing discourse in AI arms race dynamics, with debate across the literature on the existence and nature of an arms race escalation in this space. The literature broadly highlights key strategic consequences of

AI-enabled warfare as introducing a faster pace of warfare which has implications for decision-making. However, there are different views in the literature as to whether AI represents a *revolution* or an *evolution* of modern conflict. The literature overwhelmingly concludes that AI will have profound effects in leveraging other technological trends applicable to almost every aspect of military environments. While there was limited academic literature evaluating recent Chinese or Russian approaches to military AI innovation, policy-focused reports set out the understanding that AI innovation forms an aspect of great power competition. AI is therefore understood within the literature to have significant implications for strategic stability.

Looking at how AI is integrated in practice, a range of operational and technical implications were mapped out. Examining AI-enabled cyber activity, a range of literature highlights open challenges posed by AI techniques against opportunities for cyber defence teams to integrate AI-processing capabilities. Many technical and operational difficulties concerned aspects that also had an ethical lens, highlighting that AI may not be reliable or transparent, leading to negative performance and unethical outcomes. Having reviewed the range of strategic implications, a reflection on the role of alliances highlights that institutions such as NATO may have a significant role to play in addressing some of the challenges identified. With the recent adoption of the NATO AI Strategy, analysis has yet to come out on how its contents may be operationalised.

Finally, a discussion on the responsible use of AI in military contexts highlighted several debates across the literature. There is no specific ethical framework for military AI that has been adopted, bar both the five principles adopted by the U.S. DoD and the six principles agreed upon in the NATO AI strategy. While the literature contains two proposals for ethical frameworks considering AI for defence purposes, there does not appear to be an active debate on this topic – yet. Similarly, literature on LAWS has extensive open discussions on the applicability of international law to AI and autonomous systems, with several challenges still a matter of debate on how far and in what circumstances AI systems can comply with the principles of international.

In addressing the second research question relating to how actors are attempting to mitigate any challenges associated with AI, this review highlights a fragmented field in which the understanding and early responses to engage with military AI innovation are still emerging. For example, while there is a growing area of literature focusing on the range of implications of AI in military contexts, particularly in line with recent policy developments in the last five years or so, there are several gaps in the literature and AI research areas that are underdeveloped in both technical and political

disciplines (Soare and Burton, 2019). Research gaps have practical implications for decision-makers, developers, users and those on the receiving ends of AI technologies in a military setting. As AI fundamentally changes how warfare is conducted, policymakers must understand the implications of complex integrated AI systems (Jensen, Whyte and Cuomo, 2020). This conclusion has implications for doctrine and deterrence strategies, including Polyakova's (2018, 7) argument for updated Western approaches to deterrence that can better deter Russian "AI-driven asymmetric warfare".

The literature also highlighted a range of disparate communities contributing to the discussion, with literature stemming from academia, policy-focused research institutes and thinktanks, international organisations and national government materials, and industry-coordinated reporting (i.e., from MITRE). Despite acknowledging a shifting military innovation landscape, except for Morgan et al. (2019) and Tarraf et al. (2019), few publications explicitly reach out to gather expert perspectives. Except for Ahmed and Wahed (2020), there seemed to be little analysis of activity at trade shows, conferences or sites of expert community knowledge transfer. A UN summary document, "The Militarization of Artificial Intelligence", highlights the importance of connecting disparate stakeholders across communities, particularly private sector contributors, for discussions on the implications of AI (Sisson, Spindel, Scharre, and Kozyulin, 2020). Despite this recognition, the literature review shows that discussions are still taking place in silos, with little evidence of academic research surveying policymakers or industry perspectives on themes relating to the military applications of AI.

Acknowledging the state of available relevant literature at the time of research, this review highlights an increasing level of interest over time, particularly in the context of state military innovation and "great power competition" and the perception of arms race dynamics. Such evidence reveals that states are motivated to invest and adopt military AI in order to benefit from the advantages afforded by such technologies. While a more in-depth discussion of actors' engagement with military AI is described in Chapter 4 (for the U.K. context), Chapter 5 (NATO context) and Chapter 6 (U.S context), there is increasing evidence that states are dedicating funding and attention to technical R&D to mitigate some of the challenges associated with AI in military contexts. Recent years have seen the emergence of national initiatives such as dedicated military AI innovation programmes (including specific pushes to increase private-public partnerships to benefit from private sector innovation), as well as the development of international platforms for military AI discussions such as the U.S.-created AI Partnership for Defense. There are calls across the academic and policy-focused literature highlighting how NATO is beginning to approach the challenges of emerging technologies such as

AI. This review has also highlighted mitigation routes which are not being acted on but may nonetheless be productive, particularly where there are no currently accepted applicable legal frameworks or established common standards for AI in military contexts.

# Chapter Three: Methodology

## **3.1.** Introduction

This thesis employed a multiple-methods approach. This approach reflects the nature of the research, which includes the following challenges: the rapidly evolving and complex innovation landscape as highlighted in Chapter 2; the dynamics of military innovation and information classification that limited access to required resources; and the nature and placement of innovation which was typically implemented across sporadically dispersed silos. Research considerations, therefore, had to take such challenges seriously and continuously reflect upon potential impacts on the direction of the research. To reflect and analyse present trends and interact with real-world phenomena, I decided to explore and employ several qualitative methods for this thesis.

This chapter details the range of methods employed for this research alongside the justifications for critical research design choices before reflecting on broader overarching considerations, including grounded theory, ethics and responsible research, and the experience of conducting research with military and military-related communities. The latter reflects the challenges, including access and network building, and my positionality as a researcher in these spaces. The chapter also includes a reflection on how this research adapted to COVID-19 in terms of methodological changes and adaptation.

## 3.2. Grounded Theory

Ground-up research, driven by engagements with everyday field settings, is a qualitative research approach and framework that allows researchers to draw concepts from their data rather than applying pre-conceived concepts to the data to search for explanatory factors. It is thus, by its very nature, exploratory. This method allows for the creation of inductive theory through empirical study, which may involve qualitative or quantitative research methods. Glaser and Holton (2004) write that classical grounded theory avoids the restrictions on accuracy that occur via qualitative data analysis

methodologies due to the subjectivity of data gathered.[44] Grounded theory, they argue, offers a distinct, flexible methodology that enables evidence-based conceptualisation, guiding the researcher "from the first day in the field to a finished written theory" (Glaser and Holton, 2004, 4). It allows for conceptualisation from research evidence rather than applying potentially ill-fitting theories (Glaser, 2002).

Following the argument that grounded theory offers a path to theory formulation less fettered with subjective preconception, constructivist grounded theory offers a way to challenge assumptions and raise critical questions. Through "methodological self-consciousness" (Charmaz, 2016, 36), researchers engaging in constructivist grounded theory offers insight that yields new perspectives and reflects on the researcher's role and engagement with this research. For this research, this meant an awareness and conscious reflection on my role and attributes through interacting with my research subjects and research environments. This approach mirrors Hammersley and Atkinson's (2007) emphasis on the importance of reflexivity in social research, the concept via which researchers are aware of their orientations and their effect on their research environment.[45] Charmaz (2016) also raises the risk that grounded theorists may prejudice their research with inflated preconceptions about individualism, potentially failing to pick up on systemic factors and influences, ideologies, or power arrangements. Characteristics include a researcher's self-individualism, and Charmaz (2016) promotes methodical self-consciousness as a reflective case, examining their privileges, positions, and priorities as they go through the research process. My role as a civilian, white, female doctoral researcher in my 20s formed some of the basis in which interactions played out with practitioners and researchers in the field. There will have been other numerous attributes that impacted each interaction. Engaging in reflection allows for critical evaluation not only on the information gathered but on how it was interpreted, as I worked to set any preconceptions aside.

Glaser and Holton (2004) outline the virtues of grounded theory as a general method that transcends time, space, or people and instead works to uncover social patterns that the researcher may not be aware of or understand. Grounded theory works to reveal the complexity of the social world that may be applied to different research methods, including experiments, surveys, content analysis and all content methods (Glaser and Holton, 2004). In this way, the application of grounded theory can be

[44] Qualitative data analysis (QDA) refers to the procedures used on collected qualitative (non-numerical) data to organise and transform the data into a form of explanation or interpretation. For a thorough guide to QDA in research see Grbich, 2012. For further discussion on QDA and grounded theory see also Glaser (2002).
[45] Further discussion on researcher reflexivity is found within sub-section 3.2 (Narrative Literature Review) in this chapter.

utilised through each research method. I used grounded theory both in designing my research and through the observant practice and interview methods employed in my research.

Backman and Kyngäs (1999) highlight key challenges novice researchers face when using grounded theory. One main concern is how far the researcher needs to familiarise themselves with the topic before commencing a study. The researcher, they contend, must be able to understand enough to be able to outline the research phenomena without allowing this knowledge to direct the research in such a way that limits open-mindedness to new perspectives and explanations. The same research proposed that novice researchers must learn to "bracket", or self-categorise, data, and suspend any knowledge held by the researcher that may prejudice their approach to their research (Backman and Kyngäs, 1999, 19). For example, I interviewed NATO colleagues within this research while simultaneously developing a familiarity with the broader NATO-focused environment through my Visiting Scholars position at the NATO CCDCOE. I consciously aimed to "bracket" the knowledge I gained through this experience, for example, on the perceived level of awareness of emerging technologies across NATO, when designing interview questions and analysing collected data. As another example, my previous experience and education mean I have familiarity with cyber defence terms. In interviews, I nonetheless welcomed interviewees to explain the terms themselves. This approach benefited this research by ensuring I had not made incorrect assumptions about some of the concepts the interviewees mentioned. Broad questions on fundamental themes also allowed interviewees the freedom to shape their answers rather than being limited to any of my assumptions. Considering the "bracketing", or self-categorisation, of data, Backman and Kyngäs (1999) acknowledge that as researchers are social beings that engage in social processes, previous experiences are data and can be used to understand the processes being explored.

Grounded theory has previously been applied to security-related domains. Focusing on cyber security, Halawah (2012) applied grounded theory to analyse e-commerce security perceptions while noting its popularity in the Information Systems domain to investigate social phenomena. Similarly, research into early-stage technologies has included grounded theory in determining adoption rates of the semantic web (Joo, 2011), employed to understand military leadership (Larsson et al., 2006; Jennings, 2013), and to understand how army units react to the dynamic complexity of the battlefield (Kramer, 2007). Kramer describes how the grounded theory approach suits the military example of dynamic complexity (2007, 16) due to the topic of interest being a new phenomenon, with little known about associated challenges. By analysing several case studies alongside an analysis of organisational characteristics, Kramer writes that they could "add up to insights", "the result of the reflection is a

substantive theory, which is a theory specific to the field of study" (2007, 16).

This thesis uses grounded theory applied to each research method, using the data to generate themes and conceptualise theory. In the emerging, disparate study of military AI, grounded practice delivers the opportunity to locate key research themes that have not currently been highlighted, categorised, or analysed in depth. I used grounded practice to inform and drive this research, directly drawing findings from the data rather than incorrectly applying potentially outdated or inappropriate hypotheses.

In the early stages of data gathering, it was not apparent whether I was asking the right questions. I had concerns for my credibility in front of practitioners who may feel my perceived research goals to be misaligned with real-world phenomena and priorities. Glaser (1978) raises this point as a natural reaction and concern of the researcher in cases where the researcher is not initially aware of essential matters and content, which may mean the research questions change during data collection.

## 3.3 Methods: overview and application

I designed the research questions, undertook a literature review that included the analysis of relevant key government and military documents, and used observant practice and semi-structured interviews to perform data collection. The data collected focuses on the United Kingdom (U.K.), the United States (U.S.), and the North Atlantic Trade Organisation (NATO), containing policy document analysis for additional U.S.-focused insights. This research used thematic and counter-coding techniques to analyse the data and unpack the research findings.

Image 3.1 shows a timeline of this thesis: the right-hand side shows the broad time scale of research design and data collection. The left-hand side highlights significant strategic and doctrinal developments over the same period. These developments outline milestones in which the U.K, U.S. and NATO created the relevant infrastructure that focused on AI in military contexts and the agreement of U.S. and NATO defence-specific AI strategies. These developments marked a significant reference point in clarifying each actor's approach to military AI. For example, the U.S. interviewees referred to the DoD AI strategy extensively when referring to U.S activity and objectives relating to military AI. The DoD's "Joint Artificial Intelligence Centre" (JAIC) was repeatedly highlighted as a key coordinator. The U.K and NATO interviews occurred before the announcements

of a U.K. defence-focused AI Centre and NATO AI Strategy, respectively. This research, therefore, captures perspectives at a particular moment in time before the emergence of a codified approach.

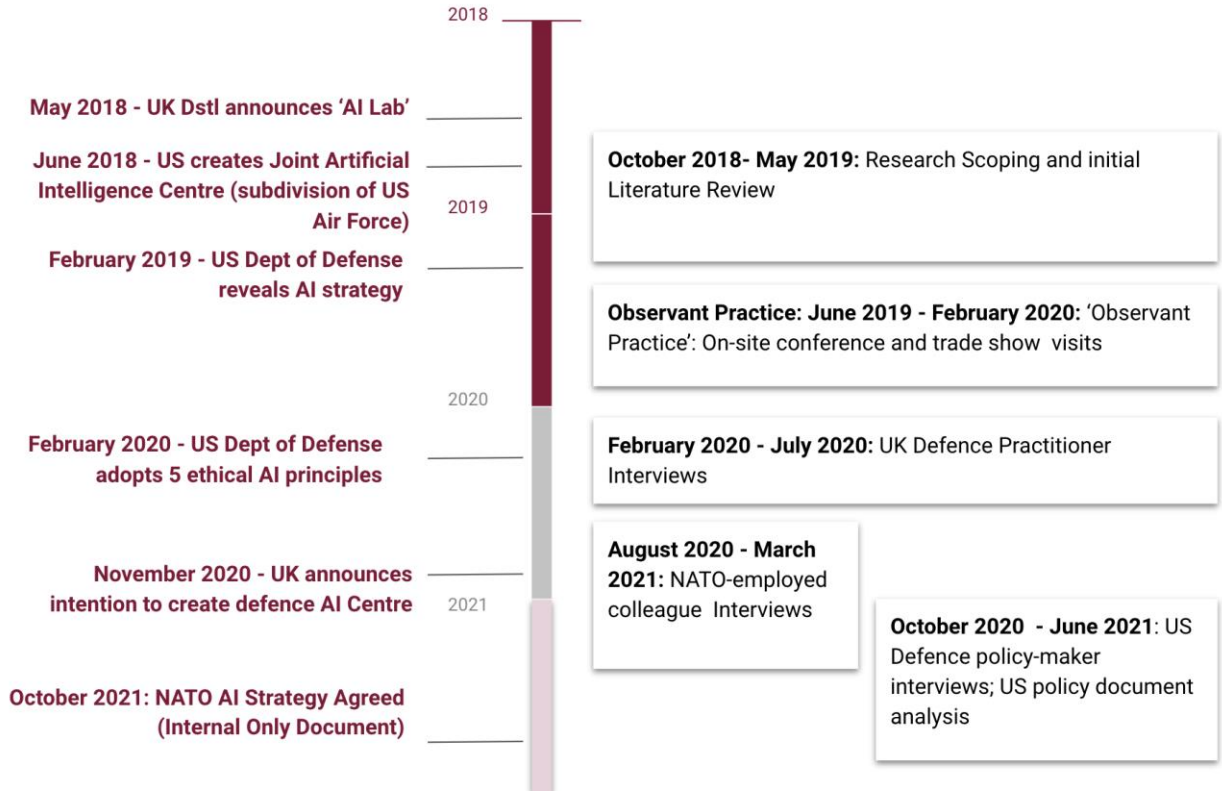


*Image 3.1: Methods and External Events timeline*

In approaching and engaging with each research stage, I employed a methodological approach that drew heavily on grounded theory principles. These methods allowed me to research the emerging field of modern military AI innovation and draw findings from analysed, minimising the risk of applying inappropriate theories in an environment that had not yet been explored in depth. The following section introduces each of these concepts and methods.

### 3.3.1 Scope and research questions

Strong research involves a well-scoped research problem and aims to answer clearly stated and important research questions (Frankel & Devers, 2000a). Bryman (2007, 13) challenges the notion that research questions wholly inform the research methods, challenging what they term a "textbook" approach that does not represent the real-world complexity when applying methodological approaches to research. For example, Bryman (2007) reflects on the responses of social researchers

who reported being steered by the research methods they know rather than the appropriateness of the research methods to the questions. This suggestion represents a bias that I consciously reflected on. I did not wish to limit my research by relying on methods I was comfortable with, but which may be less suitable than alternatives. Acknowledging the importance of straightforward research questions and this thesis' focus on emerging security challenges for the military helped me deliberately minimise any personal bias toward familiar methods. Additional research into previously unfamiliar methods, such as qualitative data analysis and "observant practice" ethnography in the field, helped inform and increase the extent to which methods were determined significantly by the research questions and broader research goals rather than by my previous experience as a researcher.

For instances where there is a rapid change in the topic area, where not much is known on a topic, it may be more appropriate to start with an exploratory research question and refine it as the research progresses (Frankel & Devers, 2000b). Researchers may even be recommended to do a literature review to clarify terms within the research question to ensure they understand the area thoroughly (Bufkin, 2006). This guidance does not go as far as Glaser & Strauss (1967), who suggest within their original proposal of grounded theory that the research question may be extremely high level to start with, firming up once data collection has begun.

Aligning somewhat with Frankel & Devers (2000b) and Glaser & Strauss (1967), the research questions for this thesis remained largely constant after initial scoping and the initial literature review. The focus was refined as initial data collection emphasised the importance of strategic and doctrinal responses to emerging military technologies. This approach impacted the formation of subsequent sub-research questions for projects undertaken as part of this research; once data collection from the U.K. interview series was underway, insights from the field shaped focused questions for the NATO and U.S. aspects of this research. Therefore, this research started with an interest in international security and AI and was refined to capture the approaches of the U.K., U.S., and NATO approaches (as described in section 3.3.2).

From initially approaching the thesis with a broad interest in AI, cyber security, and international security - in line with Glaser & Strauss' (1967) suggestions, I started with multiple possible research questions revolving around the theme of security implications of military AI-enabled technology. In pulling together my thoughts, I identified a knowledge gap in the strategic implications for AI capability-building in the military domain. I also identified a knowledge gap in how AI systems could be secured against adversarial AI, data bias, or traditional cyber-attacks against the system's

confidentiality, availability, or integrity. A few milestones helped narrow my ideas and present a final draft of the research questions. One was a conference on 'Machine Learning and Cyber Security' at Loughborough University in March 2019, in which I submitted a poster on the 'Fragmented Landscapes' of military AI discourse.

Early engagement with various communities at different conferences helped further refine my ideas. From engaging with academics in foreign affairs symposia to workshops with the developer community,[46] I continued to see different conversations happening in various fora as I thought about potential research questions in late 2018. The priorities for these groups were different again from defence-driven workshops at the Royal United Services Institute (RUSI), [47] London, at the military-led TriCyCles conference in 2019. By going out of my way to explain my research interests to different communities, I received insights from legal scholars, senior government defence staff, cyber defence practitioners, strategic researchers and developers, amongst other engaged parties. This engagement helped me narrow the research questions to include a partial explanatory lens, explaining the dynamics of military AI innovation in a way that transcended the silos I observed in 2018. Upon commencing data collection through observant practice and expert interviews, the research questions were solidified with a focus leaning toward strategic and non-technical implications.

AI technologies in military applications, alongside a corresponding range of ethical, legal, technical and geopolitical questions, represent a series of transnational challenges (Erdélyi and Goldsmith, 2018). Familiarising myself with the key themes of this research area began as a non-geographical exercise, reading about the nature of technological advancements and the capabilities of ML algorithms. I discovered early on that when considering the planned development and deployment of emerging technologies, it is challenging to talk genuinely without referring to a physical location. This aspect is true and amplified when talking about military affairs. The military refers to national armed forces and represents a force that inherently aligns with concepts of sovereignty and national security. When looking at how militaries develop, use, and deploy emerging technology, it became clear that international security theories would be necessary to make sense of diverging practices.

Furthermore, through reviewing the literature and starting data collection, it became clear that the output of some regions greatly exceeded others. Some of this phenomenon is a matter of resources.

[46] For example, at AI workshops at the Alan Turing Institute or with policymakers through Responsible Development conferences run by London's Digital Catapult.
[47] RUSI is a U.K. defence and security focused think-tank. See https://rusi.org/.

The U.S. was the most data-rich source of information when compiling my literature review, where publicly available resources showed that the U.S. was investing heavily more than any individual European nation. Secondly, the U.S. was also one of the most transparent when sharing information.[48] The sets of information released by the U.S. Department of Defense and the subsequent available research literature from other researchers in the field meant that the U.S. represented an attractive case study for this thesis. The opportunity to take up a predoctoral cybersecurity fellowship at the Belfer Center for Science and International Affairs, and the Center's Cyber Project, allowed me to develop a network that helped connect me to senior relevant experts as interviewees.

With the U.S. confirmed as one of the most transparent on the issues of AI in defence contexts, I hoped to find a comparison slightly closer to home. I ultimately settled on two other groupings: the other national state consideration of the U.K. and NATO's intergovernmental alliance.

The inclusion of the U.K. was, to a significant extent, a reflection of my position as a U.K.-based researcher. Conferences and expositions held in the U.K. were more accessible to me as a student seeking useful events close to home. U.K.-based military innovation events were also likely to offer at least one speaker with a U.K. defence perspective, whether as armed forces representatives, practitioners, or researchers. Being situated in the U.K. meant I could arrange conversations and follow-up meetings with relative ease and make the most of opportunities to travel to the Dstl sites or the sites of private contractors. In terms of the network I developed through the course of this research, at events, I frequently met U.K. government, military and ex-military personnel, and U.K.-based industry practitioners. The resulting focus on the U.K. is, therefore, an output of my physical and cultural position as a researcher based in Surrey for the first half of my doctorate – a half-hour train ride from London, with major event venues, military research sites and government departments all within reasonable commuting distance.

There were some excellent opportunities to observe and participate in a range of academic, military, and civil-military events looking at NATO (and the U.K.'s role within it), from RUSI (Whitehall, London), to CityForum working groups (BT Tower, London), to the annual CyCon event organised by the NATO CCDCOE. When looking at the future of warfare in strategic terms, exploring how states choose to share resources and commit to supranational strategic direction often led me to NATO

[48] Chapter 7 explores why the U.S. may have chosen to release so much material relating to military AI innovation publicly, relative to other national approaches.

research and resources. This aspect of my research was increasingly motivated by my initial U.K.-based interviews, in which interviewees highlighted the need for international collaboration to address a range of military AI challenges. At an Alliance level, I identified NATO as, in theory, a well-placed body to act in relation to military AI, particularly as it faces existing challenges around technological capability gaps (as examined in further detail in Chapter 6). Speaking with members of NATO's personnel is informative, even when interviewees highlighted a lack of activity or thorough institutional understanding of artificial intelligence in conflict. Highlighting how NATO approaches the issues of increasing AI in the military security context offers some insight, particularly looking at the medium-long term (2025 onwards) and how technology might influence international relations and the balance of power between actors. The opportunity to take up a Visiting Scholar's position with the CCDCOE further facilitated this research strand.

### 3.3.2 Narrative Literature Review

A narrative literature review provides an overarching point of view on a topic, linking together evidence from multiple sources, creating interconnections and building a bridge as a part of a valuable "theory-building technique" (Baumeister and Leary, 1997, 312). Compared with systematic literature reviews, which use rigorous methodological approaches to identify and solve specific and defined research questions, Baumeister and Leary (1997) argue that narrative literature reviews offer a broader, horizon-level exploration of a given research landscape.

There is a range of views on whether a literature review should be completed before commencing research (Creswell, 1994) or whether the exercise should only be undertaken as research is underway. Miller and Crabtree (1999) argue that starting to review literature only after starting data collection allows the researcher to remain open-minded to new ideas and identify which literature is most useful once they have developed an understanding of key theories and concepts. Devers and Frankel (2000b) highlight how researchers completing the bulk of their literature review before starting data collection rely heavily on prominent literature, including their own discipline or literature restricted to a specific topic, which may unnecessarily limit the scope of materials read. These views heavily influenced my approach to my research; while I attempted a literature review before data collection, I found that the vast proportion of it required re-writing as my research progressed. This update was largely due to the necessary and pragmatic aspect that much of the relevant literature was published, or released, during my research, as interest in military AI technologies exponentially increased. It was also partly

an output of conducting the review as a reflexive process and considered my increasing exposure in the field as a researcher. As McGhee, Marland and Atkinson (2007) argue, researchers engaging in grounded research must engage in reflexivity, limiting the researcher's biases towards the literature to avoid researcher distortion of collected data. This assertion was particularly relevant as literature was often published after the data collection had concluded (some examples include Konaev and Chahal, 2021; Taddeo et al., 2021; Jankowski, 2021; Johnson, 2021). I then had to ensure the scholarship included in the literature review was not unconsciously filtered to best align with the emerging findings from the data. At the same time, as I reflected on my increasing understanding of the field, I often felt that the literature review could be restructured to summarise the evolving debate better. This subsection outlines the process of drafting the narrative literature review.

Upon approaching the task of synthesising appropriate literature into a literature review in late 2018, my first challenge was to identify in which academic field the most relevant discussions were being explored. There was an element of a "chicken and egg" challenge to this approach. I could not determine which fields had produced the most relevant literature until I had searched through war and intelligence studies, security studies, international relations, geopolitics, computer science, cyber security, and critical science and technology papers. This challenge continued for about two months as the literature I worked through, while relevant, did not transform a thoroughly interdisciplinary project into a neat categorisation where, for example, war studies and cyber security could be brought into the conversation to support the questions I wanted to ask. The literature simply was not positioned or shaped that way. After a frustrating period, through which the most valuable literature came from state national security documents and secondary source think-tank analysis, I decided to alter my methods to a more grounded approach. This approach enabled the data to directly inform the literature that would eventually underpin the research and against which any research findings would be positioned.

After this scoping of the literature, I narrowed down my possible research questions to focus on two first introduced in Chapter 1:

1. What are the implications of AI innovation in military contexts?

2. How are actors attempting to mitigate challenges identified in relation to the development and use of AI in military contexts?

*3.3.22 Refining the literature review*

Literature (including academic, think-tank publications, and state strategies) was frequently released as my research progressed. The literature review became a continuous process alongside my fieldwork as the discourse on military AI expanded significantly, especially after the June 2018 creation of the Joint Artificial Intelligence Center within the U.S. Department of Defense and the February 2019 announcement of the U.S. DoD AI Strategy. The launch of the CyberAI Project in November 2019 represented the first formal group of researchers dedicated to the question of AI and national security, based at the Centre for Security and Emerging Technologies at Georgetown University. The CyberAI project produces research frequently, exploring issues at the intersection of cyber security, AI, and national security, including a focus on military contexts. In many respects, the literature review became much more intuitive to compile in part as I became comfortable with the research space, and new research became available throughout the research. By the second half of my research, I used my academic network, including my Twitter feed, to "follow" researchers and relevant institutions and remain updated on released publications. This engagement also led to the discovery of high-quality research, compared with my initial first-wave approach using search terms online.

This research includes a range of academic research, especially through the literature review, and draws on military and government reports alongside research from established relevant think tanks. Particularly where literature was not externally peer reviewed as part of an academic publishing process, I used my developing knowledge of the space to determine how reputable the publishing source (i.e., research institute) was and checked the contents of the research for quality and replicability. Particularly for technical papers, some academic pieces were listed on arXiV rather than in a journal. I checked where and how these papers had been referenced on Google Scholar to indicate quality in these cases.

All the reading cited in this thesis is unclassified and available to academics (publicly or through journals). To my awareness, I did not receive access to any classified documentation or information throughout the course of this research. This topic is examined further in section 3.5.1.

3.3.3 Observant Practice

Conferences can be valuable to the researcher in several ways. Keynote speeches set the tone of an event while networking spaces facilitate engagement with government, industry, or military figureheads. Lectures offer the chance to update oneself on recent research and perspectives, while workshops and roundtables, depending on the event, allow one to interactively work through scenarios while connecting with individuals who also have a reason to attend. Social activities outside scheduled presentations usually contain the "real" reason many delegates attend: networking over coffee, lunch, or at the bar, extending to scheduled and impromptu "after-party" events. The balance of each activity shifts on a scale between events; some conferences might stick to a more 'traditional' structure of multi-stream lectures, while others advertise corporate-sponsored drinks. Conferences are an intensely human experience: one could turn up, take notes at conferences, and leave, but in doing so, one would miss key aspects of professional connection that are often the key to information gathering in research. Hickson (2006, 465) writes that the main reason for professional conferences "has to do with becoming and remaining a professional in the discipline". According to Hickson (2006), conferences allow researchers to remain active in their field of study, with the chance to engage with the work of others, including the most current information in the discipline, explore interest groups, and become part of a relevant network. My own experiences of in-person conferences, workshops and security trade expositions, heading into and throughout the research (before COVID-19), highlighted how Hickson's (2006) points apply. The academics, NATO staff, and industry practitioners I met during this research's initial and early stages became "champions" for this thesis. They introduced me to relevant potential interviewees, directed me to relevant literature and timely policy announcements, and even co-authored with me on topics beyond this research scope but in line with my broader professional interests.

The importance of conferences goes beyond the event agenda. Having active researchers together in one place offers the chance to streamline ideas collaboratively; presentations and "after-conversations" may prompt further lectures, articles, or books contributing to the topic (Hickson, 2006, 465). Meeting relevant individuals at these events also offered a chance to gain candid insights, which provided a valuable broader context to my work; I could engage with conference presenters directly to request more detail on their presented research. For example, at CyCon 2019, Soare and Burton (2019) released the first paper that, in my eyes, approached the concept of weaponised AI in a strategic context; and I took great value in conversations with Soare to build on their approach for my research. While military-government closed-room meetings may be where many classified decisions are made, conferences offer for researchers to receive rapid and constructive feedback on their work (Hickson, 2006). Discussions are especially candid, given the protection of Chatham

House conventions designed to protect content and the right to anonymity.[49] This assurance was a theme throughout my thesis and broader cyber security research, for example, organising workshops in this way for the Offensive Cyber Working Group, an academic-led network that I co-lead, or coordinating a closed "Cyber Exercise" workshop at CyCon 2021. Organising these workshops to facilitate discussions according to Chatham House conventions[50] meant NATO, national military, and government attendees reported that they felt more comfortable expressing their views, with the latter benefitting from the reputation of the CyCon conference to encourage senior attendees.

As I progressed through my research, I was invited to events, including a closed workshop on military AI, organised by Ulrike Frank and attended by many prominent defence and security researchers actively focused on military AI themes and therefore cited through this thesis. I was also invited to present to government staff and security venues (including the Henry Jackson Society, King's Policy Institute, and a session at the European Defence Agency). Invitations to various meetings and informal gatherings offered a further, more intimate insight into the views of other attendees, including senior speakers and delegates.

At the same time, there are many other political, economic, social, and cultural dynamics occurring at events that attract government, military, industry partners, researchers, and other invested groups. Networking, public relations, technical explanations, and demonstrations occur in a comparatively small physical space. The atmosphere at these events might appear overwhelming simply because of the sheer amount of activity across different levels in one physical space. An empty exhibition hall becomes a regional hub of expertise and trade activity for one to five days, potentially attracting some of the most knowledgeable minds on a particular topic or all the relevant stakeholders involved in making valuable industry trade agreements. Jackman (2016) uses the experience of commercial drone trade shows to assess the value of events as a useful lens and to understand how a range of actors frame technology. Rech's (2015, 544) use of observant practice at military airshows highlights the benefits of exploring "visual cultures". Rech uses observation to explore spatiality, also referencing how the importance of an arranged event's physical space represents a meeting point for a range of states, militaries, and defence companies to display their technology and engage in "brokering defence deals out of sight of industrial or political rivals and civilians" (2015, 538).

[49] The Chatham House rule states: "When a meeting, or part thereof, is held under the Chatham House Rule, participants are free to use the information received, but neither the identity nor the affiliation of the speaker(s), nor that of any other participant, may be revealed." The rule is employed to encourage open dialogue.
[50] Participants' contributions were anonymised in any subsequent reports.

Effective fieldnotes can yield valuable insight through observant practice, noting the atmosphere, layout, language and social exchanges at these events (Emerson, Fretz and Shaw, 1995). How events attempt to focus their attendees' attention is also worth exploring; Rech (2015) notes that for military airshows, the overwhelming focus on technology downplays the known implication that these technologies, and the conflicts in which they are deployed, often have fatal human consequences. Rech's observation remains acutely relevant in military innovation relating to artificial intelligence, where the capabilities of algorithms or cyber-physical systems are often discussed admirably in conversations that refer to "increased lethality" as one of hundreds of metrics measuring success.

For this research, I chose to record observations on paper or a notepad app on my laptop, noting factual aspects of the events and some of my perspectives. Fieldnotes included short phrases, bullet-pointed informal observations, reflective thoughts on brief interactions, more detailed paragraphs noting details, quotes and presentation content, and more reflective interpretations of each environment and event. Given the subject matter discussed at these events and the audiences involved, I did not want to cause any disturbance by capturing photos, which was explicitly forbidden at more than one of the venues.

Image 3.2 highlights an example of the raw observations in written fieldnotes from CyCon U.S. 2019. In the notes, I generally used large side brackets '[' to signify my observations on the physical space or atmosphere and quote marks around direct quotes from speakers or other attendees.

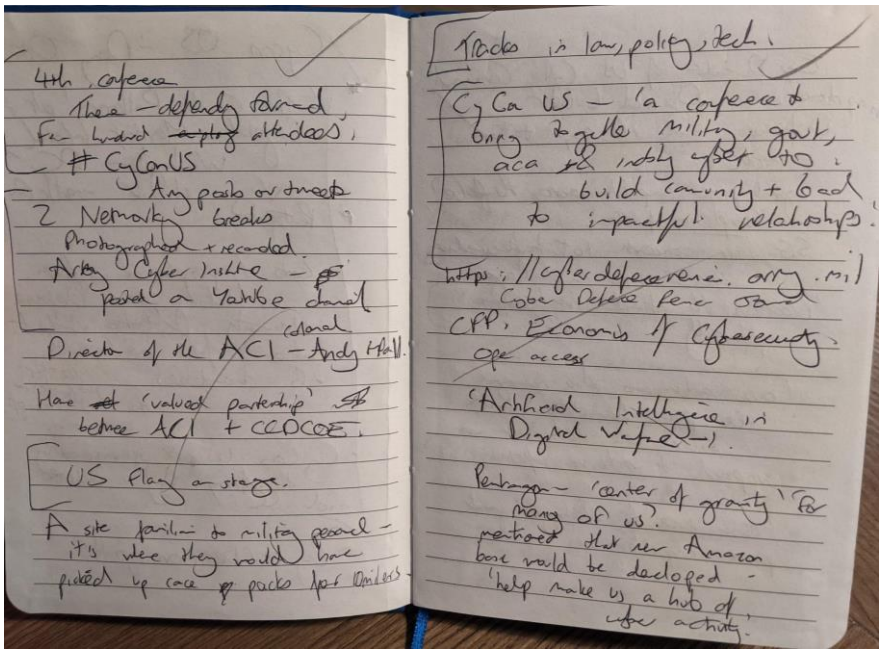

*Image 3.2: Example of written fieldnotes from CyCon U.S. (author's own image, 2019).*

Taking care to abide by Chatham House conventions, fieldnotes on the talks themselves captured the humour, attention-grabbing techniques, hype, and caution. I took notes on the entrance and entrance security procedures, the set-up of the room to note the abundance of chandeliers at some of the more formal venues, and the use of language and tone employed throughout event presentations. Notes on peripheral aspects, including the scene, sponsorship, and location, also helped contextualise the space, some of which were significantly more formal and conducive to purposeful networking than others. Rech's (2015, 537) methodology draws on Macdonald's (2006, 69) observations and research on Cold war rocketry, which explores "visual culture". What is visible or invisible in each image has direct implications for the expression of power, including geopolitical power (MacDonald, Hughes and Dodds, 2011). Showcased military technologies offer an example of the "tension between public spectacle and military secrecy", writes MacDonald (2010, 238). Representing more than their technical components, MacDonald uses the example of missiles as a source of "visual enquiry" in which the missile was seen as being more functional in peacetime than when fired in conflict (2010, 268). As I would see at trade shows like DSEI 2019, a lot of information can be gleaned from observing the spaces in which such visual objects are on display alongside their proud designers and potential owners. With attention to "visual cultures" and the "geopolitical gaze" (MacDonald, Hughes and Dodds, 2011, 273), we may quote Ó Tuathail, who stated, "The vision thing, in other words, is always more than just a vision thing" (quoted in MacDonald, Hughes and Dodds, 2011, 273).

I went to six in-person events from Summer 2019 to early Spring 2020.[51] The end of this phase of data collection was brought forward by COVID-19 - with my last physical event in February 2020 in Brussels. I continued to attend online conferences through the lockdowns. However, I found that when attending conferences via online platforms, there was much less leeway for the spontaneous interactions, observations and connections that had proved insightful at in-person events. As one example, a conversation with another conference attendee at CyCon 2019 ultimately helped me develop a relationship with this attendee as a "gate-keeper"' within the NATO "Emerging Security Challenges" research community, which connected me with several individuals who were interviewed for this research. Online conferences, by contrast, had no chance to start conversations organically, and I did not experience any online event which created a non-artificial research

[51] The details of each event are listed in Chapter Four in Table 4.1.

environment in which I could introduce myself to potential contacts. These challenges are explored in more detail further on in this chapter in section 3.5.4.

González (2012) reflects on the challenges of participant observation of secretive military organisational units. Offering a perspective from the anthropological field, González writes about the inherent challenges of fieldwork in covert or obscure environments. The author discusses their experience researching an experimental U.S. Pentagon programme to explore options for researchers who may want to research secretive organisations. Participant observation is often not an available tool in these circumstances, and the researcher must use other techniques to shed light on secretive workings. González adapts (2012, 23) Nader's (1969, 307) suggestions for "studying up, down, and sideways" as well as down, to circumvent the barriers to participation observation in government agencies or elite institutes. Nader makes three substitutes for direct participant observation: the analysis of publicly available and classified documents, interviews – face to face where possible, and "self-analysis", in which the researcher remains conscious of how they are perceived and interacted within encounters with organisations (Nader, 1969, 109). Through my research, I also draw on Nader's (1969) substitutes to provide additional sources and layers of context, interviewing participants with relevant insight, analysing U.S. policy documents, and reflecting on my positionality as a researcher. González (2012) considers document analysis to be the most important and highlights the wealth of relevant information found on library databases and the U.S. Department of Defense website, the latter of which was directly relevant to my research analysis of U.S. public documents and doctrine.

### 3.3.4 Semi-structured Interviews

A semi-structured interview elicits responses through general questions around a theme and core topics followed by more general prompts as required, allowing interviewed participants the freedom to expand on their thoughts while having a basic structure to the discussion. This gentle guidance enables participants to share their stories while remaining a general focus on topics identified by the researcher (Rabionet, 2011). A completely unstructured interview ran the risk of veering off-topic completely, for example, where a participant might want to speak about their own experiences in ways that are not relevant to my research questions. A wholly structured interview limits the potential insights of the interview to topics pre-selected by the researcher, regardless of the passions and concerns of the interviewees, and risks missing crucial aspects of the discussion around military AI.

For example, in the U.K. interviews, the U.K. Ministry of Defence's (MoD)'s approach to security risk management was a significant theme in the interviewees' responses. Several interviewees felt strongly about this factor of military innovation despite the U.K. MoD's risk appetite not being mentioned explicitly in any questions. Similarly, I would not have known to ask about specific initiatives, or procurement processes, as I was unaware of the relevant programmes until the interviewees mentioned them. In the increasingly fluid and innovative area of military AI technologies, allowing for flexibility in interview structure allowed me to put grounded theory into practice, drawing out emerging themes rather than limiting participants' responses due to overly narrow scoping. The semi-structured nature allowed each interviewee to shape their contribution to the project and led to incorporating themes and subsequent theory-generation that would not have almost definitely not been discovered through a more rigid interview design.

Tables listing each interview's dates and length are available in Appendix E. The full set of anonymised interview transcripts has been uploaded to an online data repository (restricted access) and can be accessed via https://doi.org/10.6084/m9.figshare.19672629.v1.

- The U.K.-focused interviews were held with fourteen industry practitioners working in the U.K defence sector. All but one held a seniority equivalent to 'Vice-President' or above, with roles relating to senior engineering or research positions relating to AI development. Interviewees were selected based on their career experience, specifically their "expertise in cyber security and AI technology in a national security or military defence context".
- My NATO-focused interviewees were held with seventeen interviewees with a range of technically focused, educational and training, and strategic (including policy development) roles related to AI or broader emerging security threats, including cyber security and more general defence-focused positions. Interviewees were employed at agencies across NATO or at NATO CCDCOE. Carrying out the interviews between August 2020 and March 2021, all but one of the interviews were conducted via an online platform due to pandemic restrictions.
- The U.S.-focused interviews were held with seven current or former U.S. DoD staff. Each interviewee was selected based on their professional experience directing defence AI programmes at a national policy level. All interviews were conducted online.

*3.3.41 Designing Interview Questions*

Aberbach and Rockman (2002) list three motivations for using open-ended questions: first, it allows for exploration of topics that are not yet well defined enough to allow for closed-ended questions to be sufficient; second, allowing interviewees freedom in organising and justify their thoughts leads to increased validity of data, and third, open-ended questions avoid "the straight-jacket of closed-ended questions" which are likely to be unappealing to elite participants (Aberbach and Rockman, 2002, 674). For these reasons, I overwhelmingly used open-ended questions in my research, for example, asking "to what extent"; rather than *whether* participants felt a specific dynamic was present. On occasion, particularly in the early stages of my research, I did ask a closed question such as "do you think there is an AI arms race". However, this was usually followed by a prompt, such as "why [do you think this]?". As my research progressed, I designed question sets to allow for open answers.

Of course, there are costs associated with this interview style, including the increased difficulty of coding and analysing the data and the increased length of interviews. I accepted these aspects as costs were outweighed by the additional insights that were more likely through open-ended questions.

I made sure to pilot my drafted questions on informed parties before any data collection. Pilot interviewees varied depending on the audience for each set of interviews: for the U.K. interview project, I collected feedback from personnel with similar profiles to potential interviewees. I similarly requested feedback from individuals I did not intend to interview at the NATO CCDCOE for my question-set designed for NATO-focused interviews and at the Belfer Center for my US-focused question-set. This feedback resulted in several amendments to ensure that terms were correctly explained while minimising the risks of introducing preconceptions. The feedback assured me that I had minimised the chance of simple errors such as using inappropriate language or tone in the U.S. or NATO context. For example, I wanted the interviewees to explain what "weaponised AI" meant to them, which meant I did not want to offer a definition – or possibly influence the definitions they gave me – beforehand.

The question guidance sheets for each interview set are included in Appendix A.

*3.3.42 Conducting Interviews*

In opening an interview, a researcher has an opportunity to make a first impression and build a rapport with interviewees, something that is especially crucial where, increasingly, in a pandemic environment, there has been no opportunity to meet informally beforehand. Specific protocols include

greeting the interviewee, demonstrating credibility through thorough knowledge of the topic to elicit engagement from the interviewee, and creating an environment conducive to truthful and open answers from participants (Rabionet, 2011).

In advance of each interview, every participant received a "Participant Information Sheet" (PIS) when invited to interview. The PIS explained the scope of my research and the interviewee's rights (e.g., to deny being interviewed or to withdraw at any point) and outlined how collected data would be handled and stored. All interviewees were formally invited over email with additional copies of the PIS either brought to the physical interview or attached to the online calendar invites. Each PIS for the U.K, U.S., and NATO-focused interviews can be found in Appendix D.

When conducting an interview, I first introduced myself. I then outlined the scope of the research project (drawing on the PIS where necessary) before confirming that the participant was happy with all aspects of the consent form.[52] All but one interviewee agreed to have the interview audio recorded for transcription before starting an interview. Where the interviewee did not consent to be recorded, I confirmed consent to manually take notes paraphrasing the interviewees' comments while conducting the interview.

Once commenced, the interview does not have to follow the same uniform structure or order (Longhurst, 2003). While changing the order of questions is perhaps not ideal according to some methodological purists, allowing for flexibility facilitates "conversational flow and depth", which Aberbach and Rockman (2002, 674) argue outweighs the potential drawbacks of inconsistent ordering. Longhurst (2003) highlights that questions, or topic sheets depending on the researcher's preference, may instead follow the natural discussion with the interview participant and the approach I applied to my research.

There were several techniques I used in an interview to encourage engagement. Rice (2009) offers specific recommendations on the broader theme of interviewing elite individuals, in which the interview takes place within a dynamic of evident unequal power between interviewer and elite interviewee. I noted this aspect through my research as I requested interviews with senior staff across industry, government, and departments or within military chains of command. Rice (2009)

[52] Consent was confirmed for every interview undertaken as part of this research - in written or verbal format depending on the preferences of the interviewee. Please see section 3.4 for a discussion on research ethics. The consent form template can be found in Appendix C.

recommends subtle self-positioning by the researcher that adapts to the interviewee's personality, for example, emphasising or downplaying one's identity as a neutral academic researcher, or researcher for a particular research party, depending on the interests of the interviewee. In my case, this meant stressing particular affiliations with interview participants; NATO-employed interviewees were engaged through my NATO CCDCOE email account, and I introduced this affiliation first in any conversation, while US-based interviewees were engaged with my harvard.edu email account with the Belfer Center as my primary affiliation. This introduction is also an example of "mutual self-disclosure" (Sandana, 2013, 137) that can help build rapport with interviewees in inferring aspects of "insider status" (Jowett, Peel, and Shaw, 2011). Beyond introductions, I made sure I was up to date on the required background knowledge by reading academic papers and industry reports to the extent that I was informed enough to carry out the interview effectively, given the technical nature of the technologies (Rabionet, 2011). Assessing the correct level of familiarity to conduct the interview effectively was one lesson. Making a first impression with an interviewee and then completing the interview within a short time period often gave a very short window to establish a relationship with an interviewee. Remaining within the realm of professional objectivity without risking overfriendliness or offence requires "calibrating social distances" (Kvale, 2008, 9). The shift to remote interviewing over online platforms made this task slightly more difficult, limiting the range of available social cues. I utilised additional guidance issued by Rice (2010) when approaching senior individuals in the field, particularly for my U.S. interviewees who held formal leadership roles. Adapting to the personality type of my participant, as Rice suggested, helped facilitate a space in which the interview was as productive as possible, with participants feeling positively towards the process.[53]

### *3.3.43 Analysing Military Interviews*

Castro (2018) highlights that the military is a complex category, encompassing distinct segments defined vertically, through different hierarchical levels and generations, and horizontally across different services, such as Army, Navy, Air Force, and other sub-branches, such as infantry and artillery. Across vertical and horizontal segments, Castro (2018) points out that interviewees may hold selective interpretations about events that may not reflect historical evidence. While Castro's research focuses on the particularly extreme authoritarian scenario, in examining the Brazilian military's "dirty years" in the late 20th century (Castro, 2018, 6), aspects of these diverging memories

[53] And hopefully, towards me. It was usually at the end of interviews that I mentioned an appreciation for potential connections they could assist within securing future interviewees.

are relevant when considering the allegiance of a state.[54] To avoid the pitfalls of interviewees feeling they had to conform to a viewpoint that does not represent their understandings, I made it clear that all conversations would be anonymised and that interviewees were free to refuse to answer any question with which they were uncomfortable.

*3.3.44 The shift to online interviews*

The COVID-19 pandemic put an end to face-to-face interviews. In March 2020, several interviews were cancelled, and leads were lost. Fortunately, by early June, potential interviewees were broadly comfortable with committing to interviews, which were conducted through online platforms out of necessity. Online platforms may represent an opportunity for qualitative research in several ways, with Archibald et al. (2019) finding through a survey with 16 nurses that Zoom was perceived positively (compared with other VoIP technologies and other interviewing mediums) in terms of convenience, user-friendliness, and in security terms. The reliance on remote interviewing also allows for greater accessibility. Researchers can recruit participants from a greater geographical area with fewer financial and (travel) time resources required to set up meetings (Gray et al., 2020). There are also significant limitations to engaging with research participants online. There are fewer opportunities to network or build rapport with interviewees when conducting interviews remotely (McLean et al., 2020; Jowett, Peel and Shaw, 2020), and the researcher must work harder to make online interviewing less abrupt and connect with interviewees (Jowett, Peel and Shaw, 2020). I further explore these challenges in more detail in section 3.5.4.

When organising interviews, I typically offered interviewees the choice of platform (some were happy to use their corporate audio-visual software), with the majority being organised as a zoom or MS Teams meeting. Using either of these platforms (with enterprise-licensed accounts) means conversations were encrypted. There were definite teething issues. For the first few months of the pandemic, my internet was not ideal, which often precluded sharing my video and made it harder to build a rapport through the call – a capability already curtailed by the lack of face-to-face contact. Nonetheless, over half of the interviews for my U.K. case study were conducted remotely, and I could still engage practitioners (to the extent that I could continue to operate off the "snowball effect" of

[54] As one example, in early 2019 I observed a British officer state with confidence that Britain is a world-leader in military artificial intelligence, alongside well-publicised U.S. and Chinese dominance (it may well be the case that Britain has tricks up its sleeves in this sense, however the evidence relating to relative investments, and lack of evidence suggesting the existence of world-leading military AI technology, suggest this is unlikely).

interview participant engagement). The U.S. and NATO interviews were also conducted online, with one exception.

### 3.3.5 Data Analysis

For interviews, I used a Google Pixel 3a to record each audio transcription[55] and then transcribe each file into written text word-for-word and import each transcript into NVivo 12.[56] The Google Pixel 3a was a new blank device procured by my academic department. I created a new "PhD account" Gmail account to avoid syncing the device with any non-academic accounts. With that account, I downloaded the Google "Recorder" app to record audio and subsequently assist in audio transcription.[57]

Following the suggested security practices outlined by Da Silva (2021), I then downloaded the Microsoft Outlook app, logged in to my university account, and took the device offline. Each interview was titled in sequential order (i.e., from "B1" to "B17") as part of a pseudonym-assigning process. Following an interview, I copied the auto-transcript "Pn" text from the "Recorder" app to a Microsoft Outlook email and emailed the transcription *from* my university email address *to* my university email address. I then corrected the auto-transcript (and redacted any identifiable information, including company names or role titles) and saved the corrected transcript file as "Transcript [Number] – [Date of Interview]". Once saved, I was able to delete the audio and corresponding text file.

For fieldnotes collected via my observant practice-focused research, I typed up any notes captured on paper, and placed notes from each of the six events into separate files which were then loaded into NVivo 12.

[55] All interview data was anonymised and stored in a GDPR-compliant form. See Appendix C for the consent form, and Appendix D for the participant information sheets that detail data management for each set of interviews.
[56] See Section 3.3.51 for an overview on NVivo 12.
[57] The files were deleted once the transcription was complete and the device was hard-wiped before its return to the department.

*3.3.51 Data Coding*

NVivo 12 is a qualitative data analysis software package that allows for analysing large volumes of text.[58] NVivo allows text to be "coded", allowing a researcher to distil raw text data into various forms and breakdowns. It allows for data reduction and organisation, allowing for inductive methods such as grounded theory (Cope, 2016). I employed descriptive and versus coding strategies when coding my observant practice and interview data in NVivo:

*Descriptive coding* is an initial coding of the data, also known as the "first-cycle" coding of the entire text (Saldaña, 2013). By highlighting key words or phrases at an obvious surface level (i.e., referring to direct terms mentioned by interview participants), descriptive coding is appropriate for determining "category labels" (Cope, 2016). One form of descriptive coding is "in vivo" codes, which code directly off the statements or common phrases found in the text-formatted data sets. Cope (2016) describes in vivo coding as a form of descriptive coding particularly suited to inductive reasoning. This form of descriptive coding was straightforward in highlighting concepts and phrases that came up through each transcript in a series of codes that began to take shape early in the coding process. For example, after 4-5 interviews, I highlighted the "UK approach to innovation" as a frequently used code.

*Versus-coding*: "Counter-coding" (employing a secondary, varied method), a researcher can experiment with the data to an extent, observing what data outputs occur through alternate methods to develop a richer understanding of the data.

It was necessary to begin data analysis without pre-classified categories in mind. This approach was consistent with my grounded theory position, which allows inferences to be drawn from the data rather than pre-existing assumptions and theories relating to the field. Drawing out concepts based on the insights collected through my research, this approach allows for the creation of inductive theory. It offers potentially new and novel insights into the research landscape relating to the military innovation environment. By analysing each set of data results and drawing themes from coded data, I was able to draw out the patterns and dynamics of military AI innovation that were not yet denoted in the current security discourse.

By coding as I went along with each set of interviews, I was able to determine when additional

[58] Further detail: https://www.qsrinternational.com/nvivo/what-is-nvivo

interviews were not adding any more themes (and the discussions had, in one sense, reached "saturation point"). Given that this was at least 14 interviews for the U.K. and NATO interview sets, the emerging areas of consensus or themes gave some assurance that I had captured a set of perspectives that represented practitioners in the field. For the U.K.-focused set, for example, I reached saturation at approximately 12 interviews, conducting 14 interviews in total.[59]

I used versus coding, a form of analytic coding, for my counter code, reflecting themes related to my research question (i.e. actor dynamics, security and AI, and the consequences of AI implementation). Doing so made it possible to evaluate the different dynamics between actors interacting at the intersection of AI and military innovation, following Saldaña's (2013, 115) assertions that this versus coding is appropriate for evaluation research, critical discourse analysis, and data sets suggesting conflicting and competing goals between participants. Descriptive coding highlights words or phrases that may imply certain dichotomies (i.e., national versus international or positive versus negative attitudes to rapid prototyping). Versus coding might help identify groups of stakeholders, perceptions/ actions, and issues (i.e., with stakeholders "us – U.K. military" vs "them – other militaries, governments, technology companies, investors, civil groups, academia…"). [60] The third aspect – the *issue* – is generated and emerges through analysis by the researcher and will form a major source of the theory-generation that addresses my research question (understanding the dynamics and impact of AI innovation in the military landscape).

Having performed both forms of coding on a set of each of my four sets of files (three sets of interview transcripts and the observant practice fieldnotes), I was able to view which codes came up more frequently than others. Image 3.3 shows a code table excerpt from the UK-focused transcripts, in which the second column highlights how many transcripts the code was present in, and the third (right) column shows the overall number of instances the code was applied. For example, "hype" was referenced twenty-seven times and was mentioned in twelve of the fourteen UK-focused transcripts.

[59] This approach does not apply to the US interview set, where limitations on access and available interviews meant I supplemented my data-collection to analyse two DoD documents.
[60] Saldaña states that versus coding offers three sections of analysis: 'the primary stakeholders, how each side perceives and acts toward the conflict, and the central issue at stake' (2013, 95).


*Image 3.3: UK-focused interview code excerpt (author's own screenshot, 2022).*

### *3.3.53 Data Analysis*

Once each dataset was coded and categorised, this output formed the basis of arguments addressing my research questions. I applied Clarke and Braun's (2017) form of thematic analysis (TA), a method for "identifying, analysing and interpreting patterns of meaning" ("themes") within qualitative data. Taking a flexible but systematic approach to the data proves particularly useful for qualitative data, as TA aims to move from codes (the output from NVivo coding analysis) to theory-building, using codes as a "building block" to draw out themes/patterns from the data.

For example, for my U.K.-focused set of interviews, termed "Fragmented Landscapes", after around half of the transcriptions, I was able to pull out thematic categories for my coded material, including "Military AI - Supply and Demand" and "Risk Management" (see Image 3.4).

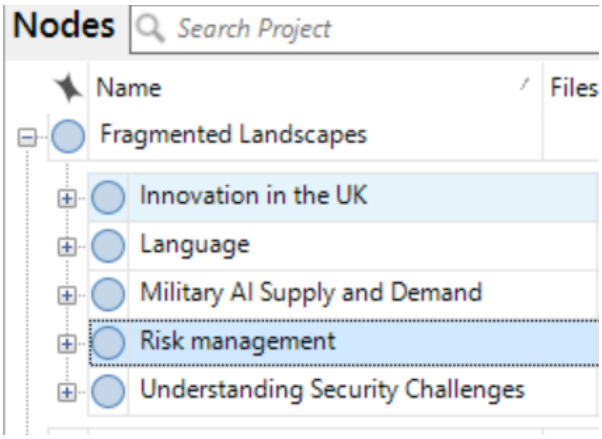


*Image 3.4: UK-focused interview top-level themes (author's own screenshot, 2022).*

## 3.4. Ethics and Responsible Research

### 3.4.1 Responsible research practices

There are broader ethical questions that need addressing when researching a field where violence is an inherent implication and focus of decision-making. Researching AI in a military and defence context is interesting for precisely this reason, in line with the view that researchers have a public responsibility to their societies which necessitates specific ethical considerations on the social impact of undertaken research. Massoumi et al. (2019) proposed a model for researchers to engage in the ethics of terrorism, nodding to the U.K. university research environment specifically. Massoumi et al. (2019) argued for enhanced ethics review processes, greater protections for researchers from powerful interests which attempt to constrain independent research, and campaigns to highlight unethical research to dissuade future unethical practices. These arguments influenced my approach and motivation for my research, contributing to my decision to refrain from seeking a security clearance via a research path that may challenge impartiality as a researcher. The concept of public responsibility was also important in shaping my research approach: focusing on the implications of current approaches to AI and AI innovation in military contexts to identify and not accept the discourse of overselling - or blunt caution - without investigating further. Especially when supporting automated processes, AI has already been highlighted to have negative social impacts (Fry, 2018; O'Neil, 2016). It would be negligent of me to ignore any identified negative consequences during my research. For this reason, a key component of my research questions focuses on the impact and implications of AI in the military and possible routes to mitigation for identified challenges.

I also held several practical ethical considerations going through the process in terms of communicating and handling any research data relating to participants as well as the broader communities of policymakers, military or government personnel, academics, and fellow event attendees. This approach meant abiding by conventions not to quote speakers or interviews by name and following measures to save and store data anonymously and securely, in line with GDPR and the privacy measures outlined in the consent forms that each participant agreed to in advance of an interview.

For each of the four aspects of my research - observant practice, U.K. interviews, U.S. interviews and NATO-focused interviews - I applied for ethical approval through the Royal Holloway Research Ethics Committee process. Further detail on these approval processes can be found in Appendix B. The PIS for each set of interviews and an example Consent Form may be viewed in Appendix D.

## 3.5. Broader Researcher Reflections

### 3.5.1 Non-Classified Research on the Military

During the period of this research I did not hold a security clearance. While I have spent time at the NATO CCDCOE throughout my PhD, this did not facilitate any access to non-public documents as part of this thesis; all cited data sources are from open-access documents or my interviews. This access represents a partial limitation on my analysis, for example, in only being able to view the public executive summaries of the U.S. DoD AI Strategy or NATO AI Strategy. Nonetheless, the decision to research without a security clearance was taken considering the potential limitations classified material might place on my work and minimise potential conflict of interest as an academic. Massoumi et al. (2019) highlighted the ethical challenges for academic researchers attempting to research security themes and described mechanisms through which researchers might be compromised by the security state through co-option, compromise or issues relating to conflict of interest with the state. In calling for more attention to the dilemmas of state involvement with academic research in security-related academic research, Massoumi et al.'s (2019) observations are particularly pertinent to my research, which saw opportunities to engage with Dstl and the wider MoD. Accessing Dstl's offices and materials may have offered exposure to projects and information that lies outside the public domain. However, by accessing this knowledge, I risked basing my research on information that cannot be shared (in published journals, conference presentations or any other form). This access might have limited how far I could publicly share my research outputs and, therefore, limited my potential public research impact. Consequently, I agreed with my supervisors that the path to publishing my research findings would be more straightforward if I had not accessed classified material along the way.

As a result of choosing to rely on publicly available resources for my research, there was information on the British development and implementation of military AI technology that I did not come across through the course of this research. Researching national security and emerging technology will inevitably touch on topics of security and state-level secrecy. One aspect of competitive advantage in

the military is technical supremacy which may be maintained longer when your competitors (and in the military context, adversaries) are unaware of the details of your capabilities and strategies. The U.K. follows this adage and has very few open-forum discussions on military AI developments compared with the U.S. Besides an AI Lab within Dstl within the MoD and the public procurement calls (Ministry of Defence Contracts Website, 2019), there was little to engage within the public sphere in terms of military AI innovation.[61] There are frequent meetings, workshops and forums occurring through and beyond Whitehall, Vauxhall, Shrivenham, Porton Down and Cheltenham that are strictly invite-only affairs. I was repeatedly warned by ex-military and ex-government personnel that publicly available information on this topic would be scarce. While I visited Dstl's AI Lab during the first year of my doctoral research, the decision to conduct research without a clearance meant the potential for future engagement was limited. Instead, I designed my data collection strategy to interview U.K. experts within the commercial defence sector. By not engaging with MoD personnel for my research, I did not need to seek approval from the MoD Research Ethics Committee, which was described informally by other U.K. academics as a complicated and lengthy process.[62]

Two approaches helped me mitigate the challenges of not holding clearance:

- *Attempting interviews:* Even where interviews failed to clarify a consensus understanding of the role of AI in the military, interviewees still contributed valuable information. The fact that interviewees may not be familiar with key terms or innovations, especially if they were selected as senior leaders, offers an insight into the preparedness of military functions for emerging technology and trends. For example, while interviewing U.K. defence experts, it became clear that interview participants often felt uncertain about providing any concrete examples of policy-led security controls for military AI. Given their positions and active involvement in the sales, design and deployment of AI systems, some may have expected staff in these fields to have been actively working on potential solutions. The fact that interviewees listed concerns and caveated any answer with the fact that they did not know, or that any answer was speculation, is valuable in highlighting the immaturity of the space. One interview participant was less established in their career than I initially anticipated, with around four years of employment relating to military technology. This interview ultimately proved beneficial as the interviewee contributed a reflection from an associate's perspective on how attitudes shifted between generations and between corporate/ military ranks. They also

[61]This might change as the Defence AI Centre comes out – though in my view this is unlikely.
[62] For more information on the MoD research approval process see: https://www.gov.uk/guidance/apply-for-ethical-approval-for-mod-research-involving-humans; this feedback was given to me in conversation.

reflected critically on assumptions that were repeated through other interviews. Finally, I drew significant insights from interviews – more than I originally anticipated – by allowing participants the space to articulate their views candidly and anonymously.

- *Remaining flexible to shift research focuses from operational to strategic-level considerations:* To date, academics have demonstrated expert analysis of the impact of AI in the military without referring to exact pieces of hardware or proprietary tools (Burton and Soare, 2019). The more I designed my research, the more I realised the value in knowing exact technical implementation models was minimal; a critique of one particular attack model would be as outdated as soon as a subsequent distinct model came along. Instead, critiquing the essence, theme and strategy became a sensible focus and one that was driven by appropriateness and research needs rather than any constraint. I did find in parts that it made sense to move away from operational analysis to look at strategic aspects, and in this way, the lack of available information did shape my work. However, this proved beneficial in opening the possibilities of discussion and research avenues on strategic implications of AI implementation, including discussions on security dilemmas, national security, and the "AI arms race". Discussions of this nature gave depth and relevance to talk of actual practical implementation and signified a key part of my overall research plan.

### 3.5.2 Access - Network Building

In presenting my initial research proposals to ex-military personnel within my existing network, their immediate and strong reactions were to tell me that any interview requests "simply would not work". Their view was that no one would be willing to talk to me on the topics of national security and emerging technologies. The effect of this feedback took me on a bit of a journey. I recognised this feedback the first two times. However, instead of changing my research goals and overall research design, I tried to tweak minor, aesthetic aspects of my research to seem more appealing to practitioners. Amendments included some changes that ultimately benefited my research, through the refinement of interview questions; the removal of pre-interview email surveys, which represented an additional onerous task for interviewees. However, after a third former military individual expressed disbelief in the project, I began to doubt my own perceptions and spent some time mapping out alternative options.

Not holding the right attributes, such as certain nationalities, also posed various challenges throughout data collection. I could not work with the Center for Security and Emerging Technology's CyberAI

project, based in Washington D.C. because I was not a U.S. National. On one occasion, I was declined interviews with U.S.-based defence-sector employees because they believed they could not discuss the topic of military AI with someone who was not a U.S. citizen. Not being a U.S. citizen, I was also unable to physically travel to the U.S. for a significant proportion of my thesis due to COVID-19 travel restrictions. Therefore, I initially struggled to develop my network remotely.

Drawing on the literature and opportunities brought to me by my network, which developed throughout my research, I found I could overcome some of the significant access challenges I experienced. The following methods assisted me from the start:

- *Open-source research:* Just because there is very little published on the British military approach to AI in cyberspace does not mean the information that exists is not of immense value. Procurement calls, various government papers, and trusted third-party analysis (e.g., academic writing or output from U.K. military defence firms) could often shed light, even indirectly, on the nature and direction of British policy.
- *Pro-active Networking with intent:* As much as others may not have had faith in the gains to be had from engaging with active experts in the field, I felt it would be disingenuous for me to give up on my "plan A" without at least trying to get the answers I wanted. Events such as the DSEI Disruptive Innovation conferences held in London had a heavy military presence and multiple opportunities to ask questions from senior personnel. By introducing myself and my research at these events, I found other attendees happy to introduce me to their colleagues that had expressed an interest in the topic.
- *Pro-active Opportunistic Networking:* I worked to build up a relevant, more comprehensive network, engaging at major security events (from expos and conferences such as the Security and Counter-Terror Expo or DSEI), attending and speaking at military-attended events (AFCEA Young Members, AFCEA Annual meetings), and engaging with military and technology practitioners wherever possible in an environment that was familiar to them, usually at a military or AI security event). The aim, in this case, was to establish myself so someone may remember me should the topic of AI and the military and national defence arise. This approach paid off reasonably early in my PhD, as after presenting at an AFCEA event, one attendee connected me with relevant individuals whom I then interviewed for this research.
- *Going International:* While Britain may prefer to maintain a largely low-profile regarding approaches to military AI, other states are more forthcoming. The U.S. Military AI Strategy

Executive Summary, released in February 2019, proved an excellent starting point for looking at a state-level strategic approach to AI. Examining subsequent international developments, we can understand what impact documents such as the DoD AI Strategy AI have on international security. Determining my regional scope and having identified multiple focus points also offered me a backup plan should a study on the U.K. not be sufficient to extrapolate and theorise on international issues. The U.S., for example, offered more insight on the subject of military AI strategy and public statements on AI in warfare. As it happened, I was able to facilitate fieldwork in Tallinn, Estonia, at the NATO Cooperative Center of Excellence. I was also able to facilitate access to senior U.S. contacts through the Belfer Center. Through the Belfer Center, I was able to invoke the "familiarity" of Harvard University, and thus the "prestigious" draw as outlined by some semi-tongue-in-cheek guidance for attracting elite interview participants, as issued by Aberback and Rockman (2002).

### 3.5.3 Positionality

There was a danger, in constant exposure to military-format environments, both through written documents and physical cultures and customs, that I subconsciously embodied and reflected the formats, language and priorities of my research environment. This tendency can happen most obviously with language. Cohn (1987) reflects on how their research in a military facility made them reflect on the use of language in framing scenarios in national security**.** While I spent a considerable amount of time focusing on military-format material for this thesis, I was aware of the balance of allowing grounded practice theory to drive my research without allowing the military environment and language surrounding me to impact my overall research framing. I must realise and identify the framing biases within my research environment. This emphasis on self-reflexivity is an essential part of ethnographic research (Göğüş, 2019; Hammersley and Atkinson, 2007), as social researchers are an active part of the world they study (Hammersley and Atkinson, 2007, 14). While a researcher cannot avoid relying on "common sense" or shaping the phenomena they study (Hammersley and Atkinson, 2007, 15), acknowledging and reflecting on the researcher's position in the research environment allows for a thorough recognition and mitigation of bias as far as possible. Working at the CCDCOE meant I was in an environment that unconsciously favoured particular language and assumptions on the security landscape. Recognising reflexivity helped me consciously reposition to take a neutral stance to approach my work without preconceptions and maintain my grounded practice approach.

My positionality as a researcher came into sharp focus during my research interview series with NATO- and NATO CCDCOE- employed staff. These interviews coincided with my Visiting Scholar's position and subsequent part-time contractor positions at the NATO CCDCOE, as discussed in Chapter 1. The blurring of the lines between being an "insider" and "outsider" through one's research risks several dilemmas. Moore (2012, 11) reflects that as a researcher's position shifts between being an insider and part of the group they are observing or researching, and an outsider, researchers face changing social dynamics and boundaries that represent "social dilemmas". There was a risk of this presenting without conscious reflection and separation through my role at the CCDCOE, particularly where the results of my research highlighted weaknesses in the current approach to security challenges. When undertaking this research, I took extra care to examine how I perceived and coded interview results, firstly to ensure objectivity and second to reflect that I was not unconsciously adopting the language and mindset of the organisation.

Perhaps ironically, the remote working practices during my position meant that my direct interactions were not necessarily different than if I had not been affiliated with the Centre. To a colleague from the NATO Defence College in Rome or Allied Command Transformation in Belgium, I remained, primarily, an outsider. I did not interview personnel I frequently interacted with through my scholars' or researcher position. I, therefore, occupied spaces in which I represented both an insider and outsider for the communities I engaged with through my research. I experienced the positionality described by Dwyer and Buckle (2009), who challenged the idea of a dichotomy between insider and outsider status. Using Dwyer and Buckle's (2009) terminology, I spent much of my research period shifting in the ambiguous "space between". This approach involved benefiting from my insider access to the security environment to make acute observations and develop trust with participants. At the same time, outsider status helped me mitigate preconceptions and biases and continually reflect on researcher objectivity and reflexivity.

Another anticipated challenge was the objectivity in reporting my research findings, where findings may not be complementary to organisations represented by research participants. In line with the motivation not to pursue a security clearance, I preferred to retain full ownership and control over my research output. This preference was clarified in my CCDCOE contract. In return for acknowledging my Visiting Scholar's position at the Centre, much like one would recognise a reviewer for a book, I was free to write my research without requiring their explicit approval.

### 3.5.4 Adapting Methods: COVID-19

In March 2020, the cancellation of ISANET's Annual Conference was the first of several cancellations, postponements, and a shift to virtual collaboration. Within 12 hours of the email notification that ISANET was cancelled, I received an email from the NATO CCDCOE informing me that my scheduled research placement, initially scheduled to start in April 2020, could no longer occur as Estonia had closed its borders. I woke that morning expecting to move to Estonia in two weeks, engaging with CCDCOE staff throughout April and May. Instead, I remained in Hertfordshire in the U.K. Very quickly, the crisis worsened, and March 13th turned out to be my last face-to-face interview. I had been scheduled to interview two staff members back-to-back, but the second interview was cancelled as it emerged that the individual was self-isolating and working from home.

Transitioning to a responsible work-from-home schedule meant a change in methods. Researchers worldwide shifted away from physical fieldwork towards virtual fieldwork or alternative methods entirely (Krause et al., 2021). For my thesis, the most prominent and pressing task was that all the interviews with military experts had to be shifted online, taking place via Zoom or Teams, for the most part, depending on the interviewees' preferences. As Woods et al. (2020) note, semi-structured interviews rely in varying degrees on a research ability to develop relationships, a task complicated by reduced social interactions through pandemic restrictions. I relied on a different approach to rapport building and network-building, as proposed by Krause et al. (2021, 4), where research students are advised not to reach out to ideal candidates directly but to contact "gatekeepers" in the field who can help facilitate connections. My experience matched these reflections in the literature in which COVID-19 enhanced the gatekeeper effect, with direct inquiries rarely leading to a confirmed interview without a gatekeeper's intervention.

"Hi Amy,

Great – 16:30 sounds great.

Can you call my mobile *redacted*. I've taken to doing my afternoon calls while walking along the local country lanes. Gets my steps in and is very pleasant…

Thanks"

(UK Interviewee, email excerpt, Spring 2020)

Perhaps the most direct effect is that an interviewer loses aspects of control when interviewing remotely. There is an element of trust that the interviewee can manage signal or technological issues (Jowett, Peel and Shaw, 2011) and make sure they are in an environment where interviewees feel comfortable speaking. Particularly in the earlier stages of the first lockdown, these challenges were unavoidable. I did not ask an interviewee to stay inside during their one opportunity for governed-sanctioned exercise a day, resulting in an interview with outside background noise. I mitigated challenges as far as possible by noting whenever interviewees hinted that they might be limiting what they said. In one example, this was when the interviewee had chosen to participate via mobile while walking outside, at a period when those in the U.K could only leave once a day for non-essential activities, and as they passed some other individuals on a hiking footpath.

Disruptions continued through the pandemic. Interviews were cancelled more than once because interviewees suspected or confirmed they had COVID-19. Interviewees rescheduled when they were forced to panic-buy before lockdown or when their childcare plans fell through once schools were closed. At times like these, I exercised empathy and spread out my interviewing to put as little pressure as possible on those who volunteered their time for my research. This choice drew on the broader need, as stated by Jowett (2020), for researchers to consider the ethical implications of their research and consider whether asking interviewees to volunteer their time places them under additional stress. Due to pandemic-driven disruption to my research, I requested an extension to my data collection period for my UK-focused interview project, which the Royal Holloway Ethical Review Board granted. I subsequently applied for and received a six-month extension for my doctoral research period from UKRI (United Kingdom Research and Innovation) to help mitigate the delays in my research due to COVID-19.

My NATO-focused interview project was delayed from March to August 2020 as my Visiting Scholar's affiliation was shifted accordingly. As infection rates increased almost as soon as I reached Estonia for my Visiting Scholar's visit to the CCDCOE, being present in Tallinn proved less directly beneficial when seeking interview participants as almost immediately, in-person meetings were discouraged. All but one of my NATO interviews were carried out online. Similarly, my Belfer Center predoctoral fellowship, which initially would have had me based in Cambridge, MA, was replaced

with a non-resident fellowship. These remote effects cumulated not only in remote interviewing but in severely limiting opportunities to network or build rapport with interviewees before a meeting, a challenge noted by McLean et al. (2020, 3), who described how remote data collection limits the working relationship between researcher and interviewee. Being based physically at the Belfer Center would have benefited from frequent high-profile visitors and proximity to active research projects at MIT and the DoD's Defense Innovation Unit. Instead, many of my requests for remote interviews ended up being 'cold' emails, which recipients often ignored. I was aware that my ability to show credibility was limited compared to previous opportunities to speak to practitioners at conferences. I mitigated the challenge of a poorer rapport-building environment by offering detailed information before the interview (as suggested by Jowett, Peel and Shaw, 2011, 361). I engaged in appropriate "mutual disclosure" and used conventional language that mirrored the interviewees' terminology to highlight credibility early on. I also secured interviews through the snowballing method, utilising current contacts and gatekeepers to suggest future interviewees, as suggested by Krause et al. (2021).

A second challenge through 2020 was the concept of working on this thesis when there were so many competing priorities for attention - particularly relating to COVID-19. Lockdown, health fears, and family concerns appear to have very little to do with whether the DoD was using machine learning to surveil targets or improve the accuracy of a missile. For all the emphasis I had placed on my research project having real-world impact, my research seemed entirely divorced from the major concerns of the day.[63] In addition to COVID-19, pressing social movements and family responsibilities all affected not just me but also loved ones around me - and to varying extents, the details of which I will never know - my research participants. My approach to justifying my attention and research interest was aided first by the nature of my research, as I continued to speak to passionate staff in the security and AI domain that reminded me of my motivations. The second was to keep abreast of the - shortly after the March escalation - violations of data protection, surveillance, and security creep taking place in efforts to keep - and the fascinating discussion around the balance of health security (preventing the spread of the disease) and cyber security.

[63] Some of these concerns are reflected, albeit one-sidedly, in The Guardian 'What does 'national defence' mean in a pandemic? It's no time to buy fighter jets', 8th April 2020, TheGuardian.com, Accessed 5 May, 2020. https://www.theguardian.com/commentisfree/2020/apr/08/national-defence-corona-pandemic-fighter-jets. I consider my view to hold more nuance - we must continue to consider the actions of adversaries, and the economic sunk costs of military investment to date.

## 3.6. Other Limitations

There are some limitations associated with this thesis. Sections 3.5.1, 3.5.2 and 3.5.3 above reflect on access challenges within this research, which may have meant that my research did not benefit from documentation or information that would offer alternative explanations for my findings. Section 3.5.4 reflected on the impact of COVID-19, which made it more challenging to engage with potential interviewees. In general, time constraints through this research meant that I focused on three different communities of experts: relevant U.K. defence industry practitioners, relevant NATO-employed personnel, and relevant U.S. policymakers. Given that my research asks different questions of each community, this research does not allow for straightforward comparative analysis between communities. The relatively small number of interviews conducted does not also allow for forms of meaningful quantitative or statistical analysis.

This thesis recognises that the activity in some regions is discussed disproportionally and is mindful of potential limitations in taking a Western-centric and largely state-centric approach. A range of other state and non-state actors could have served as central focus points of this thesis. Some of the most obvious are Russia and China, which invest heavily and distinctly in military technology. While this thesis will refer to these actors – a decision was made on balance that without the language skills, network of relevant practitioners and personnel, and access, these case studies would be less thorough than my agreed focus on the U.K., U.S. and NATO. The same pragmatic reasoning was applied during the decision not to focus on other states. Aware of the time constraints of thesis submissions, it was assessed that time would be best spent concentrating on these "accessible states" that were also the most active in applied AI research and deployment.

## 3.7. Conclusion

This thesis involves a range of methods necessitated by the context in which my research took place and the wish to corroborate findings to form a richer, more nuanced understanding of the dynamics in which AI is deployed and used in military environments. I designed my research with real-world applicability, reflection, and conscientiousness in mind, using a multiple-methods research approach to explore real-world dynamics in a way that has provided insight into discussions on the development and adoption of AI in military contexts. By approaching the issue through a grounded approach, I

have allowed my research to be evidence-led, with research questions shaped by current military AI discourse and expert perspectives.

# Chapter Four: Observing community approaches to military AI innovation at conferences and trade shows

## 4.1. Introduction and context

As part of the research underpinning this thesis, I explored relevant physical spaces in which conversations about military AI technologies occurred. This approach was particularly beneficial in the early stages of this research, where existing literature on the strategic implications of AI in the military was few and far between. Hence, grounding the research in expositions, workshops, and military trade shows all offered the opportunity to gather primary data to address the identified research gaps.

Drawing on observant practice methods as described in Chapter 3, this form of fieldwork took me to six in-person events between June 2019 and February 2020, as set out in Table 4.1.[64] Each event was attended based on how their agendas included emerging technologies in defence and military contexts.

*Table 4.1: Events attended while employing observant practice methods*

| **Event Name** | **Type** | **Organising Body** | **Location** | **Date** | **Name (ShortHand)** |
|---|---|---|---|---|---|
| CyCon 2019: the 11th International Conference on Cyber Conflict: Silent Battle | Academic / Military Strategy | NATO CCDCOE | Swissotel,Tallinn, Estonia | June 2019 (4 days) | CyCon |
| CyCon 2019 (U.S.): Defending Forward | Military Strategy | CCDCOE and U.S. Army Cyber Institute | Crystal Marriott Hotel Alexandria, DC | November 2019 (3 days) | CyConUS |
| Defence and Security Equipment International | Commercial: Trade Fair | Clarion Events Defence and Security | Excel Centre, London, U.K. | September 2019 (Attended 1 day of 3) | DSEI |
| Disruptive Technology for Defence Transformation | Military Strategy / Commercial | DefenceIQ | Millennium Gloucester Hotel, London, U.K. | September 2019 (Attended 2 days of 3) | DTDT |
| Cranfield Defence and Security Doctoral Symposium | Academic | Cranfield University & NCSC | STEAM Museum: Swindon, U.K. | November 2019 | DSDS |

[64] This aspect of my research was concluded with the onset of pandemic-related travel restrictions.

| | | | | | |
|---|---|---|---|---|---|
| AFCEA EUROPE | Military Strategy | Armed Forces Communications & Electronics | Bluepoint Conference Center, Brussels, Belgium | February 2020 | AFCEA |

Each of the six events differentiated itself based on the target audience, theme, or stated purpose.[65] For some of these events, AI-enabled technologies were advertised as the main stated focus of attention. For example, the DefenceIQ event focused on emerging technology and opportunities for militaries. At other events, AI and autonomous systems represented only a portion of the broader advertised agenda. For example, the Security and Defence Symposia at Cranfield agenda included the broad spectrum of security research, for which ML and cyber security were two subcategories (DSDS Agenda, 2019). Both categories of events were identified as helpful in capturing a broad spectrum of event spaces where attendees establish and shape their attitudes, knowledge, and procurement decisions around AI, ultimately in military spaces. The absence of AI as a key explicit area of focus, for example, at the DSEI trade fair in London, was also valuable to note to examine how technologies were, or were not, presented to attendees as crucial areas of focus. When attending events where the military was *not* a primary focus, I explored the spaces most relevant to the theme, attending relevant talks and visiting relevant stalls (e.g., AI-focused talks at the Cyber Conflict Conference (2019) or Dstl at DSEI).

## 4.2. Themes

This section explores the findings from each event, categorised by each theme. These themes were drawn from the qualitative analysis of field notes collected at each event, using NVivo[66] as outlined in Chapter 3. Five themes emerged from the analysis, discussed in turn in this chapter. My observations generally fell into two categories: the content of speaker sessions and workshop content and general observations on the environment and atmosphere. The latter constructs a set of findings and forms the first theme, section 4.2.1, "Environment and atmosphere in physical spaces", highlighting aspects such as the setting and layout of the physical venue. The second theme in section 4.2.2 focuses on "Actors in the space", examining the presence of attendees and presenters across events and analysing discussions of how various categories of actors are engaging with military AI innovation. Outlined in section 4.2.3, the third theme, "Pro-AI military sentiment", draws out how pro-military AI sentiment was communicated in each space. Observations of cautious or apprehensive

[65] See Appendix F.1 for more detail on each event
[66] As described further in Chapter 3, NVivo is a qualitative data analysis computer software package that allows for analysis on large volumes of text. See also: https://www.qsrinternational.com/nvivo/what-is-nvivo.

references to AI formed the fourth theme, "Caveats and caution", examined in section 4.2.4. In section 4.2.5, the fifth and final theme is "Military AI applications", exploring the contexts in which AI was showcased or discussed as applying to military actors. With some events spanning multiple days, some files were heavier coded and more frequently referenced than others.

The data collected through observant practice was in the form of fieldnotes. These were collected on paper or a laptop and included both fieldnotes that directly recorded my own thoughts of the events and tracked exact quotes from speakers as outlined through the following finds section. Where fieldnotes reflect my observations, rather than a quote, they are marked as "[Event], fieldnote observations", while other quotes should be understood as statements noted verbatim from speakers. Many sessions across these events adopted Chatham House rules, which prevent the attribution of a statement to a speaker or any others present at events.[67] These rules do not prevent the tracking of statements, hence why my fieldnotes included speakers' quotes.[68]

Quotes below will be attributed by event, using the shorthand names introduced in Table 4.1's far-right column: CyCon; CyConUS; DSEI; DTDT; DSDS, and AFCEA.

#### 4.4.2.1 Environment and atmosphere in physical spaces

In contrast to the themes in sections 4.2.2-4.2.5, this theme focuses primarily on general fieldwork notes rather than quotes from attendees and speakers. Observations on the physical space and atmospheres across the events tended to revolve around event location, security, event sponsors, and the event's nature (how organisers and attendees acted within the space, including in etiquette terms).

The location chosen for each event appeared to be shaped in part to attract each target audience. The AFCEA event in Brussels had several attendees from NATO Headquarters, which is based in Brussels; CyCon U.S. occurred a ten-minute walk from the U.S. Pentagon, with the opening keynote highlighting the venue as one familiar to many of the military personnel present. The DefenceIQ (DTDT) event took place in Grosvenor Square (London, U.K.) in a plush venue, a 16-minute commute from both the British Embassy and the MoD's Head Office, albeit, as one attendee remarked

[67] When a meeting, or part thereof, is held under the Chatham House Rule, participants are free to use the information received, but neither the identity nor the affiliation of the speaker(s), nor that of any other participant, may be revealed. See https://www.chathamhouse.org/chatham-house-rule (Accessed 20 April 2020).
[68] See Chapter 3 section 3.3 for additional methodical notes.

wryly, in a venue "slightly too close to the Russian Embassy" (DTDT). For the expositions, choosing a particular location is likely less relevant, though examining the city or region can highlight strategic interests relating to the event. Over 36,000 attendees attend DSEI events, with around 1,700 exhibitors (DSEI, 2019), so venue options are limited to those with sufficient space and capacity in London. Nonetheless, the choice of London and the U.K. more generally is historically determined in this case. DSEI originated out of British military equipment shows before privatisation into a format that maintained strong links with the British military and endorsement from the U.K. Ministry of Defence and the Department for International Trade (Roberts, 2017; Department for International Trade, 2021; DSEI, Online).

The settings in which the event is facilitated help shape the atmosphere at each event site. After going to a few events, I started counting chandeliers – DefenceIQ took place in the Millennium Gloucester ballroom with a rich burgundy carpet and 36 chandeliers. CyCon U.S. took place in the "grand ballroom" at the Crystal Marriott Hotel, hosting a mere six chandeliers and a prominent U.S. flag. These backdrops, coupled with the uniform worn by military personnel throughout both events, added a level of formality and an aspect of exclusivity. For Defence IQ, this exclusivity was heightened as the event was ticketed at £799+VAT for standard tickets. The deep red carpets and round seven-person tables at DTDT could not be further away from the very different form of exclusivity on the DSEI exposition floor, with attendees welcomed to inspect and hold a wide range of weapons and military equipment.

At each event, the security in place also contributed to the researcher's perceptions of exclusivity. Security varied widely between events. DSEI had the highest security levels of all events, with physical border fences blocking non-attendees from approaching the building. In addition, a high-security presence was marked by a high level of security personnel shepherding attendees and bystanders towards and away from the entrance. Three different entry points checked the researcher's face against a pre-submitted picture and ID before entry was permitted. Image 4.1 highlights the attendance rules displayed outside DSEI, located between the train station exit and the front entrance to DSEI (Excel London). The walkway between both points was limited to DSEI attendees.


*Image 4.1: The attendance rules displayed outside DSEI (Author's own image, 2019).*

The security measures at DSEI contrasted sharply with CyCon U.S., where I was welcomed with a smile and required no photo nor ID.

> CyConUS fieldnote observations: "Managed to forget my pass on the second day of the conference and was ushered in with an 'it's fine – enjoy'."

The difference in events' physical security levels often mirrored policies on the publicity and exclusivity of information beyond the event. CyCon U.S. did not have stringent security principles enforced, with talks also photographed, recorded, and later uploaded on the Army Cyber Institute Youtube channel (Army Cyber Institute, 2020b). This perceived inclusivity and transparency were at sharp odds with other events, including the AFCEA discussion workshop that stressed Chatham House conditions. At DSEI, the capture of pictures or recordings felt heavily discouraged,[69] particularly considering security restrictions and active protest activity against the event).[70] More restricted sessions seemed to allow for more candid discussions, including knowledge that was

[69] The DSEI website states that "no photography or filming of stands and/or exhibits is allowed without the expressed permission of the stand manager." https://www.dsei.co.uk/admission-policy, retrieved 8 February 2022. The additional heavy security presence at the event felt comparable to an airport security environment and I did not feel comfortable attempting to capture media once I had entered the venue, predicting verbal challenges from the event's security personnel.

[70] The campaign group "Stop the Arms Fair" has a public request for sharing image and videos of DSEI: https://stopthearmsfair.org.uk/send-dsei-images-us/, retrieved 8 February 2022.

designated to be shared between attendees only (e.g., at AFCEA sessions). In this way, the barriers to the space reveal – and contribute to how intimate the sessions might be.

I found the sessions which invoked Chatham House rules tended to hold less general content and go into more detailed discussions. An exception to this rule was where speakers used their platform to make an announcement that added to knowledge in the field. For example, Estonia's Prime Minister stated Estonia's understanding of international law in cyberspace in 2019 (ERR News, 2019). In the same year, defence companies presented at DTDT to advertise high-level current capabilities and products.

Table 4.2 summarises sponsor affiliations for that year and which organisations had stalls or advertisements for each event.

*Table 4.2: Sponsors and Partners (non-exhaustive)*

| | |
|---|---|
| CyCon | Around six commercial cybersecurity vendors, including cyber threat intelligence firms. |
| CyConUS | Sponsors included Army Cyber Institute and NATO CCDCOE, who had their own stands. Non-industry stalls included universities showing off wargaming projects. |
| DSEI | As a trade show: hundreds of vendor pitches from major established defence firms (Thales, BAE Systems, QinetiQ as some West European examples) to some less well-known entities. Half the exposition hall was split into national segments (with a separate pitch for the U.K., Turkey, Israel and so on). |
| DTDT | Media Partnership with the Defence Academy for the U.K. Various industry presenters. |
| DSDS | Sponsors: Dstl, Atomic Weapons Establishment; (Cranfield University, n.d).<br><br>Supported by: Publishes (Elsevier, Taylor and Francis), SPIE (international society for optics and photonics), and Janes (intelligence firm). (Cranfield University, n.d)<br><br>Stalls for Defence Researcher; GCHQ (including NCSC representatives). |
| AFCEA Europe | "Sustaining Partners": Samsung electronics; Blackberry, and Secusmart. Industry panellist sponsor: Commvault (AFCEA Europe, 2019). Event organised with the cooperation of the NATO C2 Centre of Excellence. |

As Table 4.2 highlights, sponsorship was a feature across all events. The less commercial the event, the less explicit focus was on free items distributed to attendees. Vendors could purchase specific

vendor packages for more trade-focused events like DTDT and DSEI. Image 4.2 shows the DTDT ticket offering as of September 2021:

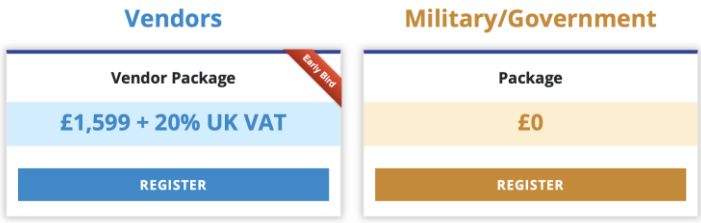


*Image 4.2: DTDT 2021 entry costs for vendors and military/ government attendees (screenshot from DTDT, Online)*

Identifying attendees by type was certainly straightforward for military personnel, who attend a range of these events in full regalia. This uniform granted them authority, as observed through events in Washington D.C., London, and Brussels, with military personnel bonding over their respective tours of duty. I also spoke with a group of eight final-year students at the United States Air Force Academy (USAFA)[71], who were in their uniforms and joked that they felt overdressed. The students commented on the division between those in and out of uniform and expressed unease when I was sidelined from conversations due to the lack of in-line attire.

> CyConUS fieldnote observations: "Elevation of military status - being interrupted in a research discussion by an older man who ignored me, thanked the in-uniform MSc student and spoke about how his son was serving. Very not-self-aware. Uniformed officer very cordial but also slightly embarrassed as he was aware I was cut out".

Except for DSDS, which primarily encouraged doctoral research attendees from across the U.K., there was a notable lack of demographic diversity across events in terms of gender, age, and an institutional norms perspective (with few critical researchers identified through the authors' interactions in each space). This representation reflects broader challenges around diversity in the military and defence contexts, as noted in U.S. (Kamarck, 2017) and U.K. (Ministry of Defence, 2021) government publications. A speaker at AFCEA was introduced as having "began his career in the 80s, around the same time as all of us". This statement applied to around 80% of the attendees in the room (about 50

[71] The USAFA is the U.S. military academy for the U.S. Air Force (and as of December 2019, the U.S. Space Force). It runs competitive-selection four-year Bachelor's of Science degrees for officer cadets and educates across a range of technical and non-technical majors in addition to military training, including opportunities to focus on cyber security systems and the cyber domain.

people in total). At CyCon U.S., it was simultaneously bemusing and frustrating to see an all-male, white, informational-warfare panel on information warfare with a Bryan, Brian, Ryan, Renny, and Robert - all of whom appeared to be over 40. The panel was informative, and each presenter was clearly qualified. However, there was no discussion nor reference to the experiences of different groups. On the contrary, speakers enjoyed what the fieldnotes call "beard banter" as presenters complimented each other's facial hair (CyConUS).

> AFCEA fieldnote observations: "mostly male - attendees are in military uniform/blazers/ formal."

These observations, coupled with remarks highlighted by attendees in the above section, raise the question of how the field can welcome inclusive engagement from a diverse range of perspectives - when younger, underrepresented individuals are less likely to enter these spaces as attendees, never mind as speakers.

> AFCEA fieldnote observations: "My thoughts - ageism and also fact that [sic] senior leaders of this field generally have this [their age] in common. In the military, you can't just come in and revolutionise, like Facebook did to tech, for example."

Finally, the emphasis on talks versus networking shifted between events. At DSEI, for example, scheduled talks appeared of secondary interest to networking, with speakers in open-plan tents on the exposition floor often barely audible over the general babble from nearby stalls. This arrangement is a deliberate prioritisation; the DSEI 2021 website promises "a range of valuable opportunities for networking…. the DSEI community can strengthen relationships, share knowledge and engage" (DSEI, 2021). The sharing of expertise through talks is highlighted only as "access to relevant content & live-action demonstrations", listed below networking and the event as a "platform for business" (DSEI, 2021). This ordering hints that engagement through talks emphasises trade and business interests rather than curiosity about the technologies' implications. Business cards were also a regular part of exchanges, particularly through networking sessions. Having a business card was the norm. I felt that having business cards, while perhaps unusual for a doctoral student, was one way I could help show credibility, and I found the cards a valuable tool to follow up with other attendees.

### 4.2.2 Actors in the space

Innovation within the private sector was a prominent theme mentioned frequently across events, in particular highlighting that cutting-edge innovation was taking place in the commercial space rather than in military institutions. Commercial firms were said to be "driving new and unexpected developments" (AFCEA), with technology sector innovation "outstripping" (AFCEA) breakthroughs in the defence domain, and commercial industry innovation was "driving [the] military" (DSDS). While the commercial sector held "stunning opportunities' for governments and militaries…" one senior military speaker warned, "industry is doing things that we don't know about" (DTDT). At CyCon, a speaker stressed that the military should be trying to keep pace with, rather than compete with, industry innovation. Commercial innovation was recognised as a trend beyond large technology corporations, with start-ups and non-traditional companies driving "disruptive innovation" (DTDT), which had "displaced established market-leading firms" (DTDT). Traditional defence firms were themselves highlighted as parties investing in emerging technologies, including quantum computing and AI (AFCEA).

There was no simple answer on the status of private-public relationships for defence innovation. When an audience member at one event asked whether it seemed that private industries did not want to work with militaries, citing Project Maven,[72] one speaker responded by highlighting that "the private sector isn't monolithic" (DTDT).

> DTDT: "the society that will be best at ML, in any form, will be the society that can best combine its civilian tech secrets with its military capabilities."

Civil-military cooperation represented "a two-way relationship – that's a relationship that should be aided and abetted" (CyCon), with consensus across events that nations should seek greater cooperation with the private sector to realise "stunning opportunities for governments" (CyConUS). There was a broader acknowledgement that the military should seek to cooperate, rather than replicate or compete with, the innovation of the private sector, utilising "horizontal innovation" in adapting

[72] See Chapter 7 for more detailed discussion on Project Maven.

existing AI tools to "avoid reinventing the wheel" (DSDS). This agreement acknowledged, speakers highlighted how "private companies [are] traditionally motivated by economic advantages, and push towards the cheapest, fastest alternative" (DTDT), highlighting some non-alignment with military priorities. At DSEI, an industry representative complained that the government did not know what they were asking for. A DTDT vendor expressed frustration while stating, "most of our [defence organisation] customers have no idea what they're going to need", mirroring the sentiment from U.K.-based industry professionals interviewed as part of this broader research (See Chapter 5).

How various governments approached technological transformation was also raised as a challenge to private sector engagement, with U.S.- and U.K.-focused speakers citing the risk-averse nature of government procurement and acquisition processes for their respective systems. The bureaucratic and organisational cultures across militaries and MoDs were raised as an overarching challenge for defence-focused innovation, not limited to the U.K. or U.S. context. As one (U.K.) government-employed speaker stated, governments in general "need to be able to accept failure" (DSDS). At DTDT, another (U.K.) military-employed speaker complained that "everything takes a million pounds - everything takes a year".

For the U.K., the difficulties in coordinating programmes across Army, Navy and RAF projects were also highlighted as a challenge, with innovation in emerging technology "opening up Pandora's box of organisational culture and service rivalry" (DTDT). Commenting on the UK MoD's initiatives for collaboration, one person presenting expressed frustration as they stated, "to be blunt, you could say it's [coordination efforts by the body in question] a waste of tax-payers money. You need some very senior colleagues to fix it" (DSDS). One speaker highlighted planning as a particularly frustrating process, arguing that the U.K. MoD wanted AI but had "no idea how to waterfall that out over the next 5-10 years" (DTDT). While industry events seemed to want to focus on delivering innovation, attitudes to innovation by decision-makers were spread across the spectrum from keen to frustrated. As noted in reflective field notes at DTDT, there was a sense that innovation was stagnating due to organisational and bureaucratic barriers: "[the] speaker wants to skip reports and consulting and just get testing rather than paying one million for 'a Yellow Pages of evidence on impact'" (DTDT).

Beyond private sector innovation, academia was highlighted as a tool that the U.K. MoD were not using as much as it should, with speakers admitting "too much time doing industry 'business of the day' with not enough time to think" (DTDT). At both academic events and expositions in the U.K., MoD stalls encouraged partnerships between industry and academics. At DSDS, the Defence and

Security Accelerator (DASA), a body within the U.K.'s MoD that "finds and funds exploitable innovation for a safer future" (DASA, 2021), pitched themselves to academics as a hub for collaboration between the U.K. government and (academic or industry-driven) external sources of innovation. Inter-governmental groups were also raised within the context of events they had contributed to organising. The NATO Cooperative Cyber Defence Center of Excellence formed a "valued partnership" with the U.S. Cyber Army Institute (CyConUS) and is described as a 'voluntary group of nations' coming together to focus on the cyber threat (CyCon).

It was common at events for presentations to be framed in the context of the strategic security landscape. Speakers described "great power competition" (CyConUS; CyCon; DTDT) and "strategic inflection point" (DTDT) as states such as China developed their technical capabilities. This rhetoric, putting realistic, security-centric language at the centre of a discussion, immediately shapes the priorities of those stating the phrase. From here, language prioritised the importance of maintaining a national competitive advantage, focusing on the potential instability that may arise should the current international order be disrupted.

DSDS: "Not everyone is as law-abiding or ethical as us [the U.K.]."

Conversations at each event focused on the perceived - or evidenced - threat from hostile international actors, again as a motivator for (usually) U.K., U.S., or Allied innovation, depending on the event. China and Russia were mentioned repeatedly across events, focusing on China for AI specifically. China's activity in intellectual property espionage had furthered their technological military advances, with this "asymmetric threat" subjecting the U.S. to the "death of a thousand cuts" when it comes to protecting military innovation (both CyConUS). Modern technologies "meant to connect the world" were used "to subvert and control" targets (CyCon), and there were emerging threats, including swarm UAVs and drones as weapons that rebel groups had used in a way that concerned the U.S. DoD (DTDT). The exception to this adversary-specific framing was at DSEI, where the field notes note a "different geopolitics - would not normally have a lot of these actors in the same room" and where the focus was more on trade facilitation than discussion.

### 4.2.3 Pro-Military AI Sentiment

The language used to frame AI-enabled technology showed that pro-innovation rhetoric was slightly more common than cautious phrasing. Presentations listed on agendas included "How AI will drive the change in the art of decision-making" (AFCEA agenda) and offered little room for a nuanced discussion of the benefits of such innovation. Conversations were often very explicit about the benefits of innovation, with one speaker describing disruption as: "nothing bad – [and] actually very cool" (AFCEA). Unmanned systems were introduced as "game-changing opportunities" (DTDT), while a future of "accelerated procurement" and "rapid integration" promised to place technology "into the hands of the warfighter" (DTDT). In one session, a presenter appealed to a room full of government and industry professionals, "you have to zoom in, and you have to fund AI, and you have to take some risk" (DSEI). This approach to risk-benefit calculations was demonstrated through another presenter's suggested motto: "the electric light did not come from the continuous improvement of candles" (AFCEA). Time pressure was utilised multiple times to stress the need for innovation. The U.K. was described as having a "time-limited opportunity" (DTDT) to innovate, with a presenter urging an audience of the "need to push relentlessly towards operational deployment" (DSEI), referring to warfighting experiments with autonomous agents.

The language used at presentations across events, particularly by industry speakers or in trade environments, demonstrated pride and excitement towards AI military technology and specifically AI-enabled weapons. At DSEI, I reflected in fieldnotes that "people were proud of their weaponry and tech - and lethal capability of this tech". One industry representative at DSEI boasted about how technology had improved the rate at which the company could manufacture bullets. A CyConUS speaker praised the U.S.' "persistent innovation" in the defence space. The fieldnotes noted the apparent assumption of presenters that the research focused on at DSDS that innovation in emerging technologies should be seen uncritically as a good thing and argued for more funding in "exploitable innovation" (DSDS). This view was agreed by everyone present without further need for any discussion on ethics or responsible innovation. Likewise, DTDT presenters highlighted AI as "the future of human-machine teaming" and a "game-changing innovation for unmanned systems" without exploring any potential limitations or challenges as part of the presentation. At DSEI, I reflected in the observational fieldnotes that the industry representatives I engaged with "were very proud of the fact their machines could make one million bullets a day - a fact that struck me - how separate colleagues [employees] see their role from the lethal consequences of their action".

The lethal consequences of using – and not using – AI technology were also highlighted across events as a motivator for increased investment in development and deployment. "Where we are going, good

enough is dead" (DTDT) was a forceful quote in one military promotional video, highlighting how any AI-enabled system must have extremely high accuracy and reliability thresholds. Another speaker promoted targeted precision weaponry with the preface "as a royal marine who likes killing people efficiently…" to audience titters (DTDT).

DTDT: "Science fiction is becoming science fact."

Language often drew on science-fiction terms and contexts, with speakers using popular culture references to frame the discussion. The dystopian show *Black Mirror* was referenced in presentations discussing the conflict between AI systems, as the speaker went on to describe the "advent of a new age in quantum computing and AI" (DTDT). A presentation by a private defence firm mentioned future battlefield studies by Ubisoft, arguing that AI was able to compete with science-fiction-inspired "real-world strategy games'" asking the audience "don't be overly sceptical" of the value of advanced multi-player games as a valuable tool for the military (DTDT). At another event (CyConUS), the speakers aligned themselves with *Homo Deus,*[73] with the speaker stating, "human technology is transforming what it means to be human". While this isn't science fiction, the speaker utilised such cultural reference points as a launchpad for a speech inspired by science fiction scenarios, using provocative futuristic scenarios to demonstrate how humans are influenced by the technology they create.

### 4.2.4 Caveats and caution

While the need to innovate with technology was noted ubiquitously across events, several speakers and presentations displayed some apprehension. There was evidence of some pushback against AI where it was deemed unnecessary and "foisted on" the military (CyConUS).

CyConUS: "We are creating technologies faster than we can civilise them."

At a technical level, some general challenges were examined at these events. The question of data was repeatedly highlighted, with one claim that defence leaders "don't need to worry about the AI side of the house [they] need to worry about the data… there are so many questions" (AFCEA). Technical terms, bias, missing context, bad quality data, and data assurance were all potential risks.

[73] Homo Deus is one of several popular science books authored by critically acclaimed author Yuval Harari. The book examines the future of humanity given the onset of trends including artificial intelligence. See: Al-Amoudi, Ismael. "Homo Deus: a Brief History of Tomorrow by Yuval Noah Harari." *Organization Studies* 39, no. 7 (2018): 995-998.

AI algorithms were still attempting to mirror the "super super complex" nature of human-based interactions (DSDS). An inadequately trained algorithm could lead to mistakes and unpredictable outcomes that risk human life. One speaker proposed an example where "if you train it [the algorithm] to recognise a trick as a target, but not distinguish it from a Red Cross truck, you have a problem" (AFCEA). Traditional cyber security concerns apply to AI systems, and any conventional security vulnerability risks the integrity, availability, and confidentiality of the algorithm and its data. While one speaker summarised supply chain concerns by stating, "if you control the patch to that piece of equipment, you own that equipment" (CyCon). Scalability was also raised as a challenge, with technical presentations highlighting how factors, including hardware or access to sufficient computing power, remain a constraint for innovators (CyCon; DTDT). Some suggestions were made on how to address these concerns, for example, collaborating on civilian-military projects to AB test repeatedly and measure the benefit (AFCEA) or conducting "appropriate experimentation" in terms of testing and validation in software "sandpits" that represent realistic conflict environments (DTDT). These conversations clashed, to an extent, with presentations that proposed more forceful innovation in "what is going to field quickly" (DTDT), though overall discussions did focus on the need for testing and validation once there was a proof-of-concept application.

Certain areas of AI-enabled technologies appeared to inspire unease, especially around trust and autonomous activities, including decision-support. A speaker at DTDT articulated that "the problem is when it [an AI application] gets truly autonomous, making decisions to engage". This unease was also raised in the lack of transparency in "algorithms talking to algorithms" (CyConUS), with speakers referring to the speed of warfare increasing to the extent to which humans would not have the time to understand or process machine-determined decisions.

DTDT: "[The] European mindset is - we would rather lose the war than use autonomous weapons against adversaries."

Trust and confidence were also concepts repeatedly discussed in written materials and speakers' sessions across events. Military technology innovation took place in a "denied environment" (AFCEA) in which conflict may often reduce visibility and force decision-making with very little information. While AI may offer assistance with decision-support algorithms, "making decisions based on info we've given [it] and that the algorithm thinks is helpful… by the time it bubbles up to decision-makers, we don't know how many filters are on it or where the information has come from"

(CyConUS). This raises the question of "how can you encourage soldiers to be positive about autonomous systems?" (DTDT), with "algorithms talking to algorithms" (CyConUS) leading to potential crises like the flash crash in financial systems and potentially catastrophic impact in a conflict in ways which threaten human safety.

At the same time as this apparent firm stance for human decision-making, there were some differences in perspectives regarding the acceptance of change. When a speaker argued, "senior leadership is willing to accept new tech", the first audience question raised "older generations… we are scared by this new technology" (both AFCEA). As one senior Army General put it, "When I make a position… I want to look a handful of people in the eye" (AFCEA).

AFCEA: "The biggest challenge today is not AI…. what's really breaking down is the old way of looking at things."

Discussions about how far AI should be trusted usually highlighted the crucial role of humans as "decision-makers in operations" (CyConUS).

CyCon: "Human beings build the future - we aren't powerless to control it."

Military-employed speakers were consistently quick to dismiss fully autonomous engagement with the decision process, stating "you need guts to do that" (AFCEA), arguing that humans will remain in the decision-making lifecycle with continuous "meaningful human control" (DTDT). This emphasis was consistent across events, with no speakers arguing for systems that could autonomously engage in conflict.

Finally, non-technical challenges to AI innovation for the U.K focused on the military and government's ability to innovate. One military presenter argued that the U.K. military is "not incentivised to do that very well…. we punish failure" (DTDT). A seminar session on government innovation, mainly with the U.K. accented attendees, highlighted how the U.K. "has to do better" to go from talking to "walking the walk" of effective technological adoption (DTDT).

### 4.2.5 Military AI applications

Several active or completed AI projects were referenced across events. DSEI demonstrations included "Army Warfighting Experiment Autonomous Warrior" by the Royal Navy's 'NavyX' programme.[74] Autonomous smart missiles were identified as an example of existing "narrow AI" (AFCEA). At the same time, project reports highlighted German air force projects in AI-assisted command and control at "green/ ongoing status" (AFCEA). The U.K.'s early-stage pilot "Minerva" represented a promising naval lab programme focusing on robotics and drones (Army Technology U.K., 2018) (DTDT). The technologies developed and deployed through the U.S. Project Maven[75] represented a potential precedent for fielding military AI innovation rapidly (CyConUS).

Events highlighted applications where AI may feasibly benefit military capabilities. In intelligence, AI can play an incredibly valuable role in processing data to "identify trusted sources, find anomalies in information, [and] filter information" (AFCEA), analytics (AFCEA/ DTDT), and simulation modelling (DSDS). Many potential applications of AI in a military context were still mentioned at a conceptual level, particularly around AI-supported decision-making. There were open, very broad questions or open challenges raised when predicting possible futures. Data processing might be vastly improved in speed and capability terms, with examples of risk management analytics to conduct impact analysis (AFCEA). DTDT speakers raised the capabilities demonstrated by Deepmind's AI system "Alphastar" in a computer game grandmaster competition.[76] This military interest in Alphastar had previously been highlighted as a risk by Professor Noel Sharkey, who emphasised that despite the game not being a realistic war simulation, "military analysts will certainly be eyeing the successful AlphaStar real-time strategies as a clear example of the advantages of AI for battlefield planning" (quoted in Sample, 2019, para 11).

Often the language focused on technology rather than humans as the referent object. Audio narrating "precision strikes [with a] lethal radius of 5m" overlaid a military promotional video of a human-anatomy dummy exploding (DTDT). Military AI innovation was described through "prototype warfare", "third dimension warfare", "machine learning warfare", and "statistics on steroids" (all DTDT). "Weaponised AI" was used at a CyCon paper presentation (CyCon, Burton and Soare 2019).

[74] "NavyX is the "Royal Navy's new Autonomy and Lethality Accelerator, which will rapidly develop, test and trial cutting-edge equipment, with the aim of getting new technology off the drawing board and into the hands of our people on operations at a pace that has not been possible before." (Online). Accessed 1 October 2021. https://www.royalnavy.mod.uk/news-and-latest-activity/operations/united-kingdom/navy-x

[75] See Chapter 7.

[76] For more information on AlphaStar's performance in the StarCraft 2 grandmaster see: https://deepmind.com/blog/article/AlphaStar-Grandmaster-level-in-StarCraft-II-using-multi-agent-reinforcement-learning

While vague, these terms highlight the phenomena brought around by AI's "disruptive innovation" (DTDT). The lack of consensus on terminology suggested a lack of clarity on how AI is discussed in warfare. It may be that the lack of widespread "weaponised AI" activity during the research period meant that there was no requirement for agreed terminology at that point. Alternatively, agreeing that terminology is considered useful, it appears more likely that discussions in these environments are too immature to have arrived at a common vocabulary, at least during the period in which these events occurred.

When describing how AI-enabled systems might be utilised by the military, there was a consistent acknowledgement of 'human-in-the-loop' dynamics as part of the discussion on human-machine teaming. There was the technical potential for AI to operate as a "synthetic human in the loop" and that "synthetic human behaviour is something to look into", though presenters considered it likely that humans would ultimately retain control over the last part of a decision-process (all AFCEA, same speaker). Some of this assumption was credited to current technical limitations that meant human assurance was still deemed superior, with modern AI techniques still struggling to represent human brains (DSDS). Others raised trust and human connection as valuable assets. As one military attendee stated, "when I make a position, before that, I want to look a handful of people in the eye!" (AFCEA). Replying, another panel member quipped, "so there'll be work for the generals", to laughter in the room. As a light-hearted comment complemented by agreement from the Chair, it was interesting as "bearing in mind the [largely military] audience the Commander is speaking to. Is he more likely to say they're needed and valuable?" (AFCEA, authors fieldnotes).

## 4.3. Discussion

Observing spaces in which key stakeholders, including the military, government, or technical partners, were likely to be exposed to semi-public discussions on military AI revealed several assumptions and a range of attitudes held on the topic. These findings highlight how the dynamics of a trade show, knowledge exchange forum or academic gathering facilitate the networks and idea-forming that drive military AI projects. McCann (2011) argued that face-to-face conference attendance is a crucial element of policy influence, drawing on Larner and Le Heron's (2002, 765) description of "global microspaces". Such spaces form in the hallways, bars, and cafes around conferences. The formation of such spaces facilitates trust-building, reputation development, the sharing of experts, and the connections between communities that may otherwise be isolated (Larner

and Le Heron, 2002). Trade shows and conferences gather a community of interest, including academic, policy, military, and industry specialists, in a way that facilitates "co-presence" and "transfer" sites (McCann, 2011, 117), with ideas shared between attending communities.

### 4.3.1 The dynamics of each event

The questions and engagement with the audience and interactions in scheduled breaks between presentations shed light on how participants influence atmospheres in event spaces. Jackman (2016, 2) highlights how such gatherings can be seen as a "barometer" that reflects perspectives on relevant topics and debates in the field. For military AI the attention and high-level discussions demonstrate an interest in the technology's implications. The language on military AI represented a range of perspectives, from pro-AI rhetoric that skipped over matters of verification and security to cautious warnings on the need for human oversight and trust-building measures. This range of views reflects different arguments reflected in the literature explored in Chapter 2 and suggests that the field is immature in assessing current capabilities and shaped by a range of actors with different interests. The material presented by industry suppliers of AI-enabled technologies was, unsurprisingly, focused on the positive benefits of their products. At the same time, conversations around private-public partnerships again held different emphases depending on whether the discussant was public sector, from a large defence firm, or the less established defence vendors.

Physical events are often an occasion where ex-personnel, or personnel in the same field, get to meet up when they may otherwise be situated across the world. They act as social gatherings and almost provide a physical reunion setting in certain circumstances, especially where the conference targets a particular community. Furthermore, there was evidence of emerging and ongoing discussions across these events, with question-and-answer sessions often converging on military acquisition, dynamic adoption, and examining the opportunities (and occasionally challenges) related to AI in military applications. Particularly considering the emerging nature of military AI, physical events represent a space to access information on activities that may not yet be published in the accessible literature. This observation aligns with Jackson's (2016) analysis of trade shows focusing on drone technology which provides a "window of access" into an "emergent landscape", and how defence deals are brokered out of sight of critics or rivals (Rech, 2015). These findings align with McCann's suggestion that travelling to conferences and meetings may be valuable forms of "policy travel", in which events form a "particular kind of social setting" to facilitate idea generation and (policy) knowledge transfer

(2011, 117).

### 4.3.2 Government procurement and "disruptive innovation."

Two prominent discussion themes repeated themselves across events. The first scenario related to discussions that revealed frustrations around procurement and innovation (usually with defence firms exasperated at the MOD, in the U.K. context, *or* between those trying to understand how innovation is being spearheaded in various national contexts – no simple feat as experts outlined). The frustrations from suppliers attempting to engage with defence sectors are well noted in the literature. The U.S. procurement process is described as cumbersome (Imbrie, Kania and Laskai, 2020), while the U.K. experiences tensions aligning its military ambitions against budget at resource constraints (Dorman and Uttley, 2020). The perspective is held by governments beyond events, as Giry and Smith (2020) outline how the two post-2010 governments have both framed defence procurement as a challenge for the U. K's defence capabilities and suggest no emerging political consensus on the issue. These findings align with the military innovation literature which outlined the organisational barriers to change. For example, the friction described by attendees highlighted that military institutions were struggling to adapt their structures to provide the appropriate expertise or flexibility, in line with Horowitz's (2010) reflection on how military innovation required organisations to have the capacity for rapid change.

The second scenario was the sponsored talks from suppliers and the military: defence firms highlighting how they are implementing autonomy into their products, highlighting the importance of human-in-the-loop dynamics when questioned but otherwise disproportionately focusing on the benefits, rather than potential risks, of employing AI-enabled technologies. This emphasis reflects the defence industry's contribution to "techno-hype in the military" in such a way that "indisputably distorts" viewers' understanding of the technology's actual capabilities (Elhefnawy, 2018, 9). Similarly, the tendency to dehumanise (human) targets while boasting of the benefits of lethal projects rather than the potential consequences for wrongly identified targets. This finding was displayed neatly in a promotional video for explosive devices against human-shaped dummies to demonstrate precise target acquisition (DTDT). Analogies can be drawn here with Cohn's (1987) observations on how the military uses abstract language to distance itself, with events mirroring her observations of military personnel's use of "collateral damage" (Cohn, 1987, 691) to refer to civilian loss of life.

The fieldnotes revealed a difference between events that highlighted AI as a key theme (AFCEA being a clear example), events that had talks specifically on AI technology (DTDT), and more traditional-style agendas which treated algorithms as a side to traditional kinetic force (DSEI). DSEI categorised their exhibition into the Aerospace, Land, Naval, Security and Joint Zones – with only a tiny minority of stands (including the MoD and NSCS) showing nods to AI in their marketing. This layout does not mean the technology was not evident in products across the fair; more that themes were not of particular general interest. A fighter jet, a major project consisting of a syndicate of defence companies working on a contract with the U.K. MOD, had elements of AI in it, enabling information processing assistance for the pilot – but was not providing any of the critical components for the plan. Rather, the AI aspects of this technology were added as an incidental software add-on. Taking a tour of the plane with one of the contracting firm's employees, the colleague emphasised how long contracts could take to fulfil, with work on the aircraft taking over a decade. Changing software requirements, such as the addition of learning algorithms, among other updates, could come with extensive full-product testing and evaluation costs to make sure the product was still secure.

While the observed events differentiated in purpose, a key theme across the data was the importance of the commercial sector and the growing role of non-traditional defence suppliers in military AI innovation. This trend reflected the literature, which argued that innovation is now largely in the commercial sector (Cummings, 2017). Sponsorships across events provided an opportunity to draw attention to a range of public and private efforts, while the events with greater industry presence encouraged attendees to engage with vendor stalls or with vendor presentations. These environments all contained opportunities to develop partnerships and industry growth, traits highlighted as a feature of security-focused trade shows (Jackman, 2016), in which trade shows provide an emphasized focus on products and market growth. Lee and Kim (2008) propose a four-dimensional model of trade shows to include sales-related, relationship-improvement, image-building and information-gathering, which I suggest extends to the non-trade show focused events attended through this research to varying degrees. DSEI was the most trade-focused environment by far; Prichard and O'Nions (2005, 477) describe the 2005 iteration as a "customer rich experience", "where negotiations are conducted, collaborations are planned, and purchases considered" for military equipment and weaponry. Looking at written resources, it appears that the language and parameters of the event have changed. The DSEI website in 2021 emphasised the event as a forum to "strengthen relationships, share knowledge and engage in the latest capabilities" rather than its 2005 proclamation of its' "important role within the selling process for defence companies" (DSEI 2005 brochure, quoted in Prichard and O'Nions, 2005,

476). Nonetheless, the observant practice approach revealed a heavy industry and networking-specific environment at DSEI and, albeit through presentations rather than equipment stands, at DTDT. Vendor presence at all events with military or government attendees highlighted how industry partners see value in utilising events to their advantage, whether this is through networking, engaging with recent discussions or through potential sales opportunities. Distinct from conferences and primarily knowledge-sharing environments, trade shows and exhibitions provide a forum for buyers to display their wares to potential buyers interested in buying such products (Situma, 2021). A survey of 32 defence sector employees highlighted that nearly all interviewees considered trade fairs the most valuable marketing tool for their industry (Cop and Kara, 2014). This perspective is likely because of the significantly fewer marketing opportunities for the defence sector compared with non-defence commercial organisations; both as the market exists as a relatively opaque environment with end-users a finite number of national armed forces, and as armed forces may be influenced by a preference for indigenous products over cheaper foreign alternatives (Cop and Kara, 2014).

### 4.3.3 Observing awareness and perceived responsibility

This research benefited from the chance to interact with the communities attending these events beyond passive observation through listening to presentations or vendor pitches. During periods of active engagement through exposition spaces or networking sessions, introducing myself to other attendees or industry representatives at vendor stands revealed that the scheduled talks rarely represented the knowledge level that some attendees held. Speaking with attending engineers, intelligence officers and civil servants involved in defence innovation, I was able to note their clear understanding that using learning algorithms, much like using any emerging technological tool, comes with a set of risks and potential unknown consequences. It was intuitive to those I spoke to that there were challenges relating to training data or potential bias. However, while individuals may find questions about "responsible AI" implementation conceptually interesting, it was not a part of their day job to apply these aspects of critical thinking to topics surrounding military innovation. Instead, the goals of their employers seemed to come to the forefront – for engineers - working to develop the technical features rather than remain concerned about implementation procedures. For decision-makers, working to determine strategic military advantage and more accurate, precise, and effective tools, not preoccupy oneself with principles around fairness, accountability and transparency (terms which researchers will come across when looking at artificial intelligence adoption across parts of the civil space). The fact that attendees and speakers were aware of these aspects yet did not focus

on them implied that this was a deliberate omission rather than a lack of awareness or understanding.

## 4.4 Closing thoughts: positionality

Through each event, I also noted my positionality as an academic researcher and observer rather than an attendee aiming to invest in, procure, or sell a particular technology. In some ways, this enabled me to take on a neutral standpoint, noting down cautious or enthusiastic language from attendees both wary of, and highly excited about, AI technology. This said, the freedom to view each event as an outsider through the lens of security likely formed a bias to pick out where an event environment, speakers, or attendees did not consider the security implications in their discussions. I also observed what I viewed to be a lack of possible critical reflection on the use of AI technology, writing that research on emerging technical opportunities tended to be granular, with 'scientists working on an algorithm rather than an end user' (DSEI). Similarly, at another event where consequences were discussed in terms of legal liability, I wrote, "so they aren't worried about doing something wrong on principle – more the consequentialism" (CyCon). This observation marked different reasoning than some civil society events I had been to where it had been more [based] on principle.[77]

My observations were shaped in part by my positionality as a non-military attendee in, for the most part, a highly securitised environment (either through physical security settings or by virtue of each event themed).

> Fieldnotes - CyConUS: "I definitely felt like an outsider at this event - the goal seemed to directly build together the military and U.S.- community."

However, this position is complicated somewhat by the fact that I would inform people about my position as an academic and articulate my research interests - and I was known to some attendees at some events. At DSEI, a private defence firm employee I had met at a previous event introduced me to their colleagues as "an influencer in this [military AI]" space, referring to an earlier presentation they had seen me give. While I may view myself as an outsider, others may not necessarily have the same view. This positioning did allow for more informal discussions, including where attendees

[77] Examples include: Stiglitz, Joseph. The Future of Work - You and AI. Speech at the Royal Society, London. 11 September, 2018. Schwarz, Elke. "Death Machine - The Ethics of Violent Technologies. Book Launch at Queen Mary University of London. 9 March, 2019.

known to me expressed their personal pride in their employers' weaponry and technical prowess in AI technology. Regarding other aspects of attendee attributes, I often made notes where I thought the phrasing used by an attendee (or presenter) revealed possible assumptions or consensus points, including age. A military commander reassuring a military audience that despite AI, "there'll be work for generals" had me comment in my fieldnotes, "bearing in mind the audience the commander is speaking to. Is he more likely to say they're needed and valuable?". During the same session, I reflected on age and the correlation between age and seniority in the military, writing: "got to do your time to have influence. Hierarchical nature".

# Chapter Five: The United Kingdom - practitioners' perspectives on military AI innovation

## 5.1. Introduction

This chapter explores the military AI innovation landscape in the U.K. Over the course of this thesis research, the U.K. has developed their approach to AI with the release of a (non-defence focused) national AI strategy and the establishment of numerous government initiatives designed to keep the U.K. "ahead of adversaries" (MoD, 2021a, 75). This chapter will focus on AI-related initiatives and statements relating to military themes. The UK Defence AI Strategy and corresponding policy paper were publicly released on June 15, 2022, after the submission date of this thesis.[78]

In the absence of formal doctrines or policy outlines at the time of research, this thesis explores the dynamics of current innovative practices and the corresponding security challenges through the perspectives of expert practitioners in the field. Fourteen interviews were conducted with senior-level defence experts at large defence organisations, all of whom had significant experience determining the direction of AI innovation within their organisations. All but two worked for established suppliers to the U.K. MoD and the U.K. military.

These interviews took place from February 2020 to June 2020, and interviewees were selected based on their professional experience. The selection process ensured all interviewees worked for commercial defence-focused organisations with responsibilities relating to AI or emerging cyber security challenges in military contexts. Most of the interviewees within this project were employed at the "defence primes", large established defence corporations representing long-standing suppliers both to the U.K. MoD and abroad. It is important to note that these interviews concluded almost a year before the U.K.'s Integrated Review in March 2021. The publication and supporting documents outlined the U.K.'s perspective on future national security and international policy concerns, as summarised in section 5.2. The publications announced the intention to release a "Defence AI Strategy" and establish an AI Centre for Defence. At the time of the interviews, there was very little

[78] See https://www.gov.uk/government/publications/defence-artificial-intelligence-strategy for the full Strategy and https://www.gov.uk/government/publications/ambitious-safe-responsible-our-approach-to-the-delivery-of-ai-enabled-capability-in-defence for the corresponding policy paper, titled "Ambitious, safe, responsible: our approach to the delivery of AI-enabled capability in Defence".

public government documentation to draw on, except knowledge of the existence of the AI Lab at Dstl.[79] Therefore, the insights gained through these interviews included diverse views in which interviewees did not necessarily have a "reference point" of a defence-focused AI strategy to refer to and draw on.[80]

The chapter first offers an overview of the U.K.'s stated approach to AI in military contexts with a review of publicly available documentation relating to strategic and policy approaches, active initiatives, and the U.K.'s position within the broader international landscape. The chapter then analyses data collected through interviews and groups findings into four themes. These themes discuss: the relationship between the private and public sectors relating to military procurement and integration of AI-enabled systems (section 5.3.1); interviewees' perspectives on, and awareness of, risk management approaches to AI development (section 5.3.2); security challenges associated with military AI (section 5.3.3); and proposed mitigations to identified security challenges (section 5.3.4). Section 5.4 contains a brief discussion that places the findings in the context of available literature to analyse how AI is understood by defence staff, public-private relationships, risk management, and the debates surrounding the use of AI applications in conflict environments. This discussion proposes how these research findings address the research questions, exploring the U.K. perceived approach to AI development for military contexts. Furthermore, the discussion highlights how at the time of interviewing, interviewees reported various security gaps where the MoD lacked the requisite knowledge to demand the level of security assurance deemed desirable by industry experts. The discussion also reflects on the second research question to discuss how to meet such security gaps. Findings highlight a role for intergovernmental actors such as NATO to act as facilitators for norms-building activity and propose potential risk management mitigations. The chapter concludes by summarising the security risks identified by interviewees, particularly in the U.K.'s military innovation landscape and reflects on how the limited awareness of AI also represents a barrier to informed AI adoption in military contexts.

## 5.2. Context

Recent years have revealed an intensifying focus on AI for national and military security. The "AI Sector Deal" (HM Government, 2019) focuses largely on the industrial strategy of developing AI

[79] See this Chapter section 2. See Chapter 3, section 5 for a brief discussion on access relating to the AI Lab and Dstl.
[80] This is in contrast with my U.S. focused interviews which took place after the release of the US DoD AI strategy among other policy documentation: see Chapter 7 section 1.

talent and an AI-focused industry sector. The U.K.'s AI Council published an "AI Road Map" in January 2021, which aimed to help guide the U.K.'s strategic direction regarding AI. The U.K. announced its intention to publish a defence-specific AI strategy in March 2021 with the publication of the *Integrated Review* (Ministry of Defence, 2021, 42; HM Government, 2021, 63).

*Note M: The broad U.K. AI Landscape*

While this chapter focuses on U.K. activity relating to AI in military contexts, it is essential to be aware of the U.K. government's broader approach to AI through various national initiatives and agencies. The U.K. has taken a "whole-of-society" approach to AI within which defence is one part (Dwyer, 2022). As such, a range of organisations and government agencies are involved in the U.K. AI innovation landscape. In 2018, the U.K. committed $1bn to promote the U.K. as a leading AI innovator through the AI Sector Deal. The Department for Business & Industrial Strategy and Department for Digital, Culture, Media and Sport Several structures coordinate several organisations, including the Office for Artificial Intelligence (OfAI), a government department tasked with approaching AI in the U.K. economy. The government also collaborates with academia and industry through initiatives including the AI Council, an expert committee set up to provide recommendations to the government, and the Alan Turing Institute, a specialist U.K. Research and Innovation (UKRI) centre dedicated to AI research. The U.K.'s National Data Strategy (2020) and National AI Strategy (2021) do not focus on defence but outline commitments to seek competitive advantage in an evolving landscape. The development of broader government-created, or supported, organisational infrastructure highlights the U.K government's intention to drive AI adoption within a vibrant, competitive domestic environment. Similarly, the creation of new government bodies in the last half-decade, from the Office for AI to the Centre for Data Ethics and Innovation, demonstrates an increasing commitment to AI adoption and integration.[81]

The U.K. has a "distributed array" of actors responsible for coordinating AI activity (Dwyer, 2022). While a range of government departments and supporting organisations support the U.K.'s approach

[81] For a general analysis of the AI Strategy see Kazim, Emre, Denise Almeida, Nigel Kingsman, Charles Kerrigan, Adriano Koshiyama, Elizabeth Lomas, and Airlie Hilliard. "Innovation and opportunity: review of the UK's national AI strategy." *Discover Artificial Intelligence* 1, no. 1 (2021): 1-10.; For discussion on how non-defence initiatives contribute to the U.K. defence and security posture see: Dwyer, Andrew. "A Foundry of Artificial Intelligence? The case of UK national security." In *Routledge Companion to Artificial Intelligence and National Security Policy*. (Accepted/ In Print). Expected 2022.

to AI, as described in note M, bodies tasked specifically with military-focused technology are primarily a part of the MoD. Table 5.1 highlights three prominent MoD organisations with tasks related to AI in military contexts.

*Table 5.1: Relevant MoD agencies/ departments responsible for AI*

| **Body** | **Responsibility and Summary Activity** |
|---|---|
| Defence Science and Technology Laboratory (Dstl) | Dstl focuses on science and technology for the defence and security field. Dstl supports funding mechanisms including:<br>• The "Defence and Security Accelerator" (DASA) "finds and funds exploitable innovation for a safer future" (DASA, Online). Committed to allocating £4 million of investment into AI defence projects (Ministry of Defence, Defence and Security Accelerator, and Heappey, 2020), split across several early-stage enterprises over two years.<br>• The "AI Lab". Within the U.K. MoD, the focal point for AI research has been the AI Laboratory or AI Lab. Created in 2018 and based just outside Salisbury, the AI lab is intended to "enhance and accelerate the U.K.'s world-class capability in applying AI-related technologies to defence and security challenges" (Ministry of Defence, 2018, para 1).<br>• In April 2021, Dstl announced the creation of a 15-person Dstl unit within the National Innovation Centre for Data (NICD) in Newcastle, focusing specifically on data analytics and collaboration with the private sector (Dstl, 2021).<br>Dstl has also released and maintained "The Biscuit Book", a public document intended to outline the fundamental concepts of AI and data science for MoD customers (Dstl, 2020). |
| Strategic Command (UKStratCom) | Provides an integrative function, supporting the MoD to develop and manage joint capabilities across the land, sea, air, space and cyber domains. The Strategic Command Strategy (2021) sets out a broad and ambitious remit of key activities, including delivering and implementing the Defence Data Strategy, establishing the Defence AI Centre, and engaging with a range of digital transformation efforts relating to experimentation and acquisition of emerging technologies (Strategic Command, 2021, 12-13). |

| | |
|---|---|
| | • Contains "jHub", a defence innovation Centre that seeks out and funds pilots for market-ready technologies, often engaging with start-ups and small enterprises to determine if the technology can be adapted for military use. jHub has indirect connections to DASA, Dstl and industry actors (Dwyer, 2022). |
| Defence Centre for Artificial Intelligence ("The Foundry")[82] | The Centre's goal is to "accelerate adoption" of AI "across the full spectrum" of capabilities and activities (HM Government, 2021, 63). The U.K. announced the creation of the DCAI in November 2020 as part of a £400m defence spending boost. The Centre will act as the "nucleus to accelerate the development and exploitation of these critical technologies from the battlespace to the back office" (Ministry of Defence, 2021, 42), coordinating defence innovation. This scope includes coordinating the development of common AI platforms, toolkits, and best practices, implementing data management techniques, and testing and validating emerging capabilities (Ministry of Defence, 2021; HM Government, 2021, 102). It appears the U.K. has prioritised operationalising the technology as opposed to any public strategy development on responsible adoption or principles for responsible use of AI. At the time of writing, few further details have been released on the scope and structure of the Centre. |

Table 5.1 is non-exhaustive. The limited information on each agency that existed during the interviews makes it difficult to assess which organisations are most prominent in shaping and implementing the U.K.'s approach to AI. Dwyer (2022) highlights the "hazier" nature regarding which agencies are responsible for which aspects of AI in military contexts and suggests an "extensive range of bodies" that may be engaging in military AI occasionally in an overlapping manner. Taylor (2019, para 14) lists additional government agencies "wrestling with AI", including the AI Centre of Expertise under Defence Digital, the Army 6 Division, the RAF's Rapid Capability Office, and the Navy's Digital Services. While the exact structure and nature of the AI Defence Centre are unknown, the organisation's establishment marks a shift to consolidating defence-focused AI activity (Dwyer, 2022; Taylor, 2019).

The U.K. is increasingly investing heavily in military AI. Some of the most prominent spending commitments relevant to defence innovation were announced after the conclusion of the interviews.

[82] The intended creation of The Foundry was announced in November 2020 after interviews had concluded.

As one example, the U.K. government has committed to increasing defence spending above the rate of inflation and exceeding manifesto promises and has ringfenced funds for military R&D and AI (Prime Minister's Office, 2020). The U.K. has promised to invest at least £6.6bn in "research, development, and experimentation over the next four years [2021-2025] so the armed forces can adapt to the threat with advanced technologies" (MoD, 2021b, 3). These investments may seem negligible compared to the declarations of huge investment sums by U.S. counterparts; the FYI2021 annual DoD budget request includes $800m to AI, complementing the $927m dedicated to the DoD JAIC in 2020 (Konaev et al., 2020, 6; Office of the Under Secretary of Defense, 2020). Indeed, the U.K. does not have the resources for the financial investment or capacity required to match U.S. or Chinese innovation in AI (Dwyer, 2022 ). Nonetheless, the U.K.'s strengths across other domains have shown that financial investment in real terms is not the only factor determining the success or failure of capability building. For example, in a comparative research study focusing on cyber intent and capabilities indicators, the U.K. was assessed as the world's third most powerful cyber power in 2020 (Voo et al., 2020). Dwyer (2022) describes a "winner-takes-all" approach the U.K. has taken to AI as the U.K. attempts to take the lead in promoting AI across U.K. national security functions. The funding dedicated to military innovation is substantial, and it remains to be seen how funded initiatives such as the Defence AI Centre will be operationalised.

In terms of written doctrine, U.K. government agencies and military institutions have published some material offering an insight into the possible U.K. approach to adopting military AI capabilities. A summary of three relevant reports: the Joint Concept Note 1/18 18 Human-Machine Teaming (2019), the Defence Technology Framework (2019) and the "Pioneering a new national security: the ethics of Artificial Intelligence" (GCHQ, 2021), can be seen in Appendix F.2.

The first detailed discussion on the U.K.'s position in the international landscape is within the GCHQ report (2021), released after the conclusion of the U.K. interviews. This publication is one example in which nuanced discussions of military AI represent recent developments. Similarly, in the period between concluding the interviews and submitting this thesis, the U.K. government has released a set of publications reflecting a strategic review on security and foreign policy: the Integrated Review of Security, Defence, Development and Foreign Policy, titled "Global Britain in a Competitive Age" (MoD, 2021), was released in March 2021 and was followed by a Command Paper, "Defence in a Competitive Age" (MoD, 2021b), and the "Integrated Operating Concept 2025" (MoD, 2021c). Together the documents set out the ambition for the U.K. to be "ready for the threats of the future" (MoD, 2021b, 12).

The Integrated Review and supporting documentation give the richest insight into the U.K.'s approach to AI in military contexts within the U.K.'s attempt to "consolidate responsibilities and strategies" within military and defence contexts (Dwyer, 2022). The Integrated Review commits to publishing a Defence AI Strategy and AI research finding for military technologies (MoD, 2021). The Command Paper sets out the intended role of the AI Defence Centre and notes the U.K. commitment to engage "with liberal-democratic partners to shape international legal, ethical and regulatory norms and standards" (MoD, 2021b, 42).[83] While these publications have been critiqued in the literature as lacking detailed policy plans and being disintegrated in nature (Hew, 2021), the three documents mark a significant development in available material than was available at the time of interviews for this research.

[83] This intended engagement is consistent with existing U.K. government publications, although Note L highlights how aspects of U.K. terminology in relation to autonomous weaponry can hamper norms development on topics relating to AI.

*NOTE L: The U.K.'s definition of Autonomous Weapons*

A Select Committee Report to the U.K. government raised significant concerns about the U.K.'s "unusual" definition of autonomous weapons (Select Committee on Artificial Intelligence, 2018, 102). While the U.K. Joint Doctrine Publication on Unmanned Aircraft System states that the U.K. does "not possess fully autonomous weapons systems and has no intention of developing them" (Ministry of Defence, 2017, 14), the U.K. definition for autonomous weapons sets the threshold for autonomous so high that it is "effectively meaningless" (Select Committee on Artificial Intelligence, 2017, Q155).[84] The U.K. definition differs from that of many other states in requiring an autonomous system to "be aware and show intention", alluding to currently non-existent third wave AI capabilities described in Note A, and the MoD acknowledges that this definition differs significantly from many other states and allies (Ministry of Defence, 2017, 20). This "semantic haze" risks the U.K. moving towards an "ill-considered drift into increasingly autonomous weaponry" with autonomous systems that are unexamined as they do not meet the MoD definition for LAWS (Select Committee on Artificial Intelligence, 2018, 102-103). The Select Committee's report (2018) recommends that the U.K. definition of autonomous systems be realigned with those used by other states to allow the U.K. to "meaningfully participate in international debates" in LAWS discussions. This recommendation does not appear to have been upheld by the U.K. government.

While there is a growing area of literature focusing on AI's implications in national security or defence, there is limited academic scholarship focusing on the U.K. approach to military AI innovation. Beyond government documents such as the Integrated Review, little additional reflection was found in the academic literature on how the U.K. perceives itself in the competitive military AI landscape. Most of the literature detailing state approaches to military AI development is weighted to an analysis of U.S. innovation (Konaev et al., 2020; Hoadley and Lucas, 2018), which is expected given the relatively higher information released by the U.S. government, including the DoD AI strategy. While the U.K. AI strategy lacks any equivalent detail relating to military contexts, Taylor (2019) reflects on the U.K. defence procurement processes. Taylor (2019) highlights how AI procurement is likely to diverge from existing Ministry of Defence procurement aspects, considering requirements, market competition, testing and verification, and the required speed of decision-making. In 2020, a RUSI report put forward the perspective that the U.K. government was not harnessing AI advancements, which were taking place primarily in the private sector or academia,

[84] This argument was contained in oral evidence to the House of Lords as provided by Professor Noel Sharkey. See reference for full transcript.

and called for the U.K. to prioritise greater integration and public-private partnerships (Babuta, Oswald and Janjeva, 2020). This call aligns with broader reflections on the broad international innovation landscape, which recognise that effective military innovation will rely on greater civil-military fusion (Burton and Soare, 2019).

## 5.3. Interview Findings

This research is grounded in interviews with fourteen private sector defence professionals, as outlined in Chapter 3. The complete list of interview questions can be viewed in Appendix A. Analysis of the interview transcripts revealed four main themes around which discussion was centred and which are discussed below. The first theme is "Military AI and market dynamics", as outlined in section 5.3.1, referring to the relationship between the private sector and UK MoD relating to military innovation. Section 5.3.2 reflects on the second theme, "risk management", exploring the awareness of risks in AI adoption in military contexts. The third theme, "security challenges", maps out the key concerns highlighted by interviewees relating to the design and use of AI, particularly in military contexts, and is discussed in section 5.3.3. Finally, section 5.3.4 discusses the fourth and final theme, "perspectives on risks mitigation," which discusses interviewee perspectives on possible approaches to military AI innovation, including international norms development, industry regulation, export controls and regulation. The interviewees are denoted as "A1" to "A14" in the following sections.

### 5.3.1 Military AI and Market Dynamics

> A7: "So, on any given day of the week, we are a partner, we're a supplier. We're in litigation with each other disputing different contracts. We're their [the MoD's] best friend, their worst enemy, late payer, you know... we'll be suing them, they'll be suing us, they'll be giving us contracts, we'll be giving them products they love, we'll be giving them what they need, sovereignty - operational sovereignty, military advantage, information, it's a very complicated relationship. It's a very complicated relationship. Because they need us to give them certain things because they can't – they don't have in-house capability to manufacture. But at the same time, they are only a small part of our customer base…."

Interviewees spoke at length about how the market model has changed regarding military technology, especially at the cutting edge. On the one hand, military budgets have fallen in real terms compared with the Cold War era, reducing the power of governments as price-makers in the defence space rather than their former monopsony status. On the other, the rise of non-traditional technology for the defence space has changed the layout of the market. In A13's view, "[the] military is in no way leading the R&D front on defence relevant capabilities". Global technology conglomerates like Google, IBM, Microsoft, and Amazon (the four firms most frequently mentioned by interviewees, followed by firms like Facebook) now have the budget and data to invest in AI applications. These firms do not contain governments of defence departments as a core part of their business model; their products may be more lucratively marketed toward a commercial audience.

> A1: "Increasingly the military domain is a shrinking economic market, so for most companies, there's no reason [*pauses*] incentive to put [a product] there."

Instead, there may be several disincentives for these commercial entities to engage in military projects. Interviewees referred to Project Maven and the public controversy that resulted in Google refusing to renew its contract to work with the U.S. Department of Defense on an image recognition programme. The interviews reported on a range of procurement issues that contributed to a "chaotic" (A3) and "frustrating" (A11) innovation landscape. The MoD was described as "a huge bureaucratic labyrinth with lots of silos" (A2), and interviewees often described that governance and scrutiny slowed down innovation. To succeed in the U.K. defence environment, A3 described how organisations must abide by "a plethora of frameworks" that limited particular workstreams and meant the MoD was "not necessarily then getting access to the best providers for the best work". Similarly, A4 described how audit functions in the MoD tend to slow down procurement processes, "so by the time you procure anything with an information system in it, it tends to be obsolete". Three interviewees (A12, A3, A4) felt that the procurement process was improving, particularly through DASA and similar MoD initiatives designed to engage smaller innovators. However, overall, interviewees felt there was significant room for improvement.

Interviewees believed that the U.K. MoD found the price a significant factor when awarding a contract between tenders, encouraging suppliers to provide competitive quotes. In turn, suppliers are incentivised to focus on the functional requirements (i.e., key purpose and aspects of the technology

to be provided) rather than non-functional requirements such as security, which add time and cost to product delivery quotes.

> A2: "[There's] Been a huge amount of pressure from MoD over where the defence supply chain represents value for money, and whether they're being screwed over on contracts, and it's led to this kind of sort of mindset that says that the people supplying MOD shouldn't really be profitable. You know, they should have real limits on how much profit they can make, they should take all of the risk, and MoD should be left with none of the risk, and – but we still want them to innovate."

There was a perceived lack of financial support available to defence firms to dedicate to non-functional requirements within a product delivery. This dynamic was explained as an effect of "requirements-based" development practices, which did not allow for emerging technologies to be introduced once the project requirements had been finalised. Interviews mentioned internal attempts to employ more flexible methodologies such as Agile and DevOps development but acknowledged that these represented a small proportion of relevant development overall. Interviewees described the procurement and supplier market as largely requirements-based; the product is built according to predefined requirements, and interviewees felt that being price competitive meant focusing only on core functional requirements. This scoping was also compounded by the fact that when payment hinges on delivering functional requirements alone, there is no incentive to provide additional security-related services.

> A11: "You'll get paid if you meet the functional requirements. The non-functional requirements are…. 'Yeah, you should really do them', but we [defence firms] play games with that area and unfortunately, security is one of them [areas]."

Interviews further mentioned the restrictions of working with the MoD. Procurement contracts with the MoD typically mean the defence supplier does not control the intellectual property for the developed product, which then limits the commercial resale of the tools. This limitation prevents adversaries from accessing the technology should it be commercially available; however, this presents a challenge for innovating actors that may view defence provision as a limitation on total sales. This challenge may lead to an incentive to focus on civilian technology, with A7 stating, "if I can only sell ten items to the MoD but I could sell it a hundred times privately - why would I restrict my market?".

A5 also described the difficulties in operationalising projects with the MoD, describing that "it's relatively easy to get people together... to come up with an idea ... to do and test a use case. The hardest step is scaling up. Getting that buy-in from enough people to make it bigger and to make people trust it more and to push it through into something that's more than just the small idea". Similarly, A2 described how the U.K. military was adept at demonstrating capabilities in sandbox environments, "but when you try to turn it into something which is a mainstream procurement for a named programme… you start to slow down, and then people worry about the verification and certification of these systems". These demos were described as costly and time-consuming by A11 given the "hit and miss" nature of progression to mainstream development, though they acknowledged demos as a useful way to test equipment and software. Both A4 and A5 described that difficulties integrating AI technology might not be due only to market dynamics or barriers to wider operationalisation but to reluctance from military staff, particularly end-users. A4 described "a natural cautiousness of various people" where there is "doubt and push back over using something new… even if it helps them". A5 highlighted the "very strong bonds" that develop between soldiers on operations and felt this led to "conservativism in terms of process and terms of using kit, because you've used it in a very stressful situation, it got you through that, why would you not use the same thing again?". A4 pointed out that cautious leadership are "easily persuaded if you get the right people in the room", stressing the importance of qualified data scientists who could advise on the capabilities of potential applications.

Given the employers of most of the interviewees, it is perhaps expected that AI innovation was viewed as "very much driven by the larger primes" (A8). The consensus among interviewees was that the MoD often struggled to onboard small-medium enterprises, including specialised AI start-ups, as part of the procurement process. Interviews reported that defence primes believed they were well-placed to partner with start-ups to facilitate non-traditional engagement with MoD projects. When asked about the impact of major technology companies refusing to engage with military innovation, most interviewees also saw the role of defence primes as preventing a degrading of product standards. The justification for this view was that defence primes have already come to terms with the nature of warfare and their role in enhancing the U.K.'s warfighting capabilities in the name of national security. Other than defence primes, A10 described a broader "cowboy culture" in the AI landscape as "little startups, people who have very variable credibilities are able to get money and potentially build themselves in without as much scrutiny as they should have" and highlighted the need for standards to drive accountability. Not all interviewees felt defence primes could be assumed to be a safer choice, as A11 believed there was no evidence to say primes or larger corporates were better in

this regard.

### 5.3.2 Risk Management

> A13: “They [MoD] send policy people well, very careful about positioning the PR space, they don't send the engineers... but really we need the engineers to talk to engineers.”

> A13: “I think that companies are failing on their responsibility here.”

*5.a. Risk Awareness*

Risk awareness was a major theme highlighted by interviewees. The interviewees were asked a series of questions that aimed to identify aspects of residual risk, defined as the total risk minus the impact of security controls. AI was clearly understood by the interviewees as a technology that could represent a threat (it may conceivably be used to facilitate offensive activity) or a vulnerability (the AI system may be compromised and subverted).

There was a consensus view that the U.K. MoD was not demanding higher AI verification assurances since decision-makers were not familiar enough with AI technology to set out their requirements in these terms. While A6 felt the U.K. has “a strong appetite [and] a strong desire to use AI in defence”, they described this as an abstract belief and said they felt no feeling of “being forced to build capabilities for ‘defending the nation’ type thing[s]”. A12 felt there was “massive pressure” to innovate in military AI, referring to their engagement with the special forces to describe, “they’re [the special forces] understanding of what they want is something slightly different, they use a lot of buzz words... [I don’t know] whether they realise what what it... what they’re asking for, and what is required to get there”.

> A10: “There’s a lot of both false optimism like overly enthusiastic people, and then there’s also a lot of people who are frightened of it [AI] for different reasons.”

In twelve of the fourteen interviews, hype, bias, and under-informed staff were highlighted as a concern. A13 felt “it is easy for the army as a buyer to be bamboozled into buying snake oil”, while

A14 felt “constantly aware is that the first thing that the customer does is they go up the hype cycle quick quickly”. With “very good marketing around” the narrative of an arms race dynamic (A11), it can be “very easy for someone to get bought, get sold on an idea” (T5). Interviewees felt that companies could use marketing to overemphasize their products to the extent that “some people could sell stuff that actually isn’t really AI in any shape” (A6) and take advantage of enthusiastic but under-informed government procurement staff.

> A11: “I think there is a general lack of understanding within the MOD about what they are asking for and what some of the consequences of it [AI technologies] are. They see the upside, they don't necessarily focus on the limitations or… the delta between what they are asking for - what is possible today, and of the other stuff that they are going to have to do to bridge the gap.”

Interviewees frequently referred to a shortage of skilled staff to explain the lack of informed decision-making they perceived happening throughout the procurement process. One interviewee described their demand-side clients as “amateurs in the neutral sense” (A11): while the interviewee had spent decades of their career developing their knowledge of data science and machine learning techniques, their MoD equivalent did not have this background. In particular, the shortage of technical staff, who can challenge the design aspects of a product, was highlighted as skill sets in short supply. As A14 summarised, “the problem, of course, is that there aren’t enough data scientists around”. While “pockets of Dstl” were seen to have the relevant expertise, A11 believed that the MoD often didn’t have Dstl with them during procurement discussions.

Risk management through the design process was referenced or alluded to in every interview, despite risk never being explicitly raised in the wording of interview questions. The current innovation process was highlighted as incredibly difficult to manage when accurately monitoring risks specifically related to AI technology. A2 reflected on how researchers did not yet know how to certify AI systems. One defence prime described the challenge of explainable AI, which was currently in the early-concept testing stages. Interviewees highlighted that risk awareness didn’t necessarily require deep technical knowledge of the area, as A14 argued that “the truth is, it doesn’t take much knowledge to know where the risks are either, it just takes a bit”.

*5.b. Operational Risk Management - Designing Security into Product Development Processes*

> A2: "Accountability, transparency...all that kind of stuff... it adds some time.... but again, if you're not worried about that sort of thing, or if you're not as worried about the side effects if you get it wrong, you can lash that kind of thing pretty quickly."

> A11: "In true human style, I think we pick and choose what suits our narratives. If you've got some very keen to do AI, they're going to find all the reasons why AI is brilliant…. I have these arguments on a daily basis with one particularly clever AI developer. He lays out how what he's doing, how he's doing it, and I point out, 'you're asking for a friendly fire incident'."

Several interviewees responded that active projects were underway within the U.K. landscape to address the challenges of explainable AI and effective validation and verification of complex software systems. At the same time, no interviewees could point to specific testing and evaluation capabilities that analysed the AI aspect of the software product beyond the standard software testing process. For example, no interviewees reported common oversight mechanisms or re-testing schedules for AI-enabled products to check that they continued to operate as intended post-deployment. This gap ties in with market supply as discussed in section 5.3.1; there was no incentive to go above and beyond to perform a more expensive service when it was generally understood that the MoD awarded contracts on price competitiveness as a priority factor. The lack of accountability for private innovation was cited as a concern by one A10 who believed that "the greatest threat with non-state actors and corporations having so much power is that they do not inhabit an organisational space where they feel responsibility for their actions the way a nation usually does…there is no International Criminal Court for them".

There was a consensus that AI was too immature to be deployed anywhere but the most banal applications across the military currently – due to the number of unanswered questions and security concerns. However, several examples of active AI projects were raised, including Project NELSON, a major active project promoting enhanced data analytics and AI across the British Royal Navy (Gov.U.K., 2018; A12). A14 believed that within their organisation, themes such as responsibility

and trust were “pushed off in the direction of human factors” (A14), with non-technical tasks being othered and understudied, compared to the staff directly building the product to specification.

> A3: “It is absolutely imperative to have, and more urgent to have, diversity in leadership and in tech.”

> A14: “If there's one thing that we need the ability to do, it is the blended diversity of teams with diverse skill sets. We're able to do reviews of projects and go, ‘okay, we see the benefits and risks of this project being here’.”

Perhaps one unexpected finding was how often interviewees mentioned the importance of individual decision-makers (given that the questions were usually strategic with state actors as reference objects). A10 felt that military AI discussions should draw in those with expertise in political theory and international relations, just war theory, and business ethics, alongside engineers and data scientists. Over half the interviews noted that different perspectives may help predict and proactively mitigate associated AI deployment challenges (A10; A12; A14; A2; A3; A5; A6; A8).

### 5.3.3 Security Challenges

The technical challenges to securing AI systems were mentioned in interviews. Interviews described the importance of basic cyber hygiene to protect AI systems (A2). Cyber defence measures should defend against data bias (A3) or inaccuracy in the data (A7). Defences should also account for existing adversarial AI techniques which aim to subvert AI systems, for example, by tricking an image recognition model into misidentifying a genuine innocuous object as a weapon (Athalye et al. 2018; A8). These defence techniques may include the algorithm and corresponding training data across the deployment lifecycle (A4). A chief concern was designing and maintaining the security infrastructure of AI systems “carefully enough that it doesn’t get hijacked by a foreign state” (A1). Interviewees also raised concerns that not enough staff understood the technology sufficiently to predict how security systems could fail (A13, A14). However, the more intractable challenges centred around complex factors, including escalating pressure to innovate, with AI arms race dynamics, fragmented innovation practices, and the perceived challenges of AI operating in incredibly unpredictable environments in which the laws of engagement are frequently changing.

Several geopolitical factors were an overarching theme of the security challenges discussed. Almost every respondent felt that there was an AI arms race, and those who were less fixed on their answer instead stressed the fluidity of international competition between military capabilities.

> A12: "It's always so complicated, isn't it…. I think it's naive to pretend that [the race] hasn't begun already."

There were many different conceptions of what an "AI arms race" might mean. The phrase represented vastly different phenomena to interviewees, despite interviewees having broadly similar roles and remits. There was no overall consensus on whether an AI arms race was a conventional arms race without traditional arms, whether there was enough evidence to prove or disprove the dynamics' existence, or whether the AI arms race was occurring simultaneously in many different spheres at once (i.e., healthcare, military, in investment terms). A4 highlighted that the most intense competition, in their view, was between the U.S. and China in "a bit battleground for the intellectual high ground in AI" and commented that at a European AI conference a few years before, they had noticed "more than half the papers were from China". Both A2 and A14 highlighted that Germany and France were engaging with military AI projects. Some interviewees raised how AI would impact the nature of warfare. With software "low-cost, ubiquitous, [and] easy to deploy", A1 felt there would be a rise in asymmetric warfare as AI affords smaller state actors a significant boost to their capabilities. Both A10 and A1 highlighted the rise in "grey" war activity where AI activity might be used maliciously to project global disinformation or scale up cyber-attacks within wider power struggles. In being deployed to enable subthreshold warfare, A6 described how AI could be used as an asymmetric capability to subvert elections. In this way, AI can adapt to a changing military environment, as described by A6, that has shifted from "a sense of a war, to being at war, to more of a Russian approach that we're just continuously at war". To some interviewees, the nature of warfare remains fundamentally unchanged, representing "a search for superiority" (A9) in which "you will try to take advantage of every opportunity you can do to disorient or misdirect the enemy where possible" (A11). Through this interpretation, AI is "a means to an end" to enhance and streamline performance in warfare (A12).

> A9: "How might it work to push out technology where the operational constraints are different between countries? I know the U.S. and the U.K. have different rules

of engagement, and then of course when you consider non-ally states, I'm not sure what the right phrase is, but states with very different thresholds…."

A10: "There are no clear guidelines at the moment."

The lack of a best practice methodology was considered an entrenched characteristic of AI. AI was described as a product of a particular culture; the AI technology developed in China will have different assumptions, characteristics, and goals than related AI applications developed in Russia, the U.K., and elsewhere. As A1 described, "We're not going to have one kind of AI. AI is going to diversify, become heterogeneous, the way it learns will be different. 'Cause it will have been created by a different culture'." This aspect posed a security problem to interviewees as it made it difficult to anticipate and allow alignment between systems. The distinction between the technologies represents distinctions at the doctrinal level between states, and interviewees believed the U.K. conducts warfare differently from its adversaries. The uncertainty on how threat actors may use AI contributed to perceived pressure for U.K. innovation. As A8 commented, "often these things are kind of driven a bit by fear". Even domestically, fielded technologies will be used differently in ways that might be problematic in an operational context. A11 referenced this had happened on a U.K. military project relating to autonomous vehicles where users "have not interacted with it [the AI/ autonomous system] necessarily as planned".

A11: "The people that we are preparing ourselves to have future battles with, to be frank, do not have the same cultural ethical or moral constraints that we set ourselves or we've developed. And we need to find a way of dealing with that."

The rules of engagement differ between alliances, states, and individual conflict scenarios and are considered "massively contextual...Every operation has a different legal framework.... ROE is a minefield" (A7). NATO has published a set of rules of engagement, MC 362/1, that defines "the circumstances, conditions, degree, and manner in which the use of force, or actions which might be construed as provocative, may be applied" (NATO, 2003; Cooper, 2019, 26). However, it has been argued that the lack of clarity of certain concepts leads to "ambiguity [which] may be detrimental for people involved and for mission accomplishment" (Cooper, 2019, back cover). The interviewees did not know how AI systems could be universally bound by international norms or clear mandates within

such a fluid operational landscape. As A1 remarked on diverting state practice and accepted use of AI: “this is a social policy issue, not a technical policy issue”.

### 5.3.4 Perspectives on Risk Mitigation

Having identified the various layers of risk that emerge when examining the military AI innovation landscape, a review of mitigations proposed by interviewees is categorised into four broad themes: international norms, expert controls, regulation, and national regulation.

*International Norms:* International agreements, perhaps facilitated by intergovernmental organisations such as NATO, the Organisation for Economic Co-operation and Development (OECD), the United Nations (U.N.) or U.N.-adjacent bodies, were raised as institutions that hold the capabilities for the development of consensus and norms building among allies. From the intelligence-sharing trust networks, such as Five Eyes, to economic mechanisms, such as the G7, G20 or World Economic Forum, interviewees saw consensus building as an activity that could take place in a variety of fora. At the time of the interviews, there had not been many precedents for multilateral strategic statements relating to military AI as A2 highlighted there was “not really a clear policy of strategy for those sorts of international partnerships”. However, interviewees caveated their hopes for norms as a fail-safe solution. States may interpret and operationalise policy differently. Interviewees highlighted that states might ignore norms and even international law, and A11 spoke of “at least two state actors who couldn’t give a monkeys about the Geneva Convention”. This challenge puts the U.K. in “an interesting position” should they enter war with adversaries who are not limited by norms or ethical/ legal frameworks relating to military AI (T11).

*Industry Standards:* Offering resources to defence companies to help them assess the risk and highlight potential mitigations associated with different applications of AI technology. Guidance may include software or risk tools or attempts to reduce silos and ensure a range of perspectives (including legal and human factors) inputs into AI development processes. A6 suggested that governance frameworks could include audit requirements to ensure AI is developed and used within set guidelines. Such guidance would represent a gap identified by interviewees, as A14 highlighted that no one had designed a systematic risk assessment process for AI to their awareness.

*Export Controls:* Export controls for cryptographic tools and systems shows how export controls may be attempted over software. The exact enforcement mechanisms were not outlined. Respondent

articulated a perceived challenge to adequate export control of software and A1 comparing AI-enabled technology to liquid helium, stating, "It's hard to contain, it's not like an aircraft carrier where you can say "I have one, you don't have one". A13 suggested that export restrictions of the *hardware* required to support AI systems might be better controlled.

> A13: "When you think about regulating in this space, you thinking [sic] about regulating hardware, software, people and data. And you might be forced to regulate hardware and people because software and data are that much harder to control".

*Regulation:* The EU General Data Protection Regulation (2018) was described as an example of how long effective regulation can take to filter through to state legislation, an undesirable characteristic given the rapid pace of innovation across AI. Legislation was highlighted as being a too slow and bureaucratic option. A1 dismissed regulation as ineffective, remarking, "you're not going to be able to regulate people to not use it [AI] for malicious purposes, that's simple". Regulation of the AI industry was also deemed undesirable in an economic and trade sense; it is assumed that organisations will move to regions with less regulation, which would quell any hopes of developing a U.K. hub for AI activity across commercial spheres.

It should be noted that these interviewees did not offer non-caveated optimism about reducing the probabilities of the above security challenges. Using the analogy of Pandora's box, more than one person argued that they did not see a way to slow the spread of arms race dynamics. For cyber-AI capabilities, the most one could do was encourage strong cyber defence practices and hope offensive AI systems would not penetrate traditional defences. A4 acknowledged significant gaps in verifying and evaluating AI systems: "and so, 'if you put machine learning into a piece of kit, how do you test it?' is the exam question, and that still being, we don't know really." A4 speculated that ongoing research efforts to explore explainable AI might offer some solutions to test AI systems. Across other domains, the answer was equally unclear.

## 5.4. Discussion

The pressure to innovate in the military space kept many of the interviewees employed, a fact many of them highlighted themselves. Their reflections on the military AI landscape in the U.K. were

pragmatic; no interviewee made grandiose claims about U.K. superiority, and no interviewee said that the U.K. was world-leading. As staff who were often coordinating AI research (or on other emerging technologies) at large defence companies, these interviewees were most of the time very technical, and many of them were not used to making strategic speeches in a spokesperson-like manner.

Interviewees' comments on national innovation efforts reflect discussions taking place in the policy-focused literature and suggested an uncertainty on how the U.K is, or might, mitigate and negative security implications of such innovation. Long-term implications of AI arms-building may result in asymmetric capabilities between allies or within alliances such as NATO, raising issues of burden-sharing as the technological capability gap grows. Valášek (2019, 45) argues that "the impact on automated warfare on alliance politics, while poorly understood, need not be destabilizing" and highlights the possible roles for NATO to encourage unity between the alliance and shift innovation responses beyond national efforts. While NATO, within multiple interviews, was seen as a body that could assist in mitigating practitioner-perceived challenges, few interviewees had any suggestions as to how NATO may be able to achieve mitigation goals. Nonetheless, the absence of suggestions on international cooperation and burden sharing, as mentioned by Valášek (2019), suggests that states are taking an inward-looking approach to innovation. This suggestion is further supported by the fact that no interviewees stated that they had explicitly looked at the innovating strategies of other states when determining their AI strategy. Instead, interviewees specifically looked at active research themes in AI, such as decision assistance or image-processing. It was perceptions of potential adversarial usage of AI, the fact that another state could use machine-speed warfare to achieve a competitive advantage, that prompted discussions on how far autonomous AI-enabled systems should be integrated into military operations. Only one respondent claimed familiarity with the U.S. DoD AI Strategy.

The interviews highlighted several themes of residual risk that exist across the U.K. military innovation landscape, specifically around AI. The narrative that emerges from the research interviews shows a lack of high-level oversight into secure AI development, particularly regarding the range of security challenges highlighted and the lack of an agreed industry or MoD top-down best-practice verification tool for military AI systems. The immaturity of complex AI systems (as described by interviewees) only highlights the opportunity to implement an oversight and monitoring system in time for periods of greater and more frequent deployments. However, the market is currently not structured to incentivise this investment in security; often, the MoD does not have the technical

skillset to demand a higher threshold of security-centred AI technology. The interviewees mirrored concerns highlighted in a review of ethical challenges associated with military AI by Morgan et al. (2019), reflecting the intense pressure to innovate rapidly without requirements to ensure any innovation takes place in a safe, reliable, or ethical manner way.

The landscape of defence innovation has also changed dramatically over the last century. This aspect was reflected on frequently throughout the interviews. Many interviewees highlighted the importance of a proactive approach. The rise of "disruptive innovation" from the commercial sector marks a change from the early 19th century (Bellais and Fiott, 2017; Cummings, 2017). Interviews described a military innovation landscape in which the U.K. government had relatively less influence as a market-maker in the modern era, which correlates with the literature highlighting the rise of commercially driven innovation (Whittaker, 2021; Christie, Buts and Du Bois, 2021). Comments from interviewees on how private-sector innovators may be tempted by the relatively less bureaucratic yet larger commercial markets echo FitzGerald and Parziale's (2017) line of reasoning: private sector commercial developers may as well focus on innovation that can be sold to hundreds of millions of commercial customers, rather the thousands within military or government users. Smaller budgets and the politicisation of defence mean that the government does not have the pick of the technology they're looking for, a challenge reflected in the literature (FitzGerald and Parziale, 2017).

A review of recent academic literature highlighted that while many publications framed AI innovation as an arms race to technical superiority, the proportion of articles framing AI as competitive is declining (Imbrie, Kania, and Laskai, 2020). This decline does not seem to be replicated across this research's small but expert group of U.K. interviewees. All interviewees felt arms race dynamics were present across the military AI landscape. Whether this race formed part of a wider economic competition between states or whether interviewees perceived elements of a "race to the bottom" dynamic, specifically where militaries felt the need to remain competitive against adversaries in conflict, external factors contributed to the need to develop AI-enabled technology. The shift through which innovation was primarily developed by the private sector, often with private-sector data, raises significant challenges around accountability and responsibility. Respondent perspectives on the AI arms race, alongside their noted views on under-informed decision-makers, align with research suggesting avoiding discussions of "AI arms race" dynamics and instead focusing on more precise language addressing the nuances of AI capabilities and competitive dynamics (Roff, 2019).

Reducing responses on the overall landscape to their simplest parts: governments are less advantaged than they may once have been. This dynamic is coupled with 21st-century hype around AI and the geopolitical pressure to maintain a comparative technical advantage, which ultimately leads to a leadership's push for greater development of AI capabilities at a potentially undetermined cost to several aspects of existing security. This search for technical superiority is not new (though we can view it explicitly through the U.S. Third Offset Strategy) (Fiott 2017), but the speed at which technology evolves certainly appears so. Bellais and Fiott (2017) comment how before World War One, a soldier could have used the same kit over decades and throughout their career; it is hard to imagine today's recruits using the same technology in thirty years. In line with the rapid rise in influence by private sector suppliers, the reduction in relative market-maker power by the militaries represents a significant shift in power dynamics compared with military innovation dynamics a century ago. Where not applied globally, regulations may encourage private actors to move operations elsewhere (McGuire, 1983). Given the relative economic clout of many larger technology companies, it may well be in a state's interest to avoid excessive regulation of companies and keep the industry's interest in the local economy. Judging from the interviewees' uncertainty and amidst wider academic and public discussions around military innovation, it appears that while the acknowledgement of security challenges around military AI is often discussed, very little is agreed upon in terms of mitigatory options. While a detailed exploration of different definitions for AI and autonomy is beyond the scope of this thesis, the U. K's criticised terminology on autonomous weapons systems risks confusing the U.K.'s attempts to engage with the international community on these themes (see Note L; Select Committee on AI, 2018). If the U.K. continues to use the terminology at odds with other states, there is the risk of actors talking past each other on AI themes, a challenge raised in the literature by Stowsky (2004).

This chapter's focus on the U.K. landscape highlighted numerous concerns raised by expert interviewees. It also highlighted areas in which the interviewees did not consider themselves sufficiently informed to advise on the best mitigation approaches or near-medium term strategies towards security management of military AI. When asked about their views on potential solutions, a range of answers was recorded from explicit pessimism to general uncertainty and speculation. This reluctance to suggest or reflect on potential solutions highlights how intractable and complex security management is in this space. The uncertainty displayed by interviewees may also reflect how underdeveloped the defence discourse space was, at least at the time of the interviews. Their responses revealed perceptions of a highly fragmented innovation landscape that lacked a clear exiting national strategy to mitigate security challenges associated with AI. In the absence of any best practice – or

even prototype best practice approaches - I considered it wise to examine other actors' activity and potential contributions in this space, exploring how the actions of international actors may contribute to a top-down approach across the U.K. military AI landscape. This insight added to the justification for this theses' focus on the role of NATO.

## 5.5. Conclusion

The sentiments revealed through these U.K. interviews reveal several security gaps and opportunities relating to military AI innovation. In drawing together various residual risks perceived by expert practitioners, we have moved from a series of unknown unknowns in which security and AI are not considered interlinked at software, economic market, and strategic levels. As we highlighted the security implications of current innovative practices, we also learned about possible speculative methods to mitigate existing risks. Regulation is not wholly dismissed as an option. Rather, the role of export controls or the development of international norms are seen as potential factors that may nudge state behaviour in this space. This argument reflects, though not in detail, the discussion points raised in the literature referring to AI arms control, which argues that while very difficult in practice due to the challenges in validating evidence of AI innovation, specific controls targeting components may have potential (Geist, 2016). Experts in the space are candid that the current innovation landscape raises more questions than answers when managing the security implications of military AI-enabled technology, highlighting the need for a deeper exploration of potential risk mitigators for identified challenges. The themes discussed within this chapter are expanded on in Chapter 8, which draws together findings from subsequent interviews with NATO and U.S.-focused staff and additional data collected for this research.

Considering the relative lack of available analysis focusing specifically on the U.K.'s evolving approach to military AI capability-building, this chapter explored the underlying drivers for developing AI technology within the U.K. commercial defence sector. The chapter is guided by the first research question, "What are the implications of AI innovation for the military context". By interviewing decision-makers, including senior staff coordinating AI research programmes in the commercial defence sector, this chapter benefited from the access and reflections of those most informed on the U.K approach to military AI innovation. By exploring interviewees' perceptions of possible mitigations to the challenges associated with current AI capabilities, these interviews captured a range of security consequences. Exploring how commercial sector staff viewed the relationships between the U.K. MoD and private sector innovators enabled me to understand the

processes and tensions regarding the potential military procurement of AI-enabled systems. The findings from these discussions on possible approaches to risk management and security helped inform responses to this thesis' second research question[85] by focusing on the U.K. approach to understanding and hopefully addressing potentially harmful implications of military AI applications. More specifically, with the interview findings highlighting significant uncertainty around how the U.K. was working to mitigate known challenges, the interviews highlighted the U.K's immature approach to military AI at the time of interviews. This research also highlights how quickly things can change; since early 2020, the U.K. has created a dedicated MoD Centre dedicated to AI, announced increased funding efforts and released the Defence AI strategy, developments showing an increasingly more active approach to investing in AI and mitigating technical and organisational challenges to reliable deployment. The U.K. has also pledged its involvement with several NATO initiatives relating to military AI adoption and responsible use, demonstrating a willingness to engage with international efforts that fulfil some of the speculative predictions from interviewees.

[85] "How are actors attempting to mitigate challenges identified in relation to the development and use of AI in military contexts".

# Chapter Six: NATO - collective defence and AI

## 6.1. Introduction

As states begin to develop their strategic approach to military AI technologies, a large part of what will shape doctrine and the chosen path forward will be the international landscape (Scharre & Horowitz, 2018). When determining national R&D priorities, states will consider how adversaries might use AI, emerging international norms, and the growing awareness that cutting-edge AI technologies can be useful in conflict (Morgan et al., 2020). Several formal alliances and bilateral partnerships have emerged when it comes to the use of AI in a military context.[86] One potential coordinator of military AI practices, both in capability building and doctrinal sense, is NATO.[87]

As an intergovernmental military alliance consisting of 30 member nations[88] initially formed to counter the Cold War threat of nuclear conflict, it might seem counterintuitive to expect NATO to be at the forefront of responding to emerging technologies. However, as Chapter 5 examined, states do not necessarily have a clear idea regarding developing AI military capabilities in AI, much less designing and deploying such technologies in a responsible and security-conscious manner. Exploring the role of NATO in the international military AI innovation landscape, the chapter is grounded in the following research questions:

(1) What are the implications of AI innovation in military contexts?

This chapter examines relevant NATO documents to draw together NATO's documented perspectives on AI and includes interviewees' perspectives on AI's strategic and operational implications within military contexts.

(2) How are actors attempting to mitigate challenges identified in relation to the development and use of AI in military contexts?

[86] One example is the 16-member strong Partnership for Defense, created by the U.S. and discussed in Chapter 8.
[87] In addition to this Chapter, see further information on NATO's posture via Gray and Ertan, 2021.
[88] As of 15 April 2022.

This chapter looks at NATO's current and possible future role in mitigating the potential insecurities created or exacerbated by AI and introduces NATO's strategic approach to AI technology. Within this chapter, reference to "NATO" refers to the Organization itself - the Enterprise structure including staff and infrastructure spanning NATO and NATO-accredited agencies. The distinction is made between the organisation and the "NATO Alliance", the grouping of member states aligned with the vision of collective defence. This chapter first introduces how NATO has approached AI based on an analysis of open-source NATO documents and current programmes before presenting the analysis of data collected through interviews. Examining how NATO engages in military innovation generally, before focusing on how NATO conceptualises and is attempting to facilitate military AI innovation within the current international landscape, findings will explore the various opportunities and challenges associated with NATO's involvement in AI. A brief discussion section proposes how these research findings address the research questions.

## 6.2. NATO Context

Created in 1949, NATO is currently the world's largest military alliance and is designed to promote the collective defence of its members, with a strict focus on matters of military security. NATO's scope is specifically military, with the organisation's goals focused on collective defence, crisis management and cooperative security. The Organisation's infrastructure provides a platform for states to discuss ideas, coordinate multilateral activity in definitions, policy, norms, standards and doctrinal development, and cooperate on capability projects (either as a common funded project or between groups of interested members, on an opt-in basis).[89] It is not within NATO's remit to protect civilian infrastructure, for example, from a cyberattack. However, there may be a common interest where massive attacks impact military operations (e.g., telecommunications disruption. Instead, NATO plays the role of facilitator in consensus-building efforts for defence and military matters. Consensus building is a 'cornerstone' of NATO (NATO, 2020b), with strategies requiring agreement from all Member states.

NATO's modern-day challenges are numerous. One theme recognised by the Organization has been the need to respond and adapt to the threat landscape, with NATO's overarching "NATO 2030 Transatlantic Agenda for the Future" reflecting NATO's desire to increase resilience considering

[89] Where it's deemed investment is to the benefit of all 30 NATO members, financial resources are allocated through a cost-share arrangement by each member through the principle of 'common funding'. For more information https://www.nato.int/cps/en/natohq/topics_67655.htm.

technological developments (Secretary-General Jens Stoltenberg, 2021).

### 6.2.1 NATO's stated approach

Unclassified documents have been written for selected distribution across NATO, and for NATO's allies, including white papers on autonomous technologies and dynamic adoption, both authored out of the Emerging Security Challenges Division at NATO HQ (Gilli, 2020, 13).[90] NATO agreed on an AI strategy in October 2021 (NATO, 2021). Similarly, Defence Ministers across the Alliance have agreed to several roadmaps and strategies, the contents of which are not public outside selected NATO readership. For details on known NATO activity relating to AI policy and strategy development and NATO projects mentioning integration of AI technology, see Appendix F Tables F.2 and F.3.[91] Below the threshold of agreement on official NATO-wide policy, NATO has been actively coordinating numerous events discussing AI (and releasing subsequent statements) on the importance of NATO's attention to emerging technologies. (Leopold, 2020; NATO ACT, 2019a; NATO, 2020e). In the past year, NATO has begun hosting more discussions on the role of collaboration in military AI and AI interoperability between member states.

### 6.2.2 Relevant activity across NATO Agencies

Beyond the development of the AI strategy, discussions relating to AI-enabled technology are already taking place across NATO. Research is underway in various agencies to identify the opportunities and challenges of integrating AI in models and simulation, enterprise systems, decision support, cyber defence and information processing. At the same time, the NATO Science and Technology Organization (STO) states more broadly that AI can boost efforts to "expose new discoveries, identify promising research areas and provide S&T tools to support further research, highlighting both AI and autonomy as two disruptive technologies" (STO, 2020, 15). This broad range is necessary from NATO's perspective. Any decisive action or approach to the broad range of emerging technologies, including AI, must consider the technology's technical and strategic implications and take in insights from specialist agencies (e.g., those looking at terminology or applied areas, including cybersecurity and data management). See Appendix F.2 Table F.3.3 for a non-exhaustive overview of key structures across NATO focusing on AI.

[90] These white papers were not publicly released and have therefore not been viewed or examined as part of this thesis.
[91] Non-exhaustive lists, drawing on publicly accessible information.

NATO's efforts on military AI draw on agencies and centres across the Alliance, drawing on the distinct scope and specialism of different bodies. With Table F.3.3 highlighting attention and activity from multiple bodies on similar themes, the NATO Enterprise's approach to AI appears relatively decentralised. Arguments have been made for a dedicated "Artificial Intelligence, Integration and Implementation-Enabling Centre (A3IC)", which could lead to AI adoption for the Alliance and connect the relevant institutions to support national and NATO efforts (Gilli, 2019, 35). Similar calls have been made for NATO to lead as an "AI champion" to deliver relevant training and help allies understand and adopt AI (Lucarelli, Marrone, & Moro, 2021, 9).

### 6.2.3 Key documents

Several press releases and statements have been made on behalf of NATO, with Secretary-General Jens Stoltenberg and Deputy Secretary-General Mircea Geoană frequently highlighting the importance of NATO adapting to meet modern security challenges like AI.[92] Looking at how NATO has articulated its understanding and proposed approach to AI, five publicly accessible documents are particularly useful resources. These documents are outlined in Table 6.1:

*Table 6.1: Relevant NATO documents*

| **Document** | **Key Points** |
|---|---|
| NATO Allied Command Transformation. Strategic Foresight Analysis (SFA) Report. 2017. | A high-level document addressing categories of security risk that will impact the Alliance up to and beyond 2035, at the political, human, technology, economics/ resources and environmental level. This document represents an early example of NATO public documentation reflecting on non-technical implications of AI.<br><br>*Contents*:<br>- Highlights AI as an emerging technology that "will expose divergent ethical and legal interpretations" (NATO, 2017, 8).<br>- Commercial innovation has outpaced defence R&D. The decline of defence R&D expertise coupled with an over-reliance on commercial solutions represents a risk to Allied defence efforts.<br>- Operational effectiveness now reliant on advanced technologies. |

[92] As one example: Speech by NATO Deputy Secretary General Mircea Geoană on NATO and innovation. 20 September 2020. Accessed 14 September, 2021. https://www.nato.int/cps/th/natohq/opinions_178354.htm. Similar styled speeches can be retrieved from opening statements at events including annual NATO Summits and the Cyber Defence Pledge conferences.

| | |
|---|---|
| NATO's Science and Technology Sub-Committee on Technology Trends and Security (STCTTS): "Artificial Intelligence: Implications for NATO's Armed Forces". October 2019. | Focuses on the opportunities, challenges, and uncertainties of AI in the armed forces. Compared with the Strategic Foresight analysis document, it focuses more on technical and non-technical challenges, including adoption challenges relating to investment and workforce constraints.<br><br>*Contents*:<br>Describes two main applied areas for opportunity, though notes several other areas that will be affected by AI:<br>- (1) Information and decision support - faster detection, analysis and reaction times. Greater insights into data, highlighting abnormalities (e.g., to more effectively discover cyber intrusions).<br>- (2) robotic autonomous systems (RAS) - presently incorporated for deployment in extreme environments (e.g. rescue missions, counter-mine operations). AI has become a "backbone technology" and may contribute to "reduce a unit's personnel number substantially" (STCTTS, 4).<br>- Militaries need to invest sufficient capital into research and development, while armed forces "must become better at adopting and integrating technologies from the non-defence commercial sector" and find ways to recruit the top AI experts, many of whom are offered much higher salaries in the private sector (STCTTS, 6). |
| NATO's Science and Technology Organization (STO): "Science & Technology Trends 2020-2040: Exploring the S&T Edge". 2020. | Offers a categorisation of critical and emerging technologies that persisted in the Emerging and Disruptive Technologies Coherent Implementation Strategy (2021), setting out NATO's approach to EDTs. Focuses on the implications of AI technology for NATO, though does not go as far as providing detailed recommendations or paths forward.<br><br>Contents:<br>- Identifies AI and autonomy as two of seven critical emerging and disruptive technologies. AI is likely to have a "revolutionary" impact on NATO operations through integration, including "combat models & simulation, enterprise systems, decision support systems, cyber defence systems...virtual/augmented reality, quantum computing, autonomy...space, materials research, manufacturing & logistics," big data analytics, and autonomous vehicles (STO, 2020, 14-15).<br>- *Predicts:* "AI will also have a significant effect on the conduct of NATO S&T efforts as meta-analyses of existing research will expose new discoveries, identify promising research areas and provide improved S&T tools to support further research."<br>- *Specifies areas where AI will impact NATO forces:* Command, Control, Communications, Computers, Intelligence, Surveillance, and Reconnaissance; Weapons and Effects; Autonomous Vehicles; Capability Planning; Chemical, biological, radiological, and nuclear defence (CBRM); Medical; Enterprise Management; Logistics; Cyber and Information Space, and Training.<br>- *Specifies areas in which adversaries may use AI systems to undermine NATO:* in the cyber domain; information warfare (deepfakes and misinformation); exploiting the brittleness or vulnerability of NATO's |

| | |
|---|---|
| | AI systems; enabling new generations of Improvised Explosive Devices (IEDs). |
| NATO Advisory Group on Emerging and Disruptive Technologies, Annual Report 2020. March 2021. | Takes a holistic approach to innovation, including considerations of the socio-technical context and talent shortages relating to EDTs. Does not propose specific initiatives or applications of AI by NATO but instead focuses on the need to engage with emerging technologies. Nonetheless, the document is useful as an indicator of NATO's approach to EDTs, with the Advisory Group report highly likely to have some influence on the agreed EDT Coherent Implementation Strategy.<br><br>Contents:<br>- Calls for NATO to act as a "convening voice" to drive development and investment in EDT adoption and for NATO to take an active role as an 'influencer and interviewee in Allied innovation ecosystems" (NATO Advisory Group on Emerging and Disruptive Technologies, 6).<br>- Recommendations include NATO involvement in facilitating the testing and integration of new military AI-enabled applications. |
| NATO Reflection Group. NATO 2030: United for a New Era. 2021. | Provides recommendations for NATO relating to the required adaptation to EDTs.<br><br>*Contents:*<br>Recommends NATO:<br>- Act "as a crucial coordinating institution for information sharing and collaboration between allies on all aspects of EDTs that have a bearing on their security." (NATO Reflection Group, 2021, 13).<br>- Hold a digital summit of governments and the private sector to identify gaps in collective cooperation in security-related AI strategies, norms, and research and development (R&D), safeguarding against the malign and aggressive use of AI. *(*NATO Reflection Group, 13*).*<br>- "Consider developing a North Atlantic equivalent of the U.S. Defence Advanced Research Projects Agency (DARPA) or European Defence Fund (EDF) charged with encouraging support for innovation in strategic areas among Allies" within a wider 'AI-focused agenda for R&D within the Alliance' (NATO Reflection Group, 2021, 31). |

The documents show a developing understanding across the Organisation from 2017. Agencies through this period released reports that demonstrate, in increasing detail, an understanding of AI's potential impact on NATO and in military contexts. From 2020, reports include recommendations and calls to take action to create an innovative environment that advantages NATO allies. While earlier reports highlight the theoretical opportunities and risks of emerging technologies on society and warfare as a phenomenon, over time, reports shifted into naming categories of applications impacted by AI capabilities. From 2020, the EDT roadmap and announcement of the AI Strategy mark another shift beyond reporting that aims to understand the implications of the technology to

activity that includes insights on NATO responses and NATO's work to operationalise AI capabilities, to begin *acting* on the information in earlier reports. The increased detail in NATO and NATO-affiliated documents on AI represents one potential proxy for interest in AI and a response to increasing demand for a focus on military AI across NATO. As another example, the NATO CCDCOE's 2019 International Conference on Cyber Conflict did not mention AI in its call for papers and had one article on AI in conflict (Burton & Soare, 2019). The respective 2021 call for papers welcomed submissions on topics including autonomous weapons systems, automated operations, artificial intelligence in military operations, strategic approaches to EDTs, use of AI in state-led cyber operations, AI and cognitive cyber security, AI training, and NATO's cyber defence concerning AI technology (CCDCOE, 2021). The swift and intense increase in focus by NATO and NATO-affiliated agencies mirrors the marked increase internationally in military AI, highlighting how actors attempt to understand and adopt AI technologies to their advantage at a pace that matches technological advancements.

### 6.2.4 Potential paths forward

NATO allies face two interrelated pillars of work. The first relates to the dynamic adoption of emerging technology, while the second relates to the responsible governing of the technology (Christie, 2020).

*NATO and AI Adoption*

Christie (2021) identifies four areas to consider regarding the cohesive, practical adoption of AI: iterative development, in which defence innovation practices must modernise across NATO's enterprise and Allied defence establishments; human talent and skills, where effective decision-making and AI development relies on technical expertise and well-informed leadership; access to data as a strategic resource, and engagement with non-defence technology actors. Meeting these goals requires a significant structural change in how NATO approaches innovation, with criticism in the literature that NATO's internal adaptation is "too slow and inflexible. NATO needs a new framework in which ambitious innovation drives greater adaptability, efficiency, and solidarity" (Soare, 2021).

To effectively train an algorithm, one needs data that is "high quality, easily accessible, and readily available" (STO, 2020, 61). This requirement poses difficulties in the NATO Alliance context, where real-world military data tends to be highly classified and often limited in size (Chahal, Fedasiuk, and

Flynn, 2020). The collection of appropriate data is crucial in developing military-specific AI applications. Hill (2020, 151) highlights the "considerable work to be done" to develop appropriate data-sharing arrangements between NATO Members. For NATO as an Organisation, work under this pillar would likely include coordinating a data policy that covers the entire lifecycle from collection and storage to quality control of the data (Christie, 2020).

While NATO cannot prescribe binding actions, NATO can facilitate allied approaches to AI challenges in what Gilli (2019; 2020) terms "NATO-mation". By introducing non-binding standards relating to AI applications and the associated data requirements, NATO could facilitate greater interoperability as members innovate to enable national tools and capabilities to interact with one another. More generally, NATO's work around adoption must also address the rapid pace of innovation observed in the commercial sector and adapt their development and innovation styles accordingly (Gilli, 2020). Shifting away from drawn-out development structures and adopting suitable development methodologies that contribute to the breakthroughs experienced by large technology corporations, increased inclusion of agile methodologies, for example, may benefit the military R&D environment (Chrisie, 2020; Fischer, 2020). Similarly, as an organisation, NATO can continue to develop cooperation with the private sector, which is spearheading the majority of modern breakthroughs relating to AI technologies (Christie, 2021). In theory, in coordinating this activity, NATO can support member states who may not yet have the resources to engage with the equivalent scope at a national level.

Finally, talent and infrastructure are crucial factors in the future potential for Allied capability-building in relation to AI (STCTTS, 2019). The NATO 2030 Reflection Group report argues that NATO must "move with alacrity to improve the technological, and specifically AI, proficiency of its leadership and technical workforce" (2020, 30). In terms of infrastructure, NATO must also ensure members invest in supporting technologies to enable AI innovation, including hardware relating to secure computation and data storage assets (Gilli, 2020).

*NATO and Responsible AI*

There are many considerations that NATO must face and engage with when it comes to principles and processes relating to military AI capabilities. With most NATO members and partners still at a relatively early stage in approaching military AI technologies (Gilli, 2020), NATO has a significant role to play in facilitating consensus on norms, principles and a united doctrinal approach to state use of AI in these contexts (van der Merwe, 2021). Speaking in the run-up to the Brussels Summit in

2021, the Assistant Secretary-General for Emerging Security Challenges stated that "we [NATO] need principles of responsible use" (Sprenger, 2021). With effective coordination, NATO's progress on the two pillars may complement one another. For example, as members agree on national principles relating to military AI technologies, these principles should guide design requirements and shape planned capability-building exercises (Christie, 2020). Not all AI-enabled systems will need the same level of scrutiny and testing, and NATO can streamline its approach to the pillars or work by considering the three categories of AI environments: enterprise AI, operational AI, and mission-support AI (Tarraf et al., 2019). Considering each category's needs and associated risks allows NATO to progress across the dynamic adoption and responsible governance pillars (Christie, 2020). For example, mission-support AI systems are likely to need enhanced technical and principle-based scrutiny, from the assurance that the system will not malfunction and potentially risk safety to the additional relating to more controversial applications (including ethical considerations) (Tarraf et al., 2020; Gilli, 2020). The deployment of enterprise AI solutions, including predictive maintenance for equipment repair or training solutions delivery, represents "low-hanging fruit" from a technical standpoint (Tarraf et al., 2020, 33). This calculation is particularly the case where requirements are similar to those in non-military environments, making it possible to repurpose and adapt existing commercial solutions to military contexts (Tarraf et al., 2020, 33).

The above overview of NATO initiatives and programmes demonstrates the sheer diversity of relevant activities undertaken by NATO in recent years. It provides one indicator of NATO's journey to date in understanding and exploring the implications of AI technologies. What is less clear is how successful each initiative has been. With strategy documents not available for public readership, researchers outside NATO are unable to see how far each strategy goes in setting out a clear and ambitious path in a way that genuinely contributes to security. Even if Members agree on that language, it is difficult to assess effectiveness from any angle. It is challenging, if not impossible, to measure the impact NATO has had in shaping national approaches. NATO has released many doctrines and documents throughout its lifetime, which did not ultimately achieve the effect promised. One can look at NATO's extensive strategic and operational activity related to the war in Afghanistan as an unfortunate example of "strategic errors" (Shea, 2021), ultimately resulting in an Allied withdrawal and swift Taliban takeover of the country.

Similarly, failures in operational and tactical pilots relating to AI are unlikely to be published publicly. With many initiatives likely happening at a classified level, it is often impossible to identify or assess each project's success as an open-source researcher. In time, research may be able to draw out themes

on which initiatives materialised into international consensus and how the Alliance adapted to face the security context in which EDTs are increasingly prominent. It will still be hard to evaluate how far NATO was the primary source of change in the context of a complex and swiftly changing international security landscape. These challenges in evaluating NATO's position in the international landscape were reflected on by interviewees who highlighted a considerable range of initiatives, goals and open challenges while remaining cautious of promising NATO success.

## 6.3. Interview Findings

This section is grounded in interviews with seventeen staff across the NATO Organisation and the NATO-affiliated CCDCOE.[93] As noted in Chapter 2, interviewees spoke at a strategic, non-classified level and had insights and employment experience relating to the Organisation. Interviewees have been labelled as "B1" to "B17" in this Chapter. Analysis of the transcripts revealed five main themes in terms of considerations for AI technology and NATO, which are discussed below. These top-level themes are: the current military AI innovation landscape, military AI innovation at NATO; implications of NATO-coordinated activity; conceptualising military AI at NATO (awareness and education), and the international geopolitical landscape.[94]

It is important to note that all interviews occurred before any public announcement of the NATO AI strategy. The majority took place before the internal adoption of NATO's EDT roadmap strategy. The findings, therefore, capture NATO employees' perspectives at a particular moment in time, in the absence of internal or publicly accessible formal Strategy relating to military AI technologies. The subsequent developments and consensus-based strategies announced by NATO will be highlighted in section 6.4 within the discussion section.

### 6.3.1. Military AI Innovation: an immature but rapidly evolving landscape

Interviewees described AI as both an immature and rapidly developing set of capabilities. Interviewees generally agreed that data-based AI was not being used heavily in operational battlefield

[93] For the complete list of questions, please see Appendix A

[94] Each theme contains codes which are referenced across files. For example, the theme 'NATO military innovation as an organisation' had 403 coded items under the theme with codes from all 17 interviews, while 'conceptualising military AI - awareness and training' has 191 coded items with codes from 16 of the 17 interview transcripts. See Chapter 2 for a greater discussion on coding and data analysis.

environments at the time of interviews. The incorporation of machine-learning algorithms was still noted as limited or in very early stages in warfare environments. However, AI could be observed in training exercises or specific cyber defence tasks. Interviewees generally rejected the idea of developed AI deployment beyond specific cyber security applications; as B10 explained: "I haven't seen weaponised AI and I mean, you can see it more stuff like, breaching networks, yeah, I personally haven't seen it weaponised…" Multiple interviewees noted that AI was not expected to replace human decision-making in the next decade.

> B3: "You know working with big masses of data and finding patterns and recognising patterns and things like that as we do it today is of course very useful, but, this is nothing compared to human decision making, which after all is the goal set for AI if you think about it. AI is useful today in its current shape, but in the future we need a far more advanced artificial intelligence."

Generally, interviewees reported that when it came to embracing capabilities based on AI, technology was perceived as being "very embryonic and immature at the moment…" (B1). As B1 noted, "I've seen roadmaps strategies for adopting AI but I think it's at the moment everything is a little bit half-cooked" (B7). While a minority of interviewees were close to strategic policy development and could illustrate an overview of coordinated NATO activity relating to military AI topics, most interviewees were unsure of the extent of NATO engagement on AI themes. The majority of interviewees reported the view that NATO was still in the early stages of considering AI capability development, mirroring the early and limited applications of existing AI technologies.

At the same time, interviewees also noted the pace of innovation as both inconsistent and fast-paced. Interviewees highlighted the nature of rapid AI innovation as a justification to pay attention and invest in the space earlier rather than later to prevent being disadvantaged. While the interviews did not specifically ask interviewees if they believed an "AI arms race" was underway, the wish to innovate ahead of adversaries was mentioned organically on multiple occasions. As B17 stated, "the trick here is, I mean just not, this isn't like normal technology, we've seen three waves of that or four ways depending on how you count it, I would say three: there's gonna be a fourth there's gonna be a fifth. So rather than pulling back now's about the time to start doubling down. Because you know, the stuff is gonna come true". Distinguishing AI from previous forms of innovation, interviewees highlighted that AI was an asymmetric capability and a "game-changer" (multiple interviewees) in the sense of potentially leveraging capabilities exponentially, in terms of speed or processing power. These

conversations mirrored the strategic discussions observed in the literature around military AI, which encourages NATO to invest strategically (Gilli, 2020) and to note how falling behind could have significant implications not just against adversaries but in terms of interoperability with allies (Dufour, 2018).

When asked about how current AI techniques may be deployed in a military environment now or in the near future, interviewees responded with a wide range of views. A number of interviewees highlighted that overall, AI systems are expected to increase stealth and rapidness, with B4 stating that AI is "an enabler - it's primarily an aggregating and speed of decision-making enabler. And maybe for situational awareness too, which ties into all of those, but it makes us understand the adversary's picture, or understand the threat more quickly as well". Interviewees also listed uses, including AI-enabled autonomous tools to enable cyber attacks, autonomous weaponry with integrated AI target recognition, logistics tasks such as predictive maintenance for military vehicles and electronic warfare (to monitor frequencies). AI was also highlighted as a valuable tool within intelligence, surveillance and reconnaissance information processing, situational awareness and decision-making, and in hardware systems from aircraft to anti-missile capabilities. It was noted by multiple interviewees that AI would have a significant impact on situational awareness and decision-making, with B15 referring to the U.S. Project Maven as an example of image analysis processing in the intelligence environment. The use of AI was also recognised as impactful in terms of enterprise-level deployment; as B15 noted, "So we'll see the proliferation not just in conflict but just in the kind of in the back office of conflict if you want to if you want to say so". Between references to logistics and personnel management, the use of AI in data analysis and data exploitation, and the reference to dual military-civilian tools like Microsoft Office technology, there was a recognition of the banal deployment of AI technologies and its importance in supporting military operations. Apart from one interviewee, those interviewed did not go as far as to focus specifically on NATO logistics or sustainment as an ideal primary focus. This tone contrasts with some strategic scholarship highlighting the relatively uncontroversial, less risky opportunities[95] to deploy AI in enterprise environments (Konaev and Chahal, 2021).

Interviewees noted that including AI capabilities were likely to have strategic consequences as decision-making cycles were shortened and staff were asked to increasingly rely on intelligent

[95] Relative to AI applications deployed in operational battlefield environments. Such applications might have safety-critical implications, including risk to human life, should it misperform.

technologies. B15 noted, “With the proliferation of AI, we will see these decision-making cycles being compressed. We'll see that conflicts will, will be escalating a lot quicker and that reactions will be also expected a lot quicker”. With AI processing and providing analysis faster than a human can comprehend the supporting information, it was recognised that within military operations AI “will not only be a tool a cognitive crutch, a cognitive tool for the commander or for ... for those involved in the conflict, but it will also enable new behaviours and new capabilities of the context of economy” (B17).

The capability for AI to provide synthesised information at vastly increased speed was considered an enabling factor for future personnel to carry out actions that would be regarded as too risky or impossible today. Looking at the operational challenges around integrating current AI technologies, interviewees' answers generally corresponded to three categories: trust in the AI system, security and verification, and non-technical focused legal or ethical concerns.

On trust, interviewees raised the nature of complex AI algorithms as black boxes, through which the calculations and decisions may not be decipherable to humans. Multiple interviewees mentioned this problem of explainability as a challenge, where personnel may not wish to risk their personnel’s safety based on a recommendation that they do not trust, especially if they cannot see the logic behind the AI’s output. Furthermore, even if there were a way to fix “explainability” technically, this would typically be an a posteriori, after-the-fact description of the recommendation, useful for audits and lessons learnt but not available at the time.

B3: “The soldier, the commander, the jet fighter pilot, we need trust in real-time.”

B17: “You can have the most incredibly sophisticated system in place. But the commander doesn't trust, or you can have a very simple system that the commander trusts, which is going to be more effective?”

Anticipated challenges around trust were also linked to broader security concerns in protecting the confidentiality, availability, and integrity of the AI-enabled system.

B3: “Then, of course, the AI itself making those decisions at the art of, for example, say software-defined networks. Cyber attacks will target those artificial intelligence mechanisms, and that's of course a way for attackers to create that

supplement of distrust in systems. So, the perfect influence is that we lose confidence in our systems. So, we have a big issue ahead of us that may limit the extent to which we can trust and rely on AI-based systems".

Broader critical infrastructure was also highlighted as a target, including the data centres and underwater sea cables that would contribute to data transfer and increasingly become strategic targets.

B15: "In the same time as we are developing an approach to AI, we are developing an approach of data as well data exploitation [sic], not just data management, but data exploitation. And here obviously security will have to be baked in from, the really, from the beginning you just, I mean, I don't want to foresee the consequences if our datasets get corrupted. I mean, that would be... quite disastrous, yeah."

B16: "Safety is one thing, security is another thing, certainty is a third. And I think a lot of the considerations about future application are more in the area of certainty than security, are more about uncertainty than they are about insecurity."

The protection of AI systems was also connected to supply chain security, ensuring high-quality data to train the algorithm (with multiple interviewees highlighting bias as a challenge). Interviewees raised the use of testing, verification, and accreditation to mitigate security risks and develop trust on an application-by-application basis.

B17: "When we have a Commander of an operation, how do we know we can trust them? Well, you look at the training, you look at their background, we look at how they've been exposed in simulated environments -training environments or exercises that they conducted. And I think, in the end, we're going to be driven in a very similar way with AI. Where it will be before you come to operation [sic], you will have to accredit your tool. You will have to do so in a virtual range of virtual sandbox. Now that has its own limitations since, by definition, those tools can only think as far as they're programmed, but they can learn over time as well

from other operations."

Reflecting on non-technical aspects of current AI innovation, interviewees almost exclusively raised the legal and ethical discussions around the development of AI and military AI systems. Interviewees highlighted the ethical challenges and divergences, particularly around controversial military AI applications such as LAWS, and underscored that states had yet to agree to a consensus approach to ethical principles. A range of views were taken on whether NATO should take a role in AI ethics and how far ethics forms a significant component of discussion on AI innovation

B15: "I think in many allied countries the ethical debate is in, is holding back, simply on the development."

B2: "At some point, we would need to... we would need to teach it to learn what's wrong. At some point, we would need to let, you know, we can't just teach it to look for patterns and all the rest of it. At some point, it needs to be able to learn, for want of a better word, you know, morals and ethics."

There was more sympathy for the view that NATO can offer the direction and vision to encourage a responsible approach to development and deployment (as highlighted in the literature by Gilli, 2020), as "the ethical impact must be thought at least already in the beginning. It's always complicated to start an ethical discussion after the fact" (B9).

Legal mechanisms were also referenced as a mechanism to shape non-technical approaches to AI innovation. Again, the legal field was highlighted as a complex and controversial area that interacts with innovation outside the military environment and is intrinsically linked with principles relating to just war, such as *jus in bello*.

B10: "So if you have any type of technology, let alone AI that doesn't take preservation of life into the conversation, that technology is gonna get sidelined, at least legislatively because people aren't going to... they're not gonna be okay with it even in warfighting".

B9: "There is also a kind of conflict between the civil society and the governments and lawmakers. Because only one side lawmakers want to regulate

it, but on the other side the service society is fearing that the rules and the laws which are derived from, from this effort are somehow too industry-leaning".

Regulation was not mentioned as much as the development of international norms between members, though one interviewee suggested regulation would limit AI innovation. Another interviewee raised the "DOTMLPFI"[96] military approach to capability development (including at NATO - which added the '-I' to the U.S.' original acronym) and highlighted that within doctrine would be the basis of legal principles for specific areas.

B1: "So, again, AI is not just a tool. AI is- AI needs to be developed within the office of DOTMLPFI in order for us to appropriately use it and understand why we're not just buying a shiny tool and plugging it in, but why we actually need this as part of our overall, you know, cyberspace approach or cyberspace strategy for NATO."

B1's view aligned with that of the former NATO ACT Commander André Lanata, who suggested the DOTMLPFI approach, with a specific focus on human dimensions such as trust and effective human-machine teaming, would be "essential to operationalize AI" (Gilli, 2020, 7, *foreword by Lanata*).

### 6.3.2. Military innovation at NATO

Interviewees highlighted a broad range of agencies and mechanisms at NATO where AI-related innovation was taking place, from low-level research and development and capability testing to operational, strategic and legal activity.

B15: "So we have actually [have] maybe too many people to be quite honest. People can be excited about it, everybody's talking about it, so in terms of policy, it's us here in the headquarters in Brussels. You have the scientists, the STO, which is talking about it over well also at headquarters and over in Paris. You have defence investment looking at it in the context also again in the headquarters in the context of a multilateral cooperation [sic] on common projects. You have allied command operations looking at it in terms of implications, what are the operational implications of it, you have allied transformation at it, you need it in

[96] Doctrine, Organisation, Training, Material, Leadership, Personnel, Facilities, and Interoperability

terms of what new capabilities we can feed into the process, we have the Office of Legal Affairs looking at principles."

Interviewee descriptions of relevant bodies included the NCI Agency, the STO, the Emerging Security Challenges Division (ESCD), the NATO Standardization Organization, and Centres of Excellence such as the NATO CCDCOE. While relevant activity was generally expected to happen at a range of NATO agencies (as B15 above alluded), there is a risk of over-fragmentation. There was also a view that different agencies and bodies all play a part in the broader NATO approach. For example, B13 outlined the distinction between NATO's two strategic command structures: where Allied Command Operations (ACO) understand the operational challenges and "need to be involved in order to explain what their problems are", Allied Command Transformation (ACT) focuses on improvements and have the funding to "have the money to have the role to set the requirements". Similarly, some agencies are best placed to focus on policy, like the ESCD, while others, like the NCI Agency, were deemed more appropriate for understanding, developing and applying AI technologies.

Relevant initiatives mentioned include NATO Smart Defense Projects (B1), the NATO Defence Planning Process (NDPP) (B10) and, more broadly, the NATO Lessons Identified and Lessons Learned processes (B11).

B1: "I think a report from analysis done on that whole process said that it takes about 16 years for NATO to adopt capability from cradle to... in-service support and that frankly is just too long of a timeline. So we've adapted to a new governance model and we are currently trying to deploy capabilities in this new model, and hopefully not a sixteen-year timeframe but more so a, you know, three to five-year time frame, which is much more palatable for the nations and then you're not, you know, deploying capability that isn't obsolete by the time it hits the ground."

The NDPP was cited by multiple interviewees as a mechanism through which NATO could shape and coordinate member approaches to capability building and investment, though it was pointed out the process did not apply neatly to consider AI-enabled technologies: "the process is there, but I don't see this….it is for the next few years this is still very theoretical." (B6). The NDPP process was considered valuable but complicated, with B14 stating, "It's [the NDPP] like five or six phases. It starts at HQ, it goes to ACT, goes back to ACO, and then I don't know where it goes, but just that's not the easiest part. But what matters for us is that introducing AI capability goals in the defence

planning process would probably help push countries to move in the direction.". With the NDPP incorporating cyber defence as an area of focus in 2012 (NATOc, online), AI may well be integrated in a similar fashion to align national efforts, whether on its own or under the wider recommended project, to "anchor EDTs in the defence planning process (NDPP) to ensure all Allies modernise appropriately and consistently (NATO Reflection Group, 2020, 30).

The National Armaments Directorate, a group of directors of member states' procurement branch, which sits under NATO's Defence Investment Division, was also listed as a relevant NATO instrument, representing a group where procurement goals can be defined.

Several interviewees (B8, B9, B14, B15, B17) highlighted the importance of NATO standardisation processes, with B8 offering the view that "NATO has been quite successful at least with standards that are basically the basis for the member countries also to develop capabilities.". Standardisation offers the chance to agree on "shared definitions of responsibility, including shared definition [sic] of safety, shared responsibility of security and explainability and so forth all the ethical standards" (B14), with this work typically carried out by the NATO Standardization Office. This view corresponds with literature describing NATO's Standardization Office as uniquely positioned to determine "a natural convening point to see which civilian standards can apply to the military realm, as well as identify niche areas where military AI standards require dedicated attention" (Stanley-Lockman, 2021, 34). With the NATO Standardization Office, the largest military standardisation body in existence, there appeared to be a consensus that standardisation was a crucial aspect in which NATO could enhance secure innovation for AI (Pepe, 2020).

More generally, NATO's operation as a consensus-building and facilitating body was stressed in 14 of the interviews. This attribute shapes the role NATO can play in relation the military AI innovation or associated activity; "NATO is not the EU, it's ultimately a consensus mechanism, is about trying to find agreement, but then the responsibility goes back to countries which have sovereignty over their defence policies and so forth" (B14). NATO provides a forum and some mechanisms to facilitate agreement but cannot sanction or punish any members; "NATO policy is ultimately the product of the consensus of all the allies" (B14), and "the mechanisms that we have at NATO are certainly consultation, consultation, consultation" (B16). As B7 described, "ideas will be brought to the table, policies will be brought to the table, and so on and so forth, and those that are supported by all of the nations, no exceptions, are the ones that will be adopted. It's a difficult process and a cumbersome process. But it means any friction or any misalignment or any disagreement at the end of the day...

the policy that is subscribed is agreed by everybody" (B7). NATO is developing standards, processes, policies, and tools that facilitate emerging military AI innovation available to members - should they choose to utilise them.

Interviewees also noted NATO's research efforts and broader cooperation with other non-state actors, namely other multinational organisations and private industry. While some of these points are discussed in section 6.3.4 and section 6.4, it is interesting to note that interviewees felt the NATO Enterprise should focus on broader research development. Proposed focuses include: intellectual property theft (B10); the near-term impact of military AI (as opposed to the horizon level analysis in NATO 2030 or STO's S&T report) (B13); utilising and tailoring open-source models to serve NATO's needs (B13 offered the NCI Agency's work adapting Google's BERT algorithm as one existing example); and strategic partnering with private industry on specific topics. It was also noted that NATO was more likely to leverage rather than compete with industries with the most advanced AI capabilities (with Google, Microsoft and Amazon mentioned by multiple interviewees). This tendency was believed to be due to resources and potentially satisfy states, in the sense that NATO funding goes back to benefit the defence industry and back into national economies. B13 stated that over half the NCI Agency's budget goes to industry, "something that the nations like and positively encourage, and ultimately it's their money".

Overall, interviewees discussed NATO's limitations at length regarding the challenges around AI innovation in the military and NATO-coordinated military AI innovation. For example, interviewees noted that NATO has a bureaucratic structure, which represents a limitation on its innovation capabilities. As B17 joked, NATO's acronym could alternatively stand for "No Action, Talk Only". As B1 highlighted, "NATO as an alliance, we tend to have not just, you know, one nation to respond to but thirty nations to respond to, and as a result, the latency introduced into all of our processes really impacts our, you know, our reflexibility and our agility". Interviewees perceived NATO's general capability development to be a slow-moving and complex process. Interviewees distinguished between common-funded projects, which include interviewees from all members, and opt-in multinational projects, which were considered more agile: "you always move at the speed of the slowest but in these multinational things rather than what we term at NATO a common-funded thing, you at least move at the speed of people who want to run. Whereas sometimes if in what we term the common funded thing, if they're spending NATO money, then it only takes one nation to say we're not sure about this or we're not ready for this, and that can slow the whole thing down" (B13). NATO's innovation processes also impact technical innovation as B11 reflected that the "project

management and the procedural framework we put on ourselves [NATO] is very limiting to innovation technology development [sic]. We're lagging behind". The need for NATO to adopt innovation mechanisms was reflected in multiple interviews, and B16 argued NATO must "look at innovative ways of financing and we need to become much more flexible organization to do trial and error a little bit and fail fast and then try again", something that required a "culture shift" at NATO. NATO exercise "TIDE SPRINT" was raised as one event that aims to encourage partnerships and interoperability relating to emergent technologies.[97]

Interviewees were generally of the view that NATO may not be able to drive military innovation compared to state militaries, lacking the instruments and available funding to do so. However, the belief that NATO could shape the strategic direction of AI emerged in over half the interviews. Interviewees acknowledged that NATO has "the opportunity of helping nations, sort of, steer their national programs" (B1) with the AI strategy setting "some direction for other nations to work to align behind" (B13) and NATO taking the role to help "create the greatest synergies between the different allies" (B8). Ultimately, the success NATO may have in shaping military AI approaches, at various levels from standards to doctrines, was seen to depend on whether states were able to agree on consensus activity and look towards NATO as a source for guidance, cooperation with members, strategy or norms. B16 provided a succinct summary mirrored in other interviewees' reflections that members must decide "how much they want to use NATO in this way" and stated, "Our [NATO's] job is to facilitate consultation amongst allies. It's not to tell them what they have to do. So we can influence that, facilitate that, provide a bit of a footing, holding for that but it depends on what allies jointly are willing to subscribe to, and consensus principle rules.". The methods and benefits of NATO-coordinated activity or facilitation are outlined in the following sub-theme.

One area where interviewees were divided was the extent to which NATO could, or should, encourage 'responsible behaviour' relating to the use of military AI. Returning to the theme of NATO as a facilitating, rather than a prescriptive, organisation, some interviewees felt ethical-related discussions must be state-driven. B15 reported their belief that "there will be some allies they will want to see extremely strong extremely precise principles, that will be others who say we'll want to see a little of a looser dealing with these principles". Multiple interviewees also answered that NATO should not

[97] Think-Tank for Information Decision and Execution (TIDE) Sprint is part of a series of NATO-focused interoperability events, including *TIDE Hackathon* in which NATO and Member nation teams develop innovative responses to multidomain challenges. For more information see: https://www.act.nato.int/articles/tide-sprint-space-creative-solutions

by trying to incentivise certain behaviours simply because the technology, and NATO's understanding of the technology, was so immature, and "the ethical questions that are not clear" (B2). While some interviewees felt NATO should play a role in directing ethical discussions given the potential consequences of escalating irresponsible use, they acknowledged that this might not necessarily play out in practice. As B9 articulated, "at the end it needs to be a political decision, the political will to decide how artificial intelligent intelligence should be used or will be used in a military, in a NATO context".

### 6.3.3 The advantages and limitations of NATO activity

An aspect highlighted consistently by most interviewees was that NATO was driven by states. As B1 highlighted, "we [NATO] don't mandate. We cannot make any nation do anything, and it is our job to facilitate and to ensure, yeah, collaboration and cooperation amongst the nations for cross-pollination of data". NATO cannot enforce binding actions on states nor set the agenda. States determine the proportion of their national budget allocated to defence and NATO and are free to invest in technologies of their choice, which may or may not include AI capability-building. The state-driven nature of NATO was highlighted as a core aspect of NATO's impact on military AI. NATO staff may have "all the strategies, desires and everything else that they want, but they are wholly reliant on their constituent countries investing it and developing it and then being willing to use it on behalf of NATO" (B2).

All interviewees also highlighted several reasons why states may choose to engage with or through NATO on the topic of military AI. As a consensus-driven body, NATO provides a platform for members to come together to form a common agreement. This consensus-building capability was noted as an invaluable tool in providing a united front across capability building and, in principal terms, with implications for international norms development.

> B12: "That we built the consensus, it takes time, but when it's consensus, we move move [sic] ahead stronger in the long term, that making compromises in, or leaving somebody out of the consensus might give short-term efficiency, but it in long-run it might you know break the the glue or the fabric of of the alliance [sic]. And I can see the point in there".

> B16: "We as an alliance, as a military defence alliance, as a Western military

> independent alliance, we have found principles in place. That's just, we need this to build trust, we need us to build trust amongst us and we need to build trust with our public and to differentiate us from our potential adversaries and competitors".

Interviews noted the advantage of interoperability with a NATO-coordinated approach to military capability building, enabling members to coordinate across national systems effectively.

> B17: "There are situations where it is in the U.S.'s best interest to say: NATO needs to make more of a leadership [sic] because especially when it comes to things that will be used together on the battlefield, the U.S. can try to dictate all it wants. It doesn't mean nations will follow. But if NATO in dialogue with the nations [sic], perhaps, prodded as little by some of their nations, not necessarily the U.S. but by some of the nations... it can force discussions and make sure that things are put into place to make sure those capabilities work because… At the end of the day, what's important is when people show up in theatre, they can work together in a way that is effective and safe."

Through common funded projects (through which all members engage) or opt-in programmes, NATO R&D was said to help lower the costs to individual states while allowing them to benefit from research outcomes collectively. In addition, through capability-building mechanisms such as the NDPP, there is the opportunity for reduced duplication and competitive advantage, where states invest in their strengths.

> B10: "We don't want every single nation, you know, getting really good ISR [intelligence, surveillance and reconnaissance] assets because there are so many other areas that NATO can use to get better, so the NDPP is going to, I think be the guide where which NATO is an organization sits down, talks about, okay who and what entity would be the most appropriate to invest in artificial intelligence so that they could exploit it and maybe use it in possible operations".

Furthermore, as an organisation that serves nations, NATO can be particularly helpful to states with a relatively less developed approach to military AI.

> B7: "The alliance operates a little bit like, over and above. It has to provide something without overlapping with the nations who will have their own policies, strategies and regulations about AI. They will have that set of capabilities, and NATO will try to either make those work together in a federation, and also provide something that the nations are not able to provide".

Interviewees also highlighted high-level relevant challenges for NATO as an Alliance facing the implications of military AI. A primary challenge not specific to AI training data or systems is limited information sharing between members. As information sharing is voluntary, members have discretion on what they choose to share with individual members and with the broad 30-strong alliance member network. Interviewees noted that sharing was limited due to challenges around trusting such a large group of states; as B5 noted: "if you really have a secret, you really won't share it with your 29 friends".

> B2: "Most of that sort of stuff [state-coordinated breakthroughs in military AI] will not be open and will not be shared because at that point you don't want your adversaries to know that you've got that capability, but as soon as you start sharing it, word will get out. NATO is not a secure environment, NATO only works up to secret, it is not a secure environment to be sharing top-end capabilities".

> B12: "In the field of innovation, that, nations are still keeping military innovation and developing new weapon systems to themselves, they're not integrating it into common NATO efforts and I'm not saying it's a bad thing. It's the reality".

The interviewees perceived the lack of information-sharing as a challenge that extended beyond AI-related data and included the broader principles of "lessons learned". Multiple interviewees commented on a culture of states not sharing failures, preventing lessons identified or lessons identified processes from being carried out by others in the alliance.

B1: “But if we were able to get on the curve earlier when it came to sharing failure stories, I think that would actually also strengthen us as an alliance because then nations independently wouldn't be spending money exploring solutions that, you know, may have already been explored somewhere else and you know, have findings or outcomes that maybe are not in line with the expectations so it would be nice for early sharing and for a little bit more trust, I suppose, because again it comes down to the politics and how nations perceive one another. Nobody wants to show a bit of vulnerability at the table, and that also applies when it comes to technology and advancements.”

The challenges in trust are affected by the sheer number of NATO members and the differences between member states. Members have different intentions regarding AI in the military context across operational and doctrinal lines.

B14: “Threat assessment within [sic] NATO allies varies dramatically. And so you can't, you know, you can have the same outcome with different causes. So probably the Baltics look at AI against Russia, the U.S. look at AI against China, and probably I don't know countries like Italy or France for a different reason, clearly also this one, but also for counterterrorism and stability in the Middle East. So you could have a plurality of different causes which may support the argument for investing in AI”.

B12: “The differences between nations are so large that trying to get everybody on board to develop cutting edge technologies... it's just not going to work.”

These diverging postures were highlighted as a potential barrier to Alliance-wide consensus or even discussing some controversial military AI capabilities applications.

B12: “… nations said: no no, automation in the military context is very very bad, we don't think... we don't want it in your promo work, so the one thing is that political sensitivity that you know, some nations feel that anything that can be interpreted as in quotation marks ‘killer robots’, they don't want to be associated with. So that oversensitivity, that atmosphere can complicate things.”

B13: “If in what we term the common funded thing, if they're spending NATO money, then it only takes one nation to say we're not sure about this or we're not ready for this, and that can slow the whole thing down.”

6.3.4. “Early days”: understanding AI at NATO

In almost every interview (16 out of 17), interviewees highlighted the need for an improved Allied understanding of AI technology. A lack of understanding was highlighted as a challenge in forming adequate regulations, security considerations, and procurement and investment, as B9 summarised that “the discussion around AI especially from lawmakers, from parliaments and governments, is not very informed. Yeah. And this makes it complicated to have an understanding of which regulations are needed, which regulations have been taken, or which regulations should be done.” Interviewees perceived “blind spots”, with staff often not realising the extent to which AI is already impacting their daily tasks (B4). This lack of recognition was perceived as a negative in limiting informed decision-making and innovation. As B13 mused, “whether you're contracting for military satellite services from other nations or whether you're contracting for an AI solution to be built, you want to have somebody who understands how the users will use it and how industry is going to provide it, otherwise, you're going to get caught. Somebody will end up very unhappy, or somebody will end up very rich, if we don't have that knowledge somewhere”. As well as echoing current NATO Defence College publications that highlight the importance of informed NATO staff in procurement roles (Gilli, 2020), these comments highlight how the success of military procurement will often depend on how educated each military is in terms of effectively understanding the technology - and subsequently procuring appropriately (Anderson et al., 2015).

Generally, it was acknowledged that NATO, as many of its members, was at the extremely early stages of exploring the potential of military AI. Examples of successful case studies were scarce, though some interviewees noted public intentions to use AI in cyber defence through research at the NATO Communications and Information Agency (NCIA).

B1: “Like most nations, NATO is also trying to explore AI options and working with industry, like I said, to try to understand the application or the best sort of use cases for AI within NATO.”

Acknowledging that military AI innovation was currently immature, interviewees nonetheless offered the sentiment that there was no time for complacency for NATO or NATO members.

> B11: “You don't want to be caught by surprise when everyone else is used, and you’re falling behind. Because as you know, technology developments is racing forward [sic], but for the users to actually learn how to use that technology that takes longer and if everyone is racing ahead, taking this more into account than we are, we will be behind.”

Over half the interviewees highlighted a lack of consensus on the definition of AI in the NATO environment. While interviewees, as experts in AI innovation at either a policy or technical level, unanimously agreed in their view that AI is used to refer to modern data-based techniques, they highlighted frustrations for those who conflate it with the concept of autonomy or with older conceptions of AI which would not today be considered “artificial intelligence”. It was noted that while NATO does include various non-definitive definitions across high-level documents, these have been drawn on working definitions from “the UK, the US, a variety of sources, and kind of amalgamated into a definition” (B17).

The lack of agreement on the scope of AI was considered to contribute to challenges in discussing the implications of AI and other emerging or disruptive technologies at NATO. B6 highlighted how this vagueness presented difficulties, stating that “there is so much insecurity about even getting the terms, right? Like, what do we talk about when we talk about AI because it can be everything?”.

> B16: “We [NATO] are still finding the language on how to couch tech for its political and military impact. We are very used to talking about military implications and political decision-making. But how technology features there is a bit of a struggle.”

Given the rapid pace of change and evolving AI techniques, the interviewees’ consensus on the lack of an agreed NATO definition for military AI is not necessarily surprising. As B17 noted, “I'll guarantee you in ten years will be talking about how to integrate AI into the battlefield and somebody will sit back and tell you: I'm sorry neural networks, that's not really AI, this is AI. Because we keep re-defining it as we go along”. Partly this is outside NATO’s control, as broader discourse on what constitutes artificial intelligence changes (knowledge-based systems no longer being captured under some modern conceptions of AI, for example). The AI strategy may go some way in establishing definitions beyond the STO’s descriptions in their reports which do not represent formally NATO-

agreed terms. As the entire document will not be publicly available, it is not possible to assess the strategies' effectiveness in facing this challenge.

The scope of what should be considered "military AI" was also highlighted as a blurred and overlapping space as both civilian and military personnel may use non-military applications.

> B9: "Yeah, and dual-use in this way that you cannot distinguish from the design. Is it now a military product or is it now a civilian product? .... Just because I have Microsoft Office on my military computer, and not only on the office computer but also on a command-and-control system. Is it a military system?"

Beyond definitions, interviewees highlighted a lack of understanding of AI both by NATO members and across NATO infrastructure at a general level, with B11 expressing the view that at NATO, "the understanding of what AI is and what it can do at a very low level currently." The lack of awareness specifically relating to potential AI capabilities and limitations was mentioned by several interviewees, and B8 offered the view that "there is a lot of misunderstanding about how much this technology are [sic] being actually used, in so many ways in our day-to-day life, but also actually also in defence." This limited understanding of the fundamental technology translated, according to interview interviewees, into misinformed views on how the technology can be leveraged, with many staff having an idea of the abstract concept but "no idea what the limitations are, they've no idea how we really operate, they've no idea of the prerequisites to be able to build effective AI or what the limitations are" (B13). An under-informed perspective of AI technologies may mean staff do not understand AI capabilities. As B7 explained, "lots of people don't necessarily understand the way AI works technically, and therefore they take two of those views: those that believe that you can do anything with AI, and those that believe that AI is so dangerous because you never know what the outcome is gonna be, right."

Interviewees highlighted the 'hype' around AI contributing to misunderstandings or misinformed non-expert views across NATO. This hype presented at both extremes: interviewees referenced marketing materials by vendors as a mechanism by which AI is "overhyped" (B12), as well as the counterview, which highlights the Terminator-style threats that may arise (B17).

B6: “We're talking about critical thinking when it comes to your media consumption, right, and the one-sided view of AI, you know, all the bad things it can do, all the good things it can do that, that is actually probably the main source of information that most people get. And when you don't balance that with analysis and, you know, sort of thinking about it for yourself, you get into these stereotypes”.

Interviewees also raised that AI will be a theme that will inevitably be increasingly relevant to NATO and should be openly addressed by NATO staff leadership.

B6: “How long are we going to pretend that it's not there is the question’”

B17: “They [NATO] could literally have every decision-maker in this building forget about AI, and it won't change the fact that AI is coming. It's going to be built into those technologies whether you recognize it as AI or not.”

This argument extended to the broader alliance, where a well-informed position was considered crucial in effectively developing and using the technology in conflict.

B11: “If you're gonna be a player, or if you're gonna be a stakeholder in a conflict, any kind of conflict, knowledge is everything.”

Having acknowledged the above, this general perceived lack of understanding at the institutional level was not necessarily an urgent fundamental challenge. Only certain parts of NATO international staff will be involved with decision-making relating to the approach to military AI innovation. As B14 explained, “...in the emergency security challenge division, they probably have a good understanding of these issues. In other divisions, probably operations, they definitely have way less understanding just because they focus on different things.’”

### 6.3.5. Geopolitics and power

Interviewees frequently mentioned the power of influential actors in the international landscape. Many actors were highlighted, including states, several of whom have already invested heavily in AI technology. Some interviewees reflected on the distinction between states in this category that may

seek to influence and shape the direction of international approaches to AI and states that had not yet publicly announced their intentions.

When it comes to states and NATO's engagement with military AI as a theme, interviewees noted that states might often have interests that do not align with the broader alliance, particularly when protecting their innovation and promoting national commercial efforts.

> B1: "You know, if nations were to just start giving away their AI or other research to nations that don't necessarily have the money or the maturity in place, then that brings about a whole set of political issues and financial issues and so really what has to happen is the two needs to be balanced. We need to encourage sharing, but we also need to understand that nations have a vested interest, obviously looking inward first and then looking outward towards the alliance, and so as part of NATO... we have to respect the boundaries that nations also have to look after their own self-interest."

It was also noted that NATO's budget was significantly smaller than some member military budgets. B13 commented that the annual NATO budget is roughly equivalent to "one service of a medium-sized NATO nation… we're [NATO] not going to lead because we've not got the money to make investments".

> B5: "The way international organisations work is that they get funding from the governments to deliver something, and then they end up doing the project management work, and they subcontract, they outsource the actual development back to the nations. So it's a circle of money. You get money from the nations, you are the front end, the NCIA, the European space agency, you name it. But then in the end it's the big defence companies that are doing the development."

Looking toward non-state-driven innovation mechanisms, some interviewees highlighted the influence of the private sector. In particular, interviewees highlighted the disruptive nature of non-traditional defence suppliers, including large technology corporations or start-up enterprises.

> B2: "You can't deliberate between public and private. You can't deliberate between the military and industry because industry has to develop all the tools or all the equipment."

Large technology companies can have a significant impact on innovation, and lack thereof, for the military sector, as offered through the example of Boston Dynamics:

> B16: "The moment that Google brought up Boston Dynamics, Boston Dynamics ceased to work in the area of defence, despite their remarkable progress in robotics... as soon as it was Alphabet rather, lost the interest and told Boston Dynamics off again puts their back in the defence sector. We lost, the defence sector lost two and a half years of Boston Dynamics, if you wish, and that has an impact. It has an impact about [sic] technological direction as well."

At the in-house development level for NATO, B10 highlighted that technically skilled staff would be better equipped than those setting the requirements (the "end-user"), offering the view that "there's a greater understanding on the side of the engineers as to what AI is capable of" (B10).

Both the United Nations (UN) and the European Union (EU) were frequently referenced as other international organisations that may, indirectly or otherwise, influence the direction of military AI innovation. The EU was mentioned as a relevant organisation when it came to economic trade and investment in AI technology, while UN working groups were mentioned more frequently concerning discussions on military AI and the application of international humanitarian law (IHL), including the Governmental Group of Experts (GGE)[98] where many NATO allies have contributed to the consensus-building discussions on lethal autonomous weapons systems, culminating in 11 non-binding principles (United Nations, 2019). As B10 outlined, "the way that NATO and the EU approach things, they're looking at different angles. You know, if you're looking at the economic developments and economic benefits of artificial intelligence, that's more of an EU function."

Interviewees acknowledged that NATO had a very clear remit in terms of military scope. However, they also acknowledged that there was still uncertainty around which organisation may be best suited to take on military AI-related decision-making, for example, concerning private sector innovation.

[98] See https://www.un.org/disarmament/group-of-governmental-experts/

B14 stated that there was a narrative through which actors could potentially complement each other when aligned correctly: "NATO] has proved to be an effective forum for discussion at least to agree whether then you would need a more, a stronger actor like the European Commission, and whether assuming that then the European Commission is the right actor, is open to debate on that... so you know, what many in Brussels, especially the European part, often do not understand is that having a more powerful central actor doesn't mean that this actor would be right". B15 raised the example of a trusted capital marketplace that could guarantee suppliers were backed by allied financing (as opposed to competitors or suppliers compromised by adversaries) as a competency that they thought fell within NATO's remit. However, they recognised that there is no consensus, and others may view this as an EU competence.

Overwhelmingly interviewees reported the view that NATO, as a defence-focused Alliance, should have an active role in coordinating military AI capability building and responsible adoption of AI capabilities, but that these efforts would not be done bilaterally and would involve working with the broader international ecosystem. Cooperation was highlighted with the UN, the EU, and the African Union (AU), with B9 highlighting "it needs to be a global coalition" beyond Europe and North America. In addition, multiple interviewees highlighted the OSCE [Organisation for Security and Co-operation in Europe].[99] They highlighted the role of "new actors that we haven't been traditionally operating with, for example, international standard-setting organisations" (B15). The concept of cooperation between international organisations recognised the need to avoid duplication:

> B16: "NATO is the only one with the defence purpose. We add always OSCE, it adds a security perspective. If we add the EU, it's somewhere in between and all over the place, but it has the most important facet: money. That none of the others have. If you add the United Nations, we have the moral high ground."

Within the NATO Alliance, a "handful of nations" are "more advanced" (B1) when it comes to military AI innovation, with interviewees usually, if referencing individual states, mentioning the U.S., U.K., France, and Germany in that order of prominence. However, interviewees generally saw nuance in how Alliance members could contribute to discussions by bringing different perspectives.

[99] The OSCE is a regional security-focused intergovernmental organisation with 57 member states (as of May 2021). The OSCE website highlights the organisation as a platform for political dialogue on issues including arms control, terrorism, energy security, media freedom, and human trafficking. For more information see osce.org.

> B17: “It's easier to think of the U.S. as, oh, they’re just going to do what they want, NATO will accept whatever and whatever the U.S. says. And sometimes that is the case. I would be lying to you if I told you that wasn't. But there are situations where it is in the U.S.'s best interest to say: NATO needs to make more of a leadership… because especially when it comes to things that will be used together on the battlefield, The U.S. can try to dictate all it wants. It doesn't mean nations will follow. But if NATO in dialogue with the nations, perhaps, prodded as little by some of their nations.”

The issue of capability differences between members was raised as a potential difficulty for collaboration through NATO, particularly through common funded projects, ‘“or the nations, advanced nations, it would be kind of playing with the lowest common denominator. And for the nations with tight resources that it would still be kind of above their head game” (B12). B14 highlighted the developing risk of a “two-speed NATO” divided into states that are more, and less, advanced at military AI innovation and offered the view that NATO should try to address any growing capability gap. NATO was seen to be able to mitigate this gap by providing guidance and resources to states that were less advanced or had fewer resources to develop capability bilaterally. B13 also highlighted that NATO already provides the opportunity to narrow the gap to the extent states choose to share their technologies, as NATO “becomes a point for many of the leading nations to come and bring what they have and then to share it with the...I don't want to say the non-leading nations, but with all the other nations who get that benefit, and it pulls everyone forward.”

Perceived military innovation by adversaries was also raised in twelve of the seventeen interviews. Most of these raised state threat actors as a significant threat, with Russia and China repeatedly offered as examples.

> B1: “Definitely nation-level threats have been a little bit more advantageous in adopting AI technologies faster than we have because a lot of the nations that we tend to think of as - I don't want to say adversary states, but states that tend to cultivate adversary behaviour - tend to move rather rapidly and don't have the same level of you know, sort of bureaucracy I think as other nations, especially as the Alliance.”

When discussing hostile state actors, interviewees were concerned that adversarial states might have an advantage in rapid military AI innovation due to their different approaches in ethical, legal, or bureaucratic terms. B10 raised the example of private-public partnerships and China's greater access to corporate data, stating, "you need really good data, so the handful of nations that can get their hands on it … it's not the U.S. government that has all the data, it's private corporations within the United States or within, you know... when you look at a country like China, they have all the data."

> B13: "This means either NATO nations will have to change and adapt to this [innovation landscape] or it will lose an advantage of decision speed against potential adversaries in the future."

A minority (five of the seventeen) interviewees referred to "AI arms race"-type dynamics. External pressure was noted as a motivator for relevant AI innovation, with one interviewee expressing the view that "most innovation is driven by a response to your neighbour doing something" (B2). It was recognised that sometimes these dynamics were unhelpful where hype was perceived as distorting the narrative, with B17 explaining that "there is also hype in the context of overblown threat: 'Well, we're going to have to go there because you know the Chinese are. So, or the Russians are whoever is there taking leadership, so we must do this as well.' So those push people in directions they may not want to go, but they are part of their thinking." This narrative aligns with the strategic literature's grappling with AI arms race challenges. These comments highlight how the competitive nature of AI innovation between states goes well beyond the realm of military defence and national security. It was also noted that beyond state-coordinated threats, AI might be utilised by a broad range of actors, including "script kiddies and people who are even just curious from their own basements" (B1), particularly as online algorithm-generation platforms make it relatively easier to develop a capability that has an impact in a military security context.

> B7: "So anybody today you can write a [sic] artificial intelligence space dis/misinformation campaign. Anyone can write an automated attack against defences, or you know, industry or the military. Practically, anybody can use artificial intelligence against, poison information on the web."

Over half of the transcripts (ten of seventeen) highlighted the need for an agreement on the applicability of international law to military AI, relevant norms building processes, and formal standards for how military AI-enabled systems should be developed. B14 described how "the legal considerations probably come on top of everything else. So, like everyone is super concerned that any capability that is developed is complying with domestic and international law". The challenge of securing consensus for international agreements was seen to be a complicated matter where states had different approaches. B15 described how "there will be some allies they will want to see extremely strong... extremely precise principles, that will be others who say we'll want to see a little of a looser dealing with these principles". These reflect the difficulties observed in the literature in which consensus-building efforts are slow and complicated - but are nonetheless in progress (Morgan et al., 2020).

## 6.4. Discussion

NATO has significantly advanced its position on AI as a disruptive technology since this research was first conceived in 2018. In 2021 alone, NATO agreed on a comprehensive EDT strategy, developed and approved a NATO AI strategy, and announced several initiatives to integrate AI to support activities across the organisation. The Brussels Communiqué agreed to the creation of the civil-military Defence Innovation Accelerator for the North Atlantic (DIANA) to "boost transatlantic cooperation on critical technologies" (Brussels Communiqué, 2021, section 36). Operational by 2023, DIANA has been described as NATO's DARPA and "will reinforce transatlantic cooperation regarding critical technologies to assure the security and defence digital literacy of NM [NATO Members]". In a potentially complementary process, the simultaneously announced NATO Innovation Fund will allow NATO members to opt in to "support start-ups working at dual-use emerging and disruptive technologies in areas key to Allied Security" (Brussels Summit Communiqué, 2021). Beyond the announcements, there is little detail on how each initiative is funded and how member states would benefit from contributing to the scheme (Tigner, 2021).[100]

Returning to our first research question, NATO has been doing a significant amount of work to determine the strategic implications of AI for the Alliance and, through the course of this research,

[100] See Tigner (2021) which quotes Secretary General Jens Stolenberg: "Exactly what kind of access [to developed EDTs] that those allies who are not part of the accelerator will have is too early to say, but of course they will not have the same access, and will not participate in exactly the same way as those who decide to be part of it and provide funding for it."

has revealed internal white papers that address these themes specifically. Drawing on the material in public NATO documentation and reports and the perspectives of the NATO-employed experts interviewed as part of this research, NATO perceives increasing challenges to the rules-based international order from actors including but not limited to China and Russia.

Our second research focus for this chapter focused on NATO's activity, capabilities and possible roles regarding AI in conflict. At a strategic level, NATO's contribution comes in many forms and directly relates to how NATO states may mitigate associated security challenges. Through the AI and other EDT strategies, NATO had helped forge high-level agreements relating to how members should incorporate AI capabilities into their operations, including AI principles to address safety and ethical concerns and considerations relating to international law and AI in conflict. In addition to increasing focus and discussions on AI and increasing nuance in reporting over the period of this research, the Organization has various initiatives including: some limited in-house R&D; the facilitation of discussion spaces for workshops and voluntary programmes; and the agreement of the NATO AI Strategy. Reviewing NATO's activity, it appears that NATO has appeared to focus on driving consensus and cooperation rather than developing innovation in-house. , Such activity aligns with perceptions described by interviewees that NATO was not able to drive innovation, whether or not it intended such, and corresponds with Soare's (2021) assessment that when it comes to EDTs, instead of driving technological progress, NATO and allied military organisations "are not the main agents of innovation and depend on effective civilian-military collaboration for their own innovation efforts". At the same time, at a strategic and norms-building level, NATO is uniquely placed as the world's largest military alliance to facilitate defence-specific consensus-building on future planning for AI. NATO can contribute to the mitigation of military AI-related challenges through its unique position facilitating norms-building and consensus development in a military context. While arriving at a consensus agreement across all nations takes time, such agreements provide a strong baseline for activity. NATO agreements can contribute to stability, or at least de-escalation of irresponsible activity, in cyberspace. NATO's Deputy Secretary-General Mircea Geoană has highlighted the importance of NATO in this regard, stating, "once NATO sets a standard, it becomes in terms of defensive security the gold standard in that respective field" (Heikkilä, 2021). While the path to achieving consensus is slow, bureaucratic, and faces significant challenges, the organisation has much of the existing structure required to shape international approaches to AI in conflict. Success in this field means addressing potential differences between members and partners, recognising capability differences, and recognising, and leveraging, the consensus-building efforts underway beyond NATO across forums, including the UN, EU, and OSCE.

Looking forwards, this research has highlighted NATO's perception of two key roles in the strategic future-planning space. First, NATO wishes to drive "dynamic adoption" of AI and other emerging and disruptive technologies among allies and partners (Christie, 2020). Part of this means facing the challenge of under-informed personnel within the organisation, which should be addressed before efforts to communicate a clear strategy across the Alliance. Interviews also included some limited discussion on how trust may prevent the sharing of data and capabilities, minimising "data advantage" relating to military AI explored in the literature (Chahal, Fedasiuk and Flynn, 2020). The NATO Data Exploitation Framework Policy, as announced in October 2021, may help bridge diverging national legislation on data storage and protection (Christie, 2021).

Furthermore, NATO is well-placed as an Alliance structure through which states can cooperate in terms of joint projects and training data exchange on less sensitive projects, including on the topic of military AI (Hill, 2020). At an Alliance level, NATO has several mechanisms to encourage capacity-building, including the NDPP, which could incorporate EDT considerations. Encouraging greater private-sector involvement is essential as part of this exercise, a perspective constant throughout this research and acknowledged by the Organization through the announcement of various investment mechanisms through the Brussels Communiqué in June 2021. As well as encouraging market participation, capability building also requires engagement with diverse talent across academia and the broader research landscape, which includes industry and research centres internationally. At a standards-setting level, the NSO has the potential to set technical requirements which can contribute to safety and consistency for AI-enabled systems.

In examining NATO interviewees' perspectives on the impact of AI in warfare and on NATO's role in shaping the use of AI in conflict, it is possible to see the potential impact NATO could have in contributing to the broader questions in this thesis. The inclusion of AI-enabled systems in battle is currently minimal, with NATO literature and interviewees reflecting an organisation still considering the practical implications of mass adoption. It is recognised that modern AI techniques still have enough technical and human-machine teaming-related challenges to make deployment across the front line largely unthinkable. Any change to this status quo must correspond with rising trust in the technology across operators and users, decision-makers, and all involved. At the time of interviews, interviewees reported a wide range of concerns relating to the use of AI in conflict, taking a cautious

approach to predict that AI would not replace human decision-making. There was an awareness of the unintended consequences of AI at an operational level. Instead, there was a greater focus on the sweeping strategic implications of AI on the nature of warfare, particularly set against the public knowledge that states including Russia and China pose (different, but both significant) threats to the international security landscape. No interviewees proposed that NATO should take responsibility for driving AI ethics-related principal development, showing alignment against one perspective in the literature that argues that NATO has a moral obligation to lead consensus development on ethical questions (Gilli, 2020). It is clear from selected interviews and the frequent announcements from NATO on policy development and agreement that NATO views AI as a significant innovation NATO must embrace as part of their wider "2030" agenda to keep NATO equipped in the 21st Century.

As outlined earlier in this chapter, NATO's scope is necessarily limited to the military context, with NATO uniquely placed to facilitate international military and political actors. However, as a generic, often termed "dual-use" set of technologies, AI will also be discussed by other international structures. On the topic of how warfare is conducted and the principles of *jus in bello*, or international humanitarian law, the UN will be a key actor facilitating discussions on how AI should be implemented in conflict and matters relating to arms control. Any AI norms development must likely engage with relevant UN working groups. The UN CCW have spent several years moving towards a consensus that the Law of Armed Conflict necessitates human involvement in military action (Morgan et al., 2020), while in a paper for the CCW, Roff and Moyes (2016) note the consensus that "no one wants weapons that operate out of human control". Outside the military domain, structures like the EU will be relevant to the development of civilian norms and markets, which may impact how the European armed forces develop new capabilities. In general, NATO efforts relating to dynamic adoption and responsible use will involve engagement with a broad range of actors. These themes will be explored further in Chapter 8.

# Chapter Seven: The United States - leading and shaping military AI

## 7.1. Introduction

This chapter looks at the U.S.' approach to military innovation, drawing on seven interviews with current or former U.S. DoD staff. These interviews took place from October 2020 to August 2021. Each interviewee was selected based on their professional experience directing defence AI programmes at a national policy level. All interviewees worked closely to design or implement aspects of the U.S. DoD AI Strategy. The findings also draw insights from the U.S. DoD AI Strategy Executive Summary (Department of Defense, 2019) and the U.S. Air Force AI Annex (U.S. Air Force, 2019) to ground and contextualise interview data.

Given the size and scale of the U.S. national security landscape, there are many AI milestones, programmes and initiatives in the U.S. DoD. While the U.S. has an extensive range of activities relating to AI innovation, this chapter focuses on the agencies and initiatives identified as the most relevant. Relevance was determined by analysing the public DoD AI executive strategy document, the public U.S. Air Force annex document, and the perspectives of the seven experts interviewed for this research. Furthermore and as discussed in Chapter 2, think-tank research from institutions including the Center for New American Security (CNAS), RAND Corporation, CSET, the independent advisory National Security Commission on AI, and U.S. government press releases all helped inform the scope of this non-exhaustive overview of relevant U.S activity.

This chapter explores U.S. perspectives and intentions relating to AI. These perspectives are explored in the context of the research questions outlined in Chapter 1: "What are the implications of AI innovation in a military context?" and "How are the actors acting to mitigate the challenges associated with military AI?". Exploring the U.S. approach to strategic challenges relating to military AI technologies helped address the second research question, as the U.S. innovation landscape reveals intense interest and increasing investment in AI adoption as well as evidence of international engagement to signal the U.S. strategic position and encourage international norms. Interviewees' interpretations of U.S. intentions and activity to date also highlight an emphasis on maintaining international stability and U.S. military superiority, both through U.S. technical supremacy as well as the attempted establishment of relevant international norms. An exploration of the U.S DoD AI Strategy and the DoD institutions dedicated to military AI-related matters contributed to

understanding the breadth of investment the U.S. had committed to AI across the DoD. Combined with the policy document analysis of both DoD and USAF public documents on AI, interviews with experienced policy-focused professionals across the DoD offered insights into how the U.S viewed AI in an unstable geopolitical landscape, which informed both research questions. In this chapter, section 7.2 will deliver a high-level overview of the U.S. approach to AI in military contexts, presenting an overview of key relevant policy developments and identification of relevant institutions and their contributions to the U.S approach. Section 7.3 will present the findings of the document analysis and U.S.-focused interviews, which are categorised into five themes. Section 7.4 concludes with a brief discussion and reflection on the U.S. position.

## 7.2. Context

The U.S. is investing significantly in military innovation and national security (Konaev et al., 2020). The SIPRI Military Expenditure Database (SIPRI, n.d.) estimates that U.S. military expenditure in the 2020 fiscal year exceeded, in real terms, the defence expenditure of the following 11 states combined. Post 9/11, the U.S. defence budget increased steadily, with significant investments in unmanned technologies (Singer, 2009). Singer notes that the Pentagon also has a "black budget", which is not released and therefore not available for public scrutiny (Singer, 2009, 61). For AI technologies specifically, the U.S. has been clear in signalling its intentions publicly, including through a specific DoD budget for AI (U.S. Department of Defense, 2020), the creation of the JAIC, and a series of statements on AI safety and ethics. Such activity represents continuity in U.S. defence priorities as the DoD maintains and strengthens U.S. global military dominance (Daniels and Chang, 2020). The DoD was one of the few U.S. federal agencies to avoid funding cuts in 2020, with the agency instead receiving an increase in its' R&D budget (Konaev et al., 2020). The 2021 U.S. defence budget requests $1.8 billion to focus on "speed of manoeuvre and lethality in contested environments", focusing on human-machine teaming, and an additional $800 million to go towards JAIC and Project Maven projects (Konaev et al., 2021, 6). In March 2022, the Pentagon proposed the largest defence budget to date, requesting $773bn for the 2023 fiscal year, of which $130.1bn is allocated to a development, test and evaluation budget (Garamone, 2022).

Looking beyond dedicated financial investments, the U.S. Department of Defense is also extensive in organisational terms, representing the U.S.' largest federal agency and consisting of over 2.9 million military and civilian staff (DoD, n.d.), of which over 813,000 are civilian as of February 2022 (Congressional Research Service, 2022). Consisting of many agencies and subdivisions, the size and

breadth of the DoD's organisational infrastructure raise significant considerations when devising coherent approaches across DoD staff and stakeholders. This structure suggests the value of clear strategic direction and leadership, which may explain the relatively early release, compared with other states, of the DoD AI strategy and related documents on AI principles. These documents set out the U.S. DoD's intended actions and assumed values and thus provide a common reference point and resource for the expansive workforce.

This section introduces a high-level timeline, strategic doctrines, and key institutions/initiatives referred to heavily across the documents and interviews.

### 7.2.1 Timeline

Image 7.1 displays a high-level timeline of U.S. doctrine development and organisational milestones relevant to military AI. This timeline is non-exhaustive, and selected items are explained in more detail in sections 7.2.2 and 7.2.3. Major strategy releases are marked in bold to distinguish them from statements that provide an update or amendment to current approaches.

| ***Date*** | ***Milestone*** |
|---|---|
| January 2015 | **Third Offset Strategy** |
| June 2018 | Joint Artificial Intelligence Center established |
| August 2018 | National Security Commission on Artificial Intelligence (NSCAI) established |
| February 2019 | **Announcement of DoD AI Strategy and release of Executive Summary**; Executive Order on "Maintaining American Leadership in Artificial Intelligence." |
| September 2019 | **The U.S. Air Force releases the "Artificial Intelligence Annex" to the DoD AI Strategy** |
| February 2020 | DoD adopts five ethical principles for defence |
| May 2021 | DoD Memorandum on Responsible Artificial Intelligence |
| June 2021 | Launch of 'AI and Data Accelerator' Initiative |
| October 2021 | NSCAI discontinued (Final report submitted to U.S. Congress March 2021, see section 7.2.2). |
| December 2021 | Announcement of Chief Digital and Artificial Intelligence Officer (CDAO) |

*Image 7.1: U.S. military AI timeline*

While high-level, image 7.1 highlights the developments between more general strategic announcements on U.S. strategic competition, from the Third Offset Strategy to the release of more granular developments, including organisational structures and accelerators. It shows the U.S. taking

the initiative when set against the international context, as the first to state release a defence AI strategy, formally adopt ethical principles for defence and publicly formalise an extensive organisational set-up including the JAIC and CDAO (discussed further in section 7.2.3).

### 7.2.2 Relevant strategies and statements

Table 7.1 highlights selected strategies and strategic documents released by the U.S. government identified as particularly influential through the literature and U.S. interviews.

*Table 7.1: Prominent U.S. strategies and documentation*

| | |
|---|---|
| Third Offset Strategy (2015) | The groundwork for DoD interest in emerging and disruptive technologies was laid partly by "The Third Offset Strategy" (Work, 2015). The strategy, delivered by Deputy Secretary of Defense Robert Work, outlined how the U.S. must invest in technological innovation to maintain its military advantage in conventional warfare capabilities (Hunter, 2017). Such motivations came from looking abroad, notably at China's investment in AI innovation and Russia's and China's use of AI in misinformation and deepfake technology (NSCAI, 2021). The Strategy recognised that AI "is expanding the window of vulnerability the United States has already entered" (Hunter, 2017).<br><br>The U.S. Third Offset Strategy aimed to mitigate the proliferation of emerging technologies that are perceived to be "blunting US military supremacy" (Fiott, 2017, 423) and widening the military-technology gap between the U.S. and its adversaries (Fiott, 2017*;* Lewis, 2017). AI represented a fundamental enabling technology for the Third Offset Strategy (Payne, 2018; MITRE, 2017). Researchers briefed by the DoD on the strategy summarise its critical elements, including autonomous learning systems, human-machine collaborative decision-making, human-assisted operations, manned and unmanned platform collaboration, and autonomous weapons capable of operating in future cyber and electronic warfare environments (MITRE, 2017 6). Fiott (2018) highlighted that the strategy lacked strategic clarity, meaning European allies may not be convinced by the need to innovate in emerging technologies within military contexts. |

| | |
|---|---|
| National Security Strategy (2017) | The National Security Strategy was "the first in history to specifically call out the importance of AI for the future of the American military" (The White House Office of Science and Technology Policy, 2018, 6). |
| National Defense Strategy (2018) | Committed to "investing broadly in military applications of autonomy, AI, and machine learning." (The White House Office of Science and Technology Policy, 2018, 6). |
| DoD AI Strategy (2018) – Announced 2019. | The U.S. was the first state actor to release a defence-specific AI strategy. The DoD published a public Executive Summary document for this AI strategy in February 2019 (U.S. Department of Defense, 2019). The DoD AI strategy commits to leading in military ethics and AI safety, listing plans that include consulting with the internet community on AI ethics, funding research into ethical AI, and sharing guidance that encourages responsible AI deployment by other states (Corn, 2019 5). The Executive Summary was analysed as part of this research and is discussed in section 7.3. |
| Executive Order 13859: "Maintaining American Leadership in Artificial Intelligence" (2019) | This Executive Order explicitly noted the need to protect American technical innovation from "attempted acquisition by strategic competitors and adversarial nations" (Trump Whitehouse Archives, 2019). Commitments listed within the Order include that the U.S.: drive the development of appropriate technical standards, develop public trust in AI technologies, train and recruit skilled AI expertise, and consider AI an R&D priority. The Order was not military-specific and did not call on the DoD specifically, focusing on broader requirements to prioritise AI R&D in the U.S. |
| DoD Cloud Data Strategy (2019) | Released as the JAIC was being built out, the Cloud Data Strategy highlighted the Center's required enterprise cloud infrastructure capability (DoD 2019b). DoD Cloud Strategy: "Data stored in an enterprise DoD cloud will be highly available, well-governed, and secure. Data will be the fuel that powers those advanced technologies, such as ML and AL." (2019b, 5, section 2.4). The document referred to the strategic landscape in which the DoD increases "readiness for AI" *(*DoD 2019b, section 1.4). |
| USAF AI Annex to the DoD AI Strategy (2019) | The DoD was followed by the U.S. Air Force Annex strategy (U.S. Air Force, 2019), which also outlined commitments and action points relating to AI technology in a military context. The document "places a priority on effectively and efficiently adopting AI across the service" (Wagemann, 2020, |

| | |
|---|---|
| | 22). The Annex document was analysed as part of this research and is discussed in section 7.3. |

The National Defense Authorisation Act (NDAA) for the fiscal year 2019 commissioned several items relating to AI, including the National Security Commission on AI (NSCAI). The NSCAI was set up to explore "the methods and means necessary to advance the development of artificial intelligence, machine learning, and associated technologies by the United States to comprehensively address the national security and defense needs of the United States" (U.S. Congress, 2018, H. R. 5515—329).[101] Set up as a temporary organisation from 2019 to 2021, the NSCAI published several in-depth reports culminating with its final report in March 2021.[102] While several arguments and recommendations from these reports are discussed further in section 7.4, a comparison of the first "Interim Report" (NSCAI, 2019) and the "Final Report" (NSCAI, 2019) demonstrates one example of how the analysis of military AI has become more detailed and nuanced over time. The NSCAI Interim Report (2019, 2) highlights the U.S. role in shaping AI technology, as the U.S. must offer "global leadership". The report states that U.S. superiority gives the U.S. "defense and security agencies access to the best technology and puts the United States in the best position to secure that technology against vulnerabilities and develop international norms and standards for responsible use" (NSCAI, 2019, 15). The report highlights the "urgent national imperative" to rapidly adoption military AI (NSCAI, 2019, 15). The report highlights how technology is almost consistently used to further power and security goals and calls for the U.S. to consider AI "through a military lens" in reaction to the actions undertaken by adversarial states. At 101 pages, it is one of the shortest and least detailed of the NSCAI's reports.

In comparison, the Final Report is set out across 756 pages (NSCAI, 2021). It goes into significantly more detail, highlighting developments in the international landscape and the implications should the U.S. not maintain military superiority. It underscores how AI will be used to pursue state power and repeatedly highlights the strategic AI competition between the U.S. and China (NSCAI, 2021). This report contains detailed recommendations and 66 pages of drafted legislative language. These include proposals to set up regional innovation networks across the U.S. for emerging technologies, streamline recruitment of technical experts, set up an AI task for governance and oversight, and certification standards (NSCAI, 2021).

[101] It was also through the 2019 NDAA that the RAND Report analysing the DoD's approach to military AI (see Morgan et al., 2020) was commissioned.
[102] The body ceased to exist in October 2021

Reading these reports and noting the various relevant U.S. DoD strategies released in recent years, sets out an understanding of the U.S. as a state that has become intensely interested in maintaining its competitive military edge in the face of global competition. The mindset displayed through these documents is that the U.S. does not shy away from military AI adoption and instead displays the view, as the NSCAI Final Report puts it, that "the U.S. government must embrace the AI competition and organize to win it [sic]" (NSCAI, 2021, 159). The following section considers how the U.S. has organised its approach to military AI through numerous agencies and initiatives.

### 7.2.3 Relevant actors/ initiatives in U.S. military innovation

Table 7.2 lists the most prominent bodies referred to in the scholarship and interviews relating to U.S. military innovation. The Joint Artificial Intelligence Center (JAIC) was the most frequently mentioned, followed by the Defense Innovation Unit (DIU) and Defense Innovation Board (DIB).

*Table 7.2: Relevant U.S. agencies*

| | |
|---|---|
| Joint Artificial Intelligence Center (JAIC). | The primary unit tasked with implementing the DoD AI Strategy is the Joint Artificial Intelligence Center (JAIC). The JAIC was created in 2018 to "seize upon the transformative potential of Artificial Intelligence technology for the benefit of America's national security" (JAICc, 2021). The JAIC has been described as the official "focal point" of the DoD AI Strategy (JAICc, 2021; DoDAI, 4) and was established by the DoD "in alignment" with the National Defense Strategy (USAFAI, 3). The JAIC's scope and responsibilities are outlined in the DoDAI document, with the promise of the JAIC's "important role in synchronizing DoD AI activities across all DoD components" (DoDAI, 10) with the DoD AI Strategy "executed" by the JAIC (USAFAI, 2). While the JAIC has launched and coordinated numerous projects in recent years, select JAIC-coordinated initiatives include:<br>• Setting up the Joint Common Foundation (JCF) to "democratize the whole process of DevSecOps for Artificial Intelligence/Machine Learning and make it easier to secure and rapidly authorize AI/ML capabilities" (JAIC, 2020). |

| | |
|---|---|
| | • The creation of the DoD AI Enterprise Infrastructure and Cybersecurity Subcommittee in 2020 (JAIC, 2020b).<br>• As a coordinator of AI adoption across the DoD in early 2021, the JAIC launched the Tradewind initiative, a procurement marketplace designed to facilitate public-private partnerships to bring innovative AI solutions for U.S. military purposes (JAIC, 2021c).<br>• Managing the Joint Enterprise Defense Infrastructure (JEDI) Cloud Program, set up to "provide the common data and infrastructure platforms that will enable Al to maximize warfighter advantage". (Corn, 2019, 13).<br>Throughout this research, the JAIC was highlighted through the literature and reaffirmed through interviews as the most prominent relevant DoD agency for military AI adoption in the U.S. |
| Defense Innovation Unit (DUI), established in 2015 | The Defense Innovation Unit (DUI) is exclusively tasked with "fielding and scaling commercial technology across the U.S. military" (Defence Innovation Unit, 2021). It was founded "because traditional approaches to defense acquisition were failing to capture innovative applications available to global consumers but out of reach for the U.S. military". (Defense Innovation Unit, 2021b). In February 2020, the U.S. DoD adopted five AI principles almost verbatim from the DIU's 2019 recommendations (DoD, 2020d; Vergun, 2019). In doing so, the DoD formally outlined that deployed AI must be responsible, equitable, tracible, reliable and governable (DoD, 2020d). |
| Defense Innovation Board (DIB), established in 2016 | The DIB represents an independent advisory board (Defense Innovation Board, 2021c) that advises and provides recommendations to the DoD on key focus areas, including AI and digital modernisation. The DIB released a report on AI Principles in 2020 (Defense Innovation Board, 2020). The DoD adopted its five proposed principles shortly after, as noted above. In September 2021, the DIB released an information sheet urging the DoD to develop their own Artificial Intelligence Test, Evaluation, Validation & Verification (AI TEV/V) mechanisms (Defense Innovation Board, 2021b; Jasper, 2020). |

| | |
|---|---|
| AI and Data Acceleration ('ADA') Initiative (2021) [103] | The ADA Initiative aims to improve tactical AI in the military's combatant commands (Gill, 2021) and focuses specifically on improving data management and adopting an interoperable array of AI tools to provide enhanced decision-support, asset-tasking, sensor and mission autonomy. (Hicks, 2021). The Initiative explicitly supports the U.S. DoD's Joint All Domain Command and Control Strategy (JADC2) announced in May 2021, setting forth a roadmap to enable U.S "leaders and warfighters to orient, decide, and act faster than our competitors...[by] providing operational commanders with data-driven technologies, including artificial intelligence, machine learning, and automation". (Hicks, 2021, para 1). |
| Chief Digital and Artificial Officer CDAO position and supporting office (OCDAO) | A December 2021 Memorandum announced the creation of the CDAO role to support the DoD in becoming "a digital and artificial intelligence (Al)-enabled enterprise capable of operating at the speed and scale necessary to preserve military advantage" (Hicks, 2021b, para 1). A February 2022 memorandum formally sets the OCDAO as the successor organisation for the JAIC and will take on the broad role of managing and approaching AI adoption for the DoD (Hicks, 2022). The CDAO is expected to achieve initial operating capacity in June 2022 (Hicks, 2021b; Hicks, 2022).<br><br>The December 2021 Memorandum was the first indication that another organisation would succeed the JAIC. At the time of submission, the JAIC is still active on social media and its home page/ "about" section does not mention details of how its responsibilities might be transferred to the OCDAO. |

A broad range of U.S. government departments will be involved in enabling military AI capabilities. An analysis of Q2 2020 funding on military AI showed that while the bulk of the funding was from the DoD Research, Development, Test, and Evaluation (RDT&E) funding, additional tasks relating to education, export controls, and strategic alignment were funded by cabinet departments and central agencies (NSCAI, 2020b, 184). Involved agencies include the National Science Foundation, the Department of the Treasury, and the Department of State (NSCAI, 2020b, 184). Analysis a quarter

[103] This initiative was announced in June 2021 after the conclusion of the U.S. interviews conducted for this research.

earlier also includes the Department of Energy, National Aeronautics and Space Agency, Office of the Director for National Intelligence, and Department for Homeland Security, among others, as funders of AI activity relating to national security (NSCAI, 2020c, 77).

### 7.2.4 The U.S. and "Responsible AI"

Beyond promoting the rapid adoption of AI-enabled capabilities, the U.S DoD has also been outspoken about AI "ethics and safety".[104] The DoD formally adopted the DoD ethical principles for AI in 2020, to state that military AI must be deployed and used in ways that are responsible, equitable, traceable, reliable and governable (DoD, 2020d; Vergun, 2019). In May 2021, the DoD issued a memorandum on "Responsible Artificial Intelligence", or "RAI" (U.S. Department of Defense, 2020b), which reiterated the Department's commitment to the five principles. The memorandum also confirmed the JAIC as the DoD's coordinator on RAI policy and guidance. The memo stated that the Director of the JAIC will "develop, assess, and report on the implementation of a DoD RAI ecosystem" (U.S. Department of Defense, 2020b). The memo further outlined that the JAIC Director, with the "RAI Working Council", would coordinate: RAI Working Council & Training; the RAI Strategy and Implementation Pathway; RAI Talent Workforce Management; and RAI Acquisition (U.S. Department of Defense, 2020b*).*

> "By leading in military ethics and AI safety, we reflect our Nation's values, encourage Responsible AI (RAI) development globally, and strengthen partnerships around the world". (U.S. Department of Defense, 2020b)

The U.S. has engaged with the international community on responsible AI, in part through the AI Partnership for Defense (PfD). The PfD is a 16-state strong consortium of states invited by the U.S to "promote the responsible use of AI, advance shared interests and best practices on AI ethics implementation, establish frameworks to facilitate cooperation, and coordinate strategic messaging on AI policy" (AI Partnership for Defense, 2020, 1).[105] The PfD includes NATO (France, Canada, Estonia, Denmark, the U.K., and Norway) and non-NATO countries (the Republic of Korea, Australia, Sweden, Finland, Israel, and Japan). In addition, the U.S. has committed to the NATO AI strategy, which includes ethical principles that are consistent with the DoD's approach.

[104] This is the terminology used in the DoD strategy and USAF Annex for the section discussing AI ethics and AI security.
[105] The PfD increased from thirteen to sixteen members at the consortium's third meeting in May 2021: https://www.ai.mil/news_05_28_21-jaic_facilitates_third_international_ai_dialogue_for_defense.html

Acknowledging the technical requirements for responsible AI, in May 2021, the JAIC's TradeWind model launched a Request for information on "Responsible AI Expertise, Products, Services, Solutions, and Best Practices". The request called for RAI expertise to help the DoD design, develop and deploy AI capabilities in a "safe, trustworthy, and responsible manner" (JAIC Public Affairs, 2021). In July 2021, it was announced that through Tradewind, the JAIC was going to "pilot a responsible AI procurement process*"* (JAIC Public Affairs, 2021b).

*NOTE K: Industry willingness to engage in military AI innovation*
"Project Maven" (more formally termed the Cross Algorithmic Warfighting team) was created in 2017 to rapidly test AI applications in a military context for deployment in intelligence, surveillance and reconnaissance (ISR) processes. In partnership with Google, the programme aimed to speed up the innovation and procurement process - and to a large extent, was successful. Within two years, the project had delivered image recognition capabilities to the U.S. DoD. The controversy around Project Maven was reported in international media and resulted in an "unprecedented wave" of technology workers dissenting about their employers' involvement with military projects (Whittaker et al., 2018, 10). Project Maven had contracted Google to work on the project, which brought international media attention once senior employees at the technology company expressed their anger at the technology firm partnering with the DoD (Simonite, 2018). Google did not renew its contract with the Pentagon and subsequently adopted a set of ethical AI principles in late 2018 (Pichai, 2018), the first major technology company to commit to AI principals (Morgan et al., 2020). Pichai's guiding principles included a section on "AI applications we will not pursue" and references weapons and technologies for which the principal purpose is to harm others, which Whittaker et al. (2018) argue is a direct response to the DoD contract non-renewal. While other organisations followed suit in setting up AI governance initiatives, as outlined in Chapter 2, it is unclear whether Google's AI principles have prompted a trend of responsible AI commitments across the U.S. commercial technology sector (Morgan et al., 2020). In technical terms, Project Maven was viewed as a success, with targets on rapid prototyping beaten and the time from proof of concept to in-the-field deployment described as *"*literally a work of magic*"* by research directors at the Center for Strategic and International Studies (Simonite, 2018, para 7). Project Maven was cited multiple times in the literature as an example of how the U.S. could rapidly innovate and field AI technologies (see Buchanan, 2020; Konaev et al., 2020; Davis, 2019).

This section has introduced some critical aspects of the U.S.'s approach to military AI. Highlighting major initiatives, relevant bodies and strategic milestones such as the release of the DoD AI strategy, this section has also noted that beyond selected examples, many related strategies and initiatives are contributing to AI adoption and national security in the U.S. The result is a complicated, shifting landscape that has continued to develop rapidly through this research.

## 7.3. Themes: The U.S. Approach to Military AI

Analysing the data from the seven interviews, the DoD AI Strategy Executive Summary, and the USAF AI Annex to the DoD AI Strategy revealed five themes around which discussion was centred. The first and most heavily discussed theme[106] was "DoD Coordination of military AI innovation", as outlined in section 7.3.1. This theme reflects on the U.S. DoD's perceptions, approach, active programmes and achievements, and peripheral U.S. state mechanisms involved in military technology development. Second, section 7.3.2 presents "U.S.' position in relation to other actors in the global landscape", which captures analysed data relating to allies, adversaries, and potential non-state partners, including industry, academia and international structures such as NATO. Third, section 7.3.3 explores the theme "perceived impact of military AI technology for the U.S." and highlights where strategies or interviewees spoke about the nature of AI in warfare, including use-cases and the relationship between AI technology and human-machine teaming. The fourth theme, "countering irresponsible use of military AI" in section 7.3.4, relates to data on AI and technical security and conversations around AI ethics and safety, including conversations on norms and legal mechanisms to prevent harm from AI tools. Finally, the fifth theme, "improving levels of understanding in AI technology," in section 7.3.5, reflects on hype and the AI talent gap and captures how far U.S. policymakers and military decision-makers understand current capabilities.

Interviewees were invited to participate based on their relevant professional experience with the DoD. The earlier interviewees provided recommendations on relevant subject matter experts to interview. Interviewees spoke under conditions of anonymity and understood their responses would not be attributed to them personally or by their role titles.[107] Through the findings section, quotes are affiliated to: 'DoDAI' (the released U.S. DoD AI Strategy Executive Summary), the USAF (the United States Air Force AI Annex to the DoD Strategy), and the seven interviewees as individually numbered (C1, C2...C7).

### 7.3.1 DoD Coordination of military AI innovation

The "tremendous utility of AI" (DoDAI) was recognised across the Department as a way to "ensure an enduring competitive military advantage" (DoDAI), with all interviewees agreeing that AI was

[106] For more information on how themes emerged from qualitative analysis see Chapter 3
[107] For more information on interviewee selection see Chapter 3.

receiving significant attention and funding overall. With the DoD posture "more, kind of, positioned through a lens of conflict, than through diplomacy" (C3), the DoD is a mission-based organisation that focuses on "the ethos of warfighting" (C6). This focus influences how the Department views and approaches AI compared to other U.S. state infrastructure, with the DoD often "more conservative because the stakes are higher" (C7) while viewing AI as "a good strategic use of money". Interviewees frequently mentioned the DoD's recognition that as the organisation "responsible for building some inherent capabilities in this space" (C4), they must become "more organised internally about how we're thinking about this" (C6).

> C2: "I think that we have some very hard-headed people in positions running artificial intelligence and think they are going to keep running at the wall until they break it down."

Clear direction on U.S. military AI coordination was cited as a significant motivator for creating the DoD AI Strategy in 2018 and the corresponding public release of the DoD AI Strategy Executive Summary. The DoD AI strategy, and the corresponding public release of the strategy's Executive Summary, were seen as a response to the recognised need for a military roadmap for AI.

> C6: "We had all these disparate efforts that were going across the government, across the Department, from a research and development perspective that were simply not tied together, now were they transcending what we affectionately called the Valley of Death."

Interviewees perceived a "stressed" (C3) DoD that saw emerging technology breakthroughs that had DoD staff realise the need for "a strategy not only to figure out how to develop the technology and how to use the technology - but also how to access the technology" (C3). A strategy represents a "starting point" to think strategically about AI investments, deployments, and implications on civil liberties (C6). The DoD reiterated this broad scope in the executive summary, recognising the U.S.' need to "adapt its culture, cultivate new skills, and streamline approaches to develop, attract, and partner with AI talent" (DoDAI, 14).

Speaking about the nature and achievements of the DoD strategy, some interviewees viewed the strategy as "reasonably influential" (C1) in terms of coordination and formally providing the U.S. approach to military AI. While it was considered "relatively usual when a technology appears to be emerging rapidly, for the Department to do this [develop and release a strategy]" (C3), this does not detract from the importance of the strategy in providing an important reference point for DoD staff.

> C3: "Because you have so many millions of people working there [the DoD] and because it is such a strict hierarchy, every Powerpoint Presentation starts with "here's the national strategy" .... no one would ever talk about what they're doing and ask for money without in fact placing themselves in the context of that strategy".

The strategy was viewed as useful for the DoD themselves, with the document's purpose formally to direct the DoD "to accelerate the adoption of AI and the creation of a force fit for our time" (DoDAI, 4) - with a publicly available executive summary being useful as, among other things, a demonstration of transparency (C3). The document itself describes four strategic focus areas: delivering AI-enabled capabilities that address key missions partnering with leading private sector technology companies; academia, and global allies; cultivating a leading AI; and workforce and leading in military AI ethics and safety (DoDAI). The consensus among interviewees was that the document contributed toward the broader U.S.' "explicit goals is [sic] to shape the norms of the international community" (C1).

Interviewees noted the importance of the JAIC in putting in place "institutional structures to enable" required reform to innovation (C1). Interviewees also noted that the JAIC created an environment that encourages innovation, including through the creation of the Joint Common Foundation (JCF), a development platform for ML software, which addressed the "specific technical aspects that make AI development hard" (C1).

Interviewees were generally positive about the JAIC: its' establishment "sends a strong message" that the DoD wants to invest in AI for defence (C3), with the acquisitions process "getting better all the time" through the JAIC's conceptual and prototyping vehicles such as Tradewind (C4). As a relatively new entity, the JAIC has "taken a lot of punches to the face" (C6) and taken criticism for not living up to the "fanfare" with which it was initially announced (C6). For example, the JAIC has faced

structural challenges in coordination in "trying to implement AI when they don't own the system… the services are the one that owns and maintains all the equipment" (C2). While a valuable focal point for AI coordination, the JAIC was not considered perfect and was accused of having "absorbed all of the oxygen in the room" by focusing on near-term incremental development, detracting attention from necessary early-stage and prototyping research and development (C3).

C7: "There is a huge difference of opinion about how the JAIC should and shouldn't operate."

The list of AI initiatives relating to U.S. defence is expansive and ever-changing. While the JAIC is the designated DoD coordinator of military AI activity, many other agencies and organisations have active AI-related initiatives, some preceding the JAIC's existence.[108] While interviewees highlighted their awareness of some oversight mechanisms within the defence procurement process, these may well have been complemented or replaced by new efforts. For example, an August 2021 government Memorandum announced and tasked two federal watchdogs to assess the use of AI in intelligence for national security (U.S. DoD Office of Inspector General, 2020). Broader AI innovation may also impact military contexts, with NASA's research being feasibly relevant, for example, through their research on autonomous systems (C3), NIST and standard-setting bodies. Other parts of U.S. state infrastructure are also running AI programmes, including the Department of State and the Department for Energy.

Given the breadth of actors involved in AI defence, it is perhaps no surprise that a major theme, from interviews and both policy documents, was the DoD's attempted approach to acquiring AI technology to avoid duplication of existing efforts. The current procurement process was viewed as a "crisis" (C1) and "overly Byzantine" (C4). One interviewee described how the processes for a private company to engage with the DoD were so complex that the DoD missed out on significant opportunities (C1). At the same time, specialised defence companies secured DoD contracts based on "their ability to navigate the bureaucratic process of defence procurement… not superior capabilities" (C1).

[108] A non-exhaustive list of bodies and vehicles for AI research include the Army AI Center at Carnegie Mellon, the Air Force Research Laboratory (with the Air Force Office of Scientific Research), the Office of Naval Research and DoD's Defense Advanced Research Projects Agency (DARPA), the Defence Intelligence Agency, Defense Innovation Board and Defense Innovation Unit. As one example, DARPA has a number of projects focusing on artificial intelligence technologies which can be explored via https://www.darpa.mil/tag-list?tag=Artificial%20Intelligence

C1: “Part of the great challenge for the DoD is, like, how do we make sure that the companies with the best stuff actually win… not just the company that’s the best at filling out forms or lobbying Congresspeople”.

Encouraging private sector engagement was seen as key, particularly with the non-traditional defence industries innovating heavily in the AI space (Cummings et al., 2018). The DoD is attempting “to bridge the relationships between the defence industrial base and these other companies” while facing the challenge that “none of those non-DoD people know how to make anything the DoD can use” (C3). Many DoD procurement mechanisms focus on long-term hardware acquisitions; the overall procurement process “is not designed to incentivise AI, or software” (C6). The Tradewind platform (announced in early Summer 2021 in time for the final two interviews) was praised by both C3 and C6 for the efforts to provide access to key DoD decision-makers in a relatively straightforward manner.

The acquisition activity in the context of the DoD’s structural constraints is one example in which the DoD’s approach to military innovation was described as inherently problematic. The number of existing acquisition policies across the DoD “make anything hard” (C1) for the relatively new JAIC, which was tasked with facilitating collaboration in a highly complex and overlapping environment. The pace of AI innovation was highlighted as a significant challenge, as “the structures supporting U.S. military innovation appear to be poorly suited to software in general” (C1). This shortfall often leads to the “valley of death” (C3), the concept in which ideas fail to progress beyond prototypes through to final stage applications (Ford and Koutsky, 2007).

C6: “The biggest problem with the Department of Defense is just by virtue of how it's organised and how enormous it is”.

Moreover, attitudes to AI innovation ranged broadly, with different development communities advocating for traditional systems engineering, waterfall style planning, versus agile development and “SecDevOps” approaches (C2). These tensions provided uncertainty to the acquisition community, who “don’t have the proper background… and are looked at as being obstructionist if they can’t do things fast” (C2), creating an environment in which projects were inherently

disadvantaged. Challenges around the AI talent gap (as discussed under Improving levels of understanding in AI tech, below) and procurement (as discussed above) also featured heavily.

Many of the challenges associated with the structural nature of DoD innovation were linked to the Department's culture across U.S. military decision-making. Discussing how to revolutionise the Department's approach to military AI, interviewees believed that distinct cultures between organisations, particularly where each organisation had different goals, prevented innovation. C6 highlighted how the authorities given to the JAIC were a disadvantage for the Center, with other DoD services and components "fearful that the JAIC would ultimately start telling them what to do… that the JAIC would ultimately take big portions of their budget" (C6).

> C6: [AI Progress] "would take the JAIC and the CDO [Chief Data Officer] and each other and other components like that to stop fight [sic] with each other and just work together on things that would be meaningful, and that's very difficult for a lot of grown men who've grown up fighting for their careers. [laughs]"

Moreover, the approach to innovation and risk was often highlighted as a challenge. According to one interviewee, the DoD's "obsession" with efficiency "has destroyed innovation…. we're so much more focused on one-size-fits-all approaches that we've stopped using our brains" (C3). This tension was noted specifically when different levels of management have different approaches to driving AI innovation, causing "confusion or just inertia problems" (C7). To change the overall approach to innovation, the Department needs "change champions" across the departments and agencies, with C7 arguing that "if you just push it [responsibility for change] off on the JAIC... the Department isn't going to make the structural and cultural investments and changes in the workforce, the Department.... for the DoD to be AI ready in 2025".

There was no clear consensus on which military branches were dominant overall regarding AI innovation. Within the military, different branches were highlighted as having different attitudes and approaches to innovation, with heterogeneous needs. C2 regarded the U.S. Army as being particularly proactive concerning AI, operating an AI Centre and taking a holistic and strategic approach, while the Air Force and Navy looked at more applied "embedded tech solutions" (C2). C6 felt that the Air Force was "particularly good at innovating and breaking down bureaucracy and, like, pushing the development and adoption of AI… with other branches, they are kind of upheld as, like, a pretty good

example" (C6). Approaches to research skill sets were also described as being different between branches, with generalists self-selecting for the Army, where technical personnel may self-select into more engineering-style positions "so that the Air Force and the Navy have more technical expertise" (C2).

The DoDAI document acknowledges the "innovative character of our [US] forces". It acknowledges that much innovation will arise "from experiments at the 'forward edge'... by the users themselves in contexts, far removed from centralized offices and laboratories" (DoDAI,7). The different environments may lead to tensions where technical personnel see an opportunity to use AI, but leadership "don't want to give ownership to something other than themselves" and prevent innovation (C2). Suggesting broader tensions between the Army and the DoD, the JAIC was described as a DoD and non-military Center "in many regards, the outside entity, trying to ask the military to work in a certain way" (C2).

> C2: The JAIC kind of works through the Army's AI Center... there's going to be less resistance in doing things in the army because the army's gonna say they did it through the Army AI Center, as opposed to the JAIC".

The NSCAI also acknowledges this tension and recommends that the JAIC director "remains a three-star general or flag officer with significant operational experience who reports directly to the Secretary of Defense or Deputy Secretary of Defense" (NSCAI, 2020, 83). This recommendation faces the challenge, as expressed by C7, that few three-star generals are relatively well informed on AI and emerging technologies to make them well-placed for the role.

Overall, a broad range of significant open questions face the DoD and senior decision-makers considering AI for defence. Not all of these are defence specific. The DoD employs almost one million personnel, uses a vast and layered ecosystem of legacy systems, and experiences organisational bureaucracy across its expansive infrastructure. As observed by interviewees, barriers to effective innovation were structural, social, technical, and organisational, with the DoD "only as fast as its slowest process" (C6). Interviewees often proposed their own concerns as questions: referring to the combination of acquisition shortfalls and the lack of informed decision-making on military AI applications, "why would you ever expect to trust an AI that's being developed by these people you don't trust?" (C2). Even when well-intentioned, the question of how the U.S. should best

prioritise investment has no clear answers, with C3 questioning that when the DoD has a $718.3bn budget overall (U.S. DoD, 2019c; Mcleary, 2020), "what's it worth… do you feel better? I mean, are you paying to feel better, you paying $700billion a year because you want oversight of every single programme so you feel better?" (C3). Other interviewees raised questions of how the DoD - and branches of the U.S. military - should best develop the required skill-sets across personnel to best drive effective and responsible innovation, while "at a much more technical level we [the DoD] still have a lot of issues just to work through with artificial intelligence more generally" (C3), citing open the need for increased testing and research on how humans work with, and rely on, AI technologies.

### 7.3.2 The U.S. position in relation to other actors in the global landscape

A dominant theme emerging from the interviews was that the U.S. had significant signalling power internationally, with every interviewee referring to international perceptions of U.S. statements and activities. In C1's view, "among all militaries... the unclassified version of the strategy is a marketing document and that the real strategy is the super-classified one". C6 referenced the DoD's perception that "there needed to be some messaging" to adversaries formed part of the motivation for the DoD AI strategy. The DoD AI document is not subtle about the U.S.' wish to explicitly signal on topics of responsible AI, stating the intent to share "aims, ethical guidelines and safety procedures to encourage responsible AI development and use by other nations" (DoDAI, 8).

> C4: "We [the DoD] needed an unclassified version [of the AI Strategy] for that signalling aspect and a classified version for that understanding of what we actually need to do here".

More generally, C4 described the DoD's language on accountability and AI in autonomous weapons to set out the U.S. view that AI technologies still fall under the law of armed conflict, with the U.S. setting out ethics frameworks as "potentially signalling devices". Signalling also has a domestic purpose, with C1 repeatedly referencing the U.S.' wish to emphasize their approach "to our citizens and to our allies". C3 described Project Maven as "the first strong signal" that the DoD was "going to work with non-traditional people to actually no-kidding solve the problem", with procurement initiatives like Tradewind another signal to potential industry suppliers. C4 highlighted a distinction between the use of public announcements to signal a perspective or intent and for which they saw evidence through DoD statements on responsible AI and signalling in terms of an actual investment.

No interviews argued that DoD investments and interest in the technology was driven by signalling; the military was described as "fairly pragmatic… we [U.S military] are looking at any technology to leverage military advantage" (C4).

Interviewees perceived the U.S.' desire to leverage its' power in military and strategic terms, and almost every transcript referenced international competition. Adversarial investments in technology "threaten to erode U.S. military advantage" (DoDAI, 17) in the context of "this return to great power competition" (USAF, C2).

> USAF, 2: "The comparative advantage currently enjoyed by the Air Force will either erode or strengthen depending on the matter in which we adopt these technologies".

The U.S.' Third Offset Strategy (Work, 2015) and Work's contribution more generally to U.S. doctrine and defence policy environment were raised as influential when approaching military AI. C1 quoted Work's view that autonomy would be "critical to the future of military power". At the same time, interviews commented that the DoD AI strategy was "an evolution" of the Third Offset strategy (C2) and that Work's work on the National Security Commission on AI was "to underscore and kind of, you know, drive a lot of the original thoughts that he had" (C6). With the Third Offset creating a "force for the future" (C7), interviewees referred to the Third Offset strategy as an implicit prelude to the DoD AI strategy.

It was also widely recognised that the U.S. cannot act unilaterally regarding the military AI landscape. The state would need to collaborate with both industry and other states. Interviewees spoke about the need to "leverage our respective strengths" as nations (C4). Cohesion with allies and partners was repeatedly stressed on the issues of normative frameworks, with the example of a NATO-wide framework being "much more powerful" compared with a national U.S. framework (C4). Several other nations have proactively acted concerning AI norms, with C7 highlighting the U.K., French, and Canadian commitments to the Global Partnership on AI (not defence-specific) and G7 discussions. It was considered "naive and productive to try to usurp the leadership roles that all the other countries have played... this [Biden] administration realises it needs to play a leadership role but doesn't want to undercut or undermine" other actors (C7).

DoDAI, 4: “We cannot succeed alone”.

The U.S. position on international engagement is not static. The transition from Trump to Biden administrations saw changes in how the U.S. engaged in multilateralism within the context of security.

C7: “I started when it [relevant project] was under the Trump administration and I ended under the Biden administration and the difference in the way other foreign governments spoke to us, it was noticeable, definitely.”

Interviews took place through both Trump and Biden administrations. The Biden administration was seen to “actually believe [in] the need to reinvigorate - that’s the actual terminology they use - reinvigorate multilateralism” (C7).

C4: “The DoD rightly understands that they don’t have all of the expertise… there’s no reason why the U.S. should do all of this alone”.

NATO was mentioned explicitly in four of the seven interviews and was not named in the DoDAI or USAFAI document. A broader phrase, “allies and partners”, was referenced frequently across the data, occurring ten times within the DoD AI executive summary. As an Alliance, the interviewees described NATO as one option but with caveats, with existing issues of insufficient interoperability (C2). NATO was described as “starting from square one” (C1) when it came to an AI framework, with many members yet to demonstrate an interest in AI (C1), which may make it difficult to agree on prioritisation and any common approach.[109] The struggle to achieve consensus on themes such as AI represents a broader identity struggle highlighted by interviewees. For example, C3 critiqued NATO as an organisation struggling to remain relevant against a common enemy in the modern era, and as an organisation that is now too “large and diverse” ... “it would be better pairing down to a smaller set of capabilities”. There was evidence that the U.S. viewed NATO as a potential mechanism to shape international norms rather than as an entirely co-creation environment between allies. C7 highlighted that, Trump administration aside, the U.S. “is very active at NATO, is really trying to shape NATO’s EDT [Emerging Disruptive Technologies] strategy”.

[109] The U.S.- focused interviews took place from October 2020 (C1) to August 2021 2020 (C7) with all interviews completed before the NATO AI Strategy was agreed in October 2021.

C1: “NATO’s military AI capabilities are basically non-existent, and the United States has a very strong incentive for a very strong NATO. And so our goal is to persuade them that this is important”.

The Five Eyes (FVEY) network was described as a more straightforward structure to work with because of a more “intimate and deep” trust (C1). The network also cooperates with the same “extremely high level of trust” (C3) through The Technical Cooperation Programme (TTCP), specifically on science and technology and defence, with an initiative “TTCP AI Strategic Challenge” as a specific active programme as of 2021 (C7). Additionally, the US-hosted AI Partnership for Defense was cited as “one of the most significant and least talked-about wins that the JAIC has had” (C6). The NSCAI report also recommends that NATO be connected with the Quadrilateral Security Dialogue to bring in India, Australia and Japan (NSCAI 2021).

Adversarial actors were also brought up more as a motivation for U.S. attention and investment in military AI innovation, reflecting “the need to articulate ways that they [the U.S.] intend to compete with the Chinese as a near-peer adversary” (C2).

DoDAI, 17: “Our adversaries and competitors are aggressively working to define the future of these powerful technologies according to their interests, values and societal models.”

The USAF document notes that U.S. adversaries “live in top-down command economies” (USAF, 6), through which the adversary state enjoys greater access to data and technologies in their markets and “millions of publicly accessible algorithms” (USAF, 5). Interviews reported perceptions that China could use AI to “leapfrog” the capabilities the U.S. had traditionally led on (C7). C4 also believed that non-state actors are a significant challenge that has not yet been addressed, with the “democratization” of technology making malicious activity more accessible, with deep fakes “the tip of the iceberg” of emerging AI-enabled threats.

The U.S. DoD AI document states that “strong partnerships are essential” (DoDAI, 7), and interviewees affirmed the strategy’s claim that “U.S private sector and academic institutions are at the forefront of modern AI advances” (DoDAI, 12). The U.S. government's in-house capability was viewed as “heavily rolled back” (C5) and “fraught” (C4) compared to 30 years ago, with the DoD no longer a market-maker in 2020 for the most advanced technology firms. While large defence

contractors still service the DoD, with C1 pointing out the 70% of DoD suppliers that exclusively serve the DoD, the U.S. needs to attract non-traditional companies "without forcing them to sacrifice what makes them special" (C1).

> C3: "For the first time in the history of, sort of World War Two, the United States, the Department of Defense, now needed an industry where it was a minority customer".

The wish to bring in new talent includes venture capital initiatives, with C7 highlighting the "huge" role of non-for-profit venture capital In-Q-Tel in servicing the U.S. intelligence communities.[110] Academia and the need for broader research were also mentioned in every interview and both documents, with the need to research longer-term and recognise that these skills must be accessible to the DoD and public sector.

### 7.3.3 The perceived impact of military AI technology

Interviewees and both strategy documents agreed that AI was posed to change the nature of warfare. This language was often connected to the need for the U.S. to rapidly innovate to hold and maintain a competitive edge in AI and best realise the potential enabled by the technology. Both strategy documents press this point heavily; AI is "posed to change the character of the future battlefield and the pace of threats we must face… the costs of not implementing this strategy are clear" (DoDAI, 4). The DoD must "adapt its culture, skills and approaches" to "lead the world in the development and adoption of transformative defense AI solutions that are safe, ethical, and secure" (DoDAI, 17). The USAF document takes an almost-identical approach, viewing AI as a cause of "exponential transformation of the entire spectrum of human life and experience" (USAF, 6). The document compares military AI potential to the development of stealth aircraft and precision-guided munitions as technologies that changed warfare.

[110] The IQT homepage states the organisation's mission to "invest in cutting-edge technologies to enhance the national security of the United States". The organisation states its focus in over 15000 early stage venture backed start-ups in the U.S. and 'select other countries', while also analysing emerging technologies that are assessed as critical for national security, For further info see https://www.iqt.org/.

> USAF, 6 (Quoting The U.S. Air Force Cross-Functional Team on Artificial Intelligence, 2018): "There are a lot of analogies in our history to AI and the changes it will bring, but no corollaries. We've had language, learning, industrial and technological revolutions before, but nothing that will so pervasively integrate all aspects of our lives and change everything about the way we interact with one another."

AI and autonomy contribute to the "speeding up" (C4) warfare with the potential to automate aspects of warfare decision-making. C4 referred to the "fundamental cross-cutting nature" of AI that can be used not only in warfighting but also in business processes and logistics. The nature of AI algorithms means that the threat landscape may face "democratisation" once AI tools are "out there and now anyone with very little skill can use it" (C4). C7 reported that an inspiration for the DoD strategy was the growing recognition that "AI, machine learning, are going to revolutionise war fighting and have already revolutionised competition - geopolitical competition and that the U.S. really needs to get its act together". Similar statements are present in the Third Offset strategy, which highlighted that autonomy would be critical to the future of military power, with modern AI technology a key enabler of autonomous systems (Work, 2015).

Interviewees raised several specific ways through which AI technology would transform warfare. C1 compared AI technology to precision-guided missiles to argue that ethical standards would increase in warfare to limit collateral damage to non-combatants, as "actually our [ U.S. military's] ethical standards for ourselves increase as our technological performance goes up". The DoD summary demonstrates the intention to use AI to enhance the implementation of the Law of War, highlighting how AI can be used to both "reduce risk to fielded forces" and "reduce the risk of civilian casualties and other collateral damage" (DoDAI, 6).

Looking at identified and near-term use cases for the U.S.' use of AI in the military, moving into specifics also shifted the focus to enterprise-level and back-office AI applications. C2 felt that the Navy and the Airforce had focused on embedded "service-centric" applications innovating in their technical groups, including aircraft designs, predictive maintenance, and logistics. C3's comments support this assessment, highlighting how logistics and maintenance held significant promise, as did optimising human resources and other enterprise processes. C3, C6 and C7 believed that enterprise-level AI had promises that deserved more attention than the "hot topic" (C7) of autonomous

weaponry. In addition to the support applications mentioned above, C6 highlighted how finance systems and transportation held significant opportunities as 'low-hanging fruit" that could also advantage military mobility. The use of AI as a "sensing and judging" tool that will assist in a broad range of intelligence functions (C7), including capabilities deployed to "predict, identify and respond to cyber and physical threats" (DoDAI, 6). The DoD AI strategy repeatedly highlights military AI capabilities' potential applications, highlighting enterprise level functions including situational awareness, logistics, and streamlining business processes (DoDAI). The USAF Annex has slightly less detail but highlights how AI will "underpin our ability to compete, deter and win" (USAF, 2).[111]

Several interviewees highlighted that many use cases that are being actively pursued are unlikely to be known about by those without clearance; as C6 observed that agencies within the DoD "do actually have some really compelling use cases and interesting applications right now, most of which are classified… people are not necessarily going to hear about them, but I think that they're making gains quietly".

One frequently highlighted aspect of AI in the military was how AI will augment staff capabilities and how staff will interact with AI technologies through human-machine teaming. C7 highlighted that while discussion disproportionately focused on autonomous weaponry, the DoD was dedicating resources to human-machine teaming capabilities, such as collecting and analysing intelligence or impacting military mobility. Trust was raised as a crucial component, as C3 argued, "the worst thing that can happen is that the algorithm screws up or degrades or changes a behaviour". The risk of overwhelming or misleading military personnel was also raised as C4 highlighted the scale of information that is now available:

> C4: "You still run the risk of, kind of cognitive overload, right? If you're getting hundreds of, tens of thousands of data points coming into a commander, can a commander really make a fully accurate decision, right?"

Oversight and accountability structures were also raised as key, with C7 recommending "meaningful human oversight" over AI-enabled weaponry. The DoD strategy highlights how AI will "empower,

[111] Full quote: "The Air Force is charged to provide the nation with Air and Space Superiority, Global Strike, Rapid Global Mobility, Intelligence, Surveillance and Reconnaissance, and Command Control. AI is a capability underpin our ability to to compete, deter and win across all five of these diverse missions." (USAFAI)

not replace, those who serve" (DoDAI, 4), with the DoD employing AI in a "human-centered manner" (DoDAI, 6). This language corroborates the interviewees' general sentiment that some critical functions will always employ human-in-the-loop AI, with humans informed and capable of overriding AI tools (C2, C4, C7). The DoDAI executive summary references DoD Directive 3000.09 as existing guidance on autonomous weaponry (DoDAI, 15).[112]

### 7.3.4 Military AI ethics and safety

The executive DoD AI Strategy repeatedly highlights the topic of responsibility and ethics though it does not explicitly distinguish between technical security and ethical questions.[113] The strategy has a combined section titled "Leading in military ethics and AI safety" (DoDAI, 15-16), covering topics on securing AI systems and designing ethical principles, and noting the technical, ethical, and social challenges posed by AI. Nonetheless, analysing the transcripts and DoD/USAF documents through descriptive coding revealed that the discussions could be split into two categories: (1) securing AI systems at the operational and theoretical level and at a higher strategic and state-policy level, (2) laws, ethics, and norms.

#### *7.3.4.1 Securing AI Systems*

The topic of security and AI spans many different subtopics, threats and techniques. C1 distinguished between the vulnerabilities of machine learning that are identical to traditional software (e.g. by the application being internet-connected) and the security vulnerabilities specific to machine learning (i.e. should an autonomous system be compromised). In line with comments made through the U.K. interviews, the interviewee expected attackers to target traditional (not machine learning specific) vulnerabilities for the near-term future due to competence levels but still held concern about the increased threats facilitated by AI vulnerabilities. For example, C4 highlighted the possible risks of adversarial attacks against AI capabilities, the brittleness of AI technology in performing best under a very narrow set of circumstances (an argument neatly summarised by Ciocca & Kahn, 2020), and the risks in protecting the datasets and in securing genuinely representative data (C4).

---

[112] For further information on DoD Directive 3000.09 see Saxon, Dan. "A Human Touch: Autonomous Weapons, DoD Directive 3000.09 and the Interpretation of 'Appropriate Levels of Human Judgment over the Use of Force.'" Chapter. In Autonomous Weapons Systems: Law, Ethics, Policy, edited by Nehal Bhuta, Susanne Beck, Robin Geiβ, Hin-Yan Liu, and Claus Kreβ, 185–208. Cambridge: Cambridge University Press, 2016. https://doi.org/10.1017/CBO9781316597873.009. For the full directive itself please see https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/300009p.pdf

[113] "Ethics" comes up 10 times; "Responsible"/ "Responsibility"/"Responsibly" 9 times, "Safety", 16; "Ethical, 6", and "Robust", "Resilient"/ "Resilience" and "Safeguard" each mentioned 4 times.

When discussing cybersecurity and the protection of AI assets, including data, the algorithm, and the overall system the AI is embedded in, interviewees noted a "checklist mindset" at the DoD (C1). The DoD was viewed as "more conservative than other agencies" (C3), potentially preventing innovation and insufficiently thorough security assurance.

> C1: "I'd say the basic DoD approach to cybersecurity is 'no you can't do anything'. I mean, it's like an internal Denial of Service attack".

> C1: "'You can be obsessively focused on security and devote time and resources to security in ways that make you less secure than alternative approaches, and I would say that is the current failure of the DoD".

There was some disagreement between interviewees on how far security was a primary concern, or a well-understood set of challenges, within the U.S. DoD. C2 felt that while "the technical people" understand security, "not everyone has come to the realisation that the security is really important", instead highlighting that the acquisition community responsible for bringing security into the Department think of it as "someone else's problem". This perspective contrasted with the description of the DoD as "obsessed" with security (C1, C3), especially when it comes to testing, evaluation, verification and validation. Indeed, awareness or recognition of AI challenges did not correlate to working solutions, with the lack of risk and security oversight for AI highlighted by several interviewees. C6 noted that while the Department has the (non-AI specific) Risk Management Framework taking a cybersecurity perspective, which "has proven almost worthless in a lot of cases" when approaching AI. In terms of a holistic approach to security, AI, and risk, the DoD does not "have the sufficient levels of flexibility, nor even awareness of all the department policies and how they interrelate to this" (C6). An exemption raised by C7 and cited in the DoDAI document was the testing requirements in Directive 3000.00, specifically related to autonomous weapons with lethal capabilities. In terms of coordination, the DoD AI strategy designates the JAIC responsible for "synchronizing DoD AI activities" (DoDAI, 10), including the development of governance and standards framework. The strategy also committed to funding research into "resilient, robust, reliable and secure AI", including "explainable AI" to "help users understand, appropriately trust, and effectively manage AI systems" (DoDAI, 5). The USAF strategy includes mention of supply chain safety in addition to trust.

Looking at current testing and assurance for AI systems, Flournoy, Haines and Chevitzl (2020, 2) highlighted several "technological and organizational barriers" to adapting the DoD's existing TEVV to approach machine learning systems sufficiently, with the technical challenge mirroring the concerns outlined by interviewees above. On the bureaucratic side, the report highlighted that at the time of writing, there were several barriers to effective TEVV for ML: shared responsibility across the Department coupled with insufficient coordination, outdated Department policy frameworks and metrics, non-Agile[114] development and testing processes, current TEVV processes ill-suited to ML systems, shortage of AI talent; lack of TEVV implementation policy and ethical guidance, and insufficient coordination with academia and the private sector (Flournoy, Haines and Chevitzl., 2020, 10-17). Regarding the potential of implementation guidance, C3 highlighted that NIST might provide industry with best practice guidance, with the Institute having a "really significant role in standard-setting". As part of the 2021 NDAA continued to dedicate provisions for AI, NIST was tasked with developing an AI risk management framework (US Government Publishing Office, 2020).

Ultimately, questions around AI and security related heavily to trust at the operational level. As C4 argued, "if something goes wrong, again, it's still going back to a human that is ultimately held accountable in that instance". Predictability in the performance of AI-enabled systems was viewed as paramount. As C3 summarised, the DoD wants "to know that it's going to do the same thing every time." The military has a lot to lose in the front lines of conflict, including in terms of risk to human life, where predictability and reliability are compromised (C5). These reflections highlight continuity from Cowan and Foray's (1995) observation that military innovation is extremely risk-averse.

*7.3.4.2 Laws, ethics, and norms*

Ethics and "responsible AI" concepts were frequently mentioned in interviews. In the DoD AI Strategy, the U.S. committed to "lead in the responsible use and development of AI", pushing "for using AI in a lawful and ethical manner" (DoDAI, 5). With China and Russia named within the strategy as actors developing applications "that raise questions for international norms and human rights" (DoDAI, 5), the Strategy repeatedly commits to responsibility-related concepts: incorporating AI responsibly, developing AI principles, and using AI to enhance the Law of War (DoDAI, 5-6). As of 2020, the U.S. DoD does have ethical principles for AI, and the DoD AI strategy advocates for a "global set of military AI guidelines" and U.S. "vision for ethical and safe military AI use" (DoDAI,

[114] The Agile software development approach promotes continuous and iterative innovation. See https://www.atlassian.com/agile

15). The USAF, similarly, committed to "engage in dialogue on the ethical, moral and legal implications" of military AI (USAF, 2) to help develop public confidence.

Beyond the promises from the DoD and USAF documents, several U.S. initiatives were highlighted by interviewees, including the work underway by the JAIC's responsible AI lead, as well as work led by White House staffers, with the general view that efforts to build out responsible AI architectures "were not actually just lip service" (C7). C6 felt that from the summer of 2020, the JAIC had achieved a positive impact on the topic of responsible AI, as they "elevated" attention and ensured testing and evaluation portfolios reported directly to senior JAIC leadership. A memo on responsible AI implementation, released May 26th, 2021, by the DoD, was also viewed as helpful in giving implementation goals and measures by which to gauge progress, though the extent to which would depend on how much budget was allocated to fund these efforts through 2022 (C6). As of Spring 2021, C1 highlighted that as part of existing procurement processes facilitated by the JAIC, "you [the vendor] have to talk about how you're going to comply with the [AI] principles". This requirement appears to mark an evolution from the days of Project Maven. As C4 explained, neither the DoD nor Google came into Project Maven with ethical frameworks, "and they both have them now" (C4).

International norms were frequently cited as a mechanism to encourage responsible AI practices. The U.S. wishes to "shape the norms of [the] international community... there is a great deal of concern around ethics." (C1). However, norms and normative frameworks cannot be mandated or upheld by the U.S. alone. C4 remarked that "if it's just the U.S. with normative frameworks, that's probably not strong enough right, but to have all of, you know, our NATO alliance also synchronized with these kind of [sic] normative frameworks, I think that becomes much more powerful". C7 highlighted the norms-development efforts attempted through the AI Partnership for Defense.

Formal standards were also raised as an option, with NIST (C3), ISO, IEC and IEEE (C7) standards-setting bodies being of interest to the DoD. According to C7, the U.S. has recognised that it "has not necessarily been showing up at those bodies, the way it needs to" as a consumer of technical standards

> C3: "I think it tends to be naive to believe that an agency [NIST] that has very little experience with or knowledge of the reality of conflict can make standards and risk assessments for conflict".

### 7.3.5 Improving levels of understanding in AI technology

The lack of agreement on what constitutes AI, or the current capabilities of modern ML technology, highlighted a lack of understanding across a significant percentage of decision-makers working on AI-related issues. C2 described the tension that might occur when overstretched senior leaders were tasked with AI-related decision-making. C4 described their professional challenge as "trying to work between the tech people that believe that AI will help solve problems, versus the leadership, where they don't want to give ownership to someone other than themselves". C6 also reported the "tenuous struggle between the kind of, people who want this to be way more advanced than it really is, and the people who are trying to put the foundational building blogs in place" (C6). In other words, "politics people are not tech people" (C5), with a lack of understanding potentially highly problematic in a hierarchical military structure that now "blinds us [the DoD] to the need to be very contrarian sometimes" (C6).

> C6: "It's very different to imagine a scenario in which your most revered military leaders, civilian or in uniform, are not right about these things."

This confusion about terms spills over into a misunderstanding of AI capabilities and the power of AI. C6 highlighted how the DoD was not immune to misunderstanding at a general level, reflecting that at the time the JAIC was set up in 2018, "the department was even less… infinitely less cognizant of what AI meant in practice or theoretically for military applications than anyone thought [laughs]". This lack of understanding has implications for senior leaders who, in C6's view, spoke consistently as "very smart, very capable, very well-respected listened to loud voices" but "without necessarily a lot of the critical thought or red-teaming or pushback on ideas".

More generally, the scope of what was considered AI varied, and interviewees noted that as the technology evolved, so did language. C3 noted how their U.S. colleagues often conflated AI with "basic machine learning", which is not necessarily due to ignorance but due to relabelling terms "to get more attention". They continued, "all that was machine learning is now AI…. honestly we've been doing machine learning for 20 or 40 years… now you're all excited about AI, and you don't know what machine learning is, so yeah, we relabel it" (C3).

Hype was often driven by external events and broader national security concerns. C1 mused, “as soon as Xi Jinping [President of the People’s Republic of China] or Vladimir Putin [President of Russia] say this is our number one thing, well, it’s so much easier for me to get a meeting, frankly, right? The broadness of people viewing it as a priority goes way up.” This form of externally driven hype was perceived as a strain by C6, who stated that as hype from the threats and adversary awareness increases, “tension becomes even greater” within the DoD. C4 raised fears that “overhyping” AI would lead to a backlash where “we’ve invested all of this money and we’re getting nothing out of it”, urging instead that AI should be considered as a “slow and steady race”.

The AI talent gap was commented on across multiple levels: within the DoD, the U.S. military, and the U.S. innovation landscape generally (including in industry). C5 highlighted how this lack of knowledge was two-fold; there was insufficient understanding of AI technology to develop an informed sense of current AI readiness and insufficient diversity of skill sets dedicated to employing AI technology.

> C5: “Too few people understand the technology still. And so not understanding how it works, how it could work, and frankly not having what I consider the sort of the more creative people working on the potential use cases….”.

Within the Department of Defense, C3 raised the lack of “technical educated policy-makers”, with few able to assess the technology's technical and political science implications. Some interviewees believed that the deepest expertise relating to AI was in the private sector and some parts of academia. For example, the nature of military career progression in the Army, C2 explained, means that few specialities require a technical background. C4 highlighted the civilian talent pool as the best source of deep expertise, sourcing “PhDs and folk that work exclusively on this topic”. C3 felt that military personnel are easier to retrain compared with the diplomatic or civilian pipelines:

> C3: “We can tell them, “You’re going to go back to school whether you like [it] or not, you’re going to take this training whether you like it or not. This is not a democracy in the military”.

At the same time, C4 highlighted that while the military has been successful at building “very narrow skill sets” that train military experts in different areas. They further believed the U.S. is “not there” yet, with the military instead just recognising the exact AI expertise required (C4). This perspective

that the U.S. was facing an “AI talent gap” was mirrored across interviews, with C6 highlighting that this also applies at senior levels, with “no three-star or tour-star officers that I know of, in the United States Department of Defense, who are digital natives or who really understands AI”. The DoD strategy also acknowledges how critical AI skills are to the strategy's success, promising to introduce comprehensive training opportunities and career progression to “recruit, train, promote, and retain a leading AI workforce” (DoDAI, 14). Similarly, the USAF document promises to “recruit, develop, upskill, and cultivate” their workforce using courses, partnerships with industry and academic partners to foster “cross-collaboration for training and tradecraft”, and “practical policies and incentives that foster talent management” (USAF, 5).

## 7.4. Discussion

The U.S. has clearly stated its goals and ambitions regarding AI in military contexts through the USAF and DoD AI Strategy executive summaries and broader national security doctrine such as the Third Offset Strategy. With significantly greater dedicated funding and institutional resources dedicated to military AI innovation relative to the rest of the NATO Alliance, there is significant evidence that the U.S. is actively working to integrate AI at a rapid pace to main military advantage. The research findings inform an answer to the second research question, highlighting the U.S.’ activity to mitigate a broad range of security challenges associated with military AI. To address the challenges relating to the secure design of AI systems, the U.S. has invested in DoD R&D efforts, emphasised leveraging private sector expertise, and highlighted the need for appropriate verification mechanisms. The creation of various AI-focused agencies across and beyond the DoD highlight an increasing institutionalisation of military AI innovation, with interviewees clearly aware of the organisational and cultural challenges that must be addressed to coordinate effectively. In addition to very active domestic activity, the U.S. is active in engaging with the international community, leveraging NATO where convenient but also setting up alternative cooperative mechanisms such as the AI Partnership for Defense or relying on separate allied and partner relationships. The interview findings show a generally positive interpretation of the U.S. AI Strategy and related doctrines as tools that signal U.S. desires and intentions to the international community that include suggested principles for responsible AI alongside a willingness to agree international norms.

This said, interviewees have highlighted some of the U.S.'s challenges regarding AI in military applications. The DoD certainly has had its drawbacks noted across the years, and AI-related matters

are no different. Deputy Secretary of Defense Deputy Secretary Bob Work (2015, para 3) described the DoD as "the most complex, bureaucratic, unwieldy organization in the world." A Congress-mandated study on AI revealed that the DoD had fallen significantly short of its promises and was "significantly challenged" in all areas (Tarraf et al., 2019, 64). In security terms, the researchers found that testing and verifications were "nowhere close to ensuring the performance and safety of AI applications", especially in the case of safety-critical systems (Tarraf et al., 2019, xiii). The DoD was judged as lacking: the metrics and baselines required to measure success, clear communication between users and builders of AI tools, effective mechanisms to bring in external innovation and collaboration, and an effective solution to source AI talent. The JAIC, researchers wrote in 2019, lacks the "visibility and authorities to carry out its present role. It also lacks a five-year strategic road map, and a precise objective allowing it to formulate one." (Tarraf et al., 2019, 47). Done in a certain way, collective responsibility is in line with the JAIC as a coordinator of DoD AI activity. The NSCAI report recommends the DoD "democratize AI development" (NSCAI, 2020, 67), with the JAIC acting as a "hub" for AI activity and recommends that the DoD "drive organizational reform through top-down leadership" through senior civilian and military officials (NSCAI, 2020, 83). Furthermore, the NSCAI final report includes several recommendations focused on empowering leadership appropriately, including the creation of a Steering Committee on Emerging technology "tri-chaired by the Deputy Secretary of Defense, the Vice Chairman of the Joint Chiefs of Staff, and the Principal Deputy Director of National Intelligence" (NSCAI, 2020, 83).

The NSCAI final report presents an "integrated national strategy for the AI era" (NSCAI, 2021, 8), dividing their recommendations to the U.S. government into two parts. The first, "Defending America in the AI era" (NSCAI, 2020, 41), urged the U.S. to better protect American society against AI-enabled attacks, prepare the U.S. military for future AI warfare to achieve "military AI readiness" by 2025 (NSCAI, 2020, 61), and to thoroughly engage in risk management practices, particularly around AI-enabled and autonomous weaponry. It highlights the potential for AI in national intelligence and across government - if the right skilled expertise is there to enable it, and if confidence at all levels (by the public, by Congress and by military operators) holds a sufficient level of trust in the technology. The second part, "Winning the Technology Competition" (NSCAI, 2020, 155), highlights how technical innovation underpins strategic ambitions, with China "organized, resourced, and determined to win this contest" (NSCAI, 2021, 11). Recommendations urge the U.S. to focus on developing home-grown innovation through R&D investment, comprehensive intellectual property regimes, and "a favourable technology order" (NSCAI, 2021, 13). Furthermore, the "United States must work hand-in-hand with allies and partners to promote the use of emerging technologies to

strengthen democratic norms and values…" (NSCAI, 2021, 13).

This chapter matches the claims of U.S. institutions with the candid realities experienced by senior leaders. Interviewees frequently pointed to collaboration with Responsible AI, the DUI's ethical principles as adopted by the DoD, and the current research engagement underway on "responsible AI" as evidence that the DoD was thinking about various implications of military AI technology. While the conversation on security may appear more fragmented, this may be due to the relatively wider applicable language. While "responsible AI", perhaps followed by "ethical AI", appears to be the language used to talk around AI ethics-related concepts, speaking about security may include language on robustness, trust, explainability, transparency and safety. As C3 highlighted, searching through DoD programmes for "AI security" brings few results, as "what's going to show up is T and E and V and V of a guidance system, oh, and by the way, if my guidance system has an algorithm in it then that's AI security….. But it's [DoD material] is never going to have the words that like… somebody from a security studies programme labels it with". By this line of argument, external claims that the DoD needs an "AI safety program" are misplaced and miss the fact that the DoD consistently aims for robustness, resilience, reliability, and predictability (C3). From these discussions, and even the occasional inconsistency between interviewees on how far the DoD recognises the need for greater focus on AI security-related concerns, it appears that understanding the actual gaps in DoD's current approach requires a nuanced and internal view of a broad swathe of programmes.

### *7.4.1 What now: The way forward for U.S. military AI Innovation*

> "Fielding AI systems before our competitors may not matter if DOD systems are brittle and break in an operational environment, are easily manipulated, or operators consequently lose faith in them."
> - (CSET, Building Trust through Testing, 2020, 4)

There is certainly no shortage of recommendations highlighted for the attention of DoD and broader U.S. decision-makers across various agencies. CSET's Building Trust through Testing report recommends "DoD leadership, working with the intelligence community, the State Department, Congress, industry, and academia" (Flournoy, Haines and Chevitzl, 2020, 2). These include the designation of a responsible coordinating body for AI/ML TEVV, the development of a risk-based framework for AI/ML testing and safety, and a proposal to develop industry/ U.S. government TEVV standards and promote them internationally (Flournoy, Haines and Chevitzl, 2020, 19-28).

Talent and the lack of skilled expertise were significant themes through analysed sources and literature and are not a challenge unique to the U.S. context. The U.S. may be relatively advantaged compared to allies with a smaller resource pool of talent and personnel. Nonetheless, there was recognition across discussions and written material that the U.S. DoD and state infrastructure was competing with the private sector for talent while also facing an absolute shortage in terms of an insufficient number of educated individuals. A huge range of policy options have been proposed here: greater DoD utilisation of resources, including federally funded research and development centres (FFRDCs), university-related, national or service labs (Flournoy, Haines and Chevitzl, 2020) and for the JAIC to produce a "best practice guide for recruiting AI talent, from developers to testers" (Flournoy, Haines and Chevitzl, 2020, 27). The NSCAI's first analysis report in 2020 pointed out that hiring practices must improve to recognise what an AI expert looks like and highlighted the lack of awareness of AI among U.S. government employees (NSCAI, 2020b).

Some of the challenges faced by the U.S. are specific to the culture within the U.S. DoD (with agencies fearing the emergence of the JAIC as a challenger to future funding and projects) or to acquisition challenges unique to different national defence set-ups. However, many of the themes and open questions raised through the research revealed common difficulties in AI innovation. AI hype and the lack of skilled expertise at all levels posed a challenge to effective decision-making. This is particularly challenging as the four-star generals deemed the best authority to direct AI programmes have not usually progressed through their careers based on any technically-minded attributes. Additionally, the discussion about ethics and lethal autonomous weapons was highlighted as a useful conversation, but one which detracted from the swathes of work applying AI-enabled capabilities to back-office, enterprise functions - mirroring the conclusions mused upon by U.K. and NATO interviewees.

# Chapter Eight: Discussion

## 8.1. Introduction

This chapter integrates the themes identified in Chapters 4 to 7 to discuss the overall research findings and contributions to the field, focusing on military innovation and security. This discussion draws on the findings revealed through interviews and observant practice-based methods to present the main themes that emerged through this research. This chapter grounds each theme within the context of the broader scholarship and conversations on military AI, connecting this research with the perspectives found across both academic and other relevant literature as examined in Chapter 2.

First, the research demonstrates that there has been a significant increase in awareness amongst defence-focused communities, including policy-focused staff, on how AI may impact military contexts. There is evidence that the attention toward military AI innovation has increased in terms of greater discussions and increased financial investments at a national and NATO level. This increased interest shows little sign of deaccelerating. Second, research findings on the expected impact of AI suggest that AI will have a far-reaching impact across all aspects of military activity, drawing parallels between insights from this research and existing scholarship. Third, research findings identify growing awareness from those interviewed and a sentiment demonstrated at national and NATO levels that the U.S., U.K., and like-minded states must adopt AI to secure or maintain perceived military advantages against adversaries. The research further suggests that such a drive to innovate is shaped significantly by perceived adversaries' actions and assumed intentions, namely Russia and China. Fourth, research findings highlight the significant challenges relating to AI in military contexts with distinctions in technical, operational, organisational, and strategic aspects. Finally, the research discusses the possible role of international military organisations as a mechanism for norms development. It explores NATO's opportunities and challenges to contribute via the provision of a forum for discussion. This discussion draws on research findings and developments in the policy landscape to reflect on what the future holds for NATO in emerging technologies such as AI and connects the research findings back to the policy-related and scholarly literature.

These findings intertwine and interact. For example, increasing awareness levels are influenced by the increasing motivation to innovate in military contexts and vice versa. The consensus that AI will have a significant impact on military contexts can be linked to increasing levels of staff awareness

and correlates with states' increasingly demonstrating commitments to invest and innovate in the space. The identified barriers facing the adoption and use of AI in military contexts may be mitigated by NATO or other supranational organisations, highlighting that NATO must understand the wide range of implications as an organisation and alliance. The following discussion consolidates this research's contributions to the interdisciplinary academic community, particularly in security studies and information security.

## 8.2. Increasing awareness and interest

The first finding discusses the significant emergence of intense interest in military AI among expert communities. While there are still substantial gaps in understanding the capabilities and implications of AI in a military context, the landscape is rapidly changing in a way that is of more interest to relevant stakeholders, including policy-focused staff. This finding speaks directly to the second research question relating to state attempts to mitigate military AI challenges, as the U.K., U.S. and NATO show both a realisation of the need to rapidly increase their level of attention and understanding of AI in military contexts. Increased attention on military AI themes also translates to the development of relevant doctrine, the creation of new government bodies dedicated to AI in defence, and increasing funding dedicated to both state-led R&D and towards greater public-private partnerships to leverage private sector-led breakthroughs

Overall, this research highlighted an increasing interest in strategic security and AI topics over this research period. The U.K.-focused interviews conducted in the first half of 2020 found that participants felt that overall discussions on military AI were significantly underdeveloped. Though in some cases conferences and trade shows considered the security implications of emerging technologies, there were few non-theoretical proposals on how operational or strategic challenges to military AI innovation may be mitigated. Particularly early in this research, most U.K. and some NATO-focused interviewees reported a lack of confidence, particularly when discussing potential mitigations at a policy level. To some extent, this is a feature of some of these interviews being private industry AI researchers and technical research team leads as opposed to policy-focused staff, and therefore not tasked with considering strategic and policy-level responses within their professional taskings.

Nonetheless, the uncertainty reported by these experts highlights that not all those involved in AI development for military contexts were able to discuss key aspects of the field. While discussions on the lack of awareness were also discussed at length in the U.S. and NATO interviews, discussions on the implications and policy approaches to AI were more detailed and contained a less disparate range of views. Over the course of this research, there was an increasing level of detail in which participants articulated how they understood AI as a set of capabilities, the dynamics of military AI innovation within the global landscape, and the security implications of AI in military contexts. This increase correlated with intensifying activity as the U.K., U.S. and NATO each began to grapple with the emerging challenges. Such activity includes the release of numerous strategies and official press releases relating to military AI and the development of agreed principles for the use of AI in defence as proposed by each actor, which if operationalised correctly, hopes to address a number of the safety and ethical implications of military AI.

This research has highlighted that the more material is available on a state's perspective or strategic approach, the more the research participants responded with consistent answers – both to each other and to the strategy. This finding implies that relevant documents, including strategies and policy statements, have the impact of contributing to coherent understandings and attitudes to military AI across expert communities. The U.S. interviews, all of which took place after the DoD Strategy announcement and public release of documents such as the DoD's principles for responsible AI, rarely argued for points that countered the U.S.-stated priorities. Generally, U.S. interviewees' answers had some consensus with other U.S. interviewees and aligned with U.S. strategic messaging as illustrated through the Third Offset strategy and the DoD strategy. In contrast, with few resources available to the U.K. participants as of 2019, this data includes a candid set of perspectives on the state of U.K. military AI innovation which does not take any formal or government-approved line.

Similarly, NATO-focused interviewees had little public material as a reference point. There were disparate views on topics, including how far NATO should be responsible for encouraging responsible use of AI in military contexts. These findings suggest that, at least for the topic of military AI, documentation is crucial.[115] A strategy allows relevant staff to refer to an agreed roadmap and set of principles. When publicly released, like the U.S. DoD AI Strategy Executive, such documents were perceived by interviewees as acting as a signalling device for allies and adversaries alike. This

[115] This may well be a finding that applies more generally across military communities for other technologies and broader topic sets, with potential research opportunities that lie beyond the scope of this thesis.

research suggests that for government and military defence communities, another significant benefit of published strategies and statements is the clear positioning of states in ways that represent a reference document and helps to develop consensus between relevant military, government and defence industry staff communities.

This change in understanding has likely influenced the perspectives of many of the experts interviewed for this research. This shift is broadly reflected in the patterns of published policy-focused literature. There is literature on the possible opportunities for AI in defence in the 20th century, referring to first wave "knowledge-based" techniques (Gilmore, 1985; Shinar, Siegel and Gold, 1988). Such historical literature highlights how militaries worldwide have been thinking about AI for decades (see Chapter 1, section 1.1.2 for how AI capabilities have evolved). Commencing this research in 2018, very little *detailed* analysis had been published on strategic security and AI (Payne, 2018), with little demonstration of a comprehensive understanding of what AI means for warfare at a strategic level. As of early 2019, the U.K.'s Dstl had an "AI lab" that two team members described had "recently" begun thinking about human-machine teaming in a meaningful way.[116] Before the U.S. DoD's AI Strategy (2019), no state nor regional military alliance had formally outlined their stance on the use of military AI. Some of the programmes for events observed through this research mentioned AI as a minor topic of interest. As observed in Chapter 4, a side-by-side comparison of materials describing the observed military trade shows, and the same shows in 2021/2022, for example, at CyCon, reveal greater attention towards military AI topics. This finding represents an important caveat to the second research question; while actors are currently undertaking a range of activities that may mitigate the security challenges associated with military AI, this is a rapidly evolving field and the mechanisms to mitigate difference challenges may change, fade away, or may yet to have emerged in full. In just over two years the U.K. has expanded from a small sub-team at Dstl (the AI Lab) to having a Defence AI Strategy, a full Defence AI Centre, proposed principles, and an increased budget to encourage technical innovation. The landscape is likely to similarly shift rapidly within the next few years as actors intensify their efforts in, for example, international norms-building or technical R&D.

Through the course of this thesis, an observation of increased awareness and interest in military AI has also corresponded with the relative proliferation of research focusing on the security implications of AI in defence and the military, and various defence AI-focused research initiatives have been

[116] Conversation with Dstl AI Lab staff, early 2019. The conversation was not an interview.

launched. CSET, and its "CyberAI" research programme, were set up in 2019. Similarly, special issues and projects are focusing on AI within existing security think-tanks: the Hague Center for Cyber Norms has commissioned a compendium of essays on AI and international security; the Center for New American Security has published several papers with a U.S. focus on AI and national security; RUSI has released papers on AI and national security, and so on. Communities of academics and researchers are actively researching the space of AI and international security. There are events taking place in 2022 that the researcher would not have seen as possible in 2018. States worldwide have released AI strategies, and as of 2022, a small number have released strategies that specifically look at AI in a defence context. This growing level of detail when approaching AI was demonstrated through each set of interview findings: while the U.K.-focused interviews, conducted in early 2020, were more likely to refer me to specific websites to see evidence of initiatives, for example, to look up Royal Navy's Project NELSON initiative,[117] there were few reports cited by interviewees. This finding was also replicated to an extent with NATO interviewees. The U.S. interviewees, in contrast, were more likely to check that the researcher had familiarity with the material, including the Third Offset Strategy and the DoD Strategy, before starting the meeting. They referred to the Offset Strategy and various DoD and JAIC initiatives, and interviews reflected on the DoD AI strategy at length. Without equivalent public materials, no similar conversation was possible in U.K. and NATO interviews. The more relevant official state-produced material was available, the greater the levels of awareness and, to a degree, confidence, were displayed by interviewees discussing AI topics.

Furthermore, the relative lack of material available to the U.K. and NATO may explain why there was often a broader range of perspectives reflected on by interviewees who often presented their (self-described and self-caveated) personal views in the absence of any official policy. The U.S. interviewees demonstrated greater coherence with less divergence of opinion between those interviewed. While also offering their perspectives with caveats on personal views, they did not propose points that contradicted the DoD strategy or associated materials, demonstrating how the strategy may have "anchored" views to some degree.

While there has been a marked increase in awareness of AI, this research highlights that there is still some way to achieve widespread military community understanding of the implications of AI in military contexts. This research has highlighted that attempts to find a consensus regarding a

[117] Project NELSON a major active project promoting enhanced data analytics and AI across the British Royal Navy. For more information see (Digitalmarketplace.service.gov.U.K, 2018).

definition for AI have not progressed beyond the same challenges Sweeney discussed in her 2003 paper on AI terminology. Both the observant practice and interviews revealed a wide array of definitions for AI in warfare and distinctions in what was considered AI. Some interviewees discussed the lack of an agreed international definition for AI in the defence space. Some interviewees struggled to define AI and were aware of their knowledge gaps. Others offered a broad range of definitions that did or (deliberately) did not conflate AI with autonomy or did or (intentionally) did not include non-ML forms of AI. Discussions across all sets of interviews (U.K., U.S., and NATO) were aware of how different understandings of AI, or a lack of decision-makers informed on AI topics, represented a challenge for respective states or alliances. In general, interviewees mirrored the frustrations of Kraaft et al. (2020), that ambiguity in what constitutes AI hinders productive discussions between communities. This ambiguity was evident through the U.K. discussions on the public and private sector relationship. Different technical and policy-focused understandings of AI terminology often meant communities were using the same words to refer to different techniques, a risk highlighted by Sweeney and explored in Chapter 2, section 2.2. While the recent emergence of policy-level documents may have provided a common reference point to discuss AI concepts, U.S. interviewees also perceived that military communities did not have a common language to discuss AI topics effectively. This ambiguity demonstrates that AI remains a contested and evolving term, as Brennan, Howard and Neilsen (2018) argued. Even as militaries increasingly converge on a consensus understanding of AI terminology, definitions may continue to evolve to match technical developments. This belief complements Kaplan and Kaenlein's (2019) statement that AI definitions have changed over time.

Beyond definitions, findings across the interviews show that more awareness is needed, especially by policymakers and decision-makers across government and military structures. The observant practice-based part of this research demonstrated the emerging discussions on the opportunities and (generally to a lesser extent) AI challenges, as interviewees presented the consensus view that AI was often misunderstood in the military context. Interviewees had a range of proposals relating to AI awareness. For example, NATO interviewees specifically highlighted a need for a greater understanding of military technologies across NATO. U.S. interviewees generally called on broader pathways to develop AI expertise within the military. These imply a continuation, to some extent, of the reflection through the observant practice research, which voiced that the field was immature when it came to understanding and assessing current AI capabilities. This also marks an open challenge that has yet to be addressed by states' activity, as this research did not find substantial efforts to standardise a common language for military terminology. On the contrary, different states definitions for

technologies like LAWS continue to prove a point of contention and barrier to addressing international norms-development (see Chapter 2 Note L). This finding informs the answer to the second research question, highlighting future opportunities for states or international organisations to focus on consensus-building and standardisation and noting that this is currently an open challenge.

Some of this confusion may be due to how information is framed to military communities. The observant practice-based findings reflected how many voices were present at military trade shows and conferences, acknowledging that many presenting vendors or institutions had an agenda that brought them to the event related to networking or sales. Observant practice revealed aspects of "hype", as first discussed in Chapter 2, section 2.5.1. Military or government decision-makers were shown presentations that often focused on the potential opportunities of AI technologies and the urgency to adopt AI technologies, with less emphasis on AI safety or de-escalation. This observation suggests some bias in the information that conference and trade show attendees receive. Significant examples of "hype" at military-focused events were observed in the data, with organisations stressing the opportunities of their products or planned innovation while focusing less on the limitations facing AI development and adoption. These dynamics show instances where industry actors contribute to "techno-hype in the military" described by Elhefnawy (2018), who argued that industry vendors could distort buyers' understanding of what is currently feasible by overemphasising the potential benefits of AI. Furthermore, the fact that industry actors are often at the cutting edge of AI innovation in some sense disadvantages state defence departments who may not have the insight or skilled expertise to make informed decisions on AI adoption or deployment, a fact reflected on frequently in UK-focused interviews. This reflection is supported by related literature highlighting how governments struggle to control the narrative on AI technical innovation spearheaded by private sector actors (FitzGerald and Parziale, 2017) as commercial sector actors increasingly dominate AI innovation (Cummings, 2017, Whittaker et al., 2021).

The "talent gap", the shortage of skilled AI expertise, cited across observed events and interviews, the published scholarship (Christie, 2021) and policy documentation, including the U.S. DoD AI strategy, may play another factor in current awareness levels. With the absence of skilled staff or detailed defence-focused education for staff in military contexts, information on AI may be received through different means and via actors with different agendas. While the literature has not explored this dynamic in a military context, previous literature has reflected on how public media coverage often treats AI as a revolutionary phenomenon without nuanced discussions of AI as a set of constantly evolving technologies (Brennan, Howard and Neilsen, 2018). Without an informed view

of AI, those coming in risk being prejudiced by a disproportionately small group of actors at events or through the media, relying heavily on industry content (Brennan, Howard and Neilsen, 2018). This thesis suggests this same dynamic applies to military trade shows or sponsored events, where attendees may develop their knowledge by drawing on vendor pitches or from other agenda-driven presentations. This finding does not negate the potential for such events also to be beneficial sites for communities to share expertise, as reflected on in the literature (McCann, 2011; Larner and Le Heron, 2020)

Reflecting on how quickly the landscape is evolving, both in terms of technological innovation and through increased reference points across published scholarship and policy statements, this finding contributes to scholarship in highlighting changing awareness over time. This finding strongly suggests that AI is a flourishing area of research that is likely to be increasingly dominant in military communities. Observing the changing awareness over time via interviewees and observant practice-based techniques has allowed this research to capture a record of key debates and perspectives in military-focused spaces over this period in ways that add to existing scholarship. This finding also highlights overarching challenges relating to how relevant communities, including military and government public staffers dedicated to defence environments, can "catch up" in a space that is evolving at such a rapid pace. The challenge of developing and maintaining an informed and knowledgeable workforce has implications for how far militaries can adopt AI in the first place and how far militaries can consider and mitigate security challenges relating to AI trends.

## 8.3. The promise of profound impact

This finding reflects on the sentiments made consistently by participants and stakeholders throughout this thesis that AI would have a significant impact within military contexts. This finding addresses the first research question, "What are the implications of AI innovation in military contexts?". To ask whether AI *will* change the nature of war and conflict almost seems naïve based on interviewees' responses and upon reflection of existing literature and national public statements. Both observant practice and interviewee-derived findings offered the consensus that AI technology is expected to fundamentally change the nature of warfare, with the consensus view that such technology would also impact almost every aspect of military activity. The "game-changing" nature of the technology was highlighted repeatedly across interviews and events attended through observant practice. Interview and observant-practice data highlighted that AI is overwhelmingly perceived as a capability

set that will speed up warfare, whether via the automation of various banal, enterprise-level functions or whether facilitating rapid processing of information for decision-assistance purposes in battlefield environments. These findings provide further detail to strategic-level statements in the scholarship discussing how AI facilitates the speeding up of warfare, from faster decision-making or processing to automated processes that remove the need for human-speed thinking (Payne, 2018; Johnson, 2019). These perspectives align with strategic overviews that claim AI will contribute to greater military efficiency (Lin-Greenberg, 2020) and vastly enhance information processing (Finlan, 2020). Overall, AI was perceived to be an innovation that would enable other technological breakthroughs such as big data and automation. This finding complements the academic scholarship (Johnson, 2021) alongside NATO publications noting that different emerging technologies, including AI, would consolidate and leverage each other's capabilities to cause exponential-type change (STO, 2020). It does so by drawing out the consensus across the interviews and events that AI is overwhelmingly seen as an impactful enabler and setting out how experts expect AI to start delivering optimisation for their militaries within the next few years.[118] These views add expert insights to build on speculation in the literature, which found enterprise AI more likely to be deployed in the near future, then mission-support AI, then operational AI (Tarraf et al., 2019).

RAND's (2020) three categorisations (as set out in Chapter 2, section 2.3) provide a helpful tool for understanding the examples collected through this research. The findings show that participants often discussed how AI might be used extensively in back-office functions, the "enterprise level" as defined by RAND's report, with interviewees supporting greater investments to improve military performance. There was some emphasis from interviewees that enterprise-level AI offers the most immediate opportunity for military advantages, deploying AI applications in an environment that can significantly increase efficiency in a less risky and less controversial space than in critical frontline equipment. This perspective aligns with policy-focused research highlighting how military logistics and sustainment can leverage AI to improve performance across the military supply chain (Konaev et al., 2021). At an operational level, information processing was raised across interviews as a capability that would increase the speed of intelligence, surveillance and reconnaissance efforts. Logistics and predictive maintenance were also seen as use-cases through which safety and efficiency could benefit militaries with relatively low associated risk levels. The U.S. DoD strategy highlighted how AI could contribute to military predictive maintenance tasks and offered the automation of

[118] The near-term benefits were thought to be applicablt to enterprise environments shortly, with significant battlefield deployment slightly further out in the future due to the identified immaturity of current AI capabilities.

routine equipment alerts as one example (DoD, 2019). Across this research, interviewees articulated how AI is posed to compress the decision-making cycle as machines increasingly process information faster than human operators may be able to comprehend. This finding relates to themes including human-machine teaming, risk, and cognitive overload explored in section 8.5. This finding reflects how interviewees demonstrated perspectives set within the academic and policy-focused literature, highlighting the correlation between the increasing use of AI in military contexts and increasingly compressed timelines for decision-makers to assess and determine the next steps (Dufour, 2018; Scharre, 2017, Singer, 2009). Discussions on AI may change the nature of warfare tended to focus on the potential implications of the technology at a general level and within the digital and cyber domain. For example, interviewees referenced several existing operational capabilities for AI-enabled cyber defence. Interviewees generally did not go into detail to discuss the impact of AI on cyber-physical systems, including LAWS, a topic heavily explored in the previous literature (Boulanin and Verbruggen, 2019; Allen and Chan, 2017). Interviews demonstrated "pockets" of awareness of *where* and *how* AI would significantly change aspects within military contexts. Regardless of the *detail* in which the mechanisms of AI were described across events or interviews, this research finds that overwhelmingly military communities perceive AI as a significant technological shift. This finding represents a contribution to scholarship by capturing community sentiments and comparing them with existing scholarship, finding that high-level observations in the scholarship are reflected in military community perspectives. This finding also potentially relates to scholarship considering international security perspectives, which views AI as a technological enabler that fits into a military's broader toolset in conflict (Johnson, 2021; Horowitz, 2018; Lin-Greenberg, 2020).

While the literature highlighted an ongoing debate as to whether AI represents an evolution or revolution in military affairs (Lucarelli, Marrone and Moro, 2021; Thornton and Miron, 2020), neither policy-focused nor industry practitioner perspectives explicitly referred to these dynamics in interviews. While some participants felt that AI represented a gamechanger in how warfare would occur, others noted that AI primarily represented an evolution of current trends. This research does not propose that AI represents either evolution or revolution within the military context, particular as this discourse did not emerge from the interview or observant-practice data. This research focuses instead on the finding that while it is an open discussion about whether AI technologies amounted to a revolution in military affairs, the fact that AI was highly significant was considered self-evident.

Interviewees also overwhelming viewed AI as a highly immature field of study. Interviewees stressed that the technology as we know it is presently too immature to be deployed at a scale beyond a limited

simple use-case and within largely enterprise-specific environments. As the impacts of AI are largely unrealised currently, it may be too early to say whether AI delivers revolutionary effects. While interviewees tended to consider AI immature in terms of the capabilities and flexibility required in an unpredictable battlefield, they also generally acknowledged that technological innovation is evolving quickly. As expressed across interviews, this argument can be linked to the discussions with the NATO-focused interviewees in particular. The interviewees also argued that now was the right time for states to start thinking about the effects of AI in military contexts. This line of reasoning highlighted that a failure to innovate in the context of such a rapidly evolving landscape meant that a state risked falling behind its allies or adversaries, a view reflected on across the scholarship, particularly concerning NATO members and the risk of a widening capability gap (Dufour, 2018, Pepe, 2021).

It is helpful to refer to hype once more to assess how far hype shaped the perspectives of those interviewed and those presenting at observed events to understand how far the impacts of AI were perceived independently of AI-related hype. Hype was a topic that was present, to varying extents, across the interview sets and the observant practice findings. With interviewees selected due to their experience and expertise relating to AI and emerging technologies, interviewees spoke of how the hype around AI affected those with less familiarity with the topic. Nonetheless, interviewees consistently demonstrated the view that the considerable impact promised by AI is not caused by hype but is to be considered evident in its own right. This argument relates to the consensus in the scholarship that AI is promising in military contexts and will deliver a significant impact within a generation, aligning with most policy-focused predictions (Finlan, 2020; Johnson, 2021, STO, 2020). NATO considers AI a disruptive technology that has emerged and will continue to deliver increasing impact at a fast pace (STO, 2020).

The agreed understanding was that AI, while currently immature, is not to be dismissed as hype. A technology that is still emerging but has shown promising use across a broad range of use cases, how a military was able to integrate AI capabilities was seen to be a crucial factor in that military's strength. These proclamations of machine-speed warfare may or not be a self-fulfilling prophecy, with decision-makers exposed to a sense of urgency that they then promote. This uncertainty may well highlight a valuable area of future research. For now, this research finding shows that to interviewees, how far the motivation to innovation came from informed or ill-informed material is now not the point of interest: military communities consider AI as a key feature of near-term warfare.

The perspectives captured throughout this research demonstrate how far arguments in the scholarship align with the experiences of expert practitioners and policy-focused researchers. The finding that experts across the military communities believe that AI will deliver large and significant changes across military contexts, notwithstanding criticisms of hype, represents a strategic contribution to the literature by highlighting that, for the most part, policy-focused scholarship agrees with expert perceptions. This finding also extends the scholarship to suggest that military-focused industry and policy communities are still determining the exact nature of the change. Grappling with related challenges, including hype and lack of broader awareness, this reflects the recent emergence of the field and may well be addressed over time.

## 8.4. The pressure to innovate

This topic demonstrates a growing awareness from those interviewed and a sentiment demonstrated at national and NATO levels that the U.S., U.K., and like-minded states must adopt AI to secure or maintain perceived military advantages against adversaries. There is a consensus across the research that such a drive to innovate is informed partially by adversaries' actions and perceived intentions, namely Russia and China, as discussed in Chapter 2, section 2.4. This finding represents a major implication of current military AI trends, highlighting how the dynamics of international competition are prompting intense investments and exploring the impact of an innovation environment dominated by private sector actors. This finding also explicitly addresses the second researching exploring how actors are responding to the challenges associated with military AI. Such activity includes U.S, U.K. and NATO commitments to strive to adopt AI rapidly to maintain military advantage, thus engaging in technological competition while publicly supporting norms-development for restraint, as well as initiatives by each actor to increase private-public collaboration and encourage private sector development of dual-use technologies.

The near-future opportunities for AI in warfare have significant implications for future planning and the nature of investment in military capability-building. Across the observant practice data, there were repeated instances of the need for rapid technological innovation, with defence departments encouraged to learn to fail fast and modernise to keep up with external factors such as private sector breakthroughs and the development of AI-enabled attack capabilities. The pressure to innovate was referenced consistently throughout interviews and across observed events. As this section explores, interviewees explained the rationale to motivate within the context of international competition, citing

perceived challenges to the international order and the idea that great power militaries did not want to be left behind. Despite the literature challenging the notion of an "AI arms race" (Roff, 2019), when explicitly asked, all U.K. interviewees felt there were arms race dynamics when it came to military AI.[119] Across interviews, policy experts repeatedly referred to Russia or China as adversaries that the U.K., U.S. and NATO alliance were warily watching regarding military AI capability-building and technological innovation. Such perceptions align with evidence across the scholarship that China and Russia are investing heavily in military AI capabilities (Kania, 2017; NSCAI, 2021; Soare, 2021), with policy literature also reporting the view that the U.S. "faces significant competition in military AI" (Morgan et al., 2020, XIII). Discussions through interviews and observed events consistently focus on Russia and China with little discussion on smaller states or non-state adversaries, mirroring trends within the broader scholarship.

Reflecting on potential trends that AI may influence in terms of the broader threat landscape, China's perceived secrecy and access to large data was a concern to interviewees, particularly U.S. interviewees. This concern is supported by evidence in the literature published during this research, revealing significant investments by the PLA into military AI capabilities, with a pronounced year-on-year increase in financial investment into military AI innovation (Fedusiak, Merlot and Murphy, 2021). With the exact nature of China's military AI investments unknown to interviewees, the concern was that China had both the technical capabilities to deploy world-leading AI in conflict and the intention to undermine what NATO calls the current "rules-based international order" (Brussels Communique, 2021). U.S. interviewees tended to discuss China as a near-peer adversary. Russia was the other state referenced concerning adversarial states across all interview sets, and relatively more so for NATO interviewees.[120] Concerns around Russian military innovation focused less on current capabilities, which were viewed as less mature than the U.S. and China, but more on the intentions of Russia on the geopolitical stage. Russia's use of disinformation to influence populations around the globe has been widely reported (DiResta et al., 2019; Whyte, 2020). There are ways in which AI-facilitated disinformation, for example, through deepfakes, can further enable psychological information operations currently employed by Russian security bodies.

[119] While the question sets across U.S. and NATO interviews meant interviewees were not asked if they perceived AI arms race dynamics, these interviews nonetheless referred to these themes.
[120] NATO may well increasingly focus on China in the near future, with the Brussels Communique highlighting China as an adversary perceived to be working "to undermine the rules-based international order" (Brussels Communique, 2021).

Importantly, interviewees described that AI arms race dynamics were not seen as something unique to the military or national security domain. Economic competitiveness and commercial competition were highlighted as a lively international environment in which China, for example, was growing AI talent across a range of industrial sectors. The use of AI on a civilian population was again raised as an example of Chinese capabilities by interviewees, highlighting the potential of AI technology that interviewees did not perceive as matching U.S. values, a perspective aligned with the language in the DoD AI Strategy.

There was a recognition at observed events and across interviews that AI is also a grouping of technologies that can be heavily leveraged, significantly boosting the capabilities of actors who may not necessarily have resources for warships or large fleets of military vehicles but who can achieve disproportionate success with the right software. Across interviews, participants reflected on the strategic importance of maintaining or developing a technological advantage against enemies. The U.S demonstrates significant activity in this regard, engaging directly with the concept of technological competition and investing heavily in the design and adoption of military AI. This approach is reflected within the U.S. Third Offset Strategy (Work, 2015). The Third Offset Strategy highlighted the critical importance of maintaining a technological edge to maintain order in the international landscape. This language is echoed in reports to the U.S. Congress, which describe how AI and autonomy will be critical to the future of military power (NSCAI, 2019). The U.K.'s Integrated Review (2021) similarly highlights the importance of technology to counter adversaries and enhance the U.K.'s security, though it contains no detail on AI competition in the military context. Interviewees across interview sets agreed that there is little room for complacency, and this view can be supported via a reflection on the pace of change in the field. As the U.S. demonstrated through Project Maven, two years is sufficient to move from proof-of-concept to a deployed operational tool (Simonite, 2018). This research also highlighted the argument for innovation that appeared one-sided overall, emphasising rapid iteration rather than restraint. This view is also reflected in the literature, which encourages innovation alongside caveats to address security implications (Gilli & Gilli, 2016, Tarraf et al., 2019). Even literature highlighting the significant challenges associated with military AI technologies did not go as far as to suggest avoiding innovation (e.g., Burton and Soare, 2019; Horowitz et al., 2018). Critical literature, such as Suchman (2020), again calls attention to problematic aspects of AI but falls short of advocating a ban or avoidance of such a technology. These discussions are distinct from the narrow scope of LAWS-focused discussions, including prominent civil

movements such as the Campaign to Stop Killer Robots.[121] Reviewed literature generally tended to highlight that while the debate on banning specific AI applications was important, there were more pressing and promising discussions to be had (Cummings et al., 2019). Promising discussion avenues included encouraging responsible design (Noorman and Johnson, 2014) or effective governance over AI integration with other technologies (Lucarelli, Marrone and Moro, 2021). These arguments assume some determinism to an extent, that states will be innovating in this area to avoid widening capability gaps with allies and, more importantly, with perceived adversaries. This research also raises the open challenge of how states engage in the dynamic adoption of military AI capabilities and finds that even states investing significant resources in military AI development have a limited understanding of the field.

A topic raised across interview sets was the dynamics of innovation in the military context, where the most significant innovators are not public sector laboratories but are private sector actors. Interviews and observant practice findings suggest the importance of commercial industry investment. In discussions across the events, U.S. and U.K. interviews highlighted the challenges of the public-private relationships at length. The nature of these discussions goes beyond conversations in the literature, which note the increasing dominance of the commercial sector concerning defence technologies (Cummings, 2017; Whittaker et al., 2021) but do not go into detail about the tensions in the procurement process for emerging technological solutions such as AI. Interview insights on the often-fraught relationship between vendors and government defence staff, as reflected on by U.K. and U.S. interviewees and across observed events, form a strategic contribution to the scholarship on the complications of AI acquisition as discussed below.

This research also builds on the findings in the literature that argue that when it comes to control of military technology, the government is less advantaged now than in the 20th century (Cummings, 2017; FitzGerald and Parziale, 2017; Christie, Buts and Du Bois, 2021; Whittaker, 2021). On several occasions, interviewees referred to Google, Microsoft and Amazon as examples of firms with colossal access to data, making most of their revenue from non-defence products. For corporations who do not need to rely on government contracts or the defence sector specifically, it may be the government who has relatively less power in the decision to engage, with some companies deliberately avoiding contracts with state DoDs. Defence procurement was noted repeatedly as a major pain point as the

[121] See https://www.stopkillerrobots.org/.

MoD and DoD struggle to overcome bureaucracy and encourage engagement with non-traditional innovators. The extent to which commercial sector innovation limited the technologies available to governments was unclear; interviewees were divided on how far private corporations *actually* avoided working with the DoD or other state bodies. Some felt that other corporations would "fill the gap" without much difficulty.

Nonetheless, interviewees reported that individual skilled personnel often avoided the defence sector. The literature highlighted how skilled AI researchers might find military-focused work ethically problematic or less well-paid than industry roles (Singer, 2009; Cummings, 2017). This research also found perceptions that skilled AI staff might find more engaging work outside the defence sector. One interviewee described the example of a data scientist who could develop new applications for large technology companies rather than doing repetitive verification-focused roles for government defence. These dynamics may provide the rationale for the consensus emerging from the U.K. interviews that MoD staff did not necessarily understand the technical implications of their requests, especially if they were policy-focused staff, causing frustrations for those involved in the broader procurement process.

On an international scale, interviewees also noted that the relationship of the Chinese state with companies based on Chinese territory was vastly different to the markets of the U.K, U.S. or many NATO members. This perspective is reflected across analysis in the literature, which argues that the PLA has benefited from their approach to civil-military fusion, due to a rapidly developing domestic market of private firms and through contracts with international suppliers (Fedusiak, Merlot and Murphy, 2021). Furthermore, policy-focused research has highlighted that the U.K, the U.S. and other states do not have China's access to/ accumulation of data (Kania, 2017) and are highly unlikely to do so (Fedusiak, Merlot and Murphy, 2021). Government procurement of external solutions was identified as a major pain point in both the U.K and U.S. interview sets. The U.K. interviewees offered a public sector perspective on how the MoD approaches technological innovation, revealing several inefficiencies. The bureaucratic nature of tendering for government contracts was highlighted by interviewees as a barrier to non-defence primes, some of which have valuable products likely to be of interest to the MoD. The "valley of death" in which AI proof of concepts rarely go to full-scale deployment was cited in the U.K., the U.S. and NATO interview sets and linked with the idea that defence departments are more risk-averse in ways that negatively impact innovation. Academic scholarship on military AI themes does not appear to significantly explore the nature of these frustrations, though Christie (2021) highlights the need for more innovative approaches to technical

development. While this research highlighted that the U.S, U.K, and NATO were all engaged in active programmes to streamline private procurement and integration, from the dedicated Tradewind DoD procurement platform for AI products to the NATO-created defence start-up accelerator (DIANA) and Innovation Fund, these findings also highlight a need for organisational and cultural change.

Interestingly, within the current research, different groups of experts emphasised different motivations for innovation. As Chapter 5 highlighted, U.K. technical-focused researchers and practitioners referred to technical challenges as motivating AI research. A few interviewees mentioned the perceived actions of adversarial states, and only one highlighted awareness of policy strategies, including the U.S. DoD AI strategy. In contrast, NATO and U.S. interviews referred to the external threat relating to how Russia and China may use AI capabilities. This discrepancy between policy-focused researchers and technical staff is not necessarily surprising, as interviewees prioritised speaking on topics on which they had the most professional experience. These differing views demonstrate how the military AI innovation landscape has different perspectives and staff with different skill sets. The interviews and observant practice findings show that conversations are often taking place in silos rather than between communities or staff with diverse expertise. While there is some overlap in what was discussed between technical and non-technical colleagues, for example, the consensus that arms race dynamics are generally at play and that there are a number of technical challenges that must be overcome before confidently deploying AI in military contexts, different other aspects of conversation revealed gaps in knowledge. U.K. industry-employed interviewees highlighted the technical opportunities of AI. At the same time, U.S. and NATO colleagues who were more policy-focused appeared to prioritise geopolitical and strategic ambitions and implications relating to AI. Observant practice events highlighted both arguments.

Finally, with the recognition that this is a rapidly evolving space is the acknowledgement that states must focus both on the near future and immediate investment in innovation but remain forward-looking and pre-empt future implications of the technology. Interviewees repeatedly stressed that agreements on an international approach to military AI and security were urgent. The more entrenched AI technologies became, the harder it became to change course or roll back innovation after the fact. This concern is neatly summarised in Singer's (2009, 10) book on this theme, with the rhetorical question, “how do you put the genie of knowledge back in the bottle?”. This feeds into a wider conceptual contribution delivered by this research which highlights that the security implications of military AI will only become more convoluted and complex to solve over time.

This research approached military AI innovation using a "top-down" approach, collecting insights on how interviewees and communities at observed events understood and perceived the dynamics of military AI innovation rather than reviewing individual applications and innovation-focused initiatives in a "bottom-up" fashion. This approach represents an empirical contribution, drawing on strategy-level perceptions of experts rather than starting with individual use cases in a way that risks artificially 'silo-ing' a topic. This research has highlighted that military AI innovation is a topic that has sweeping geopolitical implications and factors. It does necessarily make sense to reference military AI without exploring technical breakthroughs, actions by non-allied state actors, or the role of private-sector innovators. With extensive discussions on the intense competition that motivates military AI innovation and the factors involved in developing or procuring AI, this finding goes some way toward addressing the first research question in mapping out the implications of AI for military innovation. This section represents a strategic contribution to the literature in detailing the various rationales for intense innovation and exploring the dynamics of innovation, including the increasingly dominant role of commercial actors.

## 8.5. Challenges to successful adoption and use

This section sets out research findings on the security implications of AI in military contexts. It describes current and proposed approaches to mitigate various challenges and risks associated with the adoption and use of military AI technologies and highlights several open questions facing militaries, policymakers, and researchers. While many of the challenges identified are cross-cutting, for example, with broad operational, ethical and strategic implications, this section will set out findings under the simplified categorisation of technical, operational, strategic and ethical and legal security challenges. This section addresses both research questions, highlighting the various implications of AI in military contexts through the lens of security considerations, and discussing approaches to date.

### 8.4.1 Technical

Interviewees highlighted the challenges associated with developing or deploying AI systems. The consensus perspective was that AI was too immature to be deployed in operational military contexts in many cases. Their insights relate to the active research which has been undertaken concerning technical challenges with AI, examining adversarial AI (Tygar, 2011; Biggio and Roli, 2018 Athalye et al., 2011), the vulnerabilities of inadequate training data or biased algorithms (Fry, 2018; O'Neil,

2016; Osaba and Welsner, 2017). Particularly for applications unique to a military environment, interviewees highlighted the quality of training data as a crucial aspect in determining how well the tool would achieve its objectives with challenges including poor quality or insufficient data. Interviewee reflections that AI may make mistakes if not appropriately trained represent just one instance in which interviewee statements were echoed in the literature, with Suchman (2020) calling attention to the risks of poor data-labelling of training data. Interviewees repeatedly highlighted the potential for AI systems to be compromised, which is another area explored extensively across existing literature (Huang et al., 2011; Steinhardt, Koh and Liang, 2018; Geist and Lohn, 2018). Interviewees also raised several challenges relating to the verification and testing of algorithms, noting that AI products which are not "home-grown" must still be subject to appropriate testing, evaluation, verification and validation (TEVV). There is the challenge of sharing sensitive information and data with external parties, which can be complicated in terms of classified material and raise supply chain risk. This challenge is described in the literature by Lin-Greenwood (2020), who outlines how NATO must face these challenges.

Regarding mitigations, the U.K. interviewees, staff members managing technical AI research teams with a more technical background and closer to the technology itself, did not suggest *specific* proposals to mitigate some of the challenges above. This finding represents some non-alignment with policy-focused scholarship, in which there are numerous calls across the literature for states to urgently consider the ramifications of AI techniques in the military and put in measures that mitigate risks (Kania, 2017b; Burton and Soare, 2019; Christie 2020). These include technical mitigations, including robust testing and redundancies in systems (Kania, 2017b) and robust standardisation mechanisms for technical AI design and security (Pepe, 2020).

The U.K. (private sector-employed) interviewees felt that government staff did not always understand the technical risks associated with AI-enabled systems and that the MoD's expectations were often not managed by private sector partners when discussing AI development and acquisition. This finding corresponds with views across this research, via interviews and observed events, that greater education is needed to facilitate a greater understanding overall. Interviewees and observed events stressed the need for a talent pipeline in military career paths and for defence practitioners across government, industry, and academia. While this is a topic reflected on in policy-focused literature (Gilli, 2020; Christie, 2021), these findings suggest a widespread challenge in military contexts, with a need for greater efforts to educate military decision-makers and defence procurement staff in particular.

### 8.4.2 Operational and HMT

Interviewees expressed concern about the increasing complexity of AI systems, particularly where the mechanisms of an algorithm are unclear to human operators and especially in safety-critical environments and connected this to broader challenges or human-machine teaming and how human operators would trust AI systems. Across interview sets, interviewees highlighted how AI might be a "cognitive crutch" for humans (NATO-focused interviewee B17), highlighting the risks of overreliance if attempts to understand the data result in cognitive overload for operators (U.S.-focused interviewee C4). These reported concerns mirror an emerging discussion in the literature, highlighting how AI systems may reason in ways that are not comprehensible to humans (Ayoub and Payne, 2016) and the risk of potential errors as humans do not challenge AI-enabled systems' outputs (Scharre, 2018). Interviewees highlighted the open questions of responsibility and oversight, using practical examples such as misidentifying non-military objects as targets to highlight potential operational challenges with ethical and legal consequences. The broader reflection through these findings, proposing that speed offers a significant competitive edge in conflict, highlighted the associated risk in which warfare may be accelerated to the point which is incomprehensible to - and therefore uncontrollable by - human decision-makers. This reflection has been explored through literature, particularly through the analysis of friendly-fire accidents (Singer, 2009; Scharre, 2017). Some suggestions, such as Kania's (2017b) proposal to include redundancies in military AI systems to allow for secondary validation and evaluation of AI outputs, may offer the chance to minimise errors. This mitigation might operate at the expense of a speed advantage in real-time, or perhaps limited to providing feedback later after suspected errors have occurred.

Interviewees generally their agreement that humans should, and likely will, remain in the loop. Interviewees agreed on this as a critical point. However, they showed concerns that adversarial actors would benefit from a speed advantage to take humans out of the loop for certain AI systems. For LAWS at least, U.S. directive 3000.09 and U.K. doctrine as laid out by the Joint Concept of Human-Machine Teaming (Ministry of Defence, 2018) have stated that humans must remain in the loop to have oversight and control over decision-making. The U.S. is the only state with a directive on human responsibility and autonomous systems via the DoD 3000.09 directive, which mandates human intervention for autonomous weapons systems. These form a category of security concerns, not necessarily about the integrity of AI systems used by the respective state and its allies but about the

malicious use of AI by adversarial actors. The occasional fears interviewees raised about whether adversaries may nudge an actor away from meaningful human control are not necessarily unfounded when placed within strategic signalling across the policy landscape: the division chief of the U.S. Joint Counter-Unmanned Aircraft Systems Office is quoted in a U.S. DoD news piece as saying "*Right now* we don't have the authority to have a human out of the loop....*based on the existing Department of Defense policy*, you have to have a human within the decision cycle at some point to authorize the engagement" (Lopez, 2021, para 2, emphasis added). These discussions relate to the broader debates in the literature on how AI may or may not be able to meet conditions for MHC (Vignard, 2014, Horowitz, 2018; Kania, 2017b; Noorman and Johnson, 2014), often without using those terms, highlighting how practitioners were approaching these unresolved challenges. Such language suggests that concerns around MHC are broadly acknowledged by defence policy experts as open challenges, with the same debate occurring in the literature relating to legal and ethical discussions (Taddeo, McNeish and Blanchard, 2021; Boardman and Butcher, 2019). More generally, the operational challenges discussed by interviewees reflected largely the active debates occurring in research fora. Discussions included how effective human-machine teaming can operate on the battlefield (Boardman and Butcher, 2019), how far AI be able to act autonomously versus maintaining a strong "human in the loop" presence (Horowitz, 2018), and research on meaningful human control (Vignard 2014; Maas, 2019; Roff and Moyes, 2016). These discussions in the literature heavily support the perspectives collected through this research.

### 8.4.3 Strategic

The high-level geopolitical implications of military AI innovation were repeatedly cited through observant practice events and interviews. Interviewees described their perceptions of intense competition as outlined in section 8.3. The phrase "great power competition" was mentioned multiple times in discussing how states such as the U.S. or U.K. perceived the international landscape and the risks of unpredictable or undesirable behaviour by adversarial actors. Interviewees also noted that AI could reduce the barriers needed to engage in conflict (i.e., with publicly accessible tools like content creation model GPT-3, anyone can create "fake" content).[122] In this way, AI technologies can contribute to international conflict outside the realm of military conflict, in line with arguments in the literature that dual-use applications like deepfakes represent a threat to national security (Citron and Chesney, 2018; Allen and Chan, 2017; Lin-Greenberg, 2020). With AI seen as a technology that can

[122] GPT-3 is an AI language model that uses deep learning to generate human-like text. See more: https://gpt3.website/

vastly leverage an adversary's advantage, interviewees cited how China and Russia may use AI technologies in a military context as a key concern to Western states. For U.S. interviewees, these concerns formed at least a partial justification for developing the DoD AI strategy. Interviewees generally highlighted the U.S. and China as the most advanced actors in terms of military AI capabilities, a view supported by existing literature (Fedusiak, Melot and Murphy, 2021). The U.K was also highlighted across interviews, not as a world leader but as a state actively engaging in military AI. This assertion is seemingly supported by the 2021 U.K Integrated review, which included the announcement of an AI Defence Centre, and by the relatively sparse information on which other states have discussed military AI at length.

One "tool" to help de-escalate strategic competition and tensions within the international landscape, highlighted particularly in the U.S.-focused interviews, was the use of strategic messaging as a signalling tool to the international community. U.S. interviewees underscored that a public document like the DoD AI Strategy Executive Summary helps develop trust from the domestic population, thus adding public support to the state's goals and objectives. Interviews also felt that the public release of such documentation could shape other states' approaches. The policy landscape appears to prove this point, as the OECD draws significantly on the DoD responsible principles for AI (Christie and Ertan, 2022*).* At the same time, NATO's AI Strategy sets out principles of responsible use of AI, the contents of which are also considerably aligned with the U.S. (Christie et al., forthcoming*).*

Somewhat captured under the "strategic" umbrella is the consideration of how AI in military contexts complies with existing ethical and legal frameworks relating to the conduct of war and where there are gaps. A significant discussion point related to how adversaries would use AI described potential activity that would undermine the values set out by the U.S., U.K and NATO allies. The environments in which it is deemed appropriate or desirable to deploy AI systems will differ between states. There are uncertainties about how AI will influence the legal aspects of war and how AI can be incorporated into a way that abides by legislation such as the Law of Armed Conflict (LOCA) and is in line with just war principles of proportionality.

While the topic of ethics did not explicitly emerge as a major theme in the interview or observation-based datasets, this is potentially partially explained by the interview design and questions. Interviews did not explicitly steer the scope of interviews to focus on ethical challenges. Additionally, the language used to refer to similar issues of transparency and oversight tended to use language favouring discussions of risks, safety, and responsibility. This latter explanation may well be an

example of Cohn's (1987, 712) description of how military communities use "numbing language" that removes the human as a referent object and can thus dismiss human issues as relatively unimportant. While discussions on ethics and AI frameworks have been explored extensively in the civic sphere (e.g., Fry, 2018; O'Neil; Jobin, Ienca, and Vayena, 2019), the military is relatively less transparent and does not discuss ethics in the same terms. Notably, the fieldnotes collected through observing events highlighted that industry vendors had "no answers" on topics relating to responsible AI and the technical design of AI infrastructure, with the sense that ethics "wasn't their job". This finding may be partially due to the lack of resources and existing published guidance. Since observant practice data collection was completed in early 2020, the release of DoD and NATO AI principles for responsible use of AI appear to have contributed to maturing and productive discussions of ethics in AI. Interviewees referred to international norms and ethical frameworks in a vague sense and with good reason. While ethical frameworks exist that apply to civil contexts (see Jobin, 2019 as one example), Taddeo et al. (2021) point out that there are no specific ethical frameworks for military AI. Phrases such as "responsible AI" or "AI safety" allude to the same themes. A small number of interviewees did highlight the potential ethical benefits of increased precision of AI, which could minimise collateral damage and friendly fire incidents. These discussions are well underway in the academic literature (Noorman and Johnson, 2014; Scholz et al., 2020) and include the literature described in section 8.4.2, which discusses the ethical and legal implications of taking humans out of the loop. Furthermore, interviewees across interviews highlighted the use of legal frameworks and potential norms or principles as a tool that can help de-escalate arms race dynamics. As this research developed and discussions of military AI innovation became more prominent in defence communities, interviewees increasingly demonstrated an emerging consensus on the importance of international collaboration in terms of norms setting and responsible use, with more detail and conviction than in the earlier U.K. interviews. The U.S. interviewees highlighted the need for norms to encourage responsible AI in the international community, aligning with language in the DoD AI strategy, frequently highlighting U.S. collaboration with allies and parties on such themes. These perspectives add detail to broad calls in the literature for states to engage in military AI competition to do so in a way that enhances stability and minimises the risk of further escalation (Venema, 2021; Kania, 2017).

### 8.4.5 Reflections on possible mitigations

Discussions on the security implications of AI often drew out interviewee assessments of their national approaches in ways that inform a response to the second research question. Foremost, this research highlighted a fragmented approach to military AI, in the context of an immature landscape where the implications, and mitigation options, are still unfolding. Clear increases in national funding and resource towards military AI themes, the creation of coordinating bodies like the JAIC or AI Defence Centre, the release of military-focused AI strategies, and emerging enthusiasm for norms development all represent ways in which actors are hoping to mitigate negative implications of military AI. This research revealed a messy and swift-moving landscape where much of this activity is in relatively early stages. The U.K. interviews revealed broadly that as of early 2020, the U.K. did not have a known methodology or framework for defence sector partners to approach technical challenges. In contrast, U.S interviewees described a wide range of approaches, many but not all related to the JAIC. The U.K. interviewees had little awareness of U.K. government or MoD initiatives relating to military AI technologies beyond major capability-development programmes such as Project NELSON, beyond one interviewee's analysis that MoD innovation hubs were not particularly useful. This difference in national reflections is reflected in the policy landscape, where the U.K. was yet to publicly publish a national defence-AI strategy or any detailed information on its' approach to AI safety, security, or ethical issues.

There was consensus among U.S. interviewees on the DoD AI executive summary that it helped draw a line in the sand to formally state the U.S. approach to AI in warfare, including a discussion of AI safety and security. The influence of the DoD's output can be partially measured in how little public backlash it has received in scholarship and think-tank analyses and through the subsequent adoption of the NATO AI Strategy, which drew heavily on DoD wording, particularly on responsible AI principles. While policy-focused literature raises the concerns of a growing capability gap between states who can afford to invest in AI innovation (Fiott, 2017; Gilli, 2020), those interviewed focused more on the need to agree on norms early on while there was still a chance of interoperability. In this sense, the fact that many states (including the majority of NATO members) have not yet published publicly accessible material on AI was interpreted by interviewees to mean that now is the right time to proactively agree on a process for adoption and capability-building. NATO was seen as the primary venue to facilitate support to allied nations. However, the Five Eyes network was cited across interview sets as a trusted network to share information and potential capabilities. The U.S. Partnership for Defense represented another alternative grouping of nations, again on the topic of AI in defence.

Agreeing on the significant impact of AI integration into military contexts, interviews also believed that AI and other emerging technologies will only intensify increasing reliance on technology. There is, therefore, a need for personnel who understand the technologies and its implications. Interviewees perceived a shortage in relevant technical staff, such as AI researchers and data scientists. A particular need was highlighted on the public sector side within the civil service and government defence bodies, particularly the staff organising procurement or setting requirements, and for government and military decision-makers at all levels. Interviews also stressed the need for interdisciplinary teams to approach military AI innovation, which may provide advantages in the diversity of thought and leverage of knowledge from different fields.

Interestingly, some interviewees across all interview sets referenced gender as an attribute, arguing that an increased representation of women would provide advantages in capturing a broader range of views on the implications of design decisions. These research findings generally suggest that research and decision-making would benefit from the contributions of a range of experts, including but not limited to technical, legal, and public policy and from across the social sciences, including engagement with critical research. The idea that AI research and decision-making would benefit from the inclusion of a range of diverse voices is reflected in the broader literature on AI development (Fry, 2018; O'Neil, 2017; Brundage et al., 2018). However, reviewed relevant scholarship tended to focus on diversity regarding the need for greater AI understanding across defence communities (Christie, 2021) and collaboration between academia, industry, and government bodies (Lucarelli, Marrone and Moro, 2021).

The challenge in attracting talent was recognised within the context of an overall shortage of relevant personnel overall, with the private sector competition for workers with higher salaries and competitive employee packages that outstripped perceived offerings for U.S. or U.K government employees. There was also a shortage of skilled expertise within military infrastructures. Interviewees highlighted how few senior military leaders held technical acumen and how the nature of a military career with frequent and potentially unrelated deployments does not usually lead to technical specialists. These findings mirror the perceptions in the literature as Cummings (2017) outlines that the global defence industry cannot compete with the commercial sector as skilled personnel move to the private sector, discussed in more detail in section 8.4. The global defence industry is therefore falling behind its commercial counterparts in technology innovation, and the gap is only widening as the best and brightest engineers move to the commercial sphere (Cummings, 2017*).*

## 8.6. NATO's potential

The final finding discusses the possible role of international military organisations as a productive mechanism for developing norms and common approaches to the use of military organisations and explores NATO's opportunities and challenges to contribute to this domain. It represents a strategic and conceptual contribution to the literature, drawing on military experts' perspectives on why states may choose to collaborate to minimise capability gaps between allies and highlighting debates on how far NATO may be effective as a coordinator of AI activity.

The U.K. and U.S. interviews highlighted the consensus view that activity to minimise the risks associated with AI could not be unilateral. For example, a U.K. interviewee highlighted that if the U.K. put restrictions on its innovation and the rest of the world did not, the U.K would be disadvantaged, and the controls would be likely insufficient. Unilateral regulation on how AI should be designed or deployed was thus dismissed as an anti-competitive measure which would simply move innovation abroad. This rationale is reflected in economics-focused literature. McGuire (1983) concludes that such uncoordinated regulation would be an undesirable consequence to any state, particularly smaller economies, creating disadvantages in broader production and trade activity. U.K. interviews highlighted the role of international norms as potential mitigation in this regard. They suggested that multinational approaches to AI would be more effective in mitigating the technology's negative implications (McGuire, 1983). This finding informed the interview design for the U.S. and NATO interviews to further explore the role of alliances, which subsequently identified multilateral norms-building activity as the most promising mechanism for stability in an era of AI-enabled systems. In an environment where unilateral abstention risks falling behind, interviewees were pessimistic about what national initiatives could promote international behaviours. All interviewees supported the concept of international norms on how AI should be used in military contexts.

The NATO interviews, in particular, explored the possible ways in which NATO can contribute in this regard. At the time of the interviews, it was unclear to interviewees how exactly aspects of military AI innovation fall under NATO's remit. However, there was a strong consensus across interviewees that NATO may help encourage capability building among members and facilitate consensus-building efforts. On the other hand, interviewees were divided on how far NATO could act as a moral arbitrator for the technology, feeling that it was beyond the scope for NATO to provide ethical guidance. Interviewees had a clear understanding of NATO's core structure as an organisation

that requires consensus to act and highlighted that NATO would not drive innovation itself. This viewpoint aligns with Soare's (2021) broader argument that NATO is not a leading innovating agent and was unlikely to develop capabilities in-house. The interviewees' lack of consensus shows that some, but not all, of those interviewed agreed with recommendations in policy-focused literature that NATO leadership must work to instil norms and operationalise AI principles across the military alliance (Gilli, 2020; Christie, 2021).

The broader policy landscape lays out some indications of how NATO perceives AI. In strategic military terms, NATO's 2030 agenda firmly highlights AI as a critical priority area for NATO and stresses that NATO cannot afford to ignore AI, a perspective demonstrated again through the NATO AI strategy (2021). To some extent, this lies in tension with the role NATO seems set to take with the release of the NATO AI Strategy, which outlines six principles for responsible AI. The status quo seems to be careful with wording, referring to "responsible AI" rather than drawing on value-laden terms or detail on how ethics may be applied in practice.

NATO's announcement of the Innovation Fund demonstrates their intention to improve public-private partnerships across the Alliance and between the organisation and industry partners directly. The language alluding to a trusted marketplace appears again to address concerns about supply chain risk. NATO's strategy summary highlights the organisation's role as a capacity-building coordinator and a collective defence mechanism against the malicious use of AI by adversaries, highlighting the core roles NATO expects to play in this unfolding space. Finally, NATO can also contribute to the development of terminology and standards in a military-specific context through the NATO Standardization Office, highlighted in interviews and the literature (Gilli, 2020, Pepe, 2020). Pepe (2020) highlights how NATO could establish interoperability standards, including common technical standards and set out norms of use for AI in military contexts.

With AI representing a strategic disrupting force in and beyond the military context, interviewees recognised that NATO can only facilitate discussion and agreement on military matters. Interviewees highlighted other potential avenues for additional dialogues, such as the UN or EU. These perspectives align with calls in the literature and policy landscape for the UN and EU to continue their work relating to AI in military contexts. Boulanin et al. (2020) published a policy paper highlighting the benefits of EU-driven principles for AI in military contexts, and there is a considerable discussion within the literature on the UN's GGE discussions relating to LAWS in particular (e.g. Horowitz, 2016; Brundage et al., 2018; Morgan et al., 2020). Some interviewees

reflected that, particularly for technical requirements or standards for AI-enabled systems, standards-setting organisations such as the International Standards Organization might be more appropriate as a venue where processes for testing and evaluation requirements can be agreed upon across actors. These proposals appear to conflict with the argument in the scholarship that NATO may be well-placed to coordinate technical and non-technical standardisation efforts, particularly in relation to interoperability (Pepe, 2021; Gilli, 2020). An explanation may be that NATO has a specific scope as a military organisation, while ISO and equivalent structures can determine standards for fundamental technical components. Thus, standardisation efforts by NATO, focusing on the international operability of military systems, and the broader standardisation efforts by the ISO, are not mutually exclusive.

Interviewees reported a range of potential pitfalls and challenges to effective NATO activity relating to military AI. U.S. and NATO interviewees reported that NATO was a bureaucratic organisation particularly plagued by existing challenges relating to interoperability and information-sharing. The majority of interviewees that discussed NATO activity (all but one) felt there needed to be more awareness of AI technologies at NATO. The capability gap is a challenge for NATO at a strategic level but was also a factor that U.S. interviewees cited as a reason for limited engagement with NATO on AI. U.S. interviewees did not primarily highlight NATO within the international engagement on military AI, as they considered many other members too immature in their approach. Instead, they referred to the FVEY network as a more aligned trusted network. This range of challenges is reflected in varying degrees in the literature. While the fact that NATO faces difficulties in information sharing and operability is well-acknowledged in the scholarship (Lin-Greenwood, 2020; Dufour, 2018; Pepe, 2020), the perspectives of underinformed decision-makers across the NATO enterprise and at many member states and how this may limit enthusiasm for U.S. engagement on military themes via NATO mechanisms, represents a strategic contribution to the literature.

The breadth of NATO's activity to date informs an answer to the second research question and reflects emerging strategies to mitigate the challenges relating to military AI. Providing a platform for discussion and consensus-building, NATO has secured member agreement on the NATO AI strategy within the broader EDT roadmap and has several initiatives including DIANA and the Innovation Fund that can leverage the strengths of individual states to achieve collaborative capability building. This research has highlighted that while NATO is not likely to develop AI capabilities itself, there are significant opportunities to develop international norms, as well as the agreement of relevant

international standards. The realisation of these opportunities will depend on a number of factors, including the motivations and engagement of national states.

## 8.7. What now?

The duration of this research tracked a period of exponential interest in military AI tools and techniques while simultaneously highlighting a lack of awareness about the implications of military AI, including among relevant military-focused communities. This finding aligns with the literature, which sets out the need for greater decision-makers' awareness of the nuanced dynamics and consequences of military AI technologies (Chadi, 2021; Johnson, 2021). In the findings, high-level talking points were consistent as interviewees perceived a pressure to innovate. A significant number of interviewees described a “Wild West”-like innovation landscape, with innovation happening across a considerable number of actors who may or may not work with the defence sector. These findings align with calls in the literature (Soare and Burton, 2019; Kania, 2017b; West, Whittaker and Crawford, 2019) for greater research into the strategic use of AI in the military. As Kania (2017, 39) highlights, a greater nuanced understanding of the associated strategic challenges of AI capabilities would benefit states, for example, better enabling the U.S. to counter Chinese advances in the field effectively. This research suggests this argument can be extended to NATO and its allies, against adversaries, including non-state actors who will increasingly utilise AI systems for their own gain (NSCAI, 2019).

This research revealed that the information put out to decision-makers at events during this data collection period was often carefully selected and targeted at a specific audience: for example, impressing the need to innovate with fewer bureaucratic controls when facing an audience of public sector workers. This targeting risks over-enthusiasm for AI technology and under-informed defence personnel, who are then ill-equipped to outline requirements to industry in a way that facilitates effective development while minimising negative consequences of the technology. Technical personnel identified many challenges in this space and remain relevant barriers for specific applications, particularly where complex algorithms impact safety-critical applications and where personnel need to fully understand and trust the technology they are using in a battlefield environment. The U.S. DoD and the U.K. MoD have highlighted these challenges as active research areas, while challenges including adversarial AI and refinements of AI use-cases are generally actively researched across academia and industry. While these technical challenges remain relevant,

much of the defences described by interviewees rely on implementing existing risk management techniques. These included following strong cyber hygiene policies and procedures to prevent the compromise of AI systems or data and engaging with rigorous testing and evaluation processes to cover the TEVV aspects of AI-enabled systems.

While the solutions to technical challenges appear to rely mainly on technical solutions (either inventing a technological breakthrough or framework that would address challenges like explainable AI or data-cleaning), the strategic implications of AI technology are more complicated and controversial. As modern conflict relies increasingly on emerging technologies (West, Whittaker and Crawford, 2019; Payne, 2018), interviewees perceived a challenge to the current global order. The pressure to innovate draws heavily on the signalling and actions of other states – especially perceived adversaries. NATO describes the malicious use of AI by hostile actors in the NATO AI Strategy statement. At the same time, interview transcripts and event fieldwork data highlighted how staff justified their priorities according to how adversarial actors might use AI. Without insight into Russian and Chinese military AI strategies, it is difficult to assess how far AI arms race dynamics are driven by perceptions alone. This obscurity means that states must rely heavily on their perceptions and assumptions when forming and implementing policy, a topic that interviews reflected on across the research. This emphasis on adversarial interest in military AI as a motivation to act mirrors Johnson's (2021) assumption that it is the perceptions states have of each other's capabilities that matter, potentially more so than the actual capabilities of the AI itself.

Considering the strategic implications of AI in warfare, one must also consider the broader areas of how other emerging and disruptive technologies will interact or be leveraged by AI. It is not often possible to separate the impact of AI against the emergence of big data or autonomy (two other key focus areas NATO has, in addition to AI, within the EDT roadmap). AI is better understood, as one U.S. interviewee described it, as "electricity" rather than any particular set of applications, an assertion also made in the literature (Horowitz, 2016; Johnson, 2021; NCSAI, 2019). AI-enabled algorithms *can* be deployed to support almost any task, and it is practically impossible to capture all the potential implications of the technology before its widespread adoption. This characteristic highlights the need for thorough consideration of now "responsible AI" principles that can be operationalised given the challenges in identifying potential risks to mitigate. Therefore, the period of this research captures the time in which the U.K., U.S. and NATO grappled with military AI topics with varying degrees of transparency. Between 2019 and 2021, significant groundwork has occurred in scholarship and within the U.K., U.S. and NATO to understand the implications of AI in military

contexts. The last five years have demonstrated the success of niche use cases such as Project Maven, while the U.S. and U.K. both set up infrastructure (DoD JAIC and U.K. AI Defence Centre, respectively) to coordinate military AI adoption. Principles have also been agreed upon publicly by the DoD and NATO members through the NATO AI strategy, providing a starting point for further clarifications on AI's responsible use and the development of AI doctrine by non-NATO members. At a policy level, NATO and DoD published strategies represent a significant shift from 2018, potentially replacing some of the uncertainty captured through the U.K. and NATO interviews now that Strategy publications are available. Beyond the proliferation of similar policies and extension to additional actors, it is now operationalising the policy that is the next step. The U.S. has had slightly more time to do this. It is already possible to evaluate the success of various DoD initiatives as examined through the U.S interviews in Chapter 7 and emerging analysis in the literature (Tarraf et al., 2019). While this may be an advantage to the U.S. in terms of first-mover advantage (Gilli, 2020), other states may benefit from observing how early adopters succeed – or not – in their endeavours (Singer, 2009). For example, smaller NATO members, who have fewer resources to devote to AI, may be able to rely on NATO's efforts to develop capacity building, benefiting from military AI tools first commissioned for or designed by others. This dynamic raises implications for capability development as considered within the NATO interviews. Interviewees described a nascent field where many NATO members had yet to consider AI in military contexts.

Particularly as states with more capabilities take the lead in developing military AI capabilities, strategic concerns include a growing capability gap between states who can or cannot utilise AI capabilities and corresponding interoperability between allies (Fiott, 2017; Gilli, 2020; Dufour, 2018). This research adds to this current literature discussing how AI may exacerbate the NATO capability gaps and outlines the current approaches taken by NATO before the formal release of the NATO AI Strategy. This research suggests that NATO is well-placed to be a *facilitator* of technological innovation rather than a *driver*, supporting Pepe's (2020) argument that NATO is best equipped to coordinate efforts to increase interoperability between allies.

Literature has previously discussed how AI breakthroughs in research and technological innovation can be utilised by any actors (Horowitz, 2018). This situation contributes to the argument that if research is done outside a secure environment, its benefits are available to all, including states perceived as hostile threat actors (Horowitz, 2018). The extent to which information-sharing will occur between NATO members is an open question that feeds into broad discussions on military interoperability and trust (Gilli, 2020; Lin-Greenberg, 2020). While it is for the benefit of the U.S.

military that its allies have adequate defences, this benefit might be outweighed by the risk of sharing a sensitive capability among the relatively large group of 30 member nations. Such sharing raises the risk of that capability or information being compromised by hostile threat actors. This risk of information-sharing across alliances is well-known (Lin-Greenberg, 2021), and states appearing to mitigate this through the use of alternative mechanisms which may represent a more trusted network (US Partnership for Peace as one example, others highlighted through NSCAI reporting).

## 8.8. Conclusion

These findings come together to provide a rich insight into current military AI innovation. In doing so, we can provide answers to the research questions referred to throughout this research. The findings reveal a broad range of implications of military AI technologies that form an explicit answer to the first research question. Exploring aspects of the technical literature, innovation landscape, and current approaches to military integration of AI tools, this research provides a categorisation of technical, operational, strategic, and ethical and legal security challenges. Conducting this research from late 2018 to early 2022, there has been an increasing discussion of the nuances of military AI both within available literature and across expert communities, particularly relating to the strategic implications surrounding technological competition and geopolitical tensions.

Alongside the growing acknowledgement of these implications across military communities, several actors have started both investing heavily in AI, undertaking efforts that include attempted mitigation of various security challenges. In exploring such activity, this research addresses the second research question by outlining current approaches undertaken by the U.K., U.S, and NATO. First, each actor has shown an increasing interest in understanding the implications of military AI, as demonstrated through the release of public reports by each actor and the creation of government Centres within the U.K. and U.S. At the conclusion of this research, the U.K.'s Defence AI Centre and the DoD's JAIC were described as the focal points to coordinate AI activity related to military contexts, both with dedicated R&D funding. Second, an analysis of U.S. activity revealed ongoing efforts to address the technical challenges surrounding data and rapid deployment, including through contentious initiatives such as Project Maven. Next, the U.S., U.K., and NATO have each amended or created mechanisms to encourage greater collaboration with the private sector, including disruptive start-ups who are often at the forefront of AI innovation. This response reflects an acknowledgement that the integration of

military AI technology must be more rapid than previous procurement allows, and that national militaries no longer hold dominance in technological innovation. Another major activity stream has been the development of relevant doctrine and AI Strategies. This research's analysis of the DoD AI Strategy found it was perceived as a heavily influential tool by U.S-focused interviewees, in shaping the U.S. response to a range of technical, organisational and strategic challenges. The subsequent release of the NATO AI Strategy and UK Defence AI Strategy showcase a range of commitments made by each to roll out reliable AI tools. Finally, this thesis has explored the international responses to military AI trends and highlights the importance of NATO, as an example of a relevant military organisation who can act to mitigate potential instability and uncertainty. Particularly acknowledging that smaller states often will not have the resources to tackle such challenges alone, NATO's coordination function has been demonstrated through its creation of public-private innovation mechanisms, facilitation of reports, and setting out of Allied priorities through the publication of the NATO AI Strategy.

Continuing to reflect on the second research question, this research highlights open challenges where there is no obvious mitigation underway, for example, with the potential establishment of agreed legal frameworks, or the opportunities for NATO to contribute to the development of technical standards for AI in military contexts. Equally urgently, there are several challenges which consistently come through this research which remain unaddressed, including the lack of a holistic approach to combating hype and under-informed decision-making, the finding that effective military innovation often requires a change in organisational cultures to risk, as well as broader strategic challenges around mitigating AI arms race dynamics where these lead to the deployment of immature or irresponsible tech by any actor. Looking towards broader discussions on geopolitical tensions and power, questions remain on how any norms or laws on military AI can prevent misuse or hold actors accountable where AI is used, or performs, unacceptably.

This discussion of research findings in the context of policy developments and relevant literature highlights areas in which this research has contributed to existing scholarship and policy. For example, particularly for security studies and strategic-level literature, the first finding on increasing awareness levels within military-focused communities finding highlights an increasing understanding of the implications of AI in military contexts, possibly moving beyond what the limited security studies scholarship on the same topic as noted by Payne (2017). This research also adds to the literature through participant observation and the collection of expert perspectives through a period of intense re-focusing on military AI literature. Applying observational techniques to military

conference and trade show environments has revealed key topics and discussion points as military stakeholders grappled to understand and benefit from emerging technologies. These findings demonstrate Jackson's (2016, 2) argument that examining such spaces can reveal the "barometer" of sentiment across relevant communities. This research has been inspired by Jackson's (2016) approach to collecting data on how military AI topics are explored in these spaces in ways that are not currently available in published literature. The second finding highlighted the overwhelming consensus across this research that AI will impact military contexts in the near future. The belief that AI would change the nature of warfare was one represented across the collected data, as interviewees and presenters at events highlighted how AI systems' processing capabilities and speed were a game-changer. Academic literature supports this demonstrated belief (e.g. Johnson, 2021; Payne, 2018), as do think tank and research institute reports (e.g. Boulanin, 2019, Allen and Chan, 2017) and government materials, including the U.S. Department of Defense AI strategy (2019). Analysis of this finding contributes towards addressing the first research question on the implications of AI in military contexts and contributes this analysis of expert interviewee perspectives to the scholarship. Third, this chapter explores the topic of military AI innovation and the drive to motivate. The two main reasons presented by interviewees on the need to innovate first highlighted the technical opportunities afforded by AI, and second referred to the perception of how adversarial state actors, namely Russia and China, are innovating in the field. An exploration of military AI dynamics included discussing the increasing influence on private industry, particularly non-traditional defence firms such as large corporate technology firms, building on the analysis that governments face an increasingly competitive landscape in terms of resources and talent (Cumming; 2017; Lin-Greenberg, 2020). The fourth category of findings set out the key security challenges identified by the military communities engaged during this research, highlighting disparities in how such challenges may be effectively mitigated. The fifth and final main finding discussed how far NATO might contribute to the military space relating to military AI. These discussions found that NATO's narrow scope represented an opportunity to facilitate consensus-building efforts, such as the development of norms and future integration of ethical principles and capability efforts to enhance operability between members.

# Chapter Nine: Conclusion

This PhD thesis approaches the nascent, burgeoning topic of military AI. Considering current and unfolding narrow AI techniques, this work examined the question: "what are the implications of AI innovation in military contexts?". The research examined how AI, in line with other emerging technologies, holds significant implications on and off the battlefield in the near and distant future. Exploring the emerging and evolving landscape of military AI innovation required analysing how militaries and government defence departments understand and approach AI adoption and use. Doing so required an exploration of the dynamics of military AI innovation, which was undertaken by focusing on the U.K., U.S. and NATO contexts. This thesis has also addressed how actors are attempting to mitigate challenges identified in relation to the development and use of AI in military contexts. This thesis draws on literature, policy documentation, observation across relevant military conferences and trade shows, and expert interviews to understand how AI is perceived, developed, adopted, and operationalised in military contexts.

Many findings emerged from the analysis of the collected data, which can broadly be categorised under five themes. The first theme highlights the rapid growth in interest and awareness of AI-related to military contexts, describing how discussion and policy-focused material has proliferated since 2018 to include (generally) more informed understandings of the implications of military AI applications. The second theme reflects on how AI is understood as an enabling technology that represents a significant turning point for military affairs, arguing that current AI techniques will profoundly impact warfare in the near future. The third theme explores perspectives on military innovation to argue that the U.S., U.K., and NATO have all signalled their intention to intensely drive AI innovation in military contexts in ways that align with the literature on arms race dynamics. The fourth theme shows that militaries still face a broad range of barriers to the adoption and effective use of AI technologies, with some potential mitigations but significant uncertainty in many regards as to preferred ways forward. The final theme relates to the role of NATO as a facilitator for international consensus-building and argues, with caveats, that NATO may be well-placed to encourage international collaboration and norms-building related to AI in military contexts.

This thesis makes several original contributions:

- This research contributes several conceptual findings to the field. First, the findings of this research form a major contribution to the field though highlighting the need for states to adapt to a rapidly shifting but fragmented technological landscape. Drawing together the literature, emerging doctrine, and expert perspectives highlighted the trend of an early but intensifying competition between major actors that is not expected to deaccelerate soon. Second, this motivation to innovate raises several challenges related to the perceived need to dynamically adopt military AI capabilities at a rate that exceeds the activity of adversarial states – and requires addressing the numerous technical, organisational and operational challenges that currently prevent barriers to effective development and integration of military AI technologies. Effective adoption and deployment by national militaries will require a genuine shift to streamline and accelerate public-private relationships, particularly to include technical innovators that are not the traditional defence primes. National innovation will require an increased level of awareness and understanding in relevant communities including MoD procurement and across government policymaking structures, in order to prioritise accordingly. Relatedly, this research highlights that the types of high-level challenges facing innovating states are similar in nature, through the discussion of innovation within the U.K., U.S., and NATO Allied landscape. For example, all sets of interviewees highlight the realisation by expert communities that there there is an urgent need to increase know-how relating to AI and its implications in military contexts, particularly across military leadership and the talent pipeline relating to technical expertise. For all military and defence state bodies, the interaction around hype, and private sector innovation, must be considered when providing balanced information to decision-makers. Next, This research also highlighted several unsolved challenges around interoperability, and how military AI might work effectively in the field in terms of human-machine teaming considerations. While there is evidence that this is an active area of research, through the set-up of research centres and public reports, the consensus across the literature and interviews was that such implications must be clarified and mitigated before military AI tools can be reliably and safely deployed on the battlefield. Similarly, the discussion of “responsible AI” forms another conceptual contribution in highlighting the perceived utility of international norms as the most productive tool to avoid the use of AI in ways that the U.S., U.K.,

and NATO deem unacceptable. Over the course of this research, the release of national and NATO principles of responsible use of AI shows again that the field is unfolding to develop a more nuanced discussion on how to minimise the negative implications of AI. Actors must now face the challenges of operationalising these commitments and expanding their acceptance across the international landscape. The successful development of international norms relating to the use of military AI can be bolstered through broader coordination, for example tying into the development of technical standards or legal frameworks. Finally, in offering a holistic overview of the military innovation relating to military AI, this thesis also highlights challenges such as deterrence and strategic stability, which while not unique to artificial intelligence or military innovation, are complicated by the pace and virtual nature of emerging technologies. These challenges will only become more complex over time if more AI is deployed prior to the agreement of appropriate standards or mitigations, and this thesis therefore highlights that states, militaries and innovating actors should be strongly motivated to address the security implications as soon as possible.

- Empirically, this research contributes findings on several underexplored areas of military innovation relating to AI technologies in academic scholarship. Firstly, it adds to the increasingly active field of scholarship within the interdisciplinary field of military AI innovation. The data captured through interviews and observant practice contribute to discussions in areas including but not limited to security studies, information security, and other interdisciplinary endeavours, including geopolitics and work on military innovation themes. This research has explored U.K. approaches to innovation, a theme significantly underexplored in the literature that tends to focus on the U.S. and China, followed by Russia, as the leading innovating powers in strategic terms. Similarly, this research contributes important data by collecting perspectives from practitioners, military personnel at events, and policymakers. These interviewees were able to speak candidly under anonymity. They thus contributed insightful data on themes including, but not limited to, practitioner perspectives on military AI relationships, the perception of arms race dynamics by actors in the space, and attitudes towards international norms development and other possible mitigations. Aside from a RAND report (2020) and a high-level UN workshop summary report (Sisson et al., 2020), this research has found few attempts to capture the views of military experts in this way. Finally, this research has included views from experts over time. This data represents a valuable insight into perspectives within this

immature but rapidly evolving field, within which there is significant speculation. Conducting research in this time also allowed for some comparison of views between the U.S., with a DoD AI Strategy and extensive defence infrastructure designated to deal with military AI innovation, and the U.K., with no public strategy and a different, somewhat more fragmented, approach to the space at the time of interviews.

- At a policy level, this research has drawn out several strategic findings that will be of interest to policymakers and military decision-makers. Highlighting the rapidly increasing interest and activity relating to the adoption and use of AI in military contexts, this thesis has found that AI is set to have a considerable impact across almost every aspect of warfighting. Furthermore, in mapping out the range of military implications of AI in military contexts, this thesis draws together expert insights with policy developments and technical and policy-focused scholarship to articulate concerns in the field. These findings add to existing policy research on how NATO may represent a productive path to mitigate challenges. Analysing the data also revealed a series of challenges facing governments. These include the need for governments to modernise procurement mechanisms for emerging technologies and adapt to the increasingly critical role of the private sector, how AI may threaten aspects of strategic stability, and the need for innovators and militaries to enhance TEVV processes for AI-enabled systems. This thesis also demonstrates a lack of awareness across many aspects of the field, such as strategic security implications or potential policy tools to mitigate the negative consequences of immature or irresponsible AI. The findings highlight where gaps exist in terms of the open challenges relating to AI technology. These areas, including several challenges regarding the operationalising of AI principles to face challenges including MHC, trust, and reliability, highlight areas of particular interest for military communities.
- Methodologically, this research has approached an emerging and rapidly developing field. It used grounded practices methodologies to collect data within often shifting and disconnected spaces, mapping a fragmented field at a time when it was only beginning to take shape. This research utilised observant practice techniques of military/militarised spaces to observe discussions on military AI in a way that is not yet present across existing scholarship, complementing non-AI specific research which used a similar observational approach to examine military-focused environments (Jackman, 2016; Rech, 2015). Finally, this research also provides a path to conduct qualitative research with military communities in largely remote settings,

reflecting on the learning process to discuss network-building and access in this space given COVID-19 travel restrictions.

These contributions have been drawn from the data in line with the grounded practice-based methodological approach described in Chapter 3 and represent the culmination of almost four years of research, from Autumn 2018 to Spring 2022. This field has changed heavily during this period, a trend observable through the increasing number of relevant scholarship and open-source material tackling a similar scope. These include the recent proliferation of policy and security studies-focused literature described in Chapter 2 and the emergence of several key policy documents, including the DoD AI Strategy (2018) and the NATO AI Strategy (2021). Acknowledging that the field has come a long way in terms of discussions and debate on military AI, the findings of this research consistently highlight that increased awareness of the dynamics of AI innovation in military contexts is needed. Militaries must have the language and expertise to utilise emerging technologies such as AI.

The military AI landscape is rapidly changing in line with the increased prioritisation of AI for military purposes. The contributions of this thesis are based on data collected through U.K., U.S, and NATO interviews, as well as events taking a similar Western-focused perspective. Capturing these viewpoints has allowed for a nuanced understanding of how militaries and defence departments face barriers in terms of procuring AI technologies, for example, as well as the perceived drivers for military innovation within the context of the global geopolitical landscape. This thesis also examines the debates on arms race dynamics in the literature. It situates data collected within the scholarship to argue that there is an intense competition in developing and adopting AI across military contexts, driven in part by perceptions of what China and Russia are doing. These findings suggest many avenues for future research in this field. Examples include the possible exploration of additional national perspectives from additional communities or measuring the development of awareness and attitudes to military AI innovation over time.

This thesis considered the scholarship relating to AI in military contexts available for analysis as of December 2021. This thesis has also consistently referenced the rapid pace of activity across the commercial market for AI innovation, the policy landscape, and policy-focused and broader scholarship developments. Four examples demonstrate this point, referring to announcements and changes in February-March 2021. First, increasingly researchers are exploring the broader implications of AI in military contexts and doing so in a way that will engage a wider audience. For example, in March, researchers Buchanan and Imbrie (2022) released “The New Fire: War, Peace,

and Democracy in the Age of AI", a book focusing on the geopolitical themes of how democracies can harness AI. Second, the U.S. DoD approach to AI appears to have changed considerably, with the December 2021 press announcement that the JAIC would be enfolded into the Chief Digital and Artificial Intelligence Office (CDAO) as of February 2022 (Hicks, 2021). The rationale for this, and the specifics, are not yet detailed in academic or policy scholarship. Third, dynamics have changed significantly in line with geopolitical tensions, which may challenge the assumptions described by interviewees in this research. It was almost impossible to limit the spread of technology; in response to Russia's invasion of Ukraine, the U.S. has implemented extensive sanctions covering technology transfer to restrict Russian access to U.S. semiconductors and intellectual property (U.S. Treasury, 2022). Fourth, and relatedly, the war in Ukraine has seen private company Clearview AI provide its facial recognition technology to Ukraine to help distinguish between targets and non-combatants (Paresh and Dastin, 2022). This application, in this context, could be discussed at length relating to the use of commercially developed technology in a conflict scenario and the way states might be supported in conflict by private companies directly rather than via third-party state support. It also highlights various risks, including safety and ethical implications of deploying AI in conflict,[123] many of which are discussed in this thesis. The broader implications of such developments remain to be seen.

This research examined how the U.K., U.S. and NATO are currently perceiving and approaching AI, with data collection from early 2019 to mid-2021 and capturing updates until early 2022. The scope was particularly broad and was partly shaped by research participants who had the freedom to organically raise ideas and discussion points through semi-structured discussions (see Chapter 3). This research followed a grounded practice-based methodology that appears particularly well-suited to this topic, where academic literature lags behind doctrinal developments and rapid changes in the geopolitical landscape. There are many opportunities to use grounded practice-based methodologies further to build on this thesis' findings with greater exploration of observant practice opportunities and targeted interviews, with growing opportunities to review released data on documentation, patents, and developments on the international landscape.

This research has also allowed for the tracking of an emerging field over time. It is acknowledged that beyond the methodological challenges described in Chapter 3, this research recognises its

[123] Clearview's CEO has stated he does not want the product to be used to violate Geneva Conventions, but media reporting has highlighted several immediate concerns including misidentification leading to civilian harm (see Dave and Dastin, 2022).

limitations in assessing an uncertain and rapidly evolving landscape. Throughout (and particularly during the early stages of) this research, finding the best-placed people to interview or the most appropriate event spaces to observe was a challenge that I believe would be less pronounced today. The development of AI strategies and the creation of various bodies and public functions focusing on defence-specific AI has resulted in a growing community of informed expertise, in line with increasing research investments across state spending, industry-led R&D and in academia. This research collected insights from interviewees and event attendees who held various degrees of understanding. Some were aware AI would impact their security stances significantly within the next decade but felt they, and at least parts of the infrastructure they worked within, had a long way to understand the technology's potential. Others were staff who were, or had previously worked on, drafting defence AI policy or other focused non-public documentation. Some interviewees had led AI and data-science-focused programmes and represented some of the small numbers of personnel with the greatest insight into how AI policy was formed (though they still highlighted where defence department or military knowledge gaps could be addressed). Tracking the rhetoric and priority agenda of policymakers, industry actors, and military stakeholders over time would help assess how the space forms around AI and related emerging technologies such as big data and autonomy. Similarly, this research has noted how physical conferences and trade shows approached AI themes in 2020-2021 with some brief comparison to historical agendas and current forward-looking interest in AI. Focusing on how events pivot over time to address emerging and disruptive technologies would offer researchers into the perspectives of the community, using the conferences and shows as a "barometer" of sorts, as suggested by Jackson (2016, 2).

The data gathered through this research is particularly rich, including transcriptions from interviews which exceed 330,000 words total, several days' worth of observant practice event notes, and two policy documents. The discussions within this research draw on the main themes. However, discussions were wide-ranging, and chapters could just as feasibly have focused solely on a significant number of topics, including the cultural aspects of military innovation. These aspects include how diversity impacts decision-making,[124] how AI technologies were compared to other phenomena such as cybersecurity as a technological trend, the language used to describe AI drawing parallels to science fiction, AI and cyberwarfare, and AI activity at the level of sub-threshold warfare. By restricting the scope of this thesis to focus almost exclusively on the military applications of AI,

[124] For example, attributes such as the age or gender of decision-makers are all underexplored in this thesis, though several interviewees made comments that could be followed up on these themes.

there has been a layer of artificiality, segmenting AI capabilities. In reality, applications can be deployed against, or used by, military or non-military targets. In terms of taking aspects of this research further, the emerging nature of this field means that any of these aspects can be explored in great length, with exciting opportunities for researchers for the foreseeable future. The economic aspects of military innovation, or organisational innovation to reduce "valley of death" phenomena, is one related field touched on but underexplored in this thesis. Skilled researchers from economics, psychology, management, sociology, and a range of other disciplines can contribute valuable perspectives to this discussion. As interviewees often highlighted when describing the need for interdisciplinary decision-makers, this research space will only benefit from diverse exploration of the field.

How policymakers and state defence leadership approach the opportunities and challenges associated with military AI will be pivotal in the next few years. Particularly as states build out their approaches to AI, the subsequent challenge will be to operationalise national and regional policies (including the NATO AI Strategy). This research supports Christie's (2021) conclusion that there are significant unknowns in how policy will be operationalised in the context of military AI systems, particularly when it comes to AI principles and international norms.

This thesis has followed the growing understanding of AI among two "Western" state actors and one multilateral alliance as each defines its perspective on AI. At the time of submission, there is still a sense of uncertainty about the future, reflected in most interviews. Interviewees understood that AI is still an emerging technology. They understood that there will almost always be consequences of such technologies that are difficult to predict despite testing – with many implications fully realised post-deployment. There are no agreed best practices for robust testing in the absence of established AI-specific TEVV processes (Defense Innovation Board, 2020). Research into how military bodies can and should verify AI systems would prove valuable in technical security and safety terms. Given the tension between the demand for "fail fast" rapid development versus wishes for minimal risk and high confidence in the algorithm, how testing procedures are adopted is another open research question.

Focusing on the perceived priorities of each actor, and any released national strategy documents, highlighted the opportunities to research how actors *operationalise* AI policy. This opportunity applies in a practical sense in observing and analysing how states adapt their innovation methods to revitalise procurement processes and operationalise any commitments to principles for responsible

use.[125] In the year leading up to the submission of this thesis, military AI-focused bodies were announced (NATO's DIANA and Innovation Hub), created (U.K. AI Defence Centre, DoD procurement platform Tradewind) and dissolved (U.S. NSCAI). With a range of activities underway between actors, there are opportunities to evaluate the success of different approaches and attempt to capture each approach's impact. For all the above, such research will contribute to understanding how military innovation occurs and may also result in recommendations for a range of actors, including policy advisors, industry partners, standards-setting bodies, or risk functions across public and private sectors.

There are many ways in which this research might be expanded. Extensions of this research that engage with expert communities over time, alongside research extended to other communities, would be particularly enlightening. Such communities might include legal or U.K. MoD-equivalent staff and military personnel at various levels of seniority, including prospective end-users of mission-support AI systems. The existing scholarship would also benefit from research engaging policymakers and decision-makers from nations who have not yet invested in AI to the same extent as the U.K. or the U.S. A comparative study would be possible between states should the same "community type" (e.g., civil servants or military scientists) be surveyed between states. Research could also examine how perspectives differ depending on other attributes between different categories of interviewees. This research may include identifying and exploring how different communities (including potentially technical R&D, policy, government, various industry sectors and different size industry actors) approach the challenges associated with military AI. For some of these communities, research would be conditional on securing a security clearance or progressing through external approval or subject to non-disclosure agreements. In an era where travel is not restricted or complicated by factors including pandemics, being able to interview face-to-face may allow for greater rapport building and facilitate further network development, potentially resulting in greater access to relevant interviewees.

The findings of this thesis could be further enlightened by research into non-allied approaches in this field. As described in Chapter 2, section 2.5.13, research into the Russian state's understanding of AI in warfare is already underway. The gathering of perspectives from Russian-based experts would add value in revealing insights from relevant personnel themselves. Similarly, research into Chinese

[125] For further detail see: Christie, Edward H., and Amy Ertan. (Accepted/In Press). NATO and Artificial Intelligence. Routledge Companion to Artificial Intelligence and National Security Policy.

investments in military AI has helped enlighten this research. Any research that could shed further light on the PLE's approach would likely be of interest to the national security community worldwide (Kasia, 2017; Fedasiuk, Melot, and Murphy, 2021).

# Appendices

## Appendix A: Interview Questions

### A.1. U.K. Questions

1. What do you understand by the term algorithmic warfare?
a. In your perspective, is this phrase well-understood?
b. In your perspective, are there other phrases that might be more helpful to talk about these phenomena?
2. Are we seeing an AI arms race?
3. What kind of innovation is happening at the intersection of AI cybersecurity and the military?
4. Is there pressure to innovate in this AI cyber military area, and if so, why?
a. *If yes:* would this pressure to innovate be considered reactive or proactive or both?
5. Regarding AI innovation, how would you describe the relationship between private industry and the MOD in the UK?
6. What does the private industry and military engagement model look like on topics relating to AI and the cyber domain?
7. What additional actors impact how algorithmic warfare type capabilities are developed and deployed?
a. *Prompt:* So outside MOD and industry key industry access?

A.2 NATO questions:

(Interviews start by capturing participants' views on the strategic context and consequences of AI, then continue to **clarify NATO's capabilities and role** in shaping AI in conflict.)

High-level intro questions

1. **Core question:** In your view, to what extent is AI-enabled technology already shaping international conflict?
    a. Can you offer an example?

2. In your perspective, how will AI capabilities shape military conflict within the next decade? Are there emerging trends that are not yet present? If yes, which?

3. **Core question:** In your perspective, to what extent will AI military innovation lead to varying capabilities between NATO alliance members?
    a. How do differences in AI-enabled technical capability impact the responsibilities of states to others in the alliance?
        i. For example – should states with more sophisticated capabilities assume a greater protective burden for allies with fewer capabilities?

4. To what extent is knowledge being shared or copied between departments or branches of the military within a state?
        i. What information is shared internationally? What proportion of this information exchange is facilitated by NATO?

Understanding the security implications of AI in conflict

5. **Core question:** In your view, what are the main implications of military AI-enabled technology for international security?
    a. (Nudge if necessary) What are the strategic concerns/ what are the tactical and/or operational concerns?)

6. **Core question:** To what extent are non-technical concerns captured when determining AI applications in a military context? For example, how do political, economic, legal and ethical implications of AI technology shape design and implementation?
    a. In your view, to what extent is there an efficient process for developing and deploying military AI capabilities within member states? Does this process capture non-technical concerns? If yes, which? If not, which do you think that is the case?
    b. In your view, to what extent are security risks and opportunities effectively acknowledged (by decision-makers and developers) and effectively mitigated (by developers and other stakeholders)?
    c. In your view, is there a 'best-practice' approach to military innovation?
        i. If so, where, and how?
        ii. If not, is this something NATO could help design and share?

7. In your view, is AI being 'weaponised'
        i. If so, how? Is there an example that you can share?
        ii. In your view, what are the most pressing security concerns relating to 'weaponised AI' (AI being used as part of an offensive capability, or to carry out offensive activity)?
            - (Clarify, if required, that these concerns may be: political, technical, economic, human, legal, or ethical. Opportunities to steer.)

8. **Core question:** In your view, what is the best mechanism for addressing potential security concerns (including strategic security concerns) relating to military AI?

NATO's role: AI in Conflict

9. **Core question:** To what extent does NATO have the capabilities to shape the strategic direction of military AI technology? (Follow-up Prompt: Why?)

10. **Core question:** To what extent does NATO have the capabilities to mitigate potential consequences for security created by weaponised AI technology?
    a. What are the key debates that are happening between NATO allies in this space?
        i. How is NATO able to respond?

b. From your perspective, which actors do you perceive to be well-placed to influence the direction of military AI innovation, implementation and sustainability?

11. In your view, is there sufficient understanding across/within NATO departments to enable effective decision-making around military AI innovation? If no, what is missing? And why do you think that is?

12. **Core questions:** In your view, is there a role for NATO to incentivise member states to implement AI in a way that minimises negative outcomes?
    a. *Should* there be a role for NATO to do this?
        i. Which NATO functions/ institutions may be effective in shaping security strategies relating to AI? (If not mentioned: how does the CCDCOE contribute)
    b. In your view, what would be the best mechanism to incentivise responsible military AI practices?
        i. Do you think there may be other (non-NATO led) mechanisms that may effectively encourage responsible use of AI in defence contexts? Which?
    c. **Core question:** In your perspective, what are the main challenges *at a strategic level* for NATO to coordinate this kind of activity effectively?
        i. In your perspective, what challenges does NATO face when specifically considering *operational challenges* such as misuse, or unintended use, of specific applications?
        ii. In the same line, what challenges does NATO face when facing *tactical challenges* such as human-machine teaming and on-the-ground consequences.

        *End*

A.3 U.S. questions

Broad Set of Topics (not all will be asked)

Group 1: Section 1-4

Section 1: The U.S. Military AI Strategy

1. Core question: What prompted the development and release of the U.S. Dept of Defense publishing the AI Strategy?

2. Core question: How did the release of the U.S. DoD AI Strategy influence the development and deployment of military AI innovation?
   - Can you provide some example outputs of the strategy?
   - How was the DoD AI strategy designed to work with the subsequent release of branch strategies (USAF)?
   - To what extent is military AI innovation guided by other doctrines and/or published strategies?
   - How would you describe the overall reaction to the DoD AI strategy?
      - Military AI decision-makers/ from the military as a whole
      - U.S. and non-US commercial industry actors
      - Other state actors

Section 2: Terms and theory

*3.* To what extent do you see AI-enabled technology becoming a central tenet of national security capabilities?

4. Core Question: To what extent does the U.S. DoD see their military AI as contributing to the U.S.' technical competitive advantage?
   - To what extent does the U.S. intend to share military AI capabilities with allies/alliances?
   - How does the capability-building activities of other states influence U.S. decision-making relative to military AI innovation?

5. What other factors influence the direction of U.S. military AI innovation?

6. How does military AI innovation fit into the U.S. "Third Offset Strategy"?

Section 3: Fragmented Landscapes? U.S. Specific

7. From your perspective, which actors play a key role in determining AI strategy in U.S. defence?
   a. And why do you think that is?

8. In your view, how influential was Project Maven in terms of AI procurement, deployment, and awareness of U.S. capabilities?
   - To what extent have any challenges highlighted by Project Maven been addressed?

9. Core Question: In your view, how efficient is the relationship between commercial actors and the Department of Defence?
   - In your view, how far is this considered effective?
   - If there are challenges, how may these be mitigated?

10. To what extent has public perception and media attention shaped the U.S. approach to military AI innovation?
    - Why is this the case?
    - Are there other influential factors/ actors that you perceive as 'key influencers' for U.S. military innovation?

11. To what extent is military AI innovation *proactive, reactive,* or *both?*

Section 4: Institutions

12. Core Question: To what extent has the JAIC been an effective coordinator of AI activity?
    - How / how not?
    - Are which ways in which the JAIC has exceeded expectations?
    - In which ways has the JAIC fallen short of expectations?

13. In your view, which are the main government institutions coordinating military AI innovation in the US?
    - Are there other non-government actors that have significant influence in shaping AI innovation? If so, how?

Group 2: Section 5-6

Section 5: Security Management

14. To what extent does design, development and verification (testing) account for the potential security vulnerabilities in AI-enabled technology?
    - What defence-specific practices or security approaches are applied (compared with commercial AI innovation)?
    - Core: Is there a unified approach to technology-based risk management across the U.S. military? Why or why not, and to what extent?
    - § Is this more or less pronounced when it comes to military AI innovation?

15. To what extent is it possible to detect where an AI product (solution/ algorithm) may have security vulnerabilities?
    - Can you describe the process of remediating in such cases?

16. To what extent are AI algorithms monitored after deployment to test for security vulnerabilities and/or effectiveness?
    - Does this include testing if the humans are responding to the algorithm as originally intended?
    - In your view, are current AI security risk management processes effective?

17. To what extent are misuse / adversarial use cases modelled and tested automatically within the technology development /post-deployment process?

18. To what extent are the psychological aspects of human-machine teaming considered when planning AI product deployment?
    - How has the U.S. DoD addressed challenges relating to trust in AI and responsibility and accountability for actions recommended by AI systems?

- In your view, what are the effective and less-effective outcomes of how the DoD has addressed these challenges?

Section 6: Mitigating risk and misuse

19. Core: In your view, which institutions are best placed to lead the responsible development of AI in warfare?

20. To what extent is 'algorithmic warfare' or 'weaponised AI' inevitable?
    - Why/ Why not?

21. To what extent can 'weaponised AI' technology (or AI technology with the capability to be used in offensive warfare) be controlled to prevent use by adversarial actors?
    - If yes, how?

22. To what extent is there a role for:
    - *laws and/ or regulations* to prevent misuse of AI in warfare?
    - *international norms* to prevent misuse of AI in warfare?

## Appendix B: Research Ethics Approval

I planned my ethical review applications shortly after attending departmental seminars, which highlighted the importance of the ethical review process, both in terms of remaining conscientious of ethical aspects throughout the research and producing high-quality research. For projects carried out in the department, the options would be to not go through an ethical review process, go through an online submission and self-assess certification, or submit the application for a full review by the Royal Holloway Research Ethics Committee. As my research is concerned with and actively interacts with human participants, I considered it necessary to secure ethics approval to cover each part of my research.

While, in hindsight, none of my projects met a clear threshold for a full approval review, which might typically be reserved where participants are put in challenging circumstances, I erred on the side of caution for my first application. Submitting the project through the complete ethics review process with the Research Ethics Committee enabled me to ensure that my project satisfied ethical requirements at a university level. The applications involved sending a complete set of questions, confirming that the content did not overstep into any "grey zones" regarding ethical reviews. The ethics approval for the U.K.-focused interviews ethics was granted after three months. At that point, I was pursuing a fellowship in Brazil, a long way from my identified interview participants in the U.K. The project interviews did not commence until early 2020. I had, perhaps naively, not accounted for the delay; there was little precedent of full ethics approval within the information security department. Nonetheless, this was a valuable exercise as the reviewers highlighted specific requirements which I had not considered, including, for example, that the generic RHUL data storage reference link was out of date and did not refer to GDPR.

Table B.1 outlines the ethical approval forms submitted and approved throughout the thesis:

*Table B.1: Ethical approval by project*

| Project Title | Approval Type | Timeline (and expiry) |
|---|---|---|
| Fragmented Landscapes (Interviewing Practitioners) | Submitted for Full Approval | 10-April: Submission to Ethics Committee<br>**17-Apr:** Response 1 from team requesting separate information sheet<br>29-04: Author sent updated submission<br>02-May- Author requests update, reviewers are sent a reminder |

| | | |
|---|---|---|
| | | 10-May 2019: Response 2 Additional information on GDPR adherence required<br>10-May- Update submitted<br>20 May- Author requests update<br>24-May- Response 3: 'All three reviewers have approved, and it is currently with the Chair'<br>09-July- Author requests update<br>15-July- 'It is currently with Chair for sign-off'<br>15-July- Approval Confirmed (-March 2020)<br>Extension to March 2021 Granted [Justification – COVID-19] |
| Observant Practice (Expo and Conference Visits) | Self-Assessed | Instant Approval – May 2019 (-May 2021) |
| NATO Capabilities (Interviewing relevant experts) | Self-Assessed | Instant Approval – September 2019 (-September 2022) |
| US Approach (Interviewing U.S. relevant government and defence experts) | Self-Assessed | Instant Approval – September 2019 (-January 2022) |

Appendix C: Consent Form

Note: Consent form content had identical consent wording across three interview sets and was provided to all interviewees prior to interview. The only change was the sub-title under "Consent Form".

**Consent Form**

US Security Strategies in the Era of Military AI

| Initials | Statement |
|---|---|
| | I confirm that I have read and understood the Participant Information Sheet |
| | I have had the opportunity to ask questions and had them answered |
| | I understand that what I say will be treated as confidential by the researcher. |
| | I agree that data gathered in this study may be stored anonymously and securely. |
| | I understand that my name (or chosen name) will not be used in any written reports or presentations. |
| | I understand that my participation is voluntary and that I am free to withdraw at any time without giving a reason |
| | I agree to take part in this study |

**Name:** _____

**Participant signature:** _____

**Researcher signature:** Amy Ertan

**Date:** _____

# Appendix D: Personal Information Sheets (PIS)

## D.1 UK-focused PIS

**PARTICIPANT INFORMATION SHEET**

Fragmented Landscapes: Military Cyber Security in an Era of Algorithmic Warfare

**Invitation to take part**

You are being invited to take part in a research project. Before you decide whether or not you would like to take part it is important for you to understand why the research is being done and what it will involve. Please read the following information carefully and discuss it with others if you wish. If you have any questions or particular concerns, please let me know. You will find the relevant contact details on the last page of this document.

**Why is this research being done?**

This six-month study forms part of my three-year doctoral research project that aims to answer the research question: *'What are the underlying drivers for the development of 'AI' technology, with a particular focus on cyber security in the military domain?'*. It thus explores the dynamics surrounding the development of cyber-offensive capabilities, and the alleged 'arms race' that is happening in artificial intelligence.

Overall, the key objective of this project is to deliver an overview of relevant factors and drivers for AI technology in defence and security. This will be used as a scoping exercise for my broader PhD research on the topic. The overview will be developed through a series of semi-structured interviews. Interviews will be held with practitioners, who are familiar with the topics of emerging technologies, cyber security, and military innovation. Findings from the interviews will form the basis for future research directions.

**Who is doing the research?**

I am a PhD researcher in the Information Security Group, Royal Holloway. I am part of the Cyber Security Centre for Doctoral Training, which is sponsored by EPSRC grant EP/P009301/1.

**Why have I been chosen?**

You have been invited to participate as your role and/or experience is thought to position you as well-informed in this field. More specifically, you have been chosen based on your expertise in cyber security and AI technology in a national security or military defence context. All interviewees are therefore invited through current or previous career experience. This project may include practitioners from private industry, government, research organisations and the military.

It is expected that up to 20 practitioners will be interviewed as part of this project

**Do I have to take part?**

No. It is up to you to decide whether or not to take part. If you do decide to take part, you will be asked to sign a Consent Form. You can withdraw during the survey and interview process at any time without giving a reason and your data will be removed from the study. Once the survey or interview is completed you can still withdraw your data up to the point where the data has been analysed and anonymised, so that your identity cannot be determined. Your decision to take part or not to take part will involve no penalty or loss, now or in the future. Please contact Amy Ertan or Rikke Jensen to withdraw (contact details listed below).

**What will taking part involve?**
If you agree to take part you would be expected to complete a short (fewer than 6 question) survey by email, as well as complete a one hour-long interview with me in person or over the phone/equivalent audio option. Questions will be related to the research question outlined above and will be linked with the wider themes of cyber security, algorithms and military technology.

Initially you will be sent an email with a short email survey, and an invitation to schedule an interview. The interview shall be arranged at your convenience and will be audio recorded. All interviews will be anonymised and all identifiable data (including audio recordings) will be permanently deleted once the interview has been transcribed.

**What are the possible benefits and/or disadvantages of taking part?**
No personal benefits and/or disadvantages are expected as a result of taking part in this research.

**Will my taking part be kept confidential?**
All the information collected about you during the course of the research will be kept strictly confidential in accordance with current data protection regulations (for more information, please see Royal Holloway's Data Management Policy here). Data storage and access will also be managed in line with the General Data Protection Regulation Regulation (GDPR) (for more information on your rights when it comes to accessing interview-related data, please see Royal Holloway's Data Protection Policy here.) You will not be able to be identified in any reports or publications without specific consent. All data will be identified only by a code, with personal details kept in a locked file or secure computer with access only by the researcher.

**What will happen to the results of the research project?**
As any audio-recorded material will be transcribed at a later date, interview transcripts are likely to be stored in electronic format as an audiofile and a word-processed text file, for a period of time until December 2021 at the latest. The data will only be stored as long as necessary and will be permanently deleted as soon as the analysis and content has been finalised.

Results will be written up for submission as part of my PhD thesis. It is also possible that the results of the project may be submitted for academic publications or blogs, or presented to academic audiences. Results will be presented in terms of groups of individuals. If any individual data are presented or published, the data will be completely anonymous, without any means of identifying the individuals involved. No classified information will be included in any form. The data will be held privately and will not be shared with unauthorised parties.

**Ethical review of the study**
This research is not expected to raise any ethical concerns. However, as the work involves human participants, ethical approval and consent must be sought in advance. The project has been reviewed and approved by Royal Holloway University of London's Research Ethics Committee.

**Contact for further information**
PhD researcher: Amy Ertan (amy.ertan.2017@rhul.ac.uk)

PhD supervisor: Rikke Jensen (rikke.jensen@rhul.ac.uk)

ROYAL HOLLOWAY UNIVERSITY OF LONDON

D.2 NATO-focused PIS

**PARTICIPANT INFORMATION SHEET**
**Future AI in conflict: NATO's capabilities and responsibilities**

**Invitation to take part**
You are being invited to take part in a research project. Before you decide whether or not you would like to take part it is important for you to understand why the research is being done and what it will involve. Please read the following information carefully and discuss it with others if you wish. If you have any questions or particular concerns, please let me know. You will find the relevant contact details on the last page of this document.

**Why is this research being done?**
This study forms part of my three-year doctoral research project and aims to answer the research questions: '*What are the strategic security implications of AI in conflict?*' and '*What are the capabilities and possible roles for NATO in mitigating potential insecurities created by weaponised AI technology?*'

The goal of this project is to understand how artificial intelligence (AI) is perceived by NATO and NATO-affiliated colleagues. This research explores the relationship between AI innovation, on the one hand, and contemporary security and conflict questions, on the other, whilst outlining the roles (and responsibilities) NATO may hold in mitigating an 'AI arms race'.

When we talk about AI, we are referring to artificial learning systems that may be present in supporting software, or physical robotics, automating or supplementing tasks traditionally performed by humans. Through interviewing individuals, who have experience working directly with these innovations through their roles at (/ in relation or prior to) the NATO CCDCOE, we highlight sentiments and reflections on the changing practice of AI-enabled conflict, enabling us to conceptualise AI impact in (and between) these spaces. Focusing on these research questions, we explore how institutions such as NATO interact with the innovation landscape – and aim to develop a theoretical approach to how NATO may contribute to the development of agreed international norms relating to AI deployment in military contexts.

**Who is doing the research?**
I am a PhD researcher in the Information Security Group, Royal Holloway. I am part of the Cyber Security Centre for Doctoral Training, which is sponsored by EPSRC grant EP/P009301/1.

**Why have I been chosen?**
You have been invited to participate as your role and/or experience is thought to position you as well-informed of issues arising at the intersection of technical innovation and international security. More specifically, you have been chosen based on your expertise in relation to international institutions such as NATO and the CCDCOE. All interviewees are therefore invited through current or previous career experience. This project may include practitioners from private industry, government, research organisations and the military.

It is expected that up to 20 practitioners will be interviewed as part of this project.

**Do I have to take part?**
No. It is up to you to decide whether or not to take part. If you do decide to take part, you will be asked to sign a Consent Form. You can withdraw during the survey and interview process at any time without giving a reason and your data will be removed from the study. Once the survey or

interview is completed you can still withdraw your data up to the point where the data has been analysed and anonymised, so that your identity cannot be determined. Your decision to take part or not to take part will involve no penalty or loss, now or in the future. Please contact Amy Ertan or Rikke Jensen to withdraw (contact details listed below).

**What will taking part involve?**
If you agree to take part, you would be expected to *either* complete a one hour-long interview with me in person or over the phone/equivalent audio option or join a one/two-hour long focus groups with up to six other participants. Questions will be related to the research questions outlined above and will be linked with the wider themes of artificial intelligence, workplace practices, and innovation in extreme environments.

Initially you will be sent an invitation to schedule an interview. The interview shall be arranged at your convenience and will be audio recorded. All interviews will be anonymised and all identifiable data (including audio recordings) will be permanently deleted once the interview has been transcribed.

**What are the possible benefits and/or disadvantages of taking part?**
No personal benefits and/or disadvantages are expected as a result of taking part in this research.

**Will my taking part be kept confidential?**
All the information collected about you during the course of the research will be kept strictly confidential in accordance with current data protection regulations (for more information, please see Royal Holloway's Data Management Policy here). Data storage and access will also be managed in line with General Data Protection Regulation (GDPR) (for more information on your rights when it comes to accessing interview-related data, please see Royal Holloway's Data Protection Policy here.) You will not be able to be identified in any reports or publications without specific consent. All data will be identified only by a code, with personal details kept in a locked file or secure computer with access only by the researcher.

**What will happen to the results of the research project?**
As any audio-recorded material will be transcribed at a later date, interview transcripts are likely to be stored in electronic format as an audio-file and a word-processed text file, for a period of time until January 2022 at the latest. The data will only be stored as long as necessary and will be permanently deleted as soon as the analysis and content has been finalised.

Results will be written up for submission as part of my PhD thesis as well as in a research paper for the NATO Cooperative Cyber Defence Center of Excellence. It is also possible that results of the project may be submitted for academic publications or blogs or presented to academic audiences. Results will be presented in terms of groups of individuals. If any individual data are presented or published, the data will be completely anonymous, without any means of identifying the individuals involved. No classified information will be included in any form. The data will be held privately and will not be shared with unauthorised parties.

**Ethical review of the study**
This research is not expected to raise any ethical concerns. However, as the work involves human participants, ethical approval and consent must be sought in advance. The project has been reviewed and approved by Royal Holloway University of London's Research Ethics Committee.

**Contact for further information**

PhD researcher: Amy Ertan (amy.ertan.2017@rhul.ac.uk)

PhD supervisor: Rikke Jensen
(rikke.jensen@rhul.ac.uk)

CCDCOE project supervisor: Piret Pernik
(Piret.Pernik@ccdcoe.org)

D.3 U.S.-focused PIS



**PARTICIPANT INFORMATION SHEET**

US Security Strategies in the Era of Military AI: Part One

**Invitation to take part**

You are being invited to take part in a research project. Before you decide whether or not you would like to take part it is important for you to understand why the research is being done and what it will involve. Please read the following information carefully and discuss it with others if you wish. If you have any questions or particular concerns, please let me know. You will find the relevant contact details on the last page of this document.

**Why is this research being done?**

This study forms part of my three-year doctoral research project and aims to answer the research questions: '*(Cyber) Security Strategies in the Age of AI: How is AI being used as a tool within a national security context, and how can this facilitate or amplify offensive cyber-activity?*' and '*How is AI changing warfare? How does the AI development landscape contribute to military strategy, and what controls are required to manage responsible and/or de-escalatory use of 'weaponised AI' capabilities?'*

The goal of this project is to evaluate how artificial intelligence (AI) is perceived across the US military innovation landscape, and which activities relevant US agencies are undertaking to control the risks affiliated with AI-enabled technology in the military context. This research explores the opportunities provided through AI innovation, on the one hand, and security challenges and required mitigatory activity, on the other.

When we talk about AI, we are referring to artificial learning systems that may be present in supporting software, or physical robotics, automating or supplementing tasks traditionally performed by humans. Through interviewing individuals who have relevant expertise and experience working directly with these innovations, we highlight reflections on US innovation practices and broadly compare this with previous research focused on the UK landscape. . Focusing on the research questions above, we explore how the US perceives AI as a potential offensive capability – and aim to measure the effectiveness of different options that either encourage 'responsible AI' or discourage the escalatory use of 'weaponised AI'.

**Who is doing the research?**

I am a PhD researcher in the Information Security Group, Royal Holloway. I am part of the Cyber Security Centre for Doctoral Training, which is sponsored by EPSRC grant EP/P009301/1.

**Why have I been chosen?**

You have been invited to participate as your role and/or experience is thought to position you as well-informed of issues arising at the intersection of technical innovation and international security. More specifically, you have been chosen based on your expertise in relation to the US' approach to military AI innovation. All interviewees are therefore invited through current or previous career

experience. This project may include practitioners from private industry, government, research agencies and the military.

It is expected that up to 20 practitioners will be interviewed as part of this project.

**Do I have to take part?**
No. It is up to you to decide whether or not to take part. If you do decide to take part, you will be asked to sign a Consent Form. You can withdraw during the survey and interview process at any time without giving a reason and your data will be removed from the study. Once the survey or interview is completed you can still withdraw your data up to the point where the data has been analysed and anonymised, so that your identity cannot be determined. Your decision to take part or not to take part will involve no penalty or loss, now or in the future. Please contact Amy Ertan or Rikke Jensen to withdraw (contact details listed below).

**What will taking part involve?**
If you agree to take part, you would be expected to *either* complete a one hour-long interview with me in person or over the phone/equivalent audio option or join a one/two-hour long focus groups with up to six other participants. Questions will be related to the research questions outlined above and will be linked with the wider themes of artificial intelligence, workplace practices, and innovation in extreme environments.

Initially you will be sent an invitation to schedule an interview. The interview shall be arranged at your convenience and will be audio recorded. All interviews will be anonymised and all identifiable data (including audio recordings) will be permanently deleted once the interview has been transcribed.

**What are the possible benefits and/or disadvantages of taking part?**
No personal benefits and/or disadvantages are expected as a result of taking part in this research.

**Will my taking part be kept confidential?**
All the information collected about you during the course of the research will be kept strictly confidential in accordance with current data protection regulations (for more information, please see Royal Holloway's Data Management Policy here). Data storage and access will also be managed in line with General Data Protection Regulation (GDPR) (for more information on your rights when it comes to accessing interview-related data, please see Royal Holloway's Data Protection Policy here.) You will not be able to be identified in any reports or publications without specific consent. All data will be identified only by a code, with personal details kept in a locked file or secure computer with access only by the researcher.

**What will happen to the results of the research project?**
As any audio-recorded material will be transcribed at a later date, interview transcripts are likely to be stored in electronic format as an audio-file and a word-processed text file, for a period of time until January 2022 at the latest. The data will only be stored as long as necessary and will be permanently deleted as soon as the analysis and content has been finalised.

Results will be written up for submission as part of my PhD thesis as well as in various publications (policy briefs or op-ed pieces) for the Belfer Center. It is also possible that results of the project may be submitted for academic publications or blogs or presented to academic audiences. Results

will be presented in terms of groups of individuals. If any individual data are presented or published, the data will be completely anonymous, without any means of identifying the individuals involved. No classified information will be included in any form. The data will be held privately and will not be shared with unauthorised parties.

**Ethical review of the study**

This research is not expected to raise any ethical concerns. However, as the work involves human participants, ethical approval and consent must be sought in advance. The project has been reviewed and approved by Royal Holloway University of London's Research Ethics Committee.

**Contact for further information**

PhD researcher: Amy Ertan (amy.ertan.2017@rhul.ac.uk)

PhD supervisor: Rikke Jensen
(rikke.jensen@rhul.ac.uk)

Belfer Center project supervisor: Lauren Zabierek
(lauren_zabierek@hks.harvard.edu)

HARVARD Kennedy School
BELFER CENTER
for Science and International Affairs

ROYAL HOLLOWAY UNIVERSITY OF LONDON

# Appendix E: Interview Schedules

In line with the data anonymisation conditions highlighted in my ethics review applications and the PIS and consent form communications with my interviewees, Tables E.1-E.3 do not include any information that may be used to identify interviewees.[126]

*Table E.1: UK-focused Interviews*

| No. | Online/In-Person | Interviewee | Length | Date |
|---|---|---|---|---|
| 1 | In person | Private sector | 28 minutes 40 seconds | 21 February 2020 |
| 2 | In person | Private sector | 1 hour 5 minutes 2 seconds | 28 February 2020 |
| 3 | In person | Private sector | 36 minutes 1 second | 12 March 2020 |
| 4 | Online | Private sector | 55 minutes 10 seconds | 14 March 2020 |
| 5 | Online | Private sector | 36 minutes 1 second | 8 April 2020 |
| 6 | Online | Private sector | 52 minutes 9 seconds | 8 June 2020 |
| 7 | Online | Private sector | 1 hour 14 minutes 19 seconds | 11 June 2020 |
| 8 | Online | Private sector | 44 minutes 25 seconds | 17 June 2020 |
| 9 | Online | Private sector | 53 minutes 34 seconds | 18 June 2020 |
| 10 | Online | Private sector | 44 minutes 12 seconds | 24 June 2020 |
| 11 | Online | Private sector | 53 minutes 53 seconds | 25 June 2020 |
| 12 | Online | Private sector | 59 minutes 36 seconds | 2 July 2020 |
| 13 | Online | Private sector | 48 minutes 39 seconds | 6 July 2020 |
| 14 | Online | Private sector | 56 minutes 15 seconds | 28 July 2020 |

[126] For example, no details are included listing the interviews' employer, role title, or role location.

*Table E.2: NATO-focused Interviews*

| No. | Online/In-Person | Interviewee profile | Length | Date Conducted |
|---|---|---|---|---|
| 1 | Online | Recent former NATO CCDCOE, civilian staff | 45 minutes 16 seconds | 8 December 2020 |
| 2 | Online | NATO, military staff | 53 minutes 59 | 6 November 2020 |
| 3 | Online | NATO, civilian staff | 46 minutes 40 seconds | 10 December 2020 |
| 4 | Online | NATO, civilian staff, and academic | 53 minutes 45 seconds | 24 September 2020 |
| 5 | Online | NATO, civilian | 47 minutes 54 seconds | 1 December 2020 |
| 6 | Online | NATO, civilian | 54 minutes 2 seconds | 19 January 2021 |
| 7 | Online | NATO, civilian | 47 minutes 25seconds | 23 November 2020 |
| 8 | Online | Recent former NATO, civilian. Academic & defence sector. | 40 minutes 26 seconds | 9 April 2021 |
| 9 | In-Person (Masks worn, socially distanced) | NATO CCDCOE, military staff | 53 minutes 10seconds | 3 February 2021 |
| 10 | Online | NATO, civilian | 36 minutes 34 seconds | 1 February 2021 |
| 11 | Online | NATO, civilian | 31 minutes 30 seconds | 9 March 2021 |
| 12 | Online | NATO CCDCOE, military staff | 35 minutes 18 seconds | 10 March |
| 13 | Online | Recent former NATO. Academic & think-tank researcher | 49 minutes 38 seconds | 11 March |
| 14 | Online | NATO, civilian | 49 minutes 06 seconds | 18 Jan 2021 |
| 15 | Online | NATO CCDCOE, military staff | 43 minutes 55 seconds | 12 March 2021 |
| 16 | Online | NATO, civilian | 43 minutes 50 seconds | 17 March 2021 |
| 17 | Online | NATO CCDCOE, civilian | 52 minutes 02 seconds | 24 March 2021 |

*Table E.3: U.S. Interviews*

| No. | Online/In-Person | Interviewee profile | Length | Date |
|---|---|---|---|---|
| 1 | Online | Recent former DoD (Research and policy leadership). | Approx. 50 minutes (not recorded, notes taken manually) | 1 October 2020 |
| 2 | Online | DoD (AI leadership) | 51 minutes 36 seconds | 17 October 2020 |
| 3 | Online | Former DoD (Emerging security policy) | 56 minutes 22 seconds | 9 November 2020 |
| 4 | Online | Recent former DoD (AI strategy development) | 47 minutes 28 seconds | 24 November 2020 |
| 5 | Online | Former DoD and military. academic | 58 minutes 49 seconds | 25 November 2020 |
| 6 | Online | Former DoD (AI leadership). | 49 minutes 21 seconds | 22 July 2021 |
| 7 | Online | Former DoD and military (AI policy) | 53 minutes 6 seconds | 2 August 2021 |

# Appendix F: Additional Tables

## F.1 Event Self-Descriptions for Observant Practice

Table F.1 briefly introduces the way each event was described by the organisers and notes on key stakeholders organising and attending the event.

*Table F.1: Conference descriptions (self-described) and identified stakeholders*

| | |
|---|---|
| CyCon | **Event Description:** The International Conference on Cyber Conflict (CyCon)'s theme for 2019 was "Silent Battle", allowing for "diverse interpretation' ' of cyber threats (NATO CCDCOEa, Online). The four-day event had almost 100 speakers, including Estonian President Kersti Kaljulai, senior NATO leadership, and technical, strategic (policy-level) and legal researchers (NICP, 2019). The conference proceedings include 29 research papers presented at the conference, including topics like weaponised AI and emerging security challenges.<br><br>**Stakeholders:** Organised by the NATO Cooperative Cyber Defence Centre of Excellence (CCDCOE). Attracts a range of NATO, military, government, public sector and academic (student and researcher) attendees. The event revolves around panels and talks, which draw on presenters' academic papers, collated to form the IEEE International Conference on Cyber Conflict book each year. |
| CyCon U.S. (as of 2019) | **Event Description:** CyCon U.S. describes itself as "a conference to bring together military, government, academia and industry cyber to build community and lead to impactful relationships" (quoted in authors' field notes, 2019). Relatively smaller than its European counterpart, in 2019, the conference attracted 337 attendees. Over three days 6 five technical and five strategic papers were presented on the theme "Defending Forward", reflecting the U.S.' call in its 2018 cyber strategy to be actively prepared in cyberspace. The Army Cyber Institute director call's CyCon U.S. conference "the premier forum in cyber conflict" (Hall, 2020).<br><br>**Stakeholders:** Jointly organised by the Army Cyber Institute at the United States Military Academy and the NATO CCDCOE. (Army Cyber Institute, 2019). Attendees came from various sectors: "from over 40 companies, 13 universities, 17 foreign nations, congressional staffers, and various government organizations" (Hall, 2020). |
| DSEI | **Event Description:** DSEI "brings the entire defence manufacturing supply chain together with the world's key military specifiers, influencers, buyers, and end-users" (DSEI, Online).<br><br>DSEI has attracted attention as a controversial arms fair, particularly as the event typically features national stands from Israel and Saudi Arabia showcasing weapons and broader capabilities. The fair is usually subject to protesters with a dedicated "Stop the Arms Fair" campaign group coordinating opposition to the event (Stop the Arms Fair, Online). Amnesty International has criticised the fair several times (Wilcken, 2019). |
| DTDT | **Event description:** "Accelerating acquisition for Machine Speed Warfare", with sessions including AI enablement, Innovation & Military Procurement, Military Command & Control and Autonomy, AI and the OODA loop. DefenceIQ: 'will bring together senior figures from Armed Forces, Government, Industry and Academia to explore the transformation of capability for warfare in the Digital Age. It will facilitate dialogue on embedding innovation at the heart of defence, accelerating acquisition from non-traditional defence suppliers, and achieving the major dimensions of capability transformation: ISR, Command and Control, and the transition to a manned, unmanned and autonomous mix in all domains…'(DTDT, 2020) |

| | |
|---|---|
| | **Stakeholders:** The event took place "in partnership with" the Defence Academy for the United Kingdom (part of the U.K. MoD), an education provider for the British Armed Forces and MoD civil servants (U.K. MoD, Online)). |
| DSDS | **Event Description:** This annual Symposium focuses on defence and security research broadly, with agenda items ranging from biology-grounded presentations on bone disease and interdisciplinary research into gendered experience in combat, to topics closer to AI security, including "Deceptive Autonomous Agents", Autonomous Vehicles, and "Deep Learning Techniques for Missile Seeker Automatic Target Recognition" (Cranfield University, 2020). An academic outreach event primarily with presenters speaking on themes relating to government-academia collaboration.<br><br>**Stakeholders:** The Defence and Security Doctoral Symposium is hosted by Cranfield University via "Symposia at Shrivenham", a "forum to Government agencies, military and civilian, industry and research establishments for the exploration and exchange of experience and knowledge, leading to constructive questioning and a synthesis of ideas in a relaxed but professional environment." (Cranfield University, Online). The event hosted speaker sessions from a range of mostly U.K.-based and postgraduate-level academic researchers, Dstl, the Department for Transport and the National Cyber Security Centre. Presenters also included representatives from the Atomic Weapons Establishment (a Non-Departmental Public Body (NDPB) owned by the U.K. MoD) (DSDS Agenda, 2019). |
| AFCEA Europe | **Event title:** "The New Age of Computing: How will it change command and Control?". Included sessions on "The Advent of Quantum Computing and Artificial Intelligence", "Security challenges within Artificial Intelligence: How to overcome them?" and additional talks on "AI: Impact on Organisations, Processes, Culture and Leadership". AFCEA describes itself as a: 'non-profit, non-lobbying association providing a platform for ethical information exchange and education" (Banner at AFCEA event, 2020). The event was not actively publicised with a relatively closed invite list and around 50 attendees; I was invited after presenting to an AFCEA members' event in London in 2019.<br><br>**Stakeholders:** Speakers and attendees from academia, military and industry. Sessions were divided into technical and non-technically focused content. (AFCEA field notes, 2019). The event was organised with the cooperation of the NATO Command and Control Centre of Excellence. |

## F.2 U.K. Table

Table F.2 outlines several reports highlighted through this research as making notable statements on the U.K.'s position on AI in the context of military security themes.

*Table F.2: Relevant U.K. official documents*

| **Document** | **Contents** |
|---|---|
| Joint Concept Note 1/18 Human-Machine Teaming | "In a submission to the GGE in August 2018, the UK Government refers to this Joint Concept Note as 'the UK Ministry of Defence's thinking on the evolution of how humans will work with machines.'6 Among its findings, the UK identifies |

| | |
|---|---|
| Development, Concepts and Doctrine Centre, UK MoD. May 2018. | 'what [human] actors trust their machines to do' as something that will restrain 'the increasing capabilities of robotic and AI systems.' It categorises four "fundamental factors" that will determine how much human actors are prepared to trust remote and automated systems: mechanical understanding, predictability, familiarity and context" (Warren and Hillas, 2020). |
| U.K.'s Defence Technology Framework. September 2019. | Defines AI broadly as "the ability of machines to perform tasks normally requiring human intelligence" that is "expected to enable radical transformation across almost every area of Defence activity" (Ministry of Defence, 2019, 18). The Framework lists a range of potential defence applications, from AI-enabled cyber defence to intelligence analysis to streamlining logistics and back-office operations, and states that developments will need to be "safe, ethical and interoperable with other nations" (Ministry of Defence, 2019, 19). There are currently few further details on how relevant U.K. decision-makers are approaching military AI innovation at an operational or strategic level. |
| "Pioneering a new national security: the ethics of Artificial Intelligence", GCHQ (2021). *Released after the conclusion of interviews.* | This report highlights their support to UK military operations in a way that is within international legal and regulatory frameworks and cites how their operations, including through AI, are subject to independent oversight to check that all activity is legal. The report includes the ongoing development of an AI and data ethics governance system derived from consultation with external stakeholders. Nonetheless, the report acknowledges the unsolved ethical challenges presented in the development and deployment of AI in defence and articulates the need for specialist expertise and processes to minimise the risk of error – including bias and discrimination. |

## F.3 NATO Tables

Table F.3.1 provides a timeline of such documents (where their existence has been publicly confirmed) and key public milestones relating to NATO's approach to AI technology at a strategic level. In this context, "strategic level" refers to the level at which members' and NATO's policy is formed and the broad, holistic approach to the technology, including member and NATO-level security considerations.

*Table F.3.1: NATO activity relating to AI policy and strategy development*

| | | |
|---|---|---|
| December 2019 (London Leaders' Meeting) | NATO Leaders agreed Emerging and Disruptive Technology Implementation Roadmap.<br>- Acknowledges the challenges AI, autonomous systems, big data and other technologies may post to NATO.<br>- Basis for EDT strategy | Strategy Level |

| | | |
|---|---|---|
| 2020 | NATO's Innovation Unit has published AI white papers (Christie, 2020; Murray, 2020) and 2020 saw the first Innovation Board meeting with the Emerging Security Challenges Division Advisory Group (NATOa, 2020). | Strategy Level |
| July 2020 | Advisory Group on Emerging and Disruptive Technologies established.<br>(12 international experts from industry and academia. Tasked to provide advice to NATO innovation board on adoption of EDTs.) | Strategy Level Group Formation: |
| September 2020 | Advisory Group on Emerging and Disruptive Technologies communicates recommendations to NATO's Innovation Board. | Strategy Level |
| February 2021 | NATO's Coherent Implementation Strategy on Emerging and Disruptive Technologies. | Strategy Level |
| March 2021 | The NATO Advisory Group on Emerging and Disruptive Technologies published its first annual report.<br><br>This included four key recommendations for NATO: improve technology literacy throughout the Organization; establish a network of Innovation Centres; design and facilitate new financing mechanisms for innovation with private sector entities, both small and large; and create innovation partnership initiatives with external EDT stakeholders from industry and academia. | Strategy Level |
| June 2021 | At the 2021 Brussels Summit, NATO Leaders agreed:<br>(1) To establish a civil-military Defence Innovation Accelerator for the North Atlantic (DIANA)<br>(2) to establish a NATO Innovation Fund. | Cross-level |
| June 2021 | NATO 2030 Agenda agreed<br>An "ambitious" agenda to keep NATO "strong and united for a new era of increased global competition." (NATO, 2021e). Applying to the whole of NATO as an overarching initiative, NATO 2030 includes proposals including: Strengthened Deterrence and Defence; Improved Resilience; 'Preserve Our Technological Edge'; 'Uphold the Rules-Based International Order' and 'Boost training and Capacity Building' (*ibid).*<br><br>Emerging technologies are a key theme highlighted through NATO-2030 announcements and motivation. | Strategy Level |
| October 2021 | NATO AI Strategy Agreed<br>NATO members formally adopted the first NATO AI strategy, a NATO-classified document to target allied adoption of the technology and develop a responsible approach to the use of AI in conflict. A NATO press statement announcing the Strategy announced NATO's aim "to lead by example and encourage the development and use of AI in a responsible manner" (NATO, 2021). The strategy also set out six Principles of Responsible Use for AI in Defence relating to Lawfulness, Responsibility and Accountability, Explainability and Traceability, Reliability, Governability, and Bias Mitigation (Christie, 2021). | Strategy Level |

Below the threshold of agreement on official NATO-wide policy, NATO has been actively coordinating numerous events discussing AI (and releasing subsequent statements) on the importance of NATO's attention to emerging technologies. (Leopold, 2020; NATO ACTa,

2019; NATOe, 2020). In the past year, NATO has begun hosting more discussions on the role of collaboration in military AI and AI interoperability between member states. Table F.3.2 outlines several NATO projects in which AI is featured.

*Table F.3.2: NATO projects mentioning integration of AI technology*

| | |
|---|---|
| Military Uses of Artificial Intelligence, Automation, and Robotics (MUAAR) project | The MUAAR project is US-led and will investigate the uses for AI, automation, and robotics in areas such as the electromagnetic spectrum, integrated air & missile defence, logistics, and the space, cyberspace, air, land and maritime domains (NATO ACT, n.d). |
| Data Science Center | Aiming to bring NATO data science expertise under one structure (NCI Agency, 2019). Operates within the NCI Agency. (Pepe, 2020). |
| NATO Maritime Unmanned Systems Innovation Advisory Board | Established in mid-2020, the Board focuses on developing unmanned and autonomous maritime systems. |
| Other; anecdotal reporting | NATO has already experimented in incorporating AI systems into its humanitarian efforts (Hill, 2020) |

*Table F.3.3: Identified key NATO agencies involved in AI-related initiatives and planning*

| **Body** | **Summary of involvement** | **Organisational Structure at NATO** |
|---|---|---|
| Allied Command Transformation. (ACO) | The ACT has coordinated events focusing on military AI technology (STCTTS, 2019). It is coordinating a significant number of AI innovation activities, launching the Emerging and Disruptive Technologies Roadmap in 2018 and establishing an "Innovation Branch" in which innovators are "protected and nurtured" (Gilli, 2020, 4). | One of the two NATO Military Commands at NATO (the other being ACO). Reports to the Military Committee. Responsible for military transformation and innovation at NATO. |
| Allied Command Operations (ACO) | ACO has taken a "leading role" on disruptive technologies, including AI (Hill, 2020, 149). Organises events including holistic workshops (NATO, 2018) and, as reported in October 2019, discussions between Ambassadors in military representatives to discuss how best to leverage emerging technologies (NATO, 2019). | One of the two NATO military Commands at NATO. Reports to NATO Military Committee.<br><br>Responsible for planning and execution of NATO operations. |
| NATO Communications and Information Agency (NCI Agency) | In 2019 the NCI Agency announced the establishment of a NATO Data Science Centre (NCI Agency, 2019), building on the NCI Agency's decision to make AI a central topic during the 2018 NATO Information Assurance Symposium (STCTTS, 2019). The NCI Agency sponsors work on machine learning and AI research (NCI Agency, 2020) and development. The Data Science Centre will likely aim to use data to develop NATO | As NATO's cyber hub, the NCI Agency supports cyber defence for NATO HQ, NATO Commands and agencies across the organisation and has representatives from each NATO nation. The NCI Agency reports indirectly to the Agency Supervisory Board, which then reports to the North Atlantic Council. |

| | | |
|---|---|---|
| | AI capabilities. The body has also contributed expertise to machine learning hackathons (NCI Agency, 2020). | |
| Science and Technology Organisation (STO) | Produced focused research on AI and big data for military decision-making and has emphasised AI as a prominent science and technology theme through their publications (NATO STO, 2020). Their collaborative programme of work includes several AI activities, including meaningful human control and trust, and how AI may be utilised in the information environment (STCTTS, 2019). The STO's 2020 annual report highlighted that AI was considered within active programmes on maritime defence, data knowledge and operational effectiveness, AI and big data for military decision-making, and data collection for operational support (STO, 2021b). The STO also held the first Disruptive Technologies Table-Top exercise in Feb 2021, assessing the potential implications of emerging technologies across several scenarios (STO, 2021a). | The primary NATO subsidiary organisation focusing on defence science and technology. Operates under North Atlantic Council authority via a Board of Directors chaired by NATO's Chief Scientist. Supports national leaders as well as the North Atlantic Council. |
| Emerging Security Challenges Division (ESCD) | Created in 2010, the ESCD consolidates expertise at NATO headquarters in taking a holistic view of ETDs. Its remit is the delivery of policy documents to support NATO's strategic concept, and the division has taken a leading role in producing the upcoming NATO AI strategy. The ESCD contains the Innovation Unit, which has produced two (internally released) White Papers on AI and autonomous systems through 2020 (Lucarelli, Marrone, & Moro, 2021). The Department also coordinated the drafting of NATO's AI Policy. | Made up of International Staff (civilian personnel). Supports the North Atlantic Council, the principal political authority at NATO. |
| NATO Standardization Office (NSO) | The NSO designs military operational requirements and contributes to NATO's strength in initiating, supporting and administering technical standards across NATO.<br><br>There is no public information on the NSO's involvement with AI themes. However, Deputy Secretary-General Mircea Geoană highlighted NATO's "standardisation community" as a contributing strength that makes NATO a "natural platform for | An independent body that reports indirectly to the Military Committee. |

| | | |
|---|---|---|
| | transatlantic cooperation of AI" (NATOe, 2020, para 3). | |

# Bibliography


Aberbach, Joel D., and Bert A. Rockman. "Conducting and coding elite interviews." *PS: Political Science & Politics* 35, no. 4 (2002): 673-676. https://doi.org/10.1017/S1049096502001142

Acton, James M. "Reclaiming Strategic Stability." *Strategic Stability: Contending Interpretations* (2013): 117-146.

AFCEA Europe. “The New Age of Computing: How will it change Command and Control?” Workshop Event Programme. Physical Brochure. 2020.

Ahmed, Nur, and Muntasir Wahed. "The de-democratization of ai: Deep learning and the compute divide in artificial intelligence research." *arXiv preprint arXiv:2010.15581* (2020).

AI Partnership for Defense. AI Partnership for Defense (AI PfD) 15-16 September 2020 Joint Statement. Accessed October 12, 2021. https://www.ai.mil/docs/AI_PfD_Joint_Statement_09_16_20.pdf

Albert, Kendra, Jonathon Penney, Bruce Schneier, and Ram Shankar Siva Kumar. “Politics of Adversarial Machine Learning.” *SSRN Electronic Journal*, 2020, 1–6. https://doi.org/10.2139/ssrn.3547322.

Allen, Greg, and Taniel Chan. *Artificial intelligence and national security*. Cambridge, MA: Belfer Center for Science and International Affairs, 2017.

Allen, John, Philip M. Breedlove, Julian Lindley-French, and George Zambellas. "Future War NATO?: From Hybrid War to Hyper War via Cyber War." *GLOBSEC NATO adaptation initiative. The Future Tasks of the Adapted Alliance* (2017).

Alpaydin, Ethem. *Introduction to machine learning*. MIT Press, 2020.

Anderson, James M., Benoit Arbour, Roberta Arnold, Thomas Kadiofsky, Tom Keeley, Matthew R. MacLeod, Sean Bourdon et al. *Autonomous Systems: Issues for Defence*

*Policymakers*. NATO SUPREME ALLIED COMMAND TRANSFORMATION NORFOLK VA NORFOLK, 2015.

Anderson, Kenneth, and Matthew C. Waxman. "Debating Autonomous Weapon Systems, their Ethics, and their Regulation under international law." (2017).

Anderson, Kenneth, Daniel Reisner, and Matthew C. Waxman. "Adapting the law of armed conflict to autonomous weapon systems." (2014).

Archibald, Mandy M., Rachel C. Ambagtsheer, Mavourneen G. Casey, and Michael Lawless. "Using Zoom videoconferencing for qualitative data collection: perceptions and experiences of researchers and participants." *International Journal of Qualitative Methods* 18 (2019): 1609406919874596. http://dx.doi.org/10.1177/1609406919874596

Arkin, Ronald C. "Governing lethal behavior: Embedding ethics in a hybrid deliberative/reactive robot architecture." In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pp. 121-128. 2008. http://dx.doi.org/10.1145/1349822.1349839

Arkin, Ronald. "Lethal autonomous systems and the plight of the non-combatant." In *The political economy of robots*, pp. 317-326. Palgrave Macmillan, Cham, 2018. http://dx.doi.org/10.1007/978-3-319-51466-6_15

Army Cyber Institute. Conference on Cyber Conflict. Youtube playlist. 2020. Accessed 7 November, 2021. https://web.archive.org/web/20211107101739/https://www.youtube.com/playlist?list=PLtUuPz3a0Gz_1tOQNAeHsBgQVFg9BB45d

Army Cyber Institute. CVent event booking page - CyCON U.S. 2019. Accessed 21 September, 2021. https://aci.cvent.com/events/2019-international-conference-on-cyber-conflict-cycon-u-s-/archived-6fa2dad0721d43249d40ca05d3a5e94a.aspx [Page now defunct but organisational details remain along bottom banner].

Army Technology. UK DSTL-led Project Minerva tests chemical detection robots and drones. 19 September 2018. Accessed 1 October, 2021. https://www.army-technology.com/news/dstl-project-minerva-chemical-detection/

Arreguin-Toft, Ivan. "How the weak win wars: A theory of asymmetric conflict." *International security* 26, no. 1 (2001): 93-128. http://dx.doi.org/10.1162/016228801753212868

Asaro, Peter. "On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making." *International Review of the Red Cross* 94, no. 886 (2012): 687-709. http://dx.doi.org/10.1017/S1816383112000768

Athalye, Anish, Logan Engstrom, Andrew Ilyas, and Kwok Kevin. "Synthesizing Robust Adversarial Examples." *35th International Conference on Machine Learning, ICML 2018* 1 (2018): 449–68.

Ayoub, Kareem, and Kenneth Payne. "Strategy in the age of artificial intelligence." *Journal of strategic studies* 39, no. 5-6 (2016): 793-819. http://dx.doi.org/10.1080/01402390.2015.1088838

Baalen, Peter van, Paul van Fenema, and Claudia Loebbecke. "Extending the Social Construction of Technology (SCOT) Framework to the Digital World." (2016).

Babuta, Alexander, Marion Oswald, and Ardi Janjeva. "Artificial intelligence and UK national security: policy considerations." (2020).

Backman, Kaisa, and Helvi A. Kyngäs. "Challenges of the grounded theory approach to a novice researcher." *Nursing & health sciences* 1, no. 3 (1999): 147-153. http://dx.doi.org/10.1046/j.1442-2018.1999.00019.x

Bahrammirzaee, Arash. "A comparative survey of artificial intelligence applications in finance: artificial neural networks, expert system and hybrid intelligent systems." *Neural Computing and Applications 19,* no. 8 (2010): 1165119. http://dx.doi.org/10.1007/s00521-010-0362-z

Barreno, Marco, Blaine Nelson, Russell Sears, Anthony D. Joseph, and J. Doug Tygar. "Can machine learning be secure?." In *Proceedings of the 2006 ACM Symposium on Information, computer and communications security*, pp. 16-25. 2006. http://dx.doi.org/10.1145/1128817.1128824

Bathaee, Yavar. "The artificial intelligence black box and the failure of intent and causation." *Harv. JL & Tech.* 31 (2017): 889.

Baumeister, Roy F., and Mark R. Leary. "Writing narrative literature reviews." *Review of general psychology* 1, no. 3 (1997): 311-320.

Bellais, Renaud, and Daniel Fiott. "The European Defense Market: Disruptive Innovation and Market Destabilization." *The Economics of Peace and Security Journal* 12, no. 1 (2017): 37–45. https://doi.org/10.15355/epsj.12.1.37.

Bengio, Yoshua, and Yann LeCun. "Scaling learning algorithms towards AI." *Large-scale kernel machines* 34, no. 5 (2007): 1-41.

Bickel, Keith B. *Mars Learning: The Marine Corps Development of Small Wars Doctrine, 1915–1940*. Routledge, 2018.

Bidwell, C., and B. MacDonald. "Emerging Disruptive Technologies and Their Potential Threat to Strategic Stability and National Security." (2018): 21.

Biggio, Battista, and Fabio Roli. "Wild patterns: Ten years after the rise of adversarial machine learning." *Pattern Recognition* 84 (2018): 317-331. http://dx.doi.org/10.1016/j.patcog.2018.07.023

Boardman, Michael, and Fiona Butcher. *An Exploration of Maintaining Human Control in Ai Enabled Systems and the Challenges of Achieving it*. NATO: Technical report. NATO Science and Technology Organization. 2019.

Bode, Ingvild, and Hendrik Huelss. "Autonomous Weapons Systems and Changing Norms in International Relations Ingvild Bode *" 44, no. September 2017 (2018): 393–413. https://doi.org/10.1017/S0260210517000614.

Boulanin, Vincent. "The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk: Volume I: Euro-Atlantic Perspectives". SIPRI. Volume I, May (2019). https://www.sipri.org/sites/default/files/2019-05/sipri1905-ai-strategic-stability-nuclear-risk.pdf.

Boulanin, Vincent, and Maaike Verbruggen. “Mapping the Development of Autonomy in Weapon Systems.” *SIPRI*, November 2017.

Boulanin, Vincent, Lora Saalman, Petr Topychkanov, Fei Su, and Moa Peldan Carlsson. “Artificial Intelligence, Strategic Stability and Nuclear Risk.” *Stockholm International Peace Research Institute*, June (2020).

Boulanin, Vincent, Netta Goussac, Laura Bruun, and Luke Richards. “Responsible Military Use of Artificial Intelligence: Can the European Union Lead the Way in Developing Best Practice?”. Stockholm International Peace Research Institute. November 2020.

Brennen, J Scott, Philip N. Howard and Rasmus Kleis Neilsen. "An industry-led debate: How UK media cover artificial intelligence." The Reuters Institute (2018).

Brundage, Miles, Shahar Avin, Jack Clark, Helen Toner, Peter Eckersley, Ben Garfinkel, Allan Dafoe et al. "The malicious use of artificial intelligence: Forecasting, prevention, and mitigation." *arXiv preprint arXiv:1802.07228* (2018).

Brussels Summit Communiqué. Issued by the Heads of State and Government participating in the meeting of the North Atlantic Council in Brussels on 14 June 2021. 14 June 2021. Accessed 14 September 2021. https://www.nato.int/cps/en/natohq/news_185000.htm?mc_cid=41ea40d8bc&mc_eid=964a8158c6

Bryman, Alan. "The research question in social research: what is its role?." *International Journal of Social Research Methodology* 10, no. 1 (2007): 5-20. http://dx.doi.org/10.1080/13645570600655282

Buchanan, Ben. "The AI triad and what it means for national security strategy." *Center for Security and Emerging Technology. https://cset.georgetown.edu/research/the-aitriad-and-what-it-means-for-national-security-strategy* (2020).

Builder, Carl. *The masks of war: American military styles in strategy and analysis: A RAND Corporation research study*. Johns Hopkins University Press, 1989.

Bufkin, Melissa A. "Qualitative Studies: Developing Good Research Questions." *Online Submission* (2006).

Burmaoglu, Serhat, and Ozcan Sarıtas. "Changing characteristics of warfare and the future of Military R&D." *Technological Forecasting and Social Change* 116 (2017): 151-161. https://doi.org/10.1016/j.techfore.2016.10.062.

Burton, Joe, and Simona R. Soare. "Understanding the Strategic Implications of the Weaponization of Artificial Intelligence." In *2019 11th International Conference on Cyber Conflict (CyCon)*, 1–17. IEEE, 2019.https://doi.org/10.23919/cycon.2019.8756866.

Canfil, Justin Key. "Yesterday's Reach: How Legal Institutions Keep Pace with Technological Change." *Available at SSRN 3684991* (2020).

Carlo, Antonio. "Artificial Intelligence in the Defence Sector." In International Conference on Modelling and Simulation for Autonomous Systems, pp. 269-278. Springer, Cham, 2020. http://dx.doi.org/10.1007/978-3-030-70740-8_17

Castro, Celso. "Interviewing the Brazilian military: Reflections on a research experience." (2000). https://doi.org/10.4324/9781315682259-8.

Cave, Stephen, and Seán S. ÓhÉigeartaigh. "An AI Race for Strategic Advantage: Rhetoric and Risks." *AIES 2018 - Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, 2018, 36–40. https://doi.org/10.1145/3278721.3278780.

CCDCOE. Call for Papers CyCon 2022. Received via email on 7 October 2021.

Chahal, Husanjot, Ryan Fedasiuk, and Carrick Flynn. "Messier than Oil: Assessing Data Advantage in Military AI." Center for Security and Emerging Technology, July (2020).

Charmaz, K. (2016) 'The Power of Constructivist Grounded Theory for Critical Inquiry', *Qualitative Inquiry*, 23(1), pp. 34–45. https://doi.org/10.1177/1077800416657105.

Chelioudakis, Eleftherios. "Deceptive AI machines on the battlefield: Do they challenge the rules of the Law of Armed Conflict on military deception?." *Available at SSRN* 3158711 (2017). http://dx.doi.org/10.2139/ssrn.3158711

Chomiak-Orsa, Iwona, Artur Rot, and Bartosz Blaicke. "Artificial Intelligence in Cybersecurity: The Use of AI Along the Cyber Kill Chain." In International Conference on Computational Collective Intelligence, pp. 406-416. Springer, Cham, 2019. http://dx.doi.org/10.1007/978-3-030-28374-2_35

Christie, Edward H. “Artificial Intelligence at NATO: Dynamic Adoption, Responsible Use.” NATO Review. NATO Review, November 24, 2020. Accessed December 11, 2020. https://www.nato.int/docu/review/articles/2020/11/24/artificial-intelligence-at-nato-dynamic-adoption-responsible-use/index.html

Christie, Edward. H. The NATO alliance and the challenges of artificial intelligence adoption. In Lucarelli, S., Marrone, A., & Moro, F. N. (Ed.). *NATO Decision-making in the age of big data and artificial intelligence*. 2021. (pp84-93). Brussels, NATO HQ.

Christie, Edward H., Caroline Buts, and Cindy Du Bois. "America, China, and the struggle for AI supremacy." In *24th Annual International Conference on Economics and Security*. 2021.

Christie, Edward, and Amy Ertan. "NATO and Artificial Intelligence." In Romaniuk, SN, and Manjikian. M.(Eds). Routledge Companion to Artificial Intelligence and National Security Policy. Routledge. Forthcoming (2022). Available at SSRN: https://ssrn.com/abstract=4133397 or http://dx.doi.org/10.2139/ssrn.4133397

Ciocca, J., and L. Kahn. "When AI is in control, who’s to blame for military accidents." *Bulletin of the Atomic Scientists* (2020).

Citron, Danielle K., and Robert Chesney. "Deep fakes: A looming crisis for national security, democracy and privacy?." *Lawfare* (2018).

Cladi, Lorenzo. “Artificial Intelligence and the Future of Warfare: The USA, China and Strategic Stability.” *Defence Studies* 00, no. 00 (2021): 1–3. https://doi.org/10.1080/14702436.2021.2005464.

Clarke, Victoria, and Virginia Braun. "Commentary: Thematic analysis." *Journal of Positive Psychology* 12, no. 3 (2017): 297-298. http://dx.doi.org/10.1080/17439760.2016.1262613

Cohn, Carol. "Sex and death in the rational world of defense intellectuals." *Signs: Journal of women in culture and society* 12, no. 4 (1987): 687-718. http://dx.doi.org/10.1086/494362

Congressional Research Service. “Defense Primer: Department of Defense Civilian Employees - IF11510 · VERSION 6.” 2022. CRS Reports. Congressional Research Service. https://crsreports.congress.gov/product/pdf/IF/IF11510.

Cooper, Camilla Guldahl. "Rules of Engagement: Introduction, Development and Use." In *NATO Rules of Engagement*, pp. 25-88. Brill Nijhoff, 2019.

Cop, Ruziye, and Rifat Tekin Kara. “The Role of Trade Fairs in Industrial Marketing: A Research on Defence Industry Trade Fairs.” Journal of Management Marketing and Logistics 1, no. 3 (2014): 156–72.

Cope, Meghan. "Organizing and analyzing qualitative data." *Qualitative research methods in human geography* 4 (2016): 373-93.

Corn, Jeffrey D. DoD Artificial Intelligence Strategy Overview. Test Support Squadron (812th) Edwards AFB CA United States, 2019.

Council on Competitiveness (US). *Gaining new ground: Technology priorities for America's future*. Vol. 38, no. 9. Council on Competitiveness, 1991.

Cowan, Robin, and Dominique Foray. “Quandaries in the Economics of Dual Technologies and Spillovers from Military to Civilian Research and Development.” *Research Policy* 24, no. 6 (1995): 851–68. https://doi.org/10.1016/0048-7333(94)00802-7.

Cox, Jessica, and Heather Williams. “The Unavoidable Technology: How Artificial Intelligence Can Strengthen Nuclear Stability.” *Washington Quarterly* 44, no. 1 (2021): 69–85. https://doi.org/10.1080/0163660X.2021.1893019.

Crabtree, Benjamin F., and William L. Miller. *Doing qualitative research*. Sage, 1999. http://dx.doi.org/10.1097/00006199-199507000-00011

Craig, Claire. *How Does Government Listen to Scientists?*. Springer, 2018. http://dx.doi.org/10.1007/978-3-319-96086-9

Cranfield University. Defence and Security Doctoral Symposium. Webpage. https://www.cranfield.ac.uk/events/symposia/sym-doc

--- “2019 Defence and Security Doctoral Symposium (DSDS19) in Conjunction with DSTL, AWE, Department for Transport and NCSC: Symposium Outputs”. Cranfield Online Research Data (CORD), July 16, 2019. https://doi.org/10.17862/cranfield.rd.c.4578305.v15.

--- “Symposia at Shrivenham”. Webpage. Accessed 21 September 2021. https://www.cranfield.ac.uk/events/symposia-at-shrivenham (/web/20210922105154/https://www.cranfield.ac.uk/events/symposia-at-shrivenham)

Creswell, J. W. "Research Design: Qualitative and Quantitative Approaches”. Thousand Oaks, CA: SAGE." (1994).

Christensen, Clayton M. "The innovator's dilemma: when new technologies cause great firms to fail." (1997).

Cummings, Missy. Artificial intelligence and the future of warfare. London: Chatham House for the Royal Institute of International Affairs, 2017.

Cummings, Missy. L., Heather M. Roff, Kenneth Cukier, Jacob Parakilas, and Hannah Bryce. "Artificial Intelligence and International Affairs." *Chatham House Report* (2018): 7-18.

Da Silva, Joseph. "Producing ‘good enough’ automated transcripts securely: Extending Bokhove and Downey (2018) to address security concerns." *Methodological Innovations* 14, no. 1 (2021): 2059799120987766. http://dx.doi.org/10.1177/2059799120987766

Dafoe, Allan. "AI governance: a research agenda." *Governance of AI Program, Future of Humanity Institute, University of Oxford: Oxford, UK* 1442 (2018): 1443.

Daniels, Matthew, and Ben Chang. "National Power After AI." CSET Report.(2021). https://cset.georgetown.edu/publication/national-power-after-ai/

Dave, Paresh and Jeffrey Dastin. “Exclusive: Ukraine has started using Clearview AI’s facial recognition during war.” Reuters. 14 March 2022. Accessed 19 March, 2021. https://www.reuters.com/technology/exclusive-ukraine-has-started-using-clearview-ais-

facial-recognition-during-war-2022-03-13/

De Spiegeleire, Stephan, Matthijs Maas, and Tim Sweijs. *Artificial intelligence and the future of defense: strategic implications for small-and medium-sized force providers*. The Hague Centre for Strategic Studies, 2017.

Deepmind. AlphaStar: Grandmaster level in StarCraft II using multi-agent reinforcement learning. Deepmind research blog. 30 October 2019. Accessed 13 October, 2021. https://deepmind.com/blog/article/AlphaStar-Grandmaster-level-in-StarCraft-II-using-multi-agent-reinforcement-learning

DefenceIQ. Email sent to mailing list 15 April 2020. [Language previously used word for word on a now-amended webpage: Accessed January 20, 2022. https://www.defenceiq.com/events-defencetransformationweek/index.]

Defence and Security Accelerator. Homepage. Online. U.K. Ministry of Defence. https://www.gov.uk/government/organisations/defence-and-security-accelerator. Accessed 12 October, 2021.

Defence Innovation Board. Artificial Intelligence Test, Evaluation, Validation & Verification (AI TEV/V) for DoD: Introduction Sheet. September 15, 2020. Accessed 17 December 2021. https://innovation.defense.gov/Portals/63/documents/Meeting%20Documents/September%2015%202020/DIB_AI%20TEVV_Introduction%20Sheet_CLEARED.pdf?ver=2020-09-15-110907-090

--- "About". Online. Accessed October 6, 2021. https://web.archive.org/web/20211006153703/https://innovation.defense.gov/About1/

--- AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense. 2019. Accessed 13 October 2021. https://media.defense.gov/2019/Oct/31/2002204458/-1/-1/0/DIB_AI_PRINCIPLES_PRIMARY_DOCUMENT.PDF

Defense Innovation Unit. Homepage. Online. Accessed September 24, 2021. https://web.archive.org/web/20210924163519/https://www.diu.mil/

--- Team. Online. Accessed August 25, 2021. https://web.archive.org/web/20210825043051/https://www.diu.mil/team

Del Monte, Louis A. *Genius Weapons: artificial intelligence, autonomous weaponry, and the future of warfare*. Prometheus Books, 2018.

Department for Digital, Culture, Media and Sport. National AI Strategy. HM Government. September 2021.

--- National Data Strategy. HM Government. September 2020.

Department for International Trade. 'Defence and security exporting: event and exhibition support.' Online. UK.GOV. Accessed 12 October 2021. https://www.gov.uk/government/publications/defence-and-security-exporting-event-and-exhibition-support/defence-and-security-exporting-event-and-exhibition-support

Department of Defense. (2018). *2018 DoD Artificial Intelligence Strategy Fact Sheet*. 1–2.

--- 'DIRECTIVE NUMBER 3000.09, November 2012 (incorporating Change May 2017), section 4.c(1)'. https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/300009p.pdf.

--- "Summary of the 2018 Department of Defense artificial intelligence strategy: Harnessing AI to advance our security and prosperity." (2019): 4.

--- DOD Releases Fiscal Year 2021 Budget Proposal. Press release. 10 February 2020. Accessed 11 October, 2021. https://www.defense.gov/News/Releases/Release/Article/2079489/dod-releases-fiscal-year-2021-budget-proposal/

--- Raven Unmanned Systems. November 4, 2014. Accessed November 20, 2020. https://www.army.mil/article/137604/rq_11b_raven_small_unmanned_aircraft_systems_suas.

--- "Summary of the 2018 Department of Defense Artificial Intelligence Strategy: Harnessing AI to advance our security and prosperity." (2019): 4.

--- DOD Adopts Ethical Principles for Artificial Intelligence. Press Release. February 24, 2020. Accessed 14 October, 2021. https://www.defense.gov/News/Releases/Release/Article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/

--- DOD Releases Fiscal Year 2020 Budget Proposal Press Release. 12 March 2019. Accessed October 15, 2021. https://www.defense.gov/Newsroom/Releases/Release/Article/1782623/dod-releases-fiscal-year-2020-budget-proposal/

--- Implementing Responsible Artificial Intelligence in the Department of Defense. Memorandum for Senior Pentagon Leadership, Commanders of the Combatant Commands Defence Agency and DoD Field Activity Directors. 26 May 2021. Accessed 12 October, 2021. https://media.defense.gov/2021/May/27/2002730593/-1/-1/0/IMPLEMENTING-RESPONSIBLE-ARTIFICIAL-INTELLIGENCE-IN-THE-DEPARTMENT-OF-DEFENSE.PDF

--- Cloud Strategy. December 2018. Accessed 16 December 2021. https://media.defense.gov/2019/Feb/04/2002085866/-1/-1/1/DOD-CLOUD-STRATEGY.PDF

--- “Our Story”. 2022. U.S. Department of Defense. Accessed 7 March 2022. https://www.defense.gov/About/

--- “Summary of the 2018 Department of Defense Artificial Intelligence Strategy: Harnessing AI to Advance Our Security and Prosperity,” 2019.

Department of the Air Force. “The United States Air Force Artificial Intelligence Annex to the Department of Defense Artificial Intelligence Strategy,” 2019. https://www.af.mil/Portals/1/documents/5/USAF-AI-Annex-to-DoD-AI-Strategy.pdf.

Digitalmarketplace.service.gov.U.K. 2018. “Digital Marketplace > Supplier Opportunities > Royal Navy: NELSON Data Platform Product Development.” 2018. Accessed 30 March

2022. https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/6275

Din, Allan M., ed. *Arms and artificial intelligence: weapon and arms control applications of advanced computing*. Stockholm International Peace Research Institute, 1987.

DiResta, Renee, Kris Shaffer, Becky Ruppel, David Sullivan, Robert Matney, Ryan Fox, Jonathan Albright, and Ben Johnson. "The tactics & tropes of the Internet Research Agency." (2019).

Dorman, Andrew, and Matthew Uttley. "Defence spending and procurement in the United Kingdom." In Research Handbook on the Arms Trade. Edward Elgar Publishing, 2020. http://dx.doi.org/10.4337/9781789900996.00022

Dougherty, Deborah, and Trudy Heller. "The illegitimacy of successful product innovation in established firms." *Organization Science* 5, no. 2 (1994): 200-218.

Dremliuga, Roman. "General Legal Limits of the Application of the Lethal Autonomous Weapons Systems within the Purview of International Humanitarian Law." *J. Pol. & L.* 13 (2020): 115. http://dx.doi.org/10.5539/jpl.v13n2p115

DSEI. DSEI 2019 Facts and Figures. Online. Accessed 24 September 2021. https://www.dsei.co.uk/dsei-2019-facts--figures

--- Gallery - DSEI 2019. Online. Accessed 24 September 2021.

--- Homepage. Online. Accessed 21 September 2021. https://www.dsei.co.uk/welcome

Dstl. *AI And Data Science Expansion For Dstl In The North*, 7 April 2021. Press Release. Accessed 20 April, 2021. https://www.gov.uk/government/news/ai-and-data-science-expansion-for-dstl-in-the-north

--- "Assurance of Artificial Intelligence and Assurance: A Dstl Biscuit Book," 2021, 1–34.

--- *Building Blocks For Artificial Intelligence And Autonomy: A Dstl Biscuit Book*. Ministry of Defence, 2021.

Dufour, Martin. “Will Artificial Intelligence Challenge NATO Interoperability ?,” no. 6 (2018): 6–9.

Dwyer, Andrew. "A Foundry of Artificial Intelligence? The case of UK national security?" In *Routledge Companion to Artificial Intelligence and National Security Policy*. 2022.

Dwyer, Sonya Corbin, and Jennifer L. Buckle. "The space between: On being an insider-outsider in qualitative research." *International journal of qualitative methods* 8, no. 1 (2009): 54-63. http://dx.doi.org/10.1177/160940690900800105

Edmonds, Jeffrey, Samuel Bendett, Anya Fink, Mary Chesnut, Dmitry Gorenburg, Michael Kofman, Kasey Stricklin, and Julian Waller. *Artificial Intelligence and Autonomy in Russia*. Center for Naval Analyses, 2021.

Elhefnawy, Nader. "Technological Hype and the Military Balance." *Available at SSRN 3182383* (2018). http://dx.doi.org/10.2139/ssrn.3182383

Emerson, Robert M., Rachel I. Fretz, and Linda L. Shaw. *Writing ethnographic fieldnotes*. No. Sirsi) i9780226206806. 1995. http://dx.doi.org/10.7208/chicago/9780226206851.001.0001

Engers Simen Gangnæs. "Dual-Use Technology and Defence–Civilian Spillovers: Evidence from the Norwegian Defence Industry." Master's thesis, 2013.

Erdélyi, Olivia J., and Judy Goldsmith. "Regulating artificial intelligence: Proposal for a global solution." In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 95-101. 2018. http://dx.doi.org/10.1145/3278721.3278731

Erickson Jr, S. A. *Fusing AI and simulation in military modeling*. No. UCRL-93404; CONF-8502108-1. Lawrence Livermore National Lab., CA (USA), 1985.

ERR News. “President Kaljulaid at CyCon 2019: Cyber attacks should not be easy weapons.” ERR News. (2019, 29 May). Accessed 24 September 2021. https://news.err.ee/946827/president-kaljulaid-at-cycon-2019-cyber-attacks-should-not-be-easy-weapon

European Defence Agency. “Artificial Intelligence: Joint Quest for Future Defence Applications.” 28 August 2020. Accessed 30 November 2020. https://eda.europa.eu/news-and-events/news/2020/08/25/artificial-intelligence-joint-quest-for-future-defence-applications.

European Defence Agency. “Stronger Communication & Radar Systems with Help of AI.”. European Defence Agency. August 3, 2020. Accessed 30 November 2020. https://www.eda.europa.eu/info-hub/press-centre/latest-news/2020/08/31/stronger-communication-radar-systems-with-help-of-ai.

European Parliament, Resolution of 12 Sep. 2018 on autonomous weapon systems, 2018/2752(RSP).

Fedasiuk, Ryan, Jennifer Melot, and Ben Murphy. “Harnessed Lightning: How the Chinese Military Is Adopting Artificial Intelligence,” *Center for Security and Emerging Threats.* October 2021. https://doi.org/10.51593/20200089.

Fernandez, German Carro, Sergio Martin Gutierrez, Elio Sancristobal Ruiz, Francisco Mur Perez, and Manuel Castro Gil. "Robotics, the new industrial revolution." *IEEE Technology and Society Magazine* 31, no. 2 (2012): 51-58. http://dx.doi.org/10.1109/MTS.2012.2196595

Fink, GA. “Adversarial Artificial Intelligence: State of the Malpractice.” *Journal of Information Warfare* 18, no. 4 (2019): 1–23. https://www.jstor.org/stable/26894691.

Finlan, Alastair. “The Shape of Warfare to Come: A Swedish Perspective 2020–2045.” *Defense & Security Analysis* 0, no. 0 (2021): 1–20. https://doi.org/10.1080/14751798.2021.1995976.

Fiott, Daniel. “A Revolution Too Far? US Defence Innovation, Europe and NATO’s Military-Technological Gap.” *Journal of Strategic Studies* 40, no. 3 (2017): 417–37. https://doi.org/10.1080/01402390.2016.1176565.

Fischer, Sophie-Charlotte. NATO and Artificial Intelligence: the Role of Public-Private Sector Collaboration. In Lucarelli, S., Marrone, A., & Morro, F.N. (2021). NATO Decision-making in the age of big data and artificial intelligence. NATO HQ. (p.74-83).

FitzGerald, Ben, and Jacqueline Parziale. "As technology goes democratic, nations lose military control." *Bulletin of the Atomic Scientists* 73, no. 2 (2017): 102-107. http://dx.doi.org/10.1080/00963402.2017.1288445.

Flournoy, M., Avril Haines, and Gabrielle Chefitz. "Building trust through testing." CSET Report. (2020).

Ford, George S., Thomas Koutsky, and Lawrence J. Spiwak. "A valley of death in the innovation sequence: an economic investigation." *Available at SSRN 1093006* (2007). http://dx.doi.org/10.2139/ssrn.1093006

Forrest, Lt Col Christopher D. "The Competition for Critical and Emerging Technology and Its Impact on Stability." *Emergent Issues for US National Security* (2020): 32

Foy, James. "Autonomous weapons systems: Taking the human out of international humanitarian law." *Dalhousie J. Legal Stud.* 23 (2014): 47. http://dx.doi.org/10.2139/ssrn.2290995

Frankel, Richard M., and Kelly Devers. "Qualitative research: a consumer's guide." *Education for Health: Change in Learning & Practice* 13, no. 1 (2000). http://dx.doi.org/10.1080/13576280050074534

Frankel, Richard, and Kelly Devers. "Study design in qualitative research—1: Developing questions and assessing resource needs." *Education for health* 13, no. 2 (2000): 251-261.

Franklin, Stan, and Art Graesser. "Is it an Agent, or just a Program?: A Taxonomy for Autonomous Agents." In International workshop on agent theories, architectures, and languages, pp. 21-35. Springer, Berlin, Heidelberg, 1996. http://doi.org/10.1080/13576280050074534

French Ministry of Armed Forces. *Artificial Intelligence in Support of Defense*. *September 2019*.

French, Shannon E., Kiju Lee, Margaret Kibben, and Susannah Rose. "Are We Ready for Artificial Ethics: AI and the Future of Ethical Decision Making." *The International Journal of Ethical Leadership* 6, no. 1 (2019): 24-53.

Frey, Carl Benedikt, and Michael Osborne. "The future of employment." (2013)

Fry, Hannah. *Hello World: How to be Human in the Age of the Machine*. Random House, 2018.

Garamone, Jim. “Fiscal 2023 Budget Funds Military for Today, Future”. U.S Department of Defense. DoD News. 28 March 2022. https://www.defense.gov/News/News-Stories/Article/Article/2980669/fiscal-2023-budget-funds-military-for-today-future/.

Gardner, Howard E. *Intelligence reframed: Multiple intelligences for the 21st century*. Hachette UK, 2000.

Geist, Edward, and Andrew J. Lohn. "How Might Artificial Intelligence Affect the Risk of Nuclear War?." (2018). http://dx.doi.org/10.7249/PE296

German Army Concepts and Capabilities Development Center. “Artificial Intelligence in Land Forces: A Position Paper by the German Army Concepts and Capabilities Development Center.” German Army Concepts and Capabilities Development Center. Bundeswehr, November 2019. Accessed November 22, 2020. https://www.bundeswehr.de/resource/blob/156026/3f03afe6a20c35d07b0ff56aa8d04878/download-positionspapier-englische-version-data.pdf.

GGE Statement. “Final Report”. CCW/MSP/2019/9. 13 December 2019. Accessed 11 March, 2022. https://undocs.org/CCW/MSP/2019/9.

--- “Report of the 2018 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems”. CCW/GGE.1/2018/3. (2018, 23 October). Accessed 11 March, 2022. https://undocs.org/en/CCW/GGE.1/2018/3.

Gilli, Andrea. “NATO-Mation”*: Strategies for Leading in the Age of Artificial Intelligence*. NATO Defense College. 2020. 1-112

--- “Preparing for the Atlantic Alliance toward the Age of Artificial Intelligence.” Nato Defense College. NDC Policy Brief., no. 4 (2019): 1–4.

Gilmore, John F. "Military applications of expert systems." *Future Generation Computer Systems* 1, no. 6 (1985): 403-410. http://dx.doi.org/10.1016/0167-739X(85)90024-X

Giry, Benoit, and Andy Smith. "Defence capability in the UK since 2010: explaining change in procurement practices." British Politics 15, no. 4 (2020): 433-455. http://dx.doi.org/10.1057/s41293-019-00125-4

Glaser, Barney G. “Conceptualization: On Theory and Theorizing Using Grounded Theory.” *International Journal of Qualitative Methods* 1, no. 2 (2002): 23–38. https://doi.org/10.1177/160940690200100203.

--- *Theoretical sensitivity*. University of California, *Advances in the methodology of grounded theory*. 1978.

Glaser, Barney G., and Anselm L. Strauss. *Discovery of grounded theory: Strategies for qualitative research*. Routledge, 2017. http://dx.doi.org/10.4324/9780203793206

Glaser, Barney G., and Judith Holton. "Remodeling grounded theory." In *Forum qualitative sozialforschung/forum: qualitative social research*, vol. 5, no. 2. 2004. https://doi.org/10.17169/fqs-5.2.607

Glaser, Charles L. "The causes and consequences of arms races." *Annual Review of Political Science* 3, no. 1 (2000): 251-276. https://doi.org/10.1146/annurev.polisci.3.1.251

Göğüş, Sezer İdil. "'Puzzling' Moments in the Field: Dilemmas on Positionality and Self-Reflexivity." (2019): 6.

Gokhberg, Leonid, Alexander Sokolov, and Alexander Chulok. “Russian S&T Foresight 2030: Identifying New Drivers of Growth.” *Foresight* 19, no. 5 (2017): 441–56. https://doi.org/10.1108/FS-07-2017-0029.

Goldman, Emily O. *Introduction: Military Diffusion and Transformation* in Goldman, Emily, and Thomas Mahnken, eds. *The Information Revolution in Military Affairs in Asia*. Springer, 2004. http://dx.doi.org/10.1057/9781403980441_1

Goldman, Emily O., and Leslie C. Eliason. *The diffusion of military technology and ideas*. Stanford University Press, 2003.

González, Roberto J. "Anthropology and the covert: Methodological notes on researching military and intelligence." *Anthropology Today* 28, no. 2 (2012): 21-25. https://doi.org/10.1111/j.1467-8322.2012.00863.x

Goodfellow, Ian J., Jonathon Shlens, and Christian Szegedy. "Explaining and harnessing adversarial examples." *arXiv preprint arXiv:1412.6572* (2014).

Gosling, Francis George. *The Manhattan Project: making the atomic bomb*. Diane Publishing, 1999. https://doi.org/10.2172/303853

Government Communications Headquarters. (GCHQ). *Pioneering A New National Security - The Ethics Of Artificial Intelligence*. Artificial Intelligence At GCHQ. Government Communications Headquarters, 2021. https://www.gchq.gov.uk/files/GCHQAIPaper.pdf.

Gray, Lia M., Gina Wong-Wylie, Gwen R. Rempel, and Karen Cook. "Expanding qualitative research interviewing strategies: Zoom video communications." *The Qualitative Report* 25, no. 5 (2020): 1292-1301. http://dx.doi.org/10.46743/2160-3715/2020.4212

Grbich, Carol. *Qualitative data analysis: An introduction*. Sage, 2012. http://dx.doi.org/10.4135/9781529799606

Greene, Daniel, Anna Lauren Hoffmann, and Luke Stark. "Better, nicer, clearer, fairer: A critical assessment of the movement for ethical artificial intelligence and machine learning." In *Proceedings of the 52nd Hawaii international conference on system sciences*. 2019. http://dx.doi.org/10.24251/HICSS.2019.258

Grissom, Adam. "The future of military innovation studies." *Journal of strategic studies* 29, no. 5 (2006): 905-934.

Guarino, Alessandro. "Autonomous intelligent agents in cyber offence." IEEE, 2013. In 2013 5th International Conference on Cyber Conflict (CYCON 2013), pp. 1-12.

Guoning, Zhang, and Shen Shoulin. "Application of complex network theory in combat modeling." In *2010 International Conference on Computer Application and System Modeling (ICCASM 2010)*.

Guzella, Thiago S., and Walmir M. Caminhas. "A review of machine learning approaches to spam filtering." *Expert Systems with Applications* 36, no. 7 (2009): 10206-10222. http://dx.doi.org/10.1016/j.eswa.2009.02.037

Hall, A.O. The Cyber Defense Review: CyCon U.S. 2019 Conference Papers. Cyber Defense Review. Vol 5 - Spring 2020.

Hammersley, Martyn, and Paul Atkinson. *Ethnography: Principles in practice*. Third Edition. Routledge, 2007. https://doi.org/10.4324/9780203944769

Haner, Justin, and Denise Garcia. "The artificial intelligence arms race: trends and world leaders in autonomous weapons development." Global Policy 10, no. 3 (2019): 331-337. http://dx.doi.org/10.1111/1758-5899.12713

Hawley, John K. "Patriot Wars." *Center for a New American Security* (2017). Accessed 20 November 2020. https://www.cnas.org/publications/reports/patriot-wars.

Hecht, Brent, Lauren Wilcox, Jeffrey P. Bigham, Johannes Schöning, Ehsan Hoque, Jason Ernst, Yonatan Bisk et al. "It's time to do something: Mitigating the negative impacts of computing through a change to the peer review process." ACM Future of Computing Blog, March 29, 2018.

Heikkilä, M. *NATO wants to set AI standards. If only its members agreed on the basics.* Politico. March 29 2021. Accessed March 12, 2022. https://www.politico.eu/article/nato-ai-artificial-intelligence-standards-priorities/.

Hicks, Kathleen H., Andrew Hunter, Lisa Sawyer Samp, and Gabriel Coll. "Assessing the third offset strategy." *Center for Strategic & International Studies* (2017).

--- "Accelerating Data and Artificial Intelligence for the Warfighter." Memorandum for senior Pentagon Leadership Commanders of the Combatant Commands Defense Agency and DoD Field Activity Directors. U.S. Deputy Secretary for Defense. June 21 2021.

--- "Establishment of the Chief Digital and Artificial Intelligence Officer". Memorandum for senior Pentagon Leadership Commanders of the Combatant Commands Defense Agency and DoD Field Activity Directors. U.S. Deputy Secretary for Defense. December 8 2021.

--- “Initial Operating Capability of the Chief Digital and Artificial Intelligence Officer”. Memorandum for senior Pentagon Leadership Commanders of the Combatant Commands Defense Agency and DoD Field Activity Directors. U.S. Deputy Secretary for Defense. February 1 2022.

Hickson, Mark. "Raising the question# 4 why bother attending conferences?." *Communication Education* 55, no. 4 (2006): 464-468. http://dx.doi.org/10.1080/03634520600917632

Hill, Steven. “Symposium: How Will Artificial Intelligence Affect International Law? AI’s Impact on Multilateral Military Cooperation: Experience from NATO.” *AJIL Unbound* 114 (2020): 147–51. https://doi.org/10.1017/aju.2020.27.

HM Government. “Defence and Security Industrial Strategy.” *Gov.U.K.*, March 2021. Accessed 6 November, 2021. https://www.gov.U.K./government/publications/defence-and-security-industrial-strategy.

--- “Industrial Strategy Artificial Intelligence Sector Deal,” 2018. Accessed 6 November, 2021. https://assets.publishing.service.gov.U.K./government/uploads/system/uploads/attachment_data/file/702810/180425_BEIS_AI_Sector_Deal__4_.pdf.

Hoadley, Daniel S, and Nathan J Lucas. “Artificial Intelligence and National Security.” 2018.

Holland Michel, Arthur. "The Black Box, Unlocked: Predictability and Understand-ability in Military AI." *Ginebra, United Nations Institute for Disarmament Research, disponible en* (2020). https://unidir.org/publication/black-box-unlocked.

Horowitz, Michael C. "The ethics & morality of robotic warfare: Assessing the debate over autonomous weapons." *Daedalus* 145, no. 4 (2016): 25-36. http://dx.doi.org/10.1162/DAED_a_00409

--- “Artificial Intelligence, International Competition, and the Balance of Power.” *Texas National Security Review* 1, no. 3 (2018): 37–57.

--- "The diffusion of military power." In *The Diffusion of Military Power*. Princeton University Press, 2010.

Horowitz, Michael C., Gregory C. Allen, Elsa B. Kania, and Paul Scharre. *Strategic competition in an era of artificial intelligence*. Center for a New American Security., 2018.

Horowitz, Michael C., Paul Scharre, & Alexander Velez-Green. (2019). A Stable Nuclear Future? The Impact of Autonomous Systems and Artificial Intelligence.

Howell, Chuck. "Overview of relevant findings and recommendations from The National Security Commission on AI Final Report." In *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications III*, vol. 11746, p. 1174603. International Society for Optics and Photonics, 2021. http://dx.doi.org/10.1117/12.2598358

Hua, Shin-Shin. "Machine Learning Weapons and International Humanitarian Law: Rethinking Meaningful Human Control." *Geo. J. Int'l L.* 51 (2019): 117.

Hunter, Andrew. "Assessing the Third Offset Strategy". March 16, 2017. Accessed 11 October, 2021. https://www.csis.org/analysis/assessing-third-offset-strategy

IEEE Ethics in Action. "Ethically Aligned Design Version 2: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems" (New York: IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, 2018). Accessed 6 July, 2021. https://ethicsinaction.ieee.org/

Imbrie, Andrew, Elsa Kania, and Lorand Laskai. "The Question of Comparative Advantage in Artificial Intelligence: Enduring Strengths and Emerging Challenges for the United States - CSET Policy Brief," 2020.

Imbrie, Andrew, Rebecca Gelles, James Dunham, and Catherine Aiken. "Contending Frames: Evaluating Rhetorical Dynamics in AI – CSET Policy Brief" 2021.

Ipsos. "Six in Ten (61%) Respondents Across 26 Countries Oppose the Use of Lethal Autonomous Weapons Systems." Ipsos, January 21, 2019. Accessed 10 January, 2021.

https://www.ipsos.com/en-us/news-polls/human-rights-watch-six-in-ten-oppose-autonomous-weapons.

Jackman, Anna H. “Rhetorics of Possibility and Inevitability in Commercial Drone Tradescapes.” *Geogr. Helv* 71 (2016): 1–6. https://doi.org/10.5194/gh-71-1-2016.

Jackson, Aaron L., and Kristine D. Kuenzli. "Something to Believe In: Aligning The Principle Of Honor With the Modern Battlefield." *Nat'l Sec. LJ* 6 (2018): 35.

JAIC Public Affairs. JAIC lays the foundation for the Tradewind Initiative in Partnership with Indiana Innovation Institute (IN3). February 1, 2021. Accessed October 15, 2021. https://www.ai.mil/news_02_01_21-jaic_tradewind_initiative_partnership_in3.html

--- Joint Artificial Intelligence Center to Pilot a Responsible AI Procurement Process. Press Release. Accessed 14 October 2021. https://www.ai.mil/news_07_27_21-jaic_to_pilot_a_responsible_ai_procurement_process.html

--- Request for Information (RFI) Release. 23 June 2021. AI in Defense: DoD’s Artificial Intelligence Blog. Accessed 14 October 2021. https://www.ai.mil/blog_06_23_21_rfi_release.html

--- “The JAIC Launches DoD AI Enterprise Infrastructure and Cybersecurity Subcommittee”. AI in Defense: DoD’s Artificial Intelligence Blog. 7 August 2020. Accessed 13 October, 2021. https://www.ai.mil/blog_08_07_20-the_jaic_launches_dod_ai_enterprise_infrastructure_and_cybersecurity_subcommittee.html JAIC.

--- “The JAIC Pushes the Envelope with DevSecOps through the Joint Common Foundation.” AI IN Defense: DoD AI Blog. 16 July 2020. Accessed 15 October, 2021. https://www.ai.mil/blog_07_16_20-jaic_pushes_the_envelope_with_devsecops_jcf.html

--- About the JAIC. Online. Accessed October 15, 2021. https://web.archive.org/web/20211015110829/https://www.ai.mil/about.html

Jankowski, Dominik P. Russia and the Technological Race in an Era of Great Power Competition. 2021. Center for Strategic & International Studies.

Jasper, Mila. Defense Innovation Board Adopts AI Testing, Digital Workforce Recruitment Resolutions. NextGov.com. September 16, 2020. Accessed 4 October, 2021. https://webcache.googleusercontent.com/search?q=cache:Dh0n_BZo4hQJ:https://www.nextgov.com/emerging-tech/2020/09/defense-innovation-board-adopts-ai-testing-digital-workforce-recruitment-resolutions/168527/+&cd=4&hl=en&ct=clnk&gl=ee

Jeangène-Vilmer, Jean-Baptise. A French Opinion on the Ethics of Autonomous Weapons. *War on the Rocks*. 2 June 2021. Accessed 8 February, 2022. https://warontherocks.com/2021/06/the-french-defense-ethics-committees-opinion-on-autonomous-weapons.

Jennings, Peter L. "The character to lead: A grounded theory ethnography of character in U.S. army combat leaders." PhD diss., Arizona State University, 2013.

Jensen, Benjamin. *Forging the sword: Doctrinal change in the US Army*. Stanford University Press, 2016.

Jensen, Benjamin M., Christopher Whyte, and Scott Cuomo. "Algorithms at war: the promise, peril, and limits of artificial intelligence." *International Studies Review* 22, no. 3 (2020): 526-550. http://dx.doi.org/10.1093/isr/viz025

Jensen, Benjamin, Scott Cuomo, and Chris Whyte. "Wargaming with ATHENA: how to make militaries smarter, faster, and more efficient with artificial intelligence." *War on the Rocks* 5 (2018).

Jeong, Hyuk-Jin, Hyeon-Jae Lee, Chang Hyun Shin, and Soo-Mook Moon. "IONN: Incremental offloading of neural network computations from mobile devices to edge servers." In Proceedings of the ACM Symposium on Cloud Computing, pp. 401-411. 2018. http://dx.doi.org/10.1145/3267809.3267828

Jobin, Anna, Marcello Ienca, and Effy Vayena. "The global landscape of AI ethics guidelines." *Nature Machine Intelligence* 1, no. 9 (2019): 389-399. http://dx.doi.org/10.1038/s42256-019-0088-2

Johnson, James. China-US Competition in AI: Destabilising and Intensifying. Online Presentation. International Institute for Strategic Studies. 24 November 2021. Accessed 9

December 2021. https://www.iiss.org/events/2021/11/china-us-competition-in-ai-destabilising-and-intensifying

--- "Artificial intelligence & future warfare: implications for international security." *Defense & Security Analysis* 35, no. 2 (2019): 147-169. http://dx.doi.org/10.1080/14751798.2019.1600800

--- "The end of military-techno Pax Americana? Washington's strategic responses to Chinese AI-enabled military technology." *The Pacific Review* 34, no. 3 (2021): 351-378. http://dx.doi.org/10.1080/09512748.2019.1676299

--- "'Catalytic Nuclear War' in the Age of Artificial Intelligence & Autonomy: Emerging Military Technology and Escalation Risk between Nuclear-Armed States." *Journal of Strategic Studies*, 2021, 1–41. https://doi.org/10.1080/01402390.2020.1867541.

--- "Artificial Intelligence: A Threat to Strategic Stability." *Strategic Studies Quarterly 14*, no.1 2020, 16–39.

Joo, Jaehun. "Adoption of Semantic Web from the perspective of technology innovation: A grounded theory approach." *International journal of human-computer studies* 69, no. 3 (2011): 139-154. http://dx.doi.org/10.1016/j.ijhcs.2010.11.002

Jordan, Michael I., and Tom M. Mitchell. "Machine learning: Trends, perspectives, and prospects." *Science* 349, no. 6245 (2015): 255-260. http://dx.doi.org/10.1126/science.aaa8415

Jowett, Adam, Elizabeth Peel, and Rachel Shaw. "Online interviewing in psychology: Reflections on the process." *Qualitative Research in Psychology* 8, no. 4 (2011): 354-369. http://dx.doi.org/10.1080/14780887.2010.500352

Jowett, Adam. "Carrying out qualitative research under lockdown-practical and ethical considerations." *Impact of Social Sciences Blog* (2020).

Kamarck, Kristy N. "Diversity, inclusion, and equal opportunity in the armed services: Background and issues for congress." (2017).

Kania, Elisa. "Great Power Competition and the AI Revolution: A Range of Risks to Military and Strategic Stability." Lawfare. Tuesday 19 September. Accessed 1 December 2021.

--- "Battlefield singularity." *Artificial Intelligence, Military Revolution, and China's Future Military Power, CNAS* (2017).

Kaplan, Andreas, and Michael Haenlein. "Siri, Siri, in My Hand: Who's the Fairest in the Land? On the Interpretations, Illustrations, and Implications of Artificial Intelligence." *Business Horizons* 62, no. 1 (2019): 15–25. https://doi.org/10.1016/j.bushor.2018.08.004.

Kazim, Emre, Denise Almeida, Nigel Kingsman, Charles Kerrigan, Adriano Koshiyama, Elizabeth Lomas, and Airlie Hilliard. "Innovation and opportunity: review of the UK's national AI strategy." *Discover Artificial Intelligence* 1, no. 1 (2021): 1-10. http://dx.doi.org/10.2139/ssrn.3969257

Kier, Elizabeth. "Imagining war." In *Imagining War*. Princeton University Press, 2017.

King, Andrew A., and Baljir Baatartogtokh. "How useful is the theory of disruptive innovation?." *MIT Sloan management review* 57, no. 1 (2015): 77.

Klinger, Joel, Juan C. Mateos-Garcia, and Konstantinos Stathoulopoulos. "A narrowing of AI research?." *Available at SSRN 3698698* (2020). http://dx.doi.org/10.2139/ssrn.3698698

Konaev, Margarita, and Husanjot Chahal. "The Path of Least Resistance: Multinational Collaboration on AI for Military Logistics and Sustainment." *Center for Security and Emerging Threats*, 2021. https://cset.georgetown.edu/wp-content/uploads/CSET-Path-of-Least-Resistance.pdf.

Konaev, Margarita, Husanjot Chahal, Ryan Fedasiak, Tina Huang, and Ilya Rahkovsky. *U.S. Military Investments In Autonomy And AI A Budgetary Assessment*. CSET Policy Brief. Center for Security and Emerging Challenges, 2020. https://cset.georgetown.edu/wp-content/uploads/CSET-U.S.-Military-Investments-in-Autonomy-and-AI-A-Budgetary-Assessment.pdf.

Krafft, P. M., Meg Young, Michael Katell, Karen Huang, and Ghislain Bugingo. "Defining AI in policy versus practice." In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, pp. 72-78. 2020. http://dx.doi.org/10.1145/3375627.3375835

Kramer, Eric-Hans. *Organizing doubt: Grounded theory, army units and dealing with dynamic complexity*. Vol. 22. Copenhagen Business School Press DK, 2007.

Krause, Peter, Ora Szekely, Mia Bloom, Fotini Christia, Sarah Zukerman Daly, Chappell Lawson, Zoe Marks et al. "COVID-19 and Fieldwork: Challenges and Solutions." *PS: Political Science & Politics* 54, no. 2 (2021): 264-269. http://dx.doi.org/10.1017/S1049096520001754

Kugler, Mikołaj. "The United States of America's Embrace of Artificial Intelligence for Defense Purposes." In *Artificial Intelligence and Its Contexts*, pp. 183-199. Springer, Cham, 2021. http://dx.doi.org/10.1007/978-3-030-88972-2_12

Kuhlman, Caitlin, Latifa Jackson, and Rumi Chunara. "No computation without representation: Avoiding data and algorithm biases through diversity." *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '20).* Association for Computing Machinery. New York, USA, 3593. (2020). http://dx.doi.org/10.1145/3394486.3411074.

Kvale, Steinar. *Doing Interviews*. Sage. (2008) http://dx.doi.org/10.4135/9781849208963

Larsson, Gerry, Paul T. Bartone, Miepke Bos-Bakx, Erna Danielsson, Ljubicá Jelusic, Eva Johansson, Rene Moelker et al. "Leader development in natural context: A grounded theory approach to discovering how military leaders grow." *Military Psychology* 18, no. sup1 (2006): S69-S81. http://dx.doi.org/10.1207/s15327876mp1803s_6.

Laskai, Lorand, and Graham Webster. "Translation: Chinese expert group offers' governance principles' for responsible AI'." New America Foundation (2019).

LeCun, Yann & Bengio, Y. & Hinton, Geoffrey. Deep Learning. *Nature.* 521 (2015): 436-44. 10.1038/nature14539. http://dx.doi.org/10.1038/nature14539

Lee, Chang Hyun, and Sang Yong Kim. "Differential effects of determinants on multi-dimensions of trade show performance: By three stages of pre-show, at-show, and post-show activities." Industrial Marketing Management 37, no. 7 (2008): 784-796. https://doi.org/10.1016/j.indmarman.2008.01.006.

Leopold. George. "NATO Targets AI Interoperability." EnterpriseAI. November 2 2020. Accessed 22 October, 2021. https://www.enterpriseai.news/2020/11/02/nato-targets-ai-interoperability/

Lewis, Larry. *Insights for the third offset: Addressing challenges of autonomy and artificial intelligence in military operations*. Center for Naval Analyses Arlington United States, 2017.

Lin-Greenberg, Erik. "Allies and Artificial Intelligence: Obstacles to Operations and Decision-Making." *Texas National Security Review* 3, no. 2 (2020): 57–60. https://repositories.lib.utexas.edu/bitstream/handle/2152/81858/TNSRVol3Issue2Lin-Greenberg.pdf?sequence=2&isAllowed=y.

Lucarelli, S., Marrone, A., & Moro, F. N.. NATO Decision-making in the age of big data and artificial intelligence. (2021).

Maas, Matthijs M. "How viable is international arms control for military artificial intelligence? Three lessons from nuclear weapons." *Contemporary Security Policy* 40, no. 3 (2019): 285-311. http://dx.doi.org/10.1080/13523260.2019.1576464

MacDonald, Fraser, Rachel Hughes, and Klaus J. Dodds. "Observant States." *Geopolitics and Visual Culture. London and New York: IB Tauris* (2010). http://dx.doi.org/10.5040/9780755620494

MacDonald, Fraser. "Geopolitics and 'the vision thing': regarding Britain and America's first nuclear missile." *Transactions of the Institute of British Geographers* 31, no. 1 (2006): 53-71. http://dx.doi.org/10.1111/j.1475-5661.2006.00196.x

--- "Perpendicular sublime: Regarding rocketry and the Cold War." Observant States: Geopolitics and Visual Culture. London: IB Tauris (2010): 267-289. http://dx.doi.org/10.5040/9780755620494.ch-012

Machi, Vivienne. 2021. "Artificial Intelligence Leads NATO'S New Strategy For Emerging And Disruptive Tech". *C4ISRNET*. Accessed 22 April, 2021. https://www.c4isrnet.com/artificial-intelligence/2021/03/14/artificial-intelligence-leads-natos-new-strategy-for-emerging-and-disruptive-tech/

MacLean, Lauren M., Nabila Rahman, Robin L. Turner, and Jack Corbett. "Disrupted Fieldwork: Navigating Innovation, Redesign, and Ethics during an Ongoing Pandemic." *Letter from the President* (2020): 1.

Malik, Tariq H. "Defence investment and the transformation national science and technology: A perspective on the exploitation of high technology." *Technological Forecasting and Social Change* 127 (2018): 199-208. http://dx.doi.org/10.1016/j.techfore.2017.09.020

Mahnken, Thomas G. *Technology and the American way of war*. Columbia University Press, 2008.

Markotkin, Nikolai, and Elena Chernenko. "Developing artificial intelligence in Russia: Objectives and reality." *Carnegie Moscow Center* 5 (2020).

Massoumi, Narzanin, Tom Mills, and David Miller. "Secrecy, coercion and deception in research on 'terrorism' and 'extremism'." *Contemporary Social Science* 15, no. 2 (2020): 134-152. http://dx.doi.org/10.1080/21582041.2019.1616107

McCann, Eugene. "Urban policy mobilities and global circuits of knowledge: Toward a research agenda." Annals of the Association of American Geographers 101, no. 1 (2011): 107-130. http://dx.doi.org/10.1080/00045608.2010.520219

McCarthy, John, Marvin L. Minsky, Nathaniel Rochester, and Claude E. Shannon. "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955". (1955). Accessed 24 April, 2022. http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html

McDonald, Jack N. I. A.. "Drones and the European Union: Prospects for a common future." Chatham House. (2018).

McDougall, Carrie. "Autonomous weapon systems and accountability: putting the cart before the horse." *Melbourne Journal of International Law* 20, no. 1 (2019): 58-87.

McGhee, Gerry, Glenn R. Marland, and Jacqueline Atkinson. "Grounded Theory Research: Literature Reviewing and Reflexivity." *Journal of Advanced Nursing 60*, no. 3 (2007): 334–42. https://doi.org/10.1111/j.1365-2648.2007.04436.x.

Mcleary, Paul. While China Spends, Trump Budget Looks Flat For 2021. *Breaking Defense.* 8 January 2021. Accessed 15 October, 2021. https://breakingdefense.com/2020/01/while-china-spends-trump-budget-looks-flat-for-2021/

McNamara, Andrew, Justin Smith, and Emerson Murphy-Hill. "Does ACM's code of ethics change ethical decision making in software development?." In *Proceedings of the 2018 26th ACM joint meeting on European software engineering conference and symposium on the foundations of software engineering*, pp. 729-733. 2018. http://dx.doi.org/10.1145/3236024.3264833

Minárik, T., S. Alatalu, S. Biondi, M. Signoretti, I. Tolga, and G. Visky. "11th International Conference on Cyber Conflict: Silent Battle." 2019.

Ministère des Armées. "Artificial Intelligence in Support of Defense: Report of the AI Task Force," September (2019): 1–32.

Ministry of Defence Contracts Website (2019). Accessed 1 October, 2019. https://www.contracts.mod.uk/.

Ministry of Defence, Defence and Security Accelerator, and James Heappey MP. 2020. "Revolutionary Artificial Intelligence Warship Contracts Announced - GOV.U.K.." Gov.U.K. 2020. Accessed 5 April, 2021. https://www.gov.uk/government/news/revolutionary-artificial-intelligence-warship-contracts-announced.

Ministry of Defence, Defence Science and Technology Laboratory, and Stuart Andrew. 2018. "Streets Ahead: British AI Eyes Scan Future Frontline in Multinational Urban Experiment - GOV.uk." Www.Gov.uk. 2018. https://www.gov.uk/government/news/streets-ahead-british-ai-eyes-scan-future-frontline-in-multinational-urban-experiment.

Ministry of Defence. "UK armed forces biannual diversity statistics." (2021).

--- “Defence Technology Framework,” 2019. Accessed 5 April, 2022. https://assets.publishing.service.gov.uk./government/uploads/system/uploads/attachment_data/file/830139/20190829-DTF_FINAL.pdf

--- “Global Britain in a Competitive Age”: Integrated Review of Security, Defence, Development and Foreign Policy, HM Government 16 March 2021

--- “Unmanned Aircraft Systems”. Joint Doctrine Publication 0-30-2. September 2017.

--- *Defence in a Competitive Age*. CP411. Ministry of Defence. (2021)

--- *Flagship AI Lab Announced As Defence Secretary Hosts First Meet Between British And American Defence Innovators*, 2018. Accessed 14 March, 2022. https://www.gov.uk/government/news/flagship-ai-lab-announced-as-defence-secretary-hosts-first-meet-between-british-and-american-defence-innovators

--- *Integrated Operating Concept – 2025*. Development, Concepts and Doctrine Centre. August 2021.

MITRE. *Perspectives on research in artificial intelligence and artificial general intelligence relevant to DoD*. MITRE CORP MCLEAN VA MCLEAN United States, 2017.

Moore, Jenny. "A personal insight into researcher positionality." *Nurse researcher* 19, no. 4 (2012): 11-14. http://dx.doi.org/10.7748/nr2012.07.19.4.11.c9218

Morgan, Forrest, Benjamin Boudreaux, Andrew Lohn, Mark Ashby, Christian Curriden, Kelly Klima, and Derek Grossman. *Military Applications of Artificial Intelligence: Ethical Concerns in an Uncertain World*. *RAND Corporation*. RAND Corporation, 2020. https://doi.org/10.7249/RR3139-1.

Müller, Vincent C. "Ethics of artificial intelligence." *The Routledge social science handbook of AI* (2021): 122-137. http://dx.doi.org/10.4324/9780429198533-9

Murray, Rob. (2020). “Building a resilient innovation pipeline for the Alliance.” NATO Review. 1 September 2020.

https://www.nato.int/docu/review/articles/2020/09/01/building-a-resilient-innovation-pipeline-for-the-alliance/index.html.

Nader, Laura. 'Up the anthropologist: Perspectives gained from studying up', *Reinventing anthropology,* (1969): 284–311. http://dx.doi.org/10.2307/j.ctvw04j6x.6

NATO Advisory Group on Emerging and Disruptive Technologies. Annual Report 2020. March 2021.

NATO Allied Command Transformation. 2020 Fact Sheet: Military Uses of Artificial Intelligence, Automation, and Robotics (MUAAR). 2020. Accessed 10 March, 2022. https://www.act.nato.int/application/files/5515/8257/4725/2020_mcdc-muaar.pdf

--- "Artificial Intelligence - A Game Changer for the Military." NATO. October 25, 2019. Accessed 10 March, 2022. https://www.act.nato.int/articles/artificial-intelligence-game-changer-military.

--- *NATO defence ministers meeting.* 27 June 2019. Accessed 20 August, 2021. https://www.act.nato.int/articles/nato-defence-ministers-meeting

--- Strategic Foresight Analysis (SFA) 2017 Report. Accessed 10 September, 2021. https://www.act.nato.int/images/stories/media/doclibrary/171004_sfa_2017_report_hr.pdf

NATO CCDCOE. 11th International Conference on Cyber Conflict: Silent Battle. Proceedings 2019. Online. Accessed 24 September 2021 https://ccdcoe.org/library/publications/11th-international-conference-on-cyber-conflict-silent-battle-proceedings-2019/.

NATO Communications and Information Agency. "NATO Agency Contributes Expertise to Machine Learning Hackathon." NATO Communications and Information Agency, February 28, 2020. Accessed 9 January, 2021. https://www.ncia.nato.int/about-us/newsroom/nato-agency-contributes-expertise-to-machine-learning-hackathon.html.

--- "NATO Community Discusses Data, Cloud and Securing the Alliance at NIAS." NATO Communications and Information Agency, October 15, 2019. Accessed 9 January, 2021.

https://www.ncia.nato.int/about-us/newsroom/nato-community-discusses-data--cloud-and-securing-the-alliance-at-nias-.html

NATO Data Exploitation Framework Policy

NATO Reflection Group. “NATO 2030: United for a New Era - Analysis and Recommendations of the Reflection Group Appointed by the NATO Secretary-General”. 25 November 2020. Accessed 14 September, 2021. https://www.nato.int/nato_static_fl2014/assets/pdf/2020/12/pdf/201201-Reflection-Group-Final-Report-Uni.pdf

NATO Science & Technology Organization (STO). “Science & Technology Trends 2020-2040: Exploring the S&T Edge.” NATO Science & Technology Organization. NATO, March 2020. Accessed 5 November, 2020. https://apps.dtic.mil/sti/citations/AD1131124

--- First Disruptive Technologies Table-Top Exercise (D3TX). 15 February 2021. Accessed 14 September, 2021. https://www.sto.nato.int/Lists/STONewsArchive/displaynewsitem.aspx?ID=609

--- 2020 Highlights: Empowering the Alliance’s Technological Edge. (2021).

NATO. Military Decision on MC 362/1 - NATO Rules Of Engagement (2003).

NATO. Webpage – NATO Organization. (2020). Last modified 20 August 2020. https://www.nato.int/cps/en/natohq/structure.htm

--- “Alliance's Future Innovation Priorities Discussed at High-Level Meeting.” NATO, October 1, 2020. Accessed https://www.nato.int/cps/en/natohq/news_178358.htm.

--- “Cooperation on Artificial Intelligence Will Boost Security and Prosperity on Both Sides of the Atlantic, NATO Deputy Secretary-General Says.” NATO. NATO, October 28, 2020. Accessed 9 January 2021. https://www.nato.int/cps/en/natolive/news_179231.htm.

--- “Cyber defence”. Online - last updated 2 July 2021. Accessed 14 September 2021. https://www.nato.int/cps/en/natohq/topics_78170.htm

--- “NATO 2030 Factsheet”. Online. Accessed 21 September 2021. https://www.nato.int/nato_static_fl2014/assets/pdf/2021/6/pdf/2106-factsheet-nato2030-en.pdf

--- NATO Ambassadors and Military Leaders Meet to Discuss Disruptive Technologies. Oct. 2, 2019. Accessed 15 September, 2021. https://www.nato.int/cps/en/natohq/news_169264.htm?selectedLocale=en

--- SACT’s Opening Remarks to the NAC/MC Away Day. Mar. 22, 2018. Accessed 20 September, 2021. https://www.act.nato.int/images/stories/media/speeches/180319_nac-mc-awayday.pdf

Nichols, Nicole, Mark Raugas, Robert Jasper, and Nathan Hilliard. "Faster fuzzing: Reinitialization with deep neural models." *arXiv preprint arXiv:1711.02807* (2017).

Noorman, Merel, and Deborah G. Johnson. “Negotiating Autonomy and Responsibility in Military Robots.” *Ethics and Information Technology* 16, no. 1 (2014): 51–62. https://doi.org/10.1007/s10676-013-9335-0.

Nonaka, Ikujiro, and Georg Von Krogh. "Perspective—Tacit knowledge and knowledge conversion: Controversy and advancement in organizational knowledge creation theory." *Organization science* 20, no. 3 (2009): 635-652.

Nouwens, Meia, and Helena Legarda. "Emerging technology dominance: what China’s pursuit of advanced dual-use technologies means for the future of Europe’s economy and defence innovation." *International Institute for Strategic Studies/Mercator Institute for China Studies China Security Project* (2018): 2018-12. http://dx.doi.org/10.4324/9780429198533-9

Nouwens, Meia, Erica Pepe and Franz Stefan-Gady. “NATO and artificial intelligence”. IISS Podcast - Episode 70. 2020. https://www.iiss.org/blogs/podcast/2021/04/nato-artificial-intelligence

NSCAI. *Interim Report.* National Security Commission on Artificial Intelligence. November 2019.

--- *First Quarter Recommendations Memo*. National Security Commission on Artificial Intelligence. March 2020.

--- *Second Quarter Recommendations Memo*. National Security Commission on Artificial Intelligence. July 2020.

--- *Interim Report and Third Quarter Recommendations Memo*. National Security Commission on Artificial Intelligence. October 2020.

--- *Final Report - National Security Commission on Artificial Intelligence (AI)*. National Security Commission on Artificial Intelligence. March 2021.

--- *"Interim Report"*. The National Security Commission on Artificial Intelligence. November 2019.

O'Leary, Daniel E, "Expert Systems–History, Structure, Definitions, Characteristics, Life Cycle and Applications." *Marshall School of Business, University of Southern California.* (2021).

O'Rourke, Ronald. "Renewed Great Power Competition: Implications for Defense—Issues for Congress." *CRS Report* (2020).

Office of the Under Secretary of Defense. Defense Budget Overview: United States Department of Defense Fiscal Year 2021 Budget Request (Washington, DC: Department of Defense, February 2020), 1–8, https://comptroller.defense.gov/Portals/45/Documents/defbudget/fy2021/fy2021_Budget_Request_Overview_Book.pdf

--- "Budget Rollout Brief: Fiscal Year 2020 Budget Request". 2019. Accessed 25 November 2021. /web/20211125125342/https://media.defense.gov/2019/Mar/12/2002099931/-1/-1/1/FY-2020-BUDGET-ROLLOUT-BRIEF.PDF.

O'Neil, Cathy. *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown, 2016.

Osoba, Osonde A., and William Welser IV. *An intelligence in our image: The risks of bias and errors in artificial intelligence*. Rand Corporation, 2017. http://dx.doi.org/10.7249/RR1744

Payne, Kenneth. "Artificial intelligence: a revolution in strategic affairs?." *Survival* 60, no. 5 (2018): 7-32. http://dx.doi.org/10.1080/00396338.2018.1518374

--- *I, Warbot: The Dawn of Artificially Intelligent Conflict*. Hurst Publishers, 2021. http://dx.doi.org/10.1093/oso/9780197611692.001.0001

--- *Strategy, evolution, and war: from apes to artificial intelligence*. Georgetown University Press, 2018.

Pearson, Gavin, Phil Jolley, and Geraint Evans. "A Systems Approach to Achieving the Benefits of Artificial Intelligence in UK Defence." *arXiv preprint arXiv:1809.11089* (2018).

Pepe, Erica. "NATO and Collective Thinking on AI." *International Institute for Strategic Studies. IISS*. November 13, 2020. https://www.iiss.org/blogs/military-balance/2020/11/nato-artificial-intelligence.

Petrella, Stephanie, Chris Miller, and Benjamin Cooper. "Russia's artificial intelligence strategy: the role of state-owned firms." *Orbis* 65, no. 1 (2021): 75-100. http://dx.doi.org/10.1016/j.orbis.2020.11.004

Pichai, Sundar. AI at Google: our principles. Google blog. June 7 2018. Accessed October 15, 2021. https://blog.google/technology/ai/ai-principles/

Polyakova, Alina. "Weapons of the weak: Russia and AI-driven asymmetric warfare." *The Brookings Institute, November* 15 (2018).

Poulin, Chris, Brian Shiner, Paul Thompson, Linas Vepstas, Yinong Young-Xu, Benjamin Goertzel, Bradley Watts, Laura Flashman, and Thomas McAllister. "Predicting the risk of suicide by analyzing the text of clinical notes." *PloS one* 9, no. 1 (2014): e85733. http://dx.doi.org/10.1371/journal.pone.0085733

Price, Matt. "Roots of dissent: The Chicago Met Lab and the origins of the Franck Report." *Isis* 86, no. 2 (1995): 222-244.

Prichard, Ian, and James O'Nions. "Defence Systems & Equipment International: Lifeblood of the Arms Trade." *Review of African Political Economy* (2005): 475-477.

Prime Minister's Office. *PM To Announce Largest Military Investment In 30 Years*, 2020. Retrieved from https://www.gov.uk/government/news/pm-to-announce-largest-military-investment-in-30-years.

Rabionet, Silvia E. "How I learned to design and conduct semi-structured interviews: an ongoing and continuous journey." *Qualitative Report* 16, no. 2 (2011): 563-566. http://dx.doi.org/10.46743/2160-3715/2011.1070

Ramamoorthy, Anand, and Roman Yampolskiy. "Beyond mad? the race for artificial general intelligence." *ITU J* 1, no. 1 (2018): 77-84.

Rech, Matthew F. "A critical geopolitics of observant practice at British military airshows." *Transactions of the Institute of British Geographers 40*, no. 4 (2015): 536-548. http://dx.doi.org/10.1111/tran.12093

Rech, Matthew F. "A critical geopolitics of RAF recruitment." PhD diss., Newcastle University, 2012.

Rice, Gareth. (2010) 'Reflections on interviewing elites', *Area*, 42(1), pp. 70–75. https://doi.org/10.1111/j.1475-4762.2009.00898.x.

Roberts, Rachel. 'What is the DSEI arms fair taking place in London this week and why is it so controversial?'. The Independent. 14 September 2017. Accessed 12 October, 2021. https://www.independent.co.uk/news/uk/home-news/dsei-arms-fair-excel-centre-protests-arrests-saudi-arabia-yemen-conflict-a7941741.html

Roff, Heather M. "The Frame Problem: The AI "Arms Race" Isn't One." *Bulletin of the Atomic Scientists* 75, no. 3 (2019): 95–98. 2019. https://doi.org/10.1080/00963402.2019.1604836.

Roff, Heather M., and Richard Moyes. "Meaningful human control, artificial intelligence and autonomous weapons." In *Briefing Paper Prepared for the Informal Meeting of Experts on Lethal Autonomous Weapons Systems, UN Convention on Certain Conventional Weapons*. 2016.

Royal Navy, "NavyX - About". (Online). Accessed 1 October. 2021. https://www.royalnavy.mod.uk/news-and-latest-activity/operations/united-kingdom/navy-x

Sagiroglu, Seref, and Duygu Sinanc. "Big data: A review." In *2013 international conference on collaboration technologies and systems (CTS)*. IEEE, 2013. pp. 42-47. http://dx.doi.org/10.1109/CTS.2013.6567202

Saldaña, Johnny. *The Coding Manual for Qualitative Researchers (2nd Ed.), SAGE Publications Inc.* (2013). https://doi.org/10.1017/CBO9781107415324.004.

Sample, Ian. AI becomes grandmaster in 'fiendishly complex' StarCraft II. October 30, 2019. The Guardian. Accessed October 4, 2021. https://www.theguardian.com/technology/2019/oct/30/ai-becomes-grandmaster-in-fiendishly-complex-starcraft-ii

Santoni de Sio, Filippo, and Jeroen Van den Hoven. "Meaningful human control over autonomous systems: A philosophical account." *Frontiers in Robotics and AI* 5 (2018): 15. http://dx.doi.org/10.3389/frobt.2018.00015

Sassoli, Marco. "Autonomous weapons and international humanitarian law: Advantages, open technical questions and legal issues to be clarified." *International Law Studies/Naval War College* 90 (2014): 308-340.

Saxon, Dan. "A Human Touch: Autonomous Weapons, DoD Directive 3000.09 and the Interpretation of 'Appropriate Levels of Human Judgment over the Use of Force.'" Chapter. In Autonomous Weapons Systems: Law, Ethics, Policy, edited by Nehal Bhuta, Susanne Beck, Robin Geiβ, Hin-Yan Liu, and Claus Kreβ, 185–208. Cambridge: Cambridge University Press, 2016. https://doi.org/10.1017/CBO9781316597873.009.

Sayler, Kelley M. “Artificial Intelligence and National Security. R45178. Version 8.,” 2020. https://www.inss.org.il/he/publication/artificial-intelligence-and-national-security/.

Scharre, Paul, and Michael C Horowitz. “Strategic Competition in an Era of Artificial Intelligence.” *CNAS*, no. July (2018): 1–27.

--- "Autonomy in Weapon Systems." *Center for a New American Security Working Paper* (2015).

Scharre, Paul. "Autonomous weapons and operational risk." Center for a New American Security (2016): 2019.

--- "Autonomous weapons and stability." PhD diss., King's College London, 2020.

--- “Debunking the AI Arms Race Theory (Summer 2021).” *Texas National Security Review* (2021).

--- “Army of None: Autonomous Weapons and the Future of War”. March 2018.

--- “A Security Perspective: Security Concerns and Possible Arms Control Approaches.” Perspectives on Lethal Autonomous Weapon Systems. United Nations Office for Disarmament Affairs. Occasional Papers. No. 30, November 2017. http://dx.doi.org/10.18356/6b5db3ba-en

--- “Autonomous Weapons and Operational Risk - Ethical Autonomy Project.” no. February (2016).

Schneider, Benjamin, Vicente González-Romá, Cheri Ostroff, and Michael A. West. "Organizational climate and culture: Reflections on the history of the constructs in the Journal of Applied Psychology." *Journal of applied psychology* 102, no. 3 (2017): 468.

Schneier, Bruce. "Invited Talk: The Coming AI Hackers." In *International Symposium on Cyber Security Cryptography and Machine Learning*, pp. 336-360. Springer, Cham, 2021. http://dx.doi.org/10.1007/978-3-030-78086-9_26

Schuller, Alan L. "At the crossroads of control: The intersection of artificial intelligence in autonomous weapon systems with international humanitarian law." *Harv. Nat'l Sec. J.* 8 (2017): 379.

Schwarz, Elke. “Death Machine - The Ethics of Violent Technologies”. Book Launch at Queen Mary University of London. 9 March 2019. Accessed 13 October, 2021. https://www.qmul.ac.uk/politics/events/items/death-machines-the-ethics-of-violent-technologies.html

Sechser, Todd S., Neil Narang, and Caitlin Talmadge. "Emerging technologies and strategic stability in peacetime, crisis, and war." *Journal of strategic studies* 42, no. 6 (2019): 727-735. http://dx.doi.org/10.1080/01402390.2019.1626725

Select Committee on Artificial Intelligence. "AI in the UK: Ready, willing, and able?." House of Lords. HL 100. (2018).

--- “Corrected oral evidence: Artificial Intelligence.” House of Lords. 28 November 2017.

Selyanin, Yaroslav. "US Intelligence Community and Artificial Intelligence." *USA & Canada: ekonomika, politika, kultura* 6 (2021): 52-70. https://doi.org/10.31857/S268667300015219-0

Sharkey, Noel. "Automating warfare: lessons learned from the drones." *Journal of Law, Information and Science* 21, no. 2 (2011): 140-154. http://dx.doi.org/10.5778/JLIS.2011.21.Sharkey.1

Shea, Jamie. Afghanistan, Iraq, and the Future of NATO. Institut Montaigne. 8 September 2021. Accessed 12 October, 2021. https://www.institutmontaigne.org/en/blog/afghanistan-iraq-and-future-nato

Shinar, J., A. W. Siegel, and Y. I. Gold. "On the analysis of a complex differential game using artificial intelligence techniques (military systems)." In *Proceedings of the 27th IEEE Conference on Decision and Control*, pp. 1436-1441. IEEE, 1988. http://dx.doi.org/10.1109/CDC.1988.194562

Siemens, George, and Peter Tittenberger. *Handbook of emerging technologies for learning*. Canada: University of Manitoba, 2009.

Simonite, Tom. “Pentagon Will Expand AI Project Prompting Protests at Google”. Wired. 29 May 2018. Accessed 15 October, 2021. https://www.wired.com/story/googles-contentious-pentagon-project-is-likely-to-expand/

Sisson, Melanie, Jennifer Spindel, Paul Scharre, and Vadim Kozyulin. "The Militarization of Artificial Intelligence." *United Nations* (2020).

Situma, Sasaka Peter. "The effectiveness of trade shows and exhibitions as organizational marketing tool (analysis of selected companies in Mombasa)." International journal of business and social science 3, no. 22 (2012): 219-230.

Slijper, Frank, Alice Beck, and Daan Kayser. “State of AI: Artificial Intelligence, the Military and Increasingly Autonomous Weapons.” Pax for Peace, 2018.

Soare, Simona. “Innovation as Adaptation: NATO and Emerging Technologies | Strengthening Transatlantic Cooperation”. The German Marshall Fund of the United States (GMF).11 June 2021. https://www.gmfus.org/news/innovation-adaptation-nato-and-emerging-technologies

Sprenger, Sebastian. “NATO Tees up Negotiations on Artificial Intelligence in Weapons.” *C4ISRNET*, 21 April 2021. Accessed 3 April, 2022. https://www.c4isrnet.com/artificial-intelligence/2021/04/27/nato-tees-up-negotiations-on-artificial-intelligence-in-weapons/.

STCTTS. “Artificial Intelligence: Implications for NATO’s Armed Forces.” NATO Science and Technology Committee Sub-Committee on Technology Trends and Security. NATO, October 13, 2019. Accessed 2 November, 2020. https://www.nato-pa.int/download-file?filename=%2Fsites%2Fdefault%2Ffiles%2F2019-10%2FREPORT%20149%20STCTTS%2019%20E%20rev.%201%20fin-%20ARTIFICIAL%20INTELLIGENCE.pdf.

Steinhardt, Jacob, Pang Wei W. Koh, and Percy S. Liang. "Certified defenses for data poisoning attacks." *Advances in neural information processing systems* 30. pp. 3517-3529. (2017).

Stiglitz, Joseph. The Future of Work - You and AI. Speech at the Royal Society, London. 11 September, 2018. Accessed 13 October 2021. https://royalsociety.org/science-events-and-lectures/2018/09/you-and-ai/

Stodola, Petr, Jan Drozd, and Jan Nohel. "Model of Surveillance in Complex Environment Using a Swarm of Unmanned Aerial Vehicles." In *International Conference on Modelling and Simulation for Autonomous Systems*, pp. 231-249. Springer, Cham, 2020.

Stoltenberg, Jens (NATO Secretary General). Opening Speech: 2030 Brussels Forum. 14 June 2020.

Stop the Arms Fair. "About". Online. Accessed 12 October, 2021. https://www.stopthearmsfair.org.uk/about/

Stowsky, Jay. "Secrets to Shield or Share? New Dilemmas for Military R&D Policy in the Digital Age." *Research Policy* 33, no. 2 (2004): 257–69. https://doi.org/10.1016/j.respol.2003.07.002.

Strachan, Hew. "Global Britain in a competitive age: strategy of the Integrated Review." *Journal of the British Academy* (2021)

Street, Michael, Peter Lenk, Ivana Ilic Mestric, and Marc Richter. "Lessons learned from initial exploitation of big data and AI to support NATO decision making." In *STO experts meeting on "Big data and AI to support military decision making", Bordeaux*. 2018.

Suchman, Lucy. "Algorithmic warfare and the reinvention of accuracy." *Critical Studies on Security* 8, no. 2 (2020): 175-187. http://dx.doi.org/10.1080/21624887.2020.1760587

Szegedy, Christian, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. "Intriguing properties of neural networks." *arXiv preprint arXiv:1312.6199* (2013). Paper presented at 2nd International Conference on Learning Representations, ICLR 2014, Banff, Canada.

Tabassi, Elham, Kevin J. Burns, Michael Hadjimichael, Andres D. Molina-Markham, and Julian T. Sexton. "A taxonomy and terminology of adversarial machine learning." *NIST IR* (2019): 1-29. https://doi.org/10.6028/NIST.IR.8269-draft

Taddeo, Mariarosaria, and Luciano Floridi. "Regulate artificial intelligence to avert cyber arms race." *Nature.* (2018): 296-298. http://dx.doi.org/10.2139/ssrn.3198556

Taddeo, Mariarosaria, David McNeish, Alexander Blanchard, and Elizabeth Edgar. "Ethical Principles for Artificial Intelligence in National Defence." *Philosophy & Technology*, no. 0123456789 (October 13, 2021). https://doi.org/10.1007/s13347-021-00482-3.

Taddeo, Mariarosaria, Tom McCutcheon, and Luciano Floridi. "Trusting artificial intelligence in cybersecurity is a double-edged sword." *Nature Machine Intelligence* 1, no. 12 (2019): 557-560. http://dx.doi.org/10.1007/s13347-021-00482-3

Tamir, Michael. What is machine learning? *Berkeley School of Information* (UC Berkeley School of Information, June 26, 2020). Accessed January 6, 2021. https://ischoolonline.berkeley.edu/blog/what-is-machine-learning/.

Tarraf, Danielle C., William Shelton, Edward Parker, Brien Alkire, Diana Gehlhaus, Justin Grana, Alexis Levedahl, Jasmin Léveillé, Jared Mondschein, James Ryseff, Ali Wyne, Daniel Elinoff, Edward Geist, Benjamin N. Harris, Eric Hui, Cedric Kenney, Sydne Newberry, Chandler Sachs, Peter Schirmer, Danielle Schlang, Victoria M. Smith, Abbie Tingstad, Padmaja Vedula, and Kristin Warren. "The Department of Defense Posture for Artificial Intelligence: Assessment and Recommendations." Santa Monica, CA: RAND Corporation (2019). http://dx.doi.org/10.7249/RR4229

Taylor, Robert M. "Capability, Cognition and Autonomy." *RTO HFM Symposium on "The Role of Humans in Intelligent and Automated Systems,"* no. October (2002): 7–9. http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA422249.

Taylor, Trevor. "Artificial Intelligence in Defence: When AI Meets Defence Acquisition Processes and Behaviours." *The RUSI Journal* 164, no. 5-6 (2019): 72-81. http://dx.doi.org/10.1080/03071847.2019.1694229

--- "Unpacking the UK'S Newly Announced Centre On Artificial Intelligence". Blog. *RUSI Commentary*. 2020. Accessed 18 April 2022. https://www.rusi.org/explore-our-research/publications/commentary/unpacking-uks-newly-announced-centre-artificial-intelligence.

The White House Office of Science and Technology Policy. SUMMARY OF THE 2018 WHITE HOUSE SUMMIT ON ARTIFICIAL INTELLIGENCE FOR AMERICAN INDUSTRY. May 10 2018.

Thornton, Rod, and Marina Miron. "Towards the 'third revolution in military affairs' the Russian military's use of AI-enabled cyber warfare." *The RUSI Journal* 165, no. 3 (2020): 12-21. https://doi.org/10.1080/03071847.2020.1765514

Tigner, Brooks. NATO leaders set to launch new 'defence innovation accelerator' at summit. 2 June 2021. Janes. Accessed 14 March, 2021. https://www.janes.com/defence-news/news-detail/nato-leaders-set-to-launch-new-defence-innovation-accelerator-at-summit

Trump White House Archives. "Executive Order On Maintaining American Leadership In Artificial Intelligence," The White House, 2019. Accessed 2 December 2021. https://trumpwhitehouse.archives.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/.

U.S. Air Force. The United States Air Force Artificial Intelligence Annex to The Department of Defense Artificial Intelligence Strategy. 2019. Accessed October 11, 2021. https://www.af.mil/Portals/1/documents/5/USAF-AI-Annex-to-DoD-AI-Strategy.pdf

U.S. Congress. "One Hundred Fifteenth Congress of the United States of America." At the second session. Begun and held at the City of Washington on Wednesday, the third day of January, two thousand and eighteen. An Act. To authorize appropriations for fiscal year 2019 for military activities of the Department of Defense, for military construction, and for defense activities of the Department of Energy, to prescribe military personnel strengths for such fiscal year, and for other purposes. HR 5515 (2018).

U.S. Department of Defense Office of Inspector General. "Project Announcement: Announcement of the Joint Evaluation of the National Security Agency Integration of Artificial Intelligence (DoD OIG Project No. D2021-DEV0SI-01S2.000; NSA OIG Project No. EV-21-0011)". 6 August 2021. https://www.dodig.mil/reports.html/Article/2723656/project-announcement-announcement-of-the-joint-evaluation-of-the-national-secur/

U.S. Treasury. “Treasury Targets Sanctions Evasion Networks and Russian Technology Companies Enabling Putin’s War” Press Release. March 31 2022. Accessed 5 April, 2022. https://home.treasury.gov/news/press-releases/jy0692

UNIDIR. “The Weaponization of Increasingly Autonomous Technologies: Artificial Intelligence a Primer for CCW Delegates,” no. 8 (2018).

Valášek, Tomáš. “NATO at 70: Enter the Technological Age.” *NATO Policy Brief*, no. 10 (2019).

Vallor, Shannon. "The future of military virtue: Autonomous systems and the moral deskilling of the military." In *2013 5th International Conference on Cyber Conflict (CYCON 2013)*, pp. 1-15. IEEE, 2013.

van der Merwe, Joanna. “NATO Leadership on Ethical AI is Key to Future Interoperability.” February 17, 2021. CEPA. Accessed 21 September, 2021. https://cepa.org/nato-leadership-on-ethical-ai-is-key-to-future-interoperability/

Venema, Liesbeth. "Defining a role for AI ethics in national security." *Nature Machine Intelligence* 3, no. 5 (2021): 370-371. http://dx.doi.org/10.1038/s42256-021-00344-9. [quotes from Taddeo, Mariarosaria interviewed within the article]

Vergun, David. Defense Innovation Board Recommends AI Ethical Guidelines. U.S. Department of Defense. November 1, 2021. Accessed October, 2021. https://www.defense.gov/Explore/News/Article/Article/2006646/defense-innovation-board-recommends-ai-ethical-guidelines/

Vignard, Kerstin. *The Weaponization of Increasingly Autonomous Technologies: Considering how Meaningful Human Control Might Move the Discussion Forward*. United Nations Institute for Disarmament Research, 2014.

Voo, Julia, Simon Jones, Daniel Cassidy, and Anina Schwarzenbach. “National Cyber Power Index 2020.” *Belfer Center.* September 2020.

Wagemann Jr, John. *The United States Air Force and Artificial Intelligence: Moving Forward by Learning from Past Technology Implementation*. AIR FORCE FELLOWS

PROGRAM MAXWELL AFB AL MAXWELL AFB United States, 2020. Accessed 16 December 2021. 2020. https://apps.dtic.mil/sti/pdfs/AD1112360.pdf

Wagner, Ben. "Ethics as an escape from regulation. From “ethics-washing” to ethics-shopping?." In *Being Profiled*, pp. 84-89. Amsterdam University Press, 2018. http://dx.doi.org/10.2307/j.ctvhrd092

Wagner, Markus. "The dehumanization of international humanitarian law: legal, ethical, and political implications of autonomous weapon systems." *Vand. J. Transnat'l L.* 47 (2014): 1371.

Warren, Aiden, and Alek Hillas. "Friend or frenemy? The role of trust in human-machine teaming and lethal autonomous weapons systems." *Small Wars & Insurgencies* 31, no. 4 (2020): 822-850. http://dx.doi.org/10.1080/09592318.2020.1743485

Wasilow, Sherry and Thorpe, Joelle B., “Artificial Intelligence, Robotics, Ethics, and the Military: A Canadian Perspective.” AI Magazine. Spring 2019, Vol. 40 Issue 1, p 37-48. http://dx.doi.org/10.1609/aimag.v40i1.2848

Webster, Graham, Rogier Creemers, Paul Triolo, and Elsa Kania. “China’s Plan to ‘Lead’ in AI: Purpose, Prospects, and Problems.” New America Foundation, Aug. 1, 2017, https://www. newamerica.org/cybersecurity-initiative/blog/chinas-plan-lead-ai-purpose-prospects-and-problems/. Accessed 10 March 2022.

West, Sarah Myers, Meredith Whittaker, and Kate Crawford. "Discriminating systems." *AI Now* (2019).

Whittaker, Meredith, Kate Crawford, Roel Dobbe, Genevieve Fried, Elizabeth Kaziunas, Varoon Mathur, Sarah Mysers West, Rashida Richardson, Jason Schultz, and Oscar Schwartz. *AI now report 2018*. New York: AI Now Institute at New York University, 2018.

Whittaker, Meredith. "The steep cost of capture." *Interactions* 28, no. 6 (2021): 50-55. http://dx.doi.org/10.1145/3488666

Whyte, Christopher. "Deepfake news: AI-enabled disinformation as a multi-level public policy challenge." *Journal of Cyber Policy* 5, no. 2 (2020): 199-217. http://dx.doi.org/10.1080/23738871.2020.1797135

--- "Poison, Persistence, and Cascade Effects: AI and Cyber Conflict." *Strategic Studies Quarterly*, 2020, 18–46.

Wilcken, Patrik. Arms companies are hiding behind governments – it's time we held them accountable. Amnesty International. September 9, 2019. Accessed 12 October, 2021. https://www.amnesty.org/en/latest/news/2019/09/arms-companies-must-be-held-accountable/

Wohlstetter, Albert, Paul H. Nitze, Joseph Alsop, Morton H. Halperin, and Jeremy J. Stone. "Is There a Strategic Arms Race? (II): Rivals but No" Race"." *Foreign Policy 16* (1974): 48-92. http://dx.doi.org/10.2307/1147844

Wood, Elisabeth Jean, Rogers, Douglas, Sivaramakrishnan, Kalyanakrishnan, and Almeling, Rene. 2020. *Resuming Field Research in Pandemic Times*. Social Science Research Council. Available at https://items.ssrc.org/covid-19-and-the-social-sciences/social-research-and-insecurity/resuming-field-research-in-pandemic-times

Work, Bob. The Third U.S. Offset Strategy and its Implications for Partners and Allies. Deputy Secretary of Defense Speech. January 28, 2015. As Delivered by Deputy Secretary of Defense Deputy Secretary of Defense Bob Work. Willard Hotel, Washington, D.C.

Yu, Kun-Hsing, Andrew L. Beam, and Isaac S. Kohane. "Artificial intelligence in healthcare." *Nature biomedical engineering* 2, no. 10 (2018): 719-731. http://dx.doi.org/10.1038/s41551-018-0305-z