RUSI
Royal United Services Institute
for Defence and Security Studies

QINETIQ

Occasional Paper

Trust in AI
Rethinking Future Command

Christina Balis and Paul O'Neill

Trust in AI

Rethinking Future Command

Christina Balis and Paul O'Neill

RUSI Occasional Paper, June 2022

Royal United Services Institute
for Defence and Security Studies

191 years of independent thinking on defence and security

The Royal United Services Institute (RUSI) is the world's oldest and the UK's leading defence and security think tank. Its mission is to inform, influence and enhance public debate on a safer and more stable world. RUSI is a research-led institute, producing independent, practical and innovative analysis to address today's complex challenges.

Since its foundation in 1831, RUSI has relied on its members to support its activities. Together with revenue from research, publications and conferences, RUSI has sustained its political independence for 191 years.

QinetiQ is a global information, knowledge and technology-based company operating primarily in the defence and security markets. Its mission is to protect lives and secure the vital interests of its customers. QinetiQ applies cross-domain knowledge and technical expertise to solving customers' mission challenges, helping them to develop, test and deploy new and enhanced capabilities with the assurance they need to operate safely and effectively.

Published in 2022 by the Royal United Services Institute for Defence and Security Studies.

RUSI Occasional Paper, June 2022. ISSN 2397-0286 (Online).

Royal United Services Institute
for Defence and Security Studies
Whitehall
London SW1A 2ET
United Kingdom
+44 (0)20 7747 2600
www.rusi.org
RUSI is a registered charity (No. 210639)

Contents

Executive Summary

THE TRADITIONAL RESPONSE to the acceptance challenge posed by the military use of AI has been to insist on humans maintaining 'meaningful human control' as a way of engendering confidence and trust. This is no longer an adequate response when considering both the ubiquity and rapid advances of AI and related underpinning technologies. AI will play an essential, growing role in a broad range of command and control (C2) activities across the whole spectrum of operations. While less directly threatening in the public mind than 'killer robots', the use of AI in military decision-making presents key challenges as well as enormous advantages. Increasing human oversight over the technology itself will not prevent inadvertent (let alone intentional) misuse.

This paper builds on the premise that trust at all levels (operators, commanders, political leaders and the public) is essential to the effective adoption of AI for military decision-making and explores key related questions. What does trust in AI actually entail? How can it be built and sustained in support of military decision-making? What changes are needed for a symbiotic relationship between human operators and artificial agents for future command?

Trust in AI can be said to exist when humans hold certain expectations of the AI's behaviour without reference to intentionality or morality on the part of the artificial agent. At the same time, however, trust is not just a function of the technology's performance and reliability – it cannot be assured solely by resolving issues of data integrity and interpretability, important as they are. Trust-building in military AI must also address needed changes in military organisation and command structures, culture and leadership. Achieving an overall appropriate level of trust requires a holistic approach. In addition to trusting the purpose for which AI is put to use, military commanders and operators need to sufficiently trust – and be adequately trained and experienced on how to trust – the inputs, process and outputs that underpin any particular AI model. However, the most difficult, and arguably most critical, dimension is trust at the level of the organisational ecosystem. Without changes to the institutional elements of military decision-making, future AI use in C2 will remain suboptimal, confined within an analogue framework. The effective introduction of any new technology, let alone one as transformational as AI, requires a fundamental rethinking of how human activities are organised.

Prioritising the human and institutional dimensions does not mean applying more control over the technology; rather, it requires reimagining the human role and contribution within the evolving human–machine cognitive system. Future commanders will need to be able to lead diverse teams across a true 'Whole Force' that integrates contributions from across the military, government and civilian spheres. They must understand enough about their artificial teammates to be capable of both collaborating with and challenging them. This is more akin to the murmuration of starlings than the genius of the individual 'kingfisher' leader. For new concepts of command and leadership to develop, Defence must rethink its approach not only

to training and career management but also to decision-making structures and processes, including the size, location and composition of future headquarters.

AI is already transforming warfare and challenging longstanding human habits. By embracing greater experimentation in training and exercises, and by exploring alternative models for C2, Defence can better prepare for the inevitable change that lies ahead.

Introduction

AI IS CHANGING HOW humans think and make decisions. Looking ahead, it will increasingly affect how humans prioritise various cognitive processes, adapt their learning, their behaviours and their training, and more broadly transform their institutions. These changes are still not wholly evident across militaries. Despite new technologies and war's rapidly evolving character, today's armed forces do not differ dramatically in organisational structure from the professional armies of post-Napoleonic Europe. Too many people are still involved in military tasks that technology can do better and faster, and not enough attention is paid to rethinking humans' cognitive contribution to human–machine teams that will be needed to address future questions of command and control (C2).

This paper builds on an earlier report produced by QinetiQ, which looked at trust as a fundamental component of military capability and an essential requirement for military adaptability in the 2020s.[1] The current paper explores the latest trends and thinking about the growing use of AI in military decision-making. It is not directly concerned with the ethical (or indeed legal) issues of this trend, important though they are.[2] Instead, it emphasises the importance and implications of trust as a factor in military command in the age of AI.

AI's potentially profound impact on military decision-making and C2 has attracted little attention outside specialist groups. Most public attention is on the advantages and risks of the technology rather than the potential and limitations of human cognitive and institutional constructs. More than two decades ago, leading sociobiologist E O Wilson aptly captured humanity's current challenge. The real problem, according to Wilson, is that 'we have paleolithic emotions; medieval institutions; and god-like technology'.[3] Over the past few decades, technology has progressed at a much faster pace than human capacity to adapt to it. Emphasising AI's technological attributes at the expense of the human and institutional dimensions of its growing use will only compound the challenge.

---

1. QinetiQ, 'The Trust Factor: The Role of Trust in Training and in Generating Defence Capability', September 2021.
2. See, for example, the guidelines on the use of lethal autonomous weapon systems under the Convention on Certain Conventional Weapons at UN Office for Disarmament Affairs, 'Background on LAWS in the CCW', <https://www.un.org/disarmament/the-convention-on-certain-conventional-weapons/background-on-laws-in-the-ccw>, accessed 20 June 2022. See also Stop Killer Robots, <https://www.stopkillerrobots.org>, accessed 20 June 2022.
3. Quoted in *Harvard Magazine*, 'An Intellectual Entente', 9 October 2009, <https://www.harvardmagazine.com/breaking-news/james-watson-edward-o-wilson-intellectual-entente>, accessed 20 May 2022.

In many areas, military experience with AI is still limited, and more work is needed to understand the implications of AI's growing role in human decision-making. This paper aims to trigger a broader debate about the cultural and organisational changes required within the UK defence enterprise, including the role of command and commanders, to ensure the optimal use of AI in future military decision-making.

The paper's insights were drawn from the broader literature related to AI, human cognition, military decision-making and theories of trust. The research, conducted between September 2021 and February 2022, benefited substantially from interviews with a wide range of experts and users from across Defence, academia and industry.

The first two chapters provide the theoretical backdrop for the paper. Chapter I explores the concepts of AI and trust, while Chapter II analyses the role of human agency and AI's impact on humans' cognitive capacity to make choices and decisions. Combining the concepts of trust, AI and human agency, Chapter III proposes a five-dimensional framework for developing trust in AI-enabled military decision-making. Chapter IV expands the analysis to draw attention to C2, with a particular focus on AI's implications for the people and institutional structures that have traditionally underpinned the exercise of authority and direction of armed forces. The concluding chapter proposes areas for further research on future command, leadership and 'Whole Force' teams.

I. AI and Trust

THERE IS NO standard definition of AI or trust as it relates to AI. Both concepts are subject to varied interpretations and, on occasion, fierce debate. Rather than attempting to synthesise the entire literature dedicated to these two terms, this chapter establishes a baseline definition to frame the subsequent discussion on the role of trust in AI when applied to military C2.

Nature and Types of AI

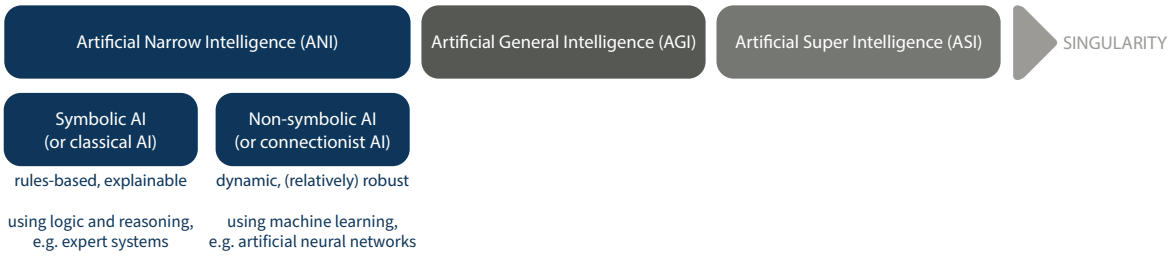
The concept of AI originates with the famous Turing test of 1950, which occurred a few years before the coining of the term.[4] It is easier to conceptualise AI by focusing on what it does rather than what it is. AI 'seeks to make computers do the sorts of things that minds can do'.[5] At its most basic, it can be understood as the capacity for virtual information processing in pursuit of a specific task. Just as 'intelligence' (or 'the mind') has many dimensions and varied uses, so does AI. Accordingly, AI draws on different ideas and techniques from a broad range of disciplines extending to not only mathematics and computer engineering, but also philosophy, economics, neuroscience, psychology and linguistics.[6]

There are, broadly speaking, three different levels of AI: artificial narrow intelligence, typically referred to as 'narrow AI'; artificial general intelligence, sometimes referred to as human-level AI; or the even more powerful artificial super intelligence that exceeds human levels of intelligence.[7] At this point, some argue that there will be a singularity in which AI either becomes self-aware or reaches an ability for continuous improvement that will enable it to

4. The term 'AI' was coined in 1956, two years after Turing's premature death, during the Dartmouth Summer Research Project on Artificial Intelligence, organised by American computer scientist John McCarthy. According to the Turing test, if based on answers to various questions, a remote (human) interrogator was unable to distinguish correctly between a human and machine, then the machine could be said to be intelligent. The machine only needed to fool the interrogator 30% of the time to pass the test. Since the test was developed, many programmes have achieved human-level performance, but the debate of whether they can be said to be human-like beyond the narrow confines of a particular task goes on. See Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, 4th Edition (Harlow: Pearson Education, 2022), p. 1035.
5. Margaret A Boden, *AI: Its Nature and Future* (Oxford: Oxford University Press, 2016), p. 1.
6. Russell and Norvig, *Artificial Intelligence*, pp. 19–35.
7. Narrow AI is designed to replicate, and in some cases surpass, human intelligence for a *specific* task or purpose. Artificial general intelligence refers to the performance of cognitive functions across a broad range of domains, achieving human ability to perform tasks, while artificial super intelligence goes even further in assuming intelligence that surpasses human ability and even reaches elements of consciousness.

evolve beyond human control.[8] The latter two levels are seen as still some way off, although how far off is contested. For now, though, it is the emergence of more advanced applications of narrow AI, such as advanced robotics, combined with the explosion in computing power, that mostly animates the current debate over the military use of AI. This paper focuses on applications of narrow AI.

Figure 1: Simplified Categorisation of AI Types



Source: Author generated.

Within narrow AI, there are further categories, although the techniques are not wholly discrete and are often used in combination. The most common distinction is between symbolic AI, often described as being based on logic, and sub-symbolic or non-symbolic AI, based on adaptation or learning.[9] Symbolic AI relies on sequential instructions and top-down control, making it particularly well suited to defined problems and rules-based processes. Non-symbolic AI, within which neural networks are a common approach, involves parallel, bottom-up processing and approximate reasoning; this is most relevant to dynamic conditions and situations in which data is incomplete. Where symbolic AI offers precision and explainability, non-symbolic AI involving, for example, neural networks is less brittle (a missing node in a network does not render the entire network inoperative) and able to recognise patterns in the absence of clear rules or consistent evidence.[10]

There are three common types of machine learning, differentiated by the type of feedback that contributes to the agent's learning process: supervised learning; unsupervised learning; and reinforcement learning. In supervised learning, the system is trained to generate hypotheses or take specific action in pursuit of target values or outputs (referred to as labels) based on specific

8. Nisha Talagala, 'Don't Worry About the AI Singularity: The Tipping Point is Already Here', *Forbes*, 21 June 2021.
9. Non-symbolic AI is sometimes referred to as parallel distributed processing or connectionist AI, a term made popular in the 1980s. Its earliest predecessor, starting in the 1940s, was known as 'cybernetics'.
10. Boden, *AI*, pp. 78–100. While non-symbolic AI is more robust than classical AI approaches in the sense that it suffers 'gradual degradation', no AI today is entirely immune to errors in the training data or to adversarial behaviour. Efforts to increase both robustness and explainability remain critical to strengthening the trustworthiness of AI systems.

inputs (for example, image recognition). Unsupervised learning has no set specifications or labels and there is no explicit feedback; rather, the system learns by finding patterns in the data (for example, DNA sequence clustering). Reinforcement learning depends on a feedback loop that steadily reinforces the system's learned behaviour through a trial-and-error or reward-and-punishment mechanism (for example, advanced robotics or driverless cars). Unlike supervised learning, the input data used in reinforcement learning is not predefined, which allows for broader exploration, but unlike unsupervised learning, it has an intended application or overall aim (linked to maximising overall reward).[11]

All three types of machine learning, regardless of the degree of supervision or self-regulation, raise important issues of trust and trustworthiness. The level and nature of trust required differ depending on the purpose for which AI is used.

The Concept of Trust

Trust describes an interaction between two or more agents. Traditional definitions of trust assume the existence of a reasonable belief on the part of the trustor in both the competency and the goodwill (or motive) of the trustee. For many, questions of moral integrity (or intention) are what differentiates trust from other concepts such as confidence. Others see trust as having a broader scope and referent than confidence, the latter seen as a discrete judgement linked to a specific event.[12] What tends to unite most definitions of trust is a sense of vulnerability.[13] Without the possibility of betrayal, without the existence of risk, there can be no trust.

It is because of the presumed moral element implied in classical conceptions of trust that some challenge the use of the term to describe the human relationship with artificial agents. At current levels of narrow AI, they argue, we cannot attribute intentionality or moral agency to AI systems, and therefore the use of the term 'trust' is misplaced.[14] Others take a less purist perspective and apply the term in a way that reflects everyday usage implying confidence in the system's reliability.

11. Boden, *AI*, pp. 47–48; Russell and Norvig, *Artificial Intelligence*, pp. 671, 840–66. Within reinforcement learning, there are still various approaches, including passive, active and inverse reinforcement learning.
12. Barbara D Adams, 'Trust vs. Confidence', Defence Research and Development Canada, 28 June 2005, <https://cradpdf.drdc-rddc.gc.ca/PDFS/unc48/p524541.pdf>, accessed 10 December 2021.
13. Stanford Encyclopedia of Philosophy, 'Trust', updated 10 August 2020, <https://plato.stanford.edu/entries/trust>, accessed 10 December 2021.
14. Others go further in arguing that there is no *need* to trust an AI because we can engineer AI for accountability and hold humans to account for any failures. See Joanna Bryson, 'AI & Global Governance: No One Should Trust AI', UN University, Centre for Policy Research, 13 November 2018, <https://cpr.unu.edu/publications/articles/ai-global-governance-no-one-should-trust-ai.html>, accessed 10 December 2021.

Trust as a term is widely used in computer science.[15] More importantly, trust remains a fundamental aspect of public and user acceptance of AI. National policies, regulations and expert advice on AI today routinely underscore the need for 'trustworthy AI'.[16] DARPA's Air Combat Evolution programme, for example, is exploring methodologies to model and objectively measure pilot trust in AI during dogfighting. Recognising these unresolved definitional issues, the authors have opted to slightly adapt the term 'trust' consistent with common practice.

The authors' adapted concept of trust entails certain expectations about the AI's performance without assuming a particular motive on the AI's part. Positive expectations of the *behaviour* of an artificial agent may thus be a sufficient condition for the existence of trust regardless of *intention*.[17]

In most current discussions of AI, the focus tends to be on the human acting as the trustor and the system behaving as the trustee, although any cognitive agent, including autonomous robots and intelligent machines, could in principle perform the role of trustor as well.[18] Understood that way, trust becomes 'a facilitator of interactions among the members of a system, whether these be human agents, artificial agents or a combination of both (a hybrid system)'.[19] Indeed, in cases of more mature applications of AI, the trustee is most likely to encompass both the AI-enabled system (artificial agent) and the provider of that system (human agent). At the current level of AI, trust appears to be a one-way relationship concerning the degree to which the human 'trusts' the AI, rather than genuinely two-way trust, where the AI takes a view on human performance.

Various factors determine (human) trust in technology, including but not limited to the trustor's level of competence and disposition to trust, and the overall environment or context (including broader cultural and institutional dynamics). Beyond such human- and environment-specific considerations, what defines the level of trust a person or organisation has in AI are

15. See, for example, Donovan Artz and Yolanda Gil, 'A Survey of Trust in Computer Science and the Semantic Web', *Journal of Web Semantics* (Vol. 5, No. 2, 2007), pp. 58–71.
16. See, for example, European Commission, 'White Paper on Artificial Intelligence: A European Approach to Excellence and Trust', 19 February 2020; US National Security Commission on Artificial Intelligence, 'Final Report', March 2021, <https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf>, accessed 10 January 2022; HM Government, *National AI Strategy*, CP 525 (London: The Stationery Office, 2021).
17. See also Steven Lockey et al., 'A Review of Trust in Artificial Intelligence: Challenges, Vulnerabilities and Future Directions', *Proceedings of the 54th Hawaii International Conference on Systems Sciences* (2021), p. 5464.
18. David-Olivier Jaquet-Chiffelle and Hans Buitelaar (eds), 'D17.4: Trust and Identification in the Light of Virtual Persons', Future of Identity in the Information Society Consortium, 25 June 2009, p. 41, <http://www.fidis.net/fileadmin/fidis/deliverables/new_deliverables/fidis-wp17-del17.4_Trust_and_Identification_in_the_Light_of_Virtual_Persons.pdf>, accessed 15 December 2021.
19. Mariarosaria Taddeo, 'Trusting Digital Technologies Correctly', *Minds and Machines* (Vol. 27, No. 4, 2017), pp. 565–68.

the technology's performance, process (how it generates specific outputs) and, importantly, purpose.[20] All three shape the design *and* deployment of AI-enabled systems.

In addition to technical robustness and safety, privacy, fairness, transparency and accountability are some of the most commonly raised issues affecting public trust in AI.[21] However, it is largely because of the difficulties of devising appropriate algorithms, understanding the internal structures of complex software systems and ascribing accountability for algorithmically based decisions that a further consideration is always added to the list of key attributes of trustworthy AI: this is interchangeably referred to as human agency, oversight or meaningful control.[22] Maintaining human oversight of the use of technology may be, in some cases, the only protection against the risk of unintentionally biased, inscrutable and/or poorly regulated AI-enabled systems.

The Concept of Control

Control is often seen as the opposite of trust. When there is trust in an agent's ability to perform a task, there is no need for supervision. However, humans will often tend to intervene even in situations where the AI is better placed to make a decision. Under-trusting can be as risky or counterproductive as over-trusting. In fact, just as absolute control is rare, so is absolute trust. A careful balance is necessary between appropriate levels of trust and adequate levels of control in the development and use of AI. This is at the heart of concepts such as 'calibrated trust' or adaptable/adaptive autonomy. Trust is calibrated according to the AI's capabilities, and expectations of what AI can or cannot do will influence the level of trust.[23] Similarly, in the case of adaptable autonomy, the user's ability to tailor the level of autonomy can support greater trust levels.[24] This is particularly critical in national security decision-making where the implications of trusting or not trusting AI have potentially the greatest consequences.

Concerns over the role of technology in human affairs are nothing new. Many see the debate over AI as no different to the arguments about technology that preceded it. According to this

20. Keng Siau and Weiyu Wang, 'Building Trust in Artificial Intelligence, Machine Learning, and Robotics', *Cutter Business Technology Journal* (Vol. 31, No. 2, 2018), pp. 47–53.
21. Questions of what constitutes 'responsible AI' are at the heart of research related to ethics in AI. Legal scholars have challenged the notion of a 'responsibility gap', which has shaped ethical debates on AI for more than a decade. See Andrea Bertolini, 'Artificial Intelligence and Civil Liability', report prepared for European Parliament, Policy Department for Citizens' Rights and Constitutional Affairs, July 2020, p. 33.
22. European Commission, High-Level Expert Group on Artificial Intelligence, 'Deliverable 1: Ethics Guidelines for Trustworthy AI', 2019, <https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai>, accessed 15 December 2021; European Parliamentary Research Service, *The Ethics of Artificial Intelligence: Issues and Initiatives* (Brussels: EU, 2020), pp. 29–36.
23. See, for example, Sue Halpern, 'The Rise of A.I. Fighter Pilots', *New Yorker*, 17 January 2022.
24. Karen Soper, 'The Future of Teams: The Move Towards Human/Machine Teams and How This Impacts on Trust', in QinetiQ, 'The Trust Factor', p. 18.

argument, AI constitutes an evolution, not a radical departure from past activities, even if humans may at times be removed from the decision-making loop in a departure from previous levels of automation. While trust remains a challenge, particularly at the institutional and societal levels, steady application of initially still limited uses of AI to support military activities can breed familiarisation and increasing confidence over time.

Others, typically outside government, question the incrementalist approach. They see the rise of AI as a paradigm shift, qualitatively different from any previous technology. No previous technology has ever combined AI's dual-use character, ease of dissemination and substantively destructive potential.[25] In the past, the most destructive technologies were kept under government control or had little application outside the military domain. Moreover, while governments previously led much of the development of new technology, the trend has almost entirely reversed; much of the investment and innovation now comes from industry. Given the blurring of military and civilian lines, and the investment in AI being made by our adversaries and competitors, it is unwise to assume that we can control the pace and extent of AI development and use. Reflecting on the advances of algorithmic technology, some go even further by claiming a reversal of roles between technology and humans, whereby people are becoming 'human artefacts' and 'agents of (the system of) technology'.[26]

If we accept limitations to exercising complete control over how AI systems will be operated (and operate) in the future, the key question is how we ensure appropriate interfaces and human judgement long after algorithms have exceeded current levels of performance. Reaction time is a key advantage in military competitions; speeding up aspects of the OODA (Observe-Orient-Decide-Act) loop will typically confer a lead to those who get there first. It only takes one side to start using AI to speed up their decision-making and response times for the other to be pressured to do so as well.

25. Henry A Kissinger, Eric Schmidt and Daniel Huttenlocher, *The Age of A.I. and Our Human Future* (London: John Murray, 2021), pp. 166–67.
26. Dionysios Demetis and Allen S Lee, 'When Humans Using the IT Artifact Becomes IT Using the Human Artifact', *Journal of the Association for Information Systems* (Vol. 19, No. 10, 2018), pp. 929–52.

II. AI and Human Agency

IN DECEMBER 2020, the US Air Force flew a military aircraft with an AI co-pilot for the first time. The algorithm, known as ARTUμ, assumed full control over sensor employment and tactical navigation, while its human teammate piloted the U2 spy plane. This was the first known occurrence of an AI controlling a military system. In the words of Will Roper, the US Air Force's former chief acquisition official, ARTUμ 'was the mission commander, the final decision authority on the human-machine team'.[27]

Even before ARTUμ's impressive demonstration, the US Department of Defense had started work on its Joint All-Domain Command Control (JADC2) initiative. Designed to connect sensors from across five military services, JADC2 promises to deliver rapid analyses of the operating environment enabling decision-making within hours or minutes. Within a future JADC2, AI will allow the rapid processing of data to inform target identification and recommend an optimal engagement weapon (whether kinetic or non-kinetic). The US Air Force's Advanced Battle Management System, the US Army's Project Convergence (referred to as a 'campaign of learning') and the US Navy's Project Overmatch are all experimenting with the use of AI in combination with autonomy to support the JADC2 effort.[28]

Other countries, including the UK through projects such as the British Army's Project Theia, and NATO have also started trialling the use of AI to support C2 and decision-making. However, such experiments are still limited in scale and scope. Unlike areas such as data mining and language translation, the use of AI in military decision-making is still nascent.

The work currently undertaken by the US Defense Advanced Research Projects Agency offers a glimpse into the future. As part of its 'AI Next' project, the agency's third wave of AI investment seeks to 'transform computers from tools into problem-solving partners' and to 'enable AI systems to explain their actions, and to acquire and reason with common sense knowledge'.[29]

Civilian vs Military Uses of AI

AI already shapes or drives many of our daily decisions. In some cases, it has transformed entire industries. This is particularly the case in highly transactional activities, such as insurance or the retail sector. Humans have delegated responsibility for critical activities to AI, allowing

27. Will Roper, 'Exclusive: AI Just Controlled a Military Plane for the First Time Ever', *Popular Mechanics*, 16 December 2020.
28. John R Hoehn, 'Joint All-Domain Command and Control (JADC2)', Congressional Research Service, 1 July 2021.
29. DARPA, 'AI Next Campaign', <https://www.darpa.mil/work-with-us/ai-next-campaign>, accessed 15 January 2022.

algorithms to make decisions with no human intervention. Today, AI shapes the content delivered by network platforms such as Google and Facebook, and also determines what content gets removed or blocked. AI-enabled decision-support systems that retain a human element are also proliferating, in use for everything from medical diagnoses to improving manufacturing processes.

In few places has AI so fundamentally changed the human–machine relationship as in finance. AI is now responsible for the vast majority of high frequency trading. Thousands of micro-decisions performed in milliseconds have the power to transform entire fortunes, sometimes with ruinous consequences as the Flash Crash of 2010 demonstrated.[30] Human decisions are no longer necessary for the efficiency of financial markets and, indeed, may even be counterproductive.[31] The invisible algorithm would seem to have overtaken the invisible hand.

As for the rest of society, potential military uses of AI cover a broad spectrum of applications. These can be usefully categorised into enterprise, mission support and operational AI applications.[32] Military applications of AI, particularly in relation to mission support and operational uses, differ in some fundamental aspects from day-to-day civilian activities. In civilian life, AI has the opportunity to train and learn against real-life examples constantly, drawing on vast amounts of easily accessible data. For the military, contact with adversaries is sporadic and lessons or 'data' from real operations are relatively low in number and frequency. In addition to the episodic nature of military confrontation, national security decisions typically rely on a far more complex set of conditions, involving multiple parameters and stakeholders (not to mention the adversary's intent) that today's algorithms are ill equipped to reproduce. Finally, and most importantly, in matters of defence and national security, lives not just fortunes are at risk.[33] Mathematical logic is not sufficient to inform decisions; moral and ethical considerations are far more prominent in the use of force than in any other human activity. When the integrity of human life is in question, the standards we set for technology will always be higher than those we set for error-prone humans.

---

30. On 6 May 2010, the Dow Jones Industrial Average plunged nearly 1,000 points in 20 minutes, triggered by an automated trading algorithm that set off a spiral of events. See, for example, Michael Mackenzie and Aline van Duyn, '"Flash Crash" Was Sparked by Single Order', *Financial Times*, 1 October 2010.
31. See, for example, Dionysios Demetis, 'Algorithms Have Already Taken Over Human Decision-Making', *The Conversation*, 8 March 2019.
32. Danielle C Tarraf et al., *The Department of Defense Posture for Artificial Intelligence: Assessment and Recommendations* (Santa Monica, CA: RAND, 2019), pp. 25–27.
33. The use of autonomous cars also puts lives at risk, but the level of technical complexity involved is relatively lower than today's modern battlefield, characterised by the so-called Five Cs: congested; cluttered; contested; connected; and constrained. Ministry of Defence (MoD), 'Future Operating Environment 2035', Development, Concepts and Doctrine Centre, 14 December 2015, p. 44. However, continued low public acceptance of self-driving cars and uncertainty over accountability underscore how difficult it is to generate trust in a technology that has the potential to harm, even if it is likely to be safer than human drivers.

As well as being current policy for the US, the UK and NATO, among others, there is a general belief that humans will retain a critical role in decisions. The US Department of Defense's AI strategy directs the use of AI 'in a human-centered manner' that has the potential to 'shift human attention to higher-level reasoning and judgment'. Weapon systems design incorporating AI should 'allow commanders and operators to exercise appropriate levels of human judgment over the use of force' and ensure 'clear human-machine interface'.[34] References to humans always being 'in the loop' and 'fully in charge of options development, solution choice, and execution'[35] – a common refrain in previous assessments of our increasingly automated future – have been replaced by a more nuanced view.

So-called supervised autonomous systems have the human sitting 'on the loop'. While humans notionally maintain oversight, some critics argue that in practice they may have no real control over automated decision-making as they may lack familiarity with the circumstances and AI processes that feed them the information on which to decide.[36] In these cases, the human ability to intervene, short of stopping the machine, is minimised and falls short of the idea of 'meaningful human control'. Only in the case of fully autonomous systems is human intervention removed entirely. Ultimately, however, attempts to define levels of autonomy can be misleading as they assume a simple separation of cognitive activities between humans and machines. A 2012 US Defense Science Board report describes how:

> there exist no fully autonomous systems, just as there are no fully autonomous soldiers, sailors, airmen or Marines. Perhaps the most important message for commanders is that all systems are supervised by humans to some degree, and the best capabilities result from the coordination and collaboration of humans and machines.[37]

Developments in two areas reveal how far governments have already gone in the direction of trusting advanced automation for critical decisions in defence and national security. One is missile defence, the other cyber defence. The effectiveness of both depends on the speed of response, which typically exceeds the capacity of the most experienced human operators.

Most defensive weapon systems, from short-range point defence to anti-ballistic missile systems, operate with advanced automation that allows them to detect and destroy incoming missiles without human intervention. Algorithms are literally calling, as well as taking, the shots. Such

34. US Department of Defense, 'Summary of the 2018 Department of Defense Artificial Intelligence Strategy', February 2019, pp. 6, 15.
35. Jean-Michel Verney and Thomas Vinçotte, 'Human-On-the-Loop', NATO Joint Air & Space Power Conference, May 2021, p. 134.
36. Paul Scharre*, Army of None: Autonomous Weapons and the Future of War* (New York, NY: W W Norton and Company, 2018), pp. 27–30; Jackson Barnett, 'AI Needs Humans "on the Loop" not "in the Loop" for Nuke Detection, General Says', *FedScoop*, 14 February 2020.
37. US Department of Defense, 'The Role of Autonomy in DoD Systems', Defense Science Board Task Force Report, July 2012, pp. 23–24, <https://irp.fas.org/agency/dod/dsb/autonomy.pdf>, accessed 7 March 2022.

systems, in which the human is said to be 'on the loop', operate within a limited design space following rigorous prior human testing, so their span of control is constrained. Though mistakes can never be fully eliminated, the risk of not responding or of responding late in most cases may exceed the risk of occasional accidents. Although accidents prompt a review of the operation of these autonomous systems and may result in the introduction of some further human checks, such interventions also introduce further complexity.[38] Defence against ever-faster, particularly hypersonic, missiles will continue to drive adoption of AI in missile defence.

Cyber warfare presents another area where AI has clear advantages over humans, which often necessitate that the human remains out of the loop. Human operators lack the algorithms' ability to rapidly detect and respond to cyber incidents and to continuously adapt systems' defences. So-called cognitive electronic warfare (EW) systems apply AI techniques to automatically detect threats to EW systems rather than relying on human operators.[39]

Human and Artificial Limitations

There are huge benefits in automating parts of the decision-making process that are highly time-consuming, labour-intensive and require low-level human reasoning. The military estimate process, a key part of the military decision-making process, has been the standard operational planning process taught at staff colleges. Part of that approach involves the collection and processing of information to inform one or more courses of action. As decisions in the information age require ever-greater speed and agility, the process by which decisions are reached will need to accelerate. AI has already proven its utility in executing rational processes rapidly based on well-defined rules, inputs and assumptions. As long as humans are responsible for setting the assumptions and defining the inputs that generate the alternatives and probability assessments, AI can enhance the overall decision-making process.[40]

Understandably, there is reluctance both within and outside government to grant AI a role that extends beyond decision support and into proper decision-making.[41] The notion of 'command

38. US Patriot missiles are known to have accidently shot down friendly aircraft. See, for example, Jamie Wilson, 'US Missile System "Misidentified" RAF Tornado', *The Guardian,* 15 May 2004. Following this incident, however, humans were reinserted into part of the decision-making loop, but it did not prevent a further shooting down of a friendly aircraft. The inclusion of a person 'on the loop' is not fool-proof and there are known issues about how ready humans are to take over at short notice if they are merely supervising a largely autonomous system. See Sydney J Freedberg Jr, 'Artificial Stupidity: Fumbling the Handoff from AI to Human Control', *Breaking Defense*, 5 June 2017.
39. Jack Browne, 'Digital Techniques Train Cognitive EW System', *Microwaves & RF*, 12 August 2020; John Keller, 'Air Force Eyes Artificial Intelligence (AI) and Machine Learning for Cognitive Electronic Warfare (EW)', *Military and Aerospace Electronics*, 14 September 2021.
40. Brad Dewees, Chris Umphres and Maddy Tung, 'Machine Learning and Life-and-Death Decisions on the Battlefield', *War on the Rocks*, 11 January 2021.
41. See, for example, James Johnson, 'Delegating Strategic Decision-Making to Machines: Dr. Strangelove Redux?'*, Journal of Strategic Studies* (Vol. 45, No. 3, 2022), pp. 439–77.

and control' is too deeply embedded in the military's psyche and structures for many to accept a future that does not involve, to some extent, a human controlling a military operation or commanding a mission. The human is required to bring their creative insights to the problem and untie the Gordian knot like a modern-day Alexander. Nothing epitomises this attachment to the image of the intuitive commander than the belief in 'the kingfisher moment'.[42] This skill, the acme of the commander's art, is limited to those few who can decide intuitively under the most demanding circumstances. AI's ability to offer unique insights not based on human logic or experience poses a profound challenge to such thinking and is likely to transform the image of the commander in the future.

Many refer to AI as a decision-support rather than a decision-making tool, the inference being that humans ultimately remain the arbiters of all decisions. Such a distinction creates the reassuring illusion of AI merely assisting in the delivery of an effect. Does a human decision for lethal action reached on the basis of data mined, sifted and interpreted by a set of algorithms entail more human agency than a decision fully executed by an intelligent machine? The obsession with 'action' – let alone lethal action – as the final element of a broader 'kill chain' conceals the growing influence AI will have in a whole range of C2 activities across the full spectrum of operations.[43]

Many experts are sceptical about humans' ability to control decisions enabled or driven by AI. Such scepticism tends to revolve around the so-called black-box problem: advanced AI such as deep learning[44] is inherently impenetrable to human understanding. This is not just due to the speed at which it works but also the way networks of algorithms interact with each other, and the size and complexity of the data on which they operate. We cannot simply interrogate the system to understand its thought process. We may know the inputs and outputs of a model but cannot appreciate what happens in between. A related, more subtle argument is that algorithms exert 'power over humans' cognitive intake'.[45] AI may determine what information humans

42. T E Lawrence's famous quote – 'nine-tenths of tactics are certain, and taught in books: but the irrational tenth is like the kingfisher flashing across the pool, and that is the test of generals' – is still cited in military doctrine documents and taught at Staff Colleges. See British Army, 'Land Operations', 31 March 2017, <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/605298/Army_Field_Manual__AFM__A5_Master_ADP_Interactive_Gov_Web.pdf>, accessed 10 February 2022.
43. Christian Brose, *The Kill Chain: Defending America in the Future of High-Tech Warfare* (New York, NY: Hachette, 2020).
44. Deep learning is a subset of machine learning using deep neural networks that allows for multilevel knowledge representation; several steps, or layers, are used to compute paths from inputs to outputs. It is significant in reinforcement learning and is widely used now in applications such as speech and visual object recognition. See Boden, *AI*, p. 89; Russell and Norvig, *Artificial Intelligence*, pp. 44–45, 801.
45. Stuart Russell, 'Reith Lectures 2021: Living with Artificial Intelligence', <https://www.bbc.co.uk/programmes/articles/1N0w5NcK27Tt041LPVLZ51k/reith-lectures-2021-living-with-artificial-intelligence>, accessed 15 January 2022.

process without revealing to them what has been omitted or rejected. It also challenges the notion that humans can exercise 'meaningful' control if their actions are being conditioned by what and how data is presented. This is the opposite argument to one of AI's benefits, namely its ability to reduce humans' cognitive load allowing for concentration of thought and action on the highest-value activities.

The typical solution to the black-box challenge is the development of explainable AI (XAI). Though AI that can explain itself may contribute to understanding, it does not lead inevitably to trust. XAI does not equate to interpretable AI; an explanation is not a decision, but a story about a decision. Thus, even a compelling explanation need not be true.[46] We are still far from being able to develop sufficiently explainable, let alone interpretable, AI for many potential use cases. Rigorous testing of more advanced AI systems may prove sufficient for their deployment, even in the absence of human ability to follow their reasoning process. Fundamentally, though, our traditional approach to testing will need to be rethought. Without adequate test and evaluation, the trust in non-explainable/interpretable AI would be 'blind trust'. We still lack a satisfactory answer to the former US Deputy Secretary of Defence Bob Work's question: 'How do you do test and evaluation of learning systems?'[47]

When there is uncertainty or lack of knowledge, humans apply a heuristic approach to approximate solutions to complex problems. Heuristics is what drives intuitive thinking; it relies on rules of thumb, typically informed by experience and experimentation. As such, it can suffer from biases and blind spots, but it can also serve as a very powerful and effective form of rapid cognition.[48] Machines lack human-like intuition, but they do rely on heuristics to solve problems. The key difference with human reasoning is that machines do not need memory or 'personal' experience to be able to 'intuit' or infer. They draw on huge databases and a superior probabilistic capacity to inform decision-making. Powerful simulations, combined with advanced computing power, offer an opportunity to test and 'train' algorithms at levels of repetition unimaginable for humans. ARTUμ had undergone more than a million training simulations in just over a month before it was declared mission ready.

Even with significant advances in the field of XAI, there will still be reasons for caution, particularly in situations requiring complex decision-making. AI is generally not good at seeing the 'big picture' or making decisions based on what is relevant.[49] Like humans, it can mistake correlation

46. Put differently, 'we say that a system is interpretable if we can inspect the source code of the model and see what it is doing, and we say it is explainable if we can make up a story about what it is doing—even if the system itself is an uninterpretable black box'. Russell and Norvig, *Artificial Intelligence*, p. 1048; see also p. 729.
47. Scharre, *Army of None*, p. 180.
48. Malcolm Gladwell, *Blink: The Power of Thinking Without Thinking* (New York, NY: Little Brown and Company, 2005); Herbert A Simon, 'Making Management Decisions: The Role of Intuition and Emotions', *Academy of Management Perspectives* (Vol. 1, No. 1, February 1987), pp. 57–64.
49. This is also referred to as the 'frame problem' whereby AI systems fail to account for anything outside their frame of reference (i.e., anything that has been left implicit in terms of assumptions

or chance events for causation.[50] Both humans and machines are bound to experience 'normal accidents' when dealing with complexity.[51] Both can fall victim to misplaced confidence[52] or mistake noise for a signal,[53] though for different reasons. Creativity is a trait commonly assigned to humans but some advanced AI can generate surprising outcomes that have eluded human ingenuity.[54] In short, many attributes often assumed to be unique to humans, such as creativity and intuition, can also be said to apply to AI systems – albeit in different ways and at speeds that exceed human ability.

What machines currently lack are the human mind's flexibility and sense of relevance (the ability to 'frame'). Humans can think laterally, reach plausible outcomes through pragmatism (a process known as abductive reasoning)[55] and reflect on their own thought processes (an ability known as metacognition).[56] These mental processes can generate amazing feats of adaptation and innovation.

---

or implications). See Boden, *AI*, pp. 43–44. Also, while AI's identification of new antibiotics for drug-resistant diseases was lauded by the press, its success depended on a multi-disciplinary human team reframing the problem from one focused on compounds with similar structural properties, to looking at existing drugs that had similar effects. See Kenneth Cukier, Viktor Mayer-Schönberger and Francis de Véricourt, *Framers: Human Advantage in an Age of Technology and Turmoil* (London: W H Allen, 2021), pp. 2, 3.

50. Marcus du Sautoy, *The Creativity Code: How AI Is Learning to Write, Paint and Think* (London: Fourth Estate, 2019), pp. 92–93; Daniel Kahneman, *Thinking, Fast and Slow* (New York, NY: Farrar, Strauss and Giroux, 2011), pp. 109–18.
51. Scharre, *Army of None*, pp. 150–54.
52. Trust in subjective (i.e., self-proclaimed) confidence is nearly always misplaced, including in the case of machines. See Kahneman, *Thinking, Fast and Slow*, pp. 240–41; Patrick Tucker, 'This Air Force Targeting AI Thought It Had a 90% Success Rate. It Was More Like 25%', *Defense One*, 9 December 2021.
53. Nate Silver, *The Signal and the Noise: Why So Many Predictions Fail – But Some Don't* (New York, NY: Penguin Books, 2015).
54. The creation of AlphaFold, able to predict the 3D structures of almost all human proteins, constitutes one of the most significant discoveries in modern biology and would have been impossible without AI. See DeepMind, 'AlphaFold: A Solution to a 50-Year-Old Grand Challenge in Biology', 20 November 2021, <https://deepmind.com/blog/article/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology>, accessed 15 December 2021.
55. Also referred to as 'inference to the best explanation', abductive reasoning generates valid hypotheses from a theoretically infinite number of explanations by applying a value system to the observation of unrelated or unexpected phenomena. This is particularly critical when trying to solve novel problems and tackling larger problem spaces.
56. Elite chess masters epitomise this capacity for shifting their thought concentration: 'Sometimes this involves probing many branches of the tree but just a couple of moves down the line; at other times, they focus on just one branch but carry out the calculation to a much greater depth. This type of trade-off between breadth and depth is common anytime that we face a complicated problem'. Silver, *The Signal and the Noise*, pp. 272–73.

The advent of AI means that future military decision-making will almost certainly require a much stronger symbiosis of human and machine, much as is seen in commercial organisations that have embraced the technology. Current discourse mostly either assumes some continued human control or seeks to apply human-like attributes to future machines. Some advocate a new 'decision manoeuvre' concept that combines 'human command with machine control'.[57] More likely, though, both command and control responsibilities will increasingly be shared between humans and AI systems in ways that may be hard to envision at present. Teaming humans with AI provides the best way of harnessing the strengths of each and mitigating the shortfalls, especially in relation to war, whose nature remains unchanged (for now) with four continuities: a political dimension; a human dimension; the existence of uncertainty; and that it is a contest of wills.[58]

57. Bryan Clark, Dan Patt and Harrison Schramm, 'Decision Maneuver: The Next Revolution in Military Affairs', *Over the Horizon*, 29 April 2019, <https://othjournal.com/2019/04/29/decision-maneuver-the-next-revolution-in-military-affairs>, accessed 10 February 2022.
58. H R McMaster, in 'America's Second-Longest War: Taking Stock: Geopolitical Lessons', Carnegie Endowment for International Peace, 21 March 2013, <https://carnegieendowment.org/files/0322ceip3-geopolitical_lessons.pdf>, accessed 20 January 2022.

III. Dimensions of Trust

TRUST IS DYNAMIC; it changes over time. Its initial formation is critical, but so is its continued development. Trust comes naturally with familiarity, so increased use of technology expands the trust boundary, assuming the experience is positive, even where the technology is not fully understood. The reverse is also true, and bad experience fosters distrust. The technological complexity of mobile phones is unknown to most users, but people's positive experience gives confidence in their use. That confidence leads to a sense of trust appropriate to the decisions shaped by the phone's use. However, phones generally do not decide on matters of life and death, although they can place in jeopardy unwary drivers who blindly follow directions. The stakes are higher in the military context, and users and policymakers are very aware of the potential consequences of their decisions – the trust bar is high.

Militaries, as contingent organisations, are not required to deliver their primary outputs routinely, which affects the pace at which the most directly relevant experience can be gained. Unlike financial services, where trading provides frequent validation of AI decision-making, the timelines in Defence are often longer and the outcomes less clearly linked in a single cause-and-effect chain. The timelag between taking a decision and observing its impact is longer and subject to multiple intervening variables.[59] While simulated exercises create opportunities to gain experience, they are only approximations of reality.

Trust Points

Building and sustaining trust involves five main 'trust points' – points at which the question of having an appropriate level of trust is crucial. These are:

- **Deployment trust**: The purpose for which AI is used.
- **Data trust**: The data inputs being used.
- **Process trust**: How the data is processed.
- **Output trust**: The outputs generated by the AI.
- **Organisational system trust**: The overall ecosystem for optimising use of AI.

Collectively, the trust points define an overall level of trust, and are multiplicative: if trust in one is 'zero', the whole will be 'zero'. The trust level for each can vary – at different times and over time – as long as overall trust is positive.

---

59. This is particularly so for commanders higher up the command chain, as Elliott Jaques predicted in his Time Span on Discretion theory. See Elliott Jaques, *Time-Span Handbook: The Use of Time-Span of Discretion to Measure the Level of Work in Employment Roles and to Arrange an Equitable Payment Structure* (London: Heinemann, 1964).

Deployment Trust

Trust in the decision to use AI in specific circumstances is essential. For military use of AI (and many civilian applications), this operates at three levels: society; organisation; and individual. The first considers whether society at large is willing to allow AI's use, which will shape how policymakers view its use. The organisation itself must also be willing to sanction this. Finally, individuals must be willing to work with AI in that role. All three levels require an acceptance of either the inevitability of the need to use AI or its desirability. Desirability might reflect AI's advantage in processing data at a speed or volume (or both) that exceeds that of a human operator, or in undertaking dull or dangerous work. And while militaries may argue that AI is both pragmatic and inevitable to avoid ceding advantage to an opponent, society appears more inclined to see lethal uses as an ethical issue in which the sanctity of human life requires a moral actor to decide to take a human life.

Society's acceptance of AI's use depends in large part on its experience, effective communication and education that will help inform AI use choices. Parts of society are likely to be more exposed to, familiar with and trusting of AI than the military in many circumstances, but the spectre of lethal autonomy is likely to remain problematic. While less directly threatening than lethal autonomy, the use of AI in decision-making raises its own challenges, not least of which is what 'meaningful human control' actually means in a world of ever more powerful algorithms and ever closer human–machine collaboration.

At an organisational level, there are important questions about how operational and mission support AI is deployed: whether in a centralised way operating at a more strategic level or in a more distributed fashion at tactical levels. In the latter case, AI will penetrate further into the organisation, becoming more diffuse and used where reaction time may limit the scope for human intervention or validation of the AI. Organisations need to be clear about the principles for determining whether to use AI, and also their approach to governing its use (see 'process trust' below). Decisions about the use of AI must consider what happens if the system fails. NASA under-used its early Mars rovers' autonomous capabilities over concerns about the consequences of system failure, micromanaging the rovers and mitigating risk through large teams of human engineers.[60] What external organisations, such as commercial technology providers, think is also important. For example, Google's employees forced the company to withdraw from a military contract in 2018 over concerns about the military's use of facial recognition technology.[61]

Individual familiarity with AI will also be important. Currently, those engaged in working on military AI are advocates for its use, but as the cohort exposed to it grows, this will change. Younger members of the military with greater exposure to the technology may be more

60. Jeffrey M Bradshaw et al., 'The Seven Deadly Myths of "Autonomous Systems"', *IEEE Intelligent Systems* (Vol. 28, No. 3, 2013), p. 3.
61. See, for example, Olivia Solon, 'When Should a Tech Company Refuse to Build Tools for the Government?', *The Guardian*, 26 June 2018.

accepting of AI's use in military decision-making than earlier generations, but in base-rank fed structures, where talent grows almost exclusively from within, resistance to its use may sit with those in positions of power; this can create problems of institutional acceptance. There is, however, a danger of over-simplifying in relation to 'generational characteristics'.[62] While younger people will have grown up with newer technology and are probably more trusting of it, the technology can be learned. Generational assumptions cannot become an excuse for not engaging with modern technology.

Deployment trust is complicated because most Western defence activity at scale assumes coalition operations and not every ally or partner may have a common view of what is an acceptable military use of AI. Defence ministries and governments need to become better at communicating their approaches, uses and safeguards in relation to the use of AI, including to allies, without giving too much away to adversaries who can develop strategies to neutralise (or worse) the advantages of AI-enabled capabilities. NATO will be crucial in this through its public outreach activity, links with member states at the political level and norm-building across militaries at different stages of technological development.[63]

Data Trust

This concerns the degree of trust in the data on which AI makes judgements that inform human decision-making. While it is relatively easy to test the hardware and software, it is more difficult to test the data, or even prepare the data that allows AI to be trained.[64] Data is essential for AI to learn effectively. Some data will be controlled, residing within existing Defence systems, or validated from reliable external sources, although Defence struggles with how data is classified (inconsistently or inaccurately), stored, accessed and shared, especially at higher levels of classification. Uncontrolled data, such as open source data, that is generated through aggregation without human knowledge or understanding is more challenging. Moreover, sophisticated adversaries will attempt to inject false data to undermine decision-making processes or swamp them with irrelevant or inaccurate data.

Armed forces need the ability to define, structure, cleanse and analyse data, as well as develop and maintain the underlying infrastructure (such as connectivity, security and storage capacity). This is a multi-disciplinary team effort requiring 'full-stack' data scientists who can work on all

62. Emma Parry and Peter Urwin, 'Generational Categories: A Broken Basis for Human Resource Management Research and Practice', *Human Resource Management Journal* (Vol. 31, No. 4, 2021).
63. Maggie Gray and Amy Ertan, 'Artificial Intelligence and Autonomy in the Military: An Overview of NATO Member States' Strategies and Deployment', NATO Cooperative Cyber Defence Centre of Excellence, 2021.
64. There are numerous instances of problems with AI, including in the safer space of enterprise AI, stemming from the data used for its training. Enterprise AI has been accused of bias on several occasions based on the biases in the training data. See Tim Kulp, 'AI and Hiring Bias: Why You Need to Teach Your Robots Well', *Human Resources Executive*, 14 April 2021; Khyati Sundaram, 'How to Embrace AI Recruitment and Avoid Bias', *People Management*, 30 September 2021.

stages of the data science lifecycle.[65] The modern battlefield will require even greater diversity of skills, including psychologists, lawyers and communications experts. Attracting and retaining these specialists in the numbers required will be difficult given the demand for such skills in the commercial world. It will require more flexible human resource practices and/or a more sophisticated understanding and use of the Whole Force,[66] including allowing non-military personnel to hold positions of influence in military headquarters.[67]

Process Trust

Process trust refers to how the AI system operates, including how data is processed (aggregated, analysed and interpreted). Current (narrow) AI decision-support systems within UK Defence attract high confidence because the algorithms are relatively simple and predictable. They are also restricted to a small group of users involved in developing, or who know those who have developed the AI systems, and understand the technology. The technology benefits from a kind of transitive trust derived from the trust people have in the human.[68] While not AI-enabled, the French Army's introduction of machines for packing parachutes led to a loss of confidence among parachute regiments. Insisting that the machine's supervisor jumped with a randomly chosen parachute packed by the machine helped restore user confidence.[69] Bringing developers closer to the user of command systems helps. The French procurement process allows certain units to engage directly with AI providers to build knowledge of and a relationship with the developers. The developer becomes a key trust point and, where not military, they must understand and be conversant with the military. This may require greater investment in inducting commercial partners into how the military works, and ensuring military personnel understand their civilian colleagues.

Demanding high levels of explainability and transparency is not a permanent solution and currently limits Defence access to more powerful, non-symbolic forms of AI. As machine learning enables technology to exceed the parameters of its initial programming, different ways will be needed to secure trust in what may appear to be a black box. As use of such AI systems proliferates, the transitive trust resulting from knowing the designers will diminish, and the more difficult it will be to overcome initial under-trust or over-trust in the process. Overreliance

65. Erich Feige, 'The Army Needs Full-Stack Data Scientists and Analytics Translators', *War on the Rocks*, 14 February 2020.
66. The 'Whole Force' comprises 'regular and reserve service personnel, MOD civil servants, contractors and other civilians'. MoD, 'Future Force Concept', Joint Concept Note 1/17, Development, Concepts and Doctrine Centre, July 2017, p. 57.
67. For example, the French Armed Forces are looking at integrating the civilian intelligence agencies more fully into their operational headquarters as a way of assuring the data and addressing the question of data trust. Author interview with Colonel (Ret.) François Villiaumey, French Army, 14 October 2021.
68. See, for example, Heather M Roff and David Danks, '"Trust but Verify": The Difficulty of Trusting Autonomous Weapons Systems', *Journal of Military Ethics* (Vol. 17, No. 1, 2018), pp. 12–13.
69. Author interview with Colonel (Ret.) François Villiaumey, French Army, 14 October 2021.

on process trust should be avoided, and other trust points strengthened, to accommodate increasingly capable AI.

Process trust must extend beyond the technology itself. It requires trust in the human processes that feed, work alongside and receive technology's outputs. Equal importance, therefore, must be placed on those other activities that together constitute the overall process. This includes the processes by which people are trained and developed, and how the teams are formed.

Output Trust

Trust in AI outputs is critical for decision-makers to act on the information they receive. Even with human-provided intelligence, it is not unknown for commanders to demand new intelligence to support their preconceptions if the original information points in a different direction (a kind of 'decision-based evidence making'). And with the proliferation of data, different interpretations will be possible, legitimately or to fit preconceptions. Questions arise, therefore, about what answers AI, or indeed human analysis, can realistically provide, and how to validate the outputs. Faster situational awareness from AI is possible in relation to the disposition of friendly forces and the physical location of the adversary. However, an adversary's actual intentions cannot be ascertained reliably, although better inferences may be drawn from available data. Predictability is often seen as a critical element in trust, but in unstable environments, AI outputs that can adapt to a fluid environment can be interpreted as unpredictable. To overcome this, Bonnie M Muir argues that human operators must be equipped with the ability to estimate the technology's predictability.[70] This predictability component impacts across deployment and process trust points too, but is most acute in trust in the outputs to reflect fluid and unpredictable environments such as military operations. In these contexts, data also has to reflect the discrete nature of most situations faced by military decision-makers and distinct cultural approaches of specific adversaries, which exacerbates the difficulty in building a large body of training data. Even where situations resemble past events, the lack of comparable historical data to account for the broad range of variables makes probabilistic reasoning difficult.

Calibration of the output, in Patricia L McDermott and Ronna N ten Brink's terms, is essential. This might be achieved by greater use of enterprise AI and simulation that expands the trust boundary and can help develop output trust.[71] Interacting with the technology and seeing its outputs in action will generate trust if the experience is positive. In the operational environment, verification will be easiest when describing things that can be known and checked (for example, data on one's own forces and, potentially, the laydown of adversary forces). It is more difficult to approximate the adversary's intent, hence higher levels of output trust will be needed. This would include greater accuracy in descriptions and more testing of inferences drawn from

70. Bonnie M Muir, 'Trust in Automation: Part I. Theoretical Issues in the Study of Trust and Human Intervention in Automated Systems', *Ergonomics* (Vol. 37, No. 11, 1994), p. 1913.
71. Patricia L McDermott and Ronna N ten Brink, 'Practical Guidance for Evaluating Calibrated Trust', *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 63, No. 1, 2019), p. 364.

big-data processing. Sharing positive stories of exercises and operations will be essential to enabling transitive trust and mitigating the slow pace at which evidence of success is amassed from relatively infrequent operations.

Organisational Ecosystem Trust

Ecosystem trust concerns the trust needed to adapt the wider organisational system to maximise the value of AI. The C2 system as a whole must be configured to exploit the benefits of AI-enabled decision-making, with appropriate checks and balances to operate within acceptable levels of risk. This is especially important where AI weaknesses or failures are in areas outside the supervisor's expertise and require calibration in different parts of the organisation.[72] Without changes at the ecosystem and organisational level, organisations will merely digitise their human systems.

Ecosystem trust is needed to ensure that the structures – including the organisation of military headquarters, the role of the commander and the balance of centralised versus more diffuse or distributive powers of decision-making – are ready to harness AI's opportunities. Without it, incremental approaches to AI adoption tend to encourage a passive or reactive approach to changes in structures and the overall ecosystem. By contrast, a dedicated strategy to realise the transformative power of AI would force an early rethink of the organisation needed to underpin such a strategy. This requires rethinking the traditional military structures, but there is no consensus on how far to go. Some envisage headquarters becoming flatter and thinner, and integrating non-military personnel in senior positions with authority in the decision-making process. For others, the ecosystem change is more profound; it requires the current staff organisation system, seen as a remnant of the industrial age, to be removed entirely. In doing so, they intend to remove the information boundaries that stifle understanding and challenge the idea of a solo commander at the pinnacle of a decision-making pyramid. Such a shift requires trust throughout the organisational ecosystem. For conservative organisations such as the military, this will be difficult, and reassurance will be needed before radical alternatives to command headquarters are acceptable. Experimentation, wargaming and simulated environments offer low-risk options to test different headquarter structures configured for specific types of missions (for example, warfighting, peace operations and capacity building).

How Much Trust Is Enough?

Trust is fundamental, but there are risks in setting a bar for technology that is impossibly high. Commanders and decision-makers have trusted fallible humans for millennia. And there are ways in which technology can assist through self-monitoring that alerts a human 'operator' when the AI begins to observe shifts in input data distributions, or operate in previously unseen ways where the risk of erroneous outputs is greater. Risk tolerance, whether concerning human or machine actors, is ultimately an expression of trust. Defence organisations need to be honest with themselves about whether they are fast adopters or fast or slow followers: the commercial

72. Muir, 'Trust in Automation: Part I', p. 1907.

world's pace of AI development makes it highly unlikely most armed forces will be 'first users' of AI for decision-making. The difference between the incrementalists and futurists, both supporters of AI, is principally one of risk and a function of the achievable trust level against the different trust points.

Generating trust through familiarity is essential, and may involve embedding military personnel in commercial organisations using sophisticated AI, or bringing civilians into Defence. Such changes are needed at a senior enough level to foster ecosystem trust. Simulation, experimentation and exercises are important tools, and must be sufficiently widespread so as to not be confined to a small group of zealots. BT's project to replace the UK's telephony network with an AI-enabled decision-support tool was most effective when the longest-serving and most knowledgeable engineers, who might have been expected to be suspicious of AI, worked with the AI and data experts.[73] Introducing enterprise AI to reform business processes, such as finance and human resources, is another way of spreading familiarity beyond the small cadres currently directly involved in AI's development and use.

Once something is familiar, it is human nature to trust, but habits of trust entail their own risks. Humans are known to be poor at setting the right objectives and, when presented with 'expert opinion' (whether derived artificially or otherwise), are more prone to consent than to doubt. 'Trust and forget' dynamics must be avoided.[74] One consideration is to develop a notion of 'continuing trustworthiness' for AI-enabled systems, similar in concept to continuing airworthiness for air platforms to ensure they remain fit for use.[75] Hence, efforts to build trust in AI (and avoid over-trusting) must address all of the trust points, and include the whole human–machine team in which the human operators are effective collaborators with and constructive critics to their digital counterparts.

73. Author interview with Steven Cassidy, Chief Researcher Future Organisation, BT, March 2022.
74. Taddeo, 'Trusting Digital Technologies Correctly', pp. 565–68.
75. Patrick Baker et al., 'We Need to Talk About Trust', King's College London, 28 September 2021, <https://www.kcl.ac.uk/we-need-to-talk-about-trust>, accessed 15 December 2021.

IV. Implications for Command and Commanders

AI'S IMPACT ON how decisions are informed, made and implemented will be profound. By processing vastly more data at speeds that defy current human-based processes, AI can improve understanding of the operating environment and reduce the cognitive load on decision-makers. This is more than an evolution of today's ways of working. Merely speeding up current C2 systems would be impractical. A car designed to be driven at 70mph is configured for operating at that speed. Tweaking the engine to work at 100mph may be possible, but places strain on the vehicle systems and driver that cannot be sustained. The discontinuity represented by AI-enabled decision-making demands a new approach. As the Multinational Capability Development Campaign (MCDC) described:

> Whatever our C2 models, systems and behaviours of the future will look like, they must not be linear, deterministic and static. They must be agile, autonomously self-adaptive and self-regulating, and at least as contingent and emergent as the context within which they are formed and operate.[76]

Militaries must reorganise for tomorrow's C2, and develop their commanders and staff differently. Without such changes, 'ecosystem trust' may prove impossible to attain.

Command and Control

C2 incorporates two distinct elements: command, usually associated with creativity, flexibility and leadership; and control, associated with rules, predictability and standardisation. AI will affect the control function first, with command remaining, for now, a mainly human activity. AI data-processing ability will remove substantial burdens to control by, for example, providing commanders with better understanding of their own forces (such as disposition, status, levels of equipment and supply) that currently consumes a lot of attention and time. It will also change the way in which the information is available to commanders. Currently, much of this data is available on a 'pull' basis – requested or gathered sporadically in response to headquarters' reporting processes. AI, however, can continuously monitor situations and push information to commanders through living documents that highlight relevant changes – akin to a 24-hour newsroom feed. However, by moving further into control, AI will inevitably affect the exercise of command and shape command decisions; this challenges the rather too neat distinction between command and control described above. In future C2 systems, it is conceivable that AI could constrain the exercise of command, rather as anti-lock braking systems, traction control

76. Multinational Capability Development Campaign, 'Final Study Report on Information Age Command and Control Concepts', 8 February 2019, p. 20.

and electronic stability allow human drivers to command a vehicle until control is lost, at which point the systems take over until the situation is stabilised.

AI presents a paradox for human command. It simultaneously enables greater knowledge to be held centrally, allowing headquarters to see and interact with what is going on at the 'front', and diffuses knowledge throughout the command system, providing lower-level formations with access to information that was previously the preserve of senior commanders. Delegating more authority to local commanders allows greater responsiveness, which matters when events unfold unpredictably requiring rapid reaction. Western armed forces tend to adopt (to a greater or lesser extent) the concept of mission command, where a commander's intent describes the desired effects and allows subordinate commanders freedom to deliver appropriate to the circumstances they face. Military learning and development systems and exercises embed this approach – commanders will need to be confident that AI can implement that intent in its operating. AI use at tactical and operational levels is likely to be more effective than at strategic levels of command given strategic complexity and ambiguity in the data and reward functions, although the levels are not discrete and cannot be compartmentalised easily in reality.[77] AI and greater network connectivity would provide a network of structures, processes and technologies connecting multiple small, dispersed forward headquarters and distributed (and hardened) rear functions that are harder to detect and counter, even in more transparent battlespaces. This would enhance resilience should enemies target the C2 system.

The capacity to process greater data volumes at every level must be carefully channelled. People should have access to the information relevant to their position and relative ability to influence developments within their environment. W Ross Ashby described this as the problem of 'requisite variety': a viable (eco)system is one that can handle the variability of its environment.[78] Actors should operate at a level of abstraction suited to their task. A brigade headquarters cannot cope with nor needs detailed information about individual soldiers; it requires a good general understanding of the physical and moral state of its subordinate units. At more tactical levels, NCO commanders should be alert to the state of individuals in their team. Strategic and operational commanders may need to relax control, allowing tactical commanders to exploit emerging opportunities closer to the fight. While mission command already allows this, the temptation to 'take control' will be greater as senior-level commanders gain unprecedented access to information about what is happening at tactical levels.

AI, too, will need to use abstractions, approximations and calibrated levers to avoid drowning headquarters in a data deluge. This requires 'process trust' in the use of those abstractions and approximations. It is also likely that headquarters will require access to different AI

77. Kenneth Payne, 'Artificial Intelligence: A Revolution in Strategic Affairs?', *Survival* (Vol. 60, No. 5, 2018), pp. 8–10. However, Payne also notes the advantages of AI in processing vast amounts of information and eliminating many of the human biases, individual or collective, such as excessive optimism, groupthink, confirmation bias and poor risk judgement.
78. W Ross Ashby, 'Requisite Variety and Its Implications for the Control of Complex Systems', *Cybernetica* (Vol. 1, No. 2, 1958), pp. 83–99.

systems whose capabilities are better or less suited to different scenarios on different time horizons. Decision-making may also include an element of determining which AI model to trust (deployment and process trust) in specific situations.

Automation in joint human–machine systems will enhance human performance and, in some cases, change the nature of the task itself.[79] At whatever level it is deployed, AI will affect not just how but what tasks humans perform. Current approaches typically start by looking at which human processes can be automated – namely, the digitisation of human work. It is possible to start with a presumption for using AI, putting humans into the system only where humans are essential (for legal, policy or ethical reasons) or desirable (better suited to the task) – determining what should not, rather than what can, be digitised. Such an approach challenges current conceptions of how headquarters should be sized, organised, staffed and operated.

Impact on the Structure of Future Headquarters

Joint Concept Note (JCN) 2/17 notes that C2 is likely to vary at the different levels of warfare (strategic, operational and tactical) and in response to the changing character of the operating environment, which is no longer just physical.[80] The blurring of war and peace – underscored by the need to be effective across the continuum between 'operate' and 'warfight'[81] – and the UK's shift to force structures that enable persistent engagement will require approaches that go beyond what is needed for combat. However, there is probably no single headquarters archetype; a headquarters configured for fighting will, therefore, be different from one dealing with upstream engagement and capacity building. While it is too soon to be definitive about AI's impact on military headquarters, commercial organisations have found that flatter structures with more horizontal information sharing are better suited to exploiting AI's advantages than more traditional vertical hierarchies where each layer assures and authorises data before it is released. It is likely, therefore, that military headquarters – no matter what their specific form – will be smaller and flatter than today, able to work faster and along horizontal lines.

Exploring alternative headquarters concepts can be achieved through greater use of experimentation and simulation. This should challenge the classic J1-9 staff branches,[82] perhaps with new groupings that reflect AI's capacity for replacing human-intensive data processing and sharing tasks. This is particularly the case in the J3/5 area, the boundary between plans and operations; a faster-paced conflict brought about by faster decision-making renders such a boundary obsolete. Alternative approaches to organising the headquarters may include

79. Bradshaw et al., 'Seven Deadly Myths of "Autonomous Systems"', p. 6.
80. MoD, 'Future of Command and Control', Joint Concept Note 2/2017, Development, Concepts and Doctrine Centre, September 2017.
81. MoD, 'Integrated Operating Concept', Development, Concepts and Doctrine Centre, August 2021.
82. NATO describes the staff branches as: J1, personnel and administration; J2, intelligence; J3, operations; J4, logistics; J5, plans; J6, communications and information systems; J7, training; J8, budget and finance; J9, civil–military cooperation. See NATO, 'Allied Joint Doctrine for the Conduct of Operations', February 2019, pp. A2–A5.

those that are outcome focused. The UK Standing Joint Force Headquarters (SJFHQ) construct described in JCN 2/17 was organised around four functions: understand; design; operate; and enable.[83] SJFHQ subsequently reverted to the traditional J1-9 staff branches. However, Exercise *Joint Protector 2021*, a complex sub-threshold operation in which AI decision-support tools were used, revealed weaknesses in the J1-9 construct. The headquarters started the exercise configured for high-intensity warfare, but subsequently adapted to one better suited to working with other agencies. Work is being done within SJFHQ to apply the lessons from *Joint Protector 2021* and determine what that means for the headquarters structure. It is unlikely, however, that there is a single perfect headquarters model that works for all operation types. Further experimentation is needed, not limited to the SJFHQ. It is telling that more than four years since the release of JCN 2/17, little progress towards implementing some of its propositions has been made. Even the relatively slow pace of technology adoption in UK Defence is outstripping the MoD's capacity to explore changes to structures beyond small groups of enthusiasts. 'Ecosystem trust' is essential and requires opportunities for testing alternative approaches, across the range of mission types, in simulated or real environments and involving a wider cohort who will be essential to effective adoption of new technologies, structures and processes.

Existing processes need to change to connect and optimise the new structures. This will probably require changes to the military estimate, which forms the basis of armed forces' planning processes. While a sophisticated and logical planning tool, it is rather linear, deterministic and heavily dependent on the commander, especially in the 'commander-led' UK approach. In other countries, the staffs play a larger role in driving towards solutions, which may be better suited to AI-enabled approaches. AI offers the opportunity for a more iterative and collaborative process that responds better to the MCDC demand for a shift to more agile models. The new approaches should place less pressure on commanders to ask for information (commander's critical information requirements), and requests for information. AI can also construct, analyse and compare courses of action, allowing scenarios to be modelled, tested and refined before choices are made to commit forces on a large scale.

A thought-experiment in automating the intelligence assessment process of the UK Permanent Joint Headquarters (PJHQ) identified opportunities to replace large numbers of staff, accelerate the headquarters' battle rhythm and allow horizontal sharing of information using automatic summarisation and natural language processing. Testing this in an operational deployment, the UK's 20th Armoured Infantry Brigade Combat Team shortened parts of the planning process tenfold.[84] However, there may be limits to how much faster decision-loops can be made while humans remain in the loop. At some point, human decision-makers will be unable to keep up, becoming decision-monitors. This will be problematic if humans are still needed to make decisions that AI cannot do for itself, which will probably be the most difficult ones.[85]

83. MoD, 'Future of Command and Control', p. 17.
84. Author interview with Brigadier Stefan Crossfield, Head Information Exploitation and Chief Data Officer, British Army, 12 November 2021.
85. Author interview with Björn Johansson, Co-Chair NATO SAS-143, 1 March 2022.

Despite clear advantages, it is unlikely that headquarters will be reduced as far as the technology would permit. Current headquarters compensate for human fragilities through redundancy in size and assurance processes, and this is likely to remain true for mitigating the fragility of AI team members. Moreover, the increased tempo may drive up demand in some areas as the battle rhythm morphs into a continuous 24-hour planning cycle. These pressures may not be confined to the headquarters themselves; it may drive increased activity in front-line units who have to process data and respond to the direction that they are given. Human actors will still need time to rest, even if the technology does not. Additionally, unlike commercial organisations, militaries need redundancy to cope with competitors deliberately seeking to destroy or disrupt their decision-making apparatus and lack the security of fixed infrastructure on which to build their networks. In short, the need for resilience and mobility affects the robustness and efficiency of military C2 systems. Militaries will, therefore, need to retain structures that are not wholly reliant on AI for effective operations, and ensure reversionary processes are available in the event of failures of the AI, or the deliberate erosion of trust in the AI.

Growing the Commanders

Traditionally, the commander is the apex of vertical decision-making structures, the point at which all information comes together. While not all military cultures emphasise the genius of the individual, as epitomised in the notion of the 'kingfisher moment', the commander's privileged access to information is denied to those at lower levels of the headquarters. AI's potential to democratise information will change this; command is likely to become a more collegiate and iterative activity, involving not just those in uniform but a more eclectic mix including intelligence agencies and contractors with expertise in multi-faceted aspects of data science – a 'whole force' contribution. Facing a complex and adaptive battlespace, another bird perhaps provides a better analogy for future command: the starling. Their collective, highly adaptive murmuration offers a better image for the UK's Development, Concepts and Doctrine Centre C2 concept as 'a dynamic and adaptive socio-technical system configured to design and execute joint action'.[86]

Commanders must continue to be able to handle dynamic environments; the saying 'no plan survives contact with the enemy' remains true.[87] Dealing with complex, fast-evolving problems will be especially important given technology's ability to increase both pace (reducing response times) and complexity (through a more transparent battlespace). Military organisations are experimenting with how AI will change C2, including the NATO Command and Control Centre

86. MoD, 'Future of Command and Control', p. 11.

87. The original quote attributed to the 19th-century Prussian Field Marshal Helmuth von Moltke the Elder states: 'No plan of operations extends with any certainty beyond the first encounter with the main enemy forces. Only the layman believes that in the course of a campaign he sees the consistent implementation of an original thought that has been considered in advance in every detail and retained to the end'. Subsequent comments about the primacy of planning over plans, attributed to Eisenhower and Churchill, are abridged versions of the original quote.

of Excellence,[88] US JADC2,[89] and the British Army's Digital Readiness Experiment.[90] Early indications are that commanders will have to focus more on framing the problem and ensuring unity of understanding and purpose among more diverse teams in smaller, flatter structures. This suggests a different kind of commander and a different kind of staff; people who can integrate the work of diverse teams comprising members from different disciplines, and often from outside the military.

Ensuring that commanders can frame the problem properly is essential. AI is very good at operating within a frame, but is currently at least poor at 'reading between the lines' or extrapolating from poorly defined data sets – a fragility that still depends on having a human to set the frame. Having framed the issue, commanders must be able to judge that the outputs make sense within that frame.[91] This requires people who can see the big picture, and armed forces need to grow future commanders through staff experience in headquarters so they are familiar with the environment and processes and thus able to command at increasingly senior levels. Simulation can facilitate exposure to headquarters, as can ensuring that smaller headquarters still retain roles for people to gain experience through which the requisite command skills can be acquired.

While commanders need to know how to interact with the technology, they must remain focused on the operational requirement that AI is intended to serve, and be appropriately sceptical of it so they are informed actors in the process, not passive recipients of the algorithms' outputs. Commanders need to resemble industry's 'pi-shaped leaders' with digital and data awareness alongside their military specialisation.[92] They do not need to become experts in the technology, but should know enough to appreciate its limitations, be able to work with the specialists in their teams and be satisfied enough to allow trust in the data, processes and output to flourish.[93]

---

88. The NATO Command and Control Centre of Excellence, including the library of publications, is available at <https://c2coe.org>.
89. For an overview of JADC2, see Hoehn, 'Joint All-Domain Command and Control (JADC2)'.
90. British Army, 'Army and Crack Artificial Intelligence Experts Go Full-Throttle with Digital Transformation', 18 January 2022, <https://www.army.mod.uk/news-and-events/news/2022/01/army-and-artificial-intelligence-experts-go-full-throttle-with-digital-transformation>, accessed 10 February 2022.
91. Pressure-testing assumptions behind any model, challenging how the model is defined and where it is used are essential parts of effective decision-making. An example of a failed attempt to frame the question correctly occurred during efforts to track the AIDS epidemic in Africa; the recognised World Health Organization model was based on the total number of sexual contacts rather than the number of different sexual contacts, which was a more critical determinant. John Kay and Mervyn King, *Radical Uncertainty*: *Decision-Making for an Unknowable Future* (London: W W Norton, 2020), p. 375.
92. See, for example, Paul O'Neill and Alison Gregory, '21st-Century Assistance Dogs? Harnessing Data and Technology', *RUSI Conference Report*, March 2021, p. 8.
93. As already highlighted, there is a danger of over-trusting AI and what it can deliver. As Eliot A Cohen and John Gooch describe, intelligence can tell you where you are and with what, and what the enemy has where, but cannot be authoritative about the enemy's intentions. Eliot A Cohen

Collectively, the headquarters team needs the skills, with individual team members able to speak to and understand each other. This extends beyond intelligence analysts and includes a wide range of operations, technology and data experts from within and outside the armed forces.[94] It also includes a more sophisticated understanding of, and ability to communicate, risk.[95] War is fundamentally a matter of risk management, which requires empirical ways of understanding and communicating risk. Understanding probabilities and confidence levels is, therefore, a crucial command skill, but one-off decisions such as those in conflict also require judgement built over time.

Military education needs to respond by bringing data and technology awareness earlier in a career. Moreover, how militaries value different aptitudes may also need to change. Anecdotally, British Army career management processes often direct those who score well on numeracy towards areas such as procurement rather than operations, with majors selected to attend staff college frequently in the lower quartile for numeracy. The challenge is not just for the military: countries that wish to compete successfully require national education systems that recognise the value of data and technology literacy skills and nurture them from an early age. The authors are not advocating turning education into pre-employment training; while STEM skills are needed (in larger numbers than today), the humanities and social sciences remain important, producing graduates who are adaptable, able to solve complex problems and communicate with influence and empathy.[96] National success depends on academic as much as other forms of diversity, and preparing people to thrive in the digital world requires not only technical competency but also (human) traits, such as creativity and emotional intelligence. Commanders and staffs will need both sets of skills in the future, perhaps even more so than today. Current good practice needs to become more common.

Alongside analysis, intuition is a complementary component in information processing. It is an important part of human cognition in a dual-track approach to decision-making that commanders need to exercise. Effective decisions combine the strengths of the intuitive and the analytical. Where data and intuition agree, decision-makers can act with confidence. Where they disagree, further exploration is needed before acting.[97] In 1983, Russian Lieutenant Colonel Stanislav Petrov averted potential nuclear war. His missile detection systems reported the launch of five ICBMs from the US but rather than report this immediately, he decided to wait because the information did not feel right. His (subconscious) dual-mode decision-making enabled him

---

and John Gooch, *Military Misfortunes: The Anatomy of Failure in War* (New York: Vintage, 1992), pp. 42–43.

94. Maria Korolov, 'What a Successful AI Team Really Looks Like', *CIO*, 8 June 2021.
95. See, for example, David Spiegelhalter, 'Risk and Uncertainty Communication', *Annual Review of Statistics and Its Application* (Vol. 4, No. 1, March 2017), pp. 31–60.
96. Valerie Strauss, 'Why We Still Need to Study the Humanities in a STEM World', *Washington Post,* 18 October 2017. See also Australian National University, 'The Value of a Humanities, Arts and Social Sciences Degree', <https://cass.anu.edu.au/study/value-of-hass>, accessed 10 February 2022.
97. Author interview with Professor Eugene Sadler-Smith, University of Surrey, 3 March 2021. See also Eugene Sadler-Smith, *Inside Intuition* (Abingdon: Routledge, 2007).

to make the right decision.[98] AI's greater data-processing and analysis capacity can enhance the analysis element of the decision-making process, but it needs commanders to recognise the value as well as limitations of intuition. Professional military education needs to reflect a balanced approach towards the twin components of data and intuition.

Managing the Whole Force

Commanders in the future will command teams that are necessarily more diverse than today, leading interdisciplinary teams that bring fresh insights to complex problems. The human capacity to frame effectively and develop intuition is enhanced by being exposed to different ways of seeing the world.[99] This goes beyond improving diversity in terms of protected characteristics, important though that is, and includes ensuring a breadth of education, experience and perspectives within Whole Force teams. The different elements of the Whole Force are part of this diversity.

Increasingly, integrated activity across the military domains requires that the military components of the Whole Force work together effectively. For regular military personnel, progress has been made with 'jointery', but more work is needed. Introducing joint training earlier in a military career is a way to enable this; this might require a rethink about when military personnel access professional military education, currently in the mid to late 30s in the UK. In contrast, the Australian Defence Forces provide a largely joint military syllabus in initial officer training for those attending the Australian Defence Force Academy, with navy, army and air force specialists receiving single service training as well. This provides an interdisciplinary, 'joint' model for growing future commanders from early in their military careers.[100] The progress with the regular military needs to be extended to integrating the Reserves, where it is likely that more technological expertise will reside in future.

Integrating the non-military elements of the Whole Force has proven more difficult. It was noted in a Serco Institute report that 'despite progress in operationalising the Whole Force over the last decade, efforts have stalled in achieving a seamless partnership between the military and industry'.[101] While armed forces have become better at bringing non-military people into their headquarters, there is a big difference between being present and being included. Exercises, such as *Joint Protector 2021*, often invite international partners and civilian subject matter

98. Alex Lockie, 'The Real Story of Stanislav Petrov, the Soviet Officer Who "Saved" the World from Nuclear War', *Business Insider*, 26 September 2018.
99. Cukier et al., *Framers*, pp. 207–08.
100. While not all Australian Defence Force officers enter through the Australian Defence Force Academy, a significant percentage do. The programme builds understanding, relationships and networks across the Defence Force much earlier in a career than in the UK. For further details of the three-year degree programme, see Australian Defence Force Academy, <https://defence.gov.au/ADFA/>, accessed 7 June 2022.
101. John Gearson et al., 'The Whole Force by Design: Optimising Defence to Meet Future Challenges', Serco Institute, October 2020, p. 110.

experts to help the planning process, but too often they are invited to comment on plans only after military planners have completed their work. The lack of flexibility in many headquarters' planning cycles means that by the time the plan is offered for review it can be too late for changes to be made.

This is not just an observation on the military; civilian experts often lack familiarity with military processes and wait to be invited to contribute, which weakens their influence. Military personnel who do not instinctively understand the full range of the contribution their non-military colleagues can make therefore do not include them. AI will force a need to build Whole Force diversity into the planning processes *ab initio*, so that plans are genuinely collaborative.

With AI capabilities, technology will increasingly be an actor in the Whole Force. Chess grandmaster Gary Kasparov has noted how the combination of good technology and good human players is often more successful than either superior technology or better human players working on their own.[102] In some situations, people and machines may be so tightly integrated in shared tasks that they become interdependent, where the idea of task handoffs becomes incongruous. This is already apparent in the design of work supporting cyber sense-making, where human analysts are combined with software agents in understanding, anticipating and responding to unfolding events in near real-time.

Getting the most from these human–machine Whole Force teams goes beyond merely effective task distribution. It involves finding ways to support and enhance the performance of each member, human or machine, so that the collective output is greater than the sum of the individual parts.[103] The right behaviours and ability to create an inclusive culture will be essential to get the most from such teams. Rather than focusing on trying to manage 'emergence' – a concept that seeks to describe how simple things can, in interaction, lead to complex and unpredictable outcomes[104] – or the activities of team members, commanders will need to invest more in shaping the team and fostering the relationships within it.[105]

---

102. Francis Churchill, 'The AI Takeover Is Happening "Too Slow", Says Chess Grandmaster', *People Management*, 12 June 2019.
103. Matthew Johnson et al., 'Autonomy and Interdependence in Human-Agent-Robot Teams', *IEEE Intelligent Systems* (Vol. 27, No. 2, 2012), pp. 43–51.
104. The concept derives from Complex Adaptive Systems research at the Santa Fe Institute. It seeks to describe 'how simple things interacting in simple ways yield complex outcomes' that may not be predicted or even predictable. See Paul Grobstein, 'From Complexity to Emergence and Beyond: Towards Empirical Non-Foundationalism as a Guide for Enquiry', *Soundings* (Vol. 90, No. 1/2, 2007), pp. 9–31. The commander, the staff and the headquarters as a collective agent are part of emergence, as are the murmurations of starlings.
105. Stanley McChrystal argues that his success as the commander of the Joint Special Operations Task Force in Iraq in 2003/04 was due to a shift from vertical command with him at the pinnacle to horizontal flows enabled by trust. See Stanley McChrystal, *Team of Teams: New Rules of Engagement for a Complex World* (London: Penguin, 2015). AI's impact on human teams is

While AI currently serves as a tool, as the technology advances, it warrants consideration as a genuine member of the team, with rights impacting on its human teammates and duties to them. Irrespective of its eventual status, however, AI is likely to change team dynamics and what is expected of the human team members. Introducing AI to a team changes the team dynamics, and its difference to a human team member makes team formation harder. Moving through Bruce W Tuckman's classic stages of forming, storming, norming and performing requires compromise and accommodation.[106] AI is currently less able to do this, requiring greater flexibility from the human participants, making it harder to build human–machine teams and more difficult to recover trust that has been lost.

Advanced AI, if it can be said to have a motivation or bias, is likely to be logical and task-oriented (in Strength Deployment Inventory terms, Green and Red). A balanced team will increasingly need humans who can sustain team relationships, both internally and across teams.[107] Human–machine teams, therefore, will be different although they might have some analogies with purely human teams that include neurodiverse colleagues for whom empathy or understanding of emotional cues are difficult. As with neurodiverse teams, human–machine teams will benefit from the value the diversity of team members bring to the whole, but also need adjustments to be made to maximise the opportunities for team performance.[108] Exactly how the concept of AI as a team member will develop remains unclear but there are calls for organisations to consider the needs of advanced technologies on a more equal footing.[109] Enhancing use of enterprise AI in business support activity will offer opportunities for exploring how human–machine teams can work together most effectively, as well as potentially delivering the hoped-for reduction in running costs and moving humans up the value chain to undertake more meaningful work.

Career Management

The new styles of leadership, new skills and enhanced understanding of technology, data and risk that are needed will also require new approaches to career management. Military career management systems move people (too) frequently, yet it takes time to form effective teams with the requisite levels of trust. Militaries might slow down movement of key people, and perhaps even teams, so that a senior headquarters team is managed as a collective entity rather

---

potentially more far-reaching and its incorporation into what have traditionally been human-centric organisation models far harder to implement.

106. Bruce W Tuckman, 'Developmental Sequence in Small Groups', *Psychological Bulletin* (Vol. 63, No. 6, 1965), pp. 384–99.
107. For a description of the Strength Deployment Inventory, see Elis H Porter, 'Theory Overview', SDI, <https://personalstrengthsuk.com/theory-overview-2/>, accessed 20 May 2022.
108. Monika Mahto et al., 'A Rising Tide Lifts All Boats', *Deloitte Insights*, 18 January 2022.
109. See, for example, Christy Pettey, 'The Rise of the Chief Robotics Officer', Gartner, 15 March 2017; Narendra Mulani, 'On-Boarding Your New Co-Worker – AI', *CIO*, 8 May 2018; Daphne Leprince-Ringuet, 'The Growing Robot Workforce Means We'll Need A Robot HR Department, Too', *ZDNet*, 5 February 2020.

than as individuals. However, current HR practices make it unlikely that either the armed forces, or indeed industry, will be willing to hold people in positions indefinitely in anticipation of future requirements. In Raphael Pascual and Simon Bowyer's terms, this gives rise to 'hybrid teams', those whose membership is fluid, and for whom the ability to build team trust quickly is essential. Even permanent headquarters will be affected by this, especially where they become 'Whole Force'. The matter is more acute for 'ad hoc teams', such as temporary headquarters created for a specific mission.[110] Mechanisms are needed to accelerate the development of trust, which experience suggests can be done by the early practice of behaviours including demonstrating 'technical competence, openness with information, reciprocity of support and perceived integrity in decision-making'.[111]

Slowing down the rate of movement of people in senior appointments in headquarters will help but will not be enough. In the absence of being able to guarantee pre-established teams ready for whenever a mission is needed, there needs to be a way of reducing the time taken to form new Whole Force teams. Simulation offers a way of preparing a newly formed team by condensing the time for mission rehearsal and providing different components of the Whole Force shared experiences of working together. The military does this well for regular personnel; the military process of socialisation creates strong bonds, including sending people to partners for training, exercises and assignments. This investment in cross-cultural understanding is absent in relation to other parts of the Whole Force. Building understanding of the other and thus trust is just as important for the civilian component. Militaries could do more to offer their staff experience of working with the commercial sector, including with technologists, data specialists and coders, while civilians also need to develop a better understanding of the military, its language, processes and values. Armed forces can assist this process by offering exchange appointments and modularising and/or shortening their courses to make it possible for civilians to attend. The coronavirus pandemic has introduced new ways of working and accelerated changes in military training and education that could offer the foundations for trust on which new teams and types of headquarters can emerge.

In short, AI-enabled decision-making is not just a question of technology; it requires changes to command structures, processes and people skills if it is to live up to its potential as a revolution in how armed forces operate across the full range of missions. It is essential, however, that in adapting to the changing character of war, armed forces do not lose sight of war's enduring nature: commanders must remain leaders and warriors who can inspire ordinary people to do extraordinary things in the most difficult of circumstances, not merely people who are good managers of battle. In the military context, AI is a tool to maximise the chances of the armed forces winning in a deeply competitive environment.

110. Raphael Pascual and Simon Bowyer, 'The Importance of Trust within Teams', in QinetiQ, 'The Trust Factor', p. 9.

111. Castleton Partners and TCO International Diversity Management, 'Building Trust in Diverse Teams', Scoping Study Report, February 2007, quoted in *ibid.*, p. 10.

Conclusion

AI IS FAST BECOMING a core part of our national security fabric. Militaries and intelligence agencies are experimenting with algorithms to make sense of large amounts of data, shorten processing times, and accelerate and improve their decision-making. Growing use of and familiarity with AI can foster trust in it, but as the debates among experts indicate, there are serious challenges to building and sustaining trust in a technology as transformational as AI.

This paper has focused on operational and mission support applications of AI and explored the importance and implications of the evolving human–AI relationship for future military decision-making and command. When the military commander's role shifts from one of controller to that of teammate, when we can no longer ascribe only a supporting function to artificial agents, then we need to fundamentally rethink the role of humans and the structures of our institutions. In short, we need to reassess the conditions for and implications of trust in human–machine decision-making. Without that trust, effective adoption of AI will continue to advance more slowly than the technology and, importantly, behind the rate at which it is being adopted by some of our adversaries.

A slightly modified notion of trust – one that need not impute intentionality or morality on the part of the artificial agent – can and does apply to AI. So long as we entrust machines to do things that can have severe, even deadly consequences for humans, we make ourselves vulnerable. So long as there is a risk that AI's performance falls short of our expectations, any use of it becomes essentially an act of trust.

In all but the rarest cases, trust in AI will never be total; in some cases, users may consciously consent to lower levels of trust. That trust needs to be considered across five different elements, which the authors call 'trust points'. We should never rely on any single point to generate overall trust. Indeed, the areas that tend to get most of the attention – questions regarding the quality of data or the explainability of AI outputs – are bound to deliver unsatisfying answers in the longer term, and potentially a misplaced sense of reassurance about the technology.

Most often overlooked is the need for trust at the level of the organisational ecosystem. This requires rethinking the organisation of armed forces and their C2 structures. If the growing role of machines was once a key driver behind the rise of the bureaucratic army structure to concentrate the means of management,[112] AI is challenging this feature of standing armies in fundamental ways. If its use is to become more than the digitisation of analogue ways of working, Defence has to change its decision-making structures across the spectrum of 'operate' and 'warfight'. It

112. H H Gerth and C Wright Mills (eds), *From Max Weber: Essays in Sociology* (New York, NY: Oxford University Press, 1946), pp. 221–23.

will also need to engage with and involve far more intimately all aspects of a true Whole Force, including its under-utilised Reserve force as well as industry and broader government.

Leadership as an enduring element of the military profession also requires reconsideration. There is a tendency to think of leadership as an abstract or immutable quality of military command. In the age of AI, commanding a mission or leading teams requires both new skills (such as the ability to 'speak digital') and a more diverse set of traits (for example, the ability to think laterally, frame problems, and apply critical judgement where data and intuition are in conflict). Even more so than before, AI requires commanders who can make sense of complexity, frame problems and ask the right questions suited to the circumstances. These 'deliberate amateurs' eschew early narrow specialisation and opt for range and an experimental mindset; they can build expert teams and draw on the input of specialists so that the collective talent is both broad and deep.[113] These Whole Force teams will include humans and machines, all of whom will contribute according to their expertise in shaping and making decisions.

In seeking to answer how trust affects the evolving human–AI relationship in military decision-making, this paper has exposed several key issues requiring further research:

- How we build the trust necessary to reconfigure the organisation of command headquarters, their size, structure, location and composition, at tactical, operational and strategic levels.
- How we adapt military education to better prepare commanders for the age of AI.
- How we optimise and transform collective training across all domains to improve command involving greater collaboration with artificial agents.
- How we operationalise the concept of 'Whole Force' to make better use of the extensive talent within our societies, industries and research establishments.
- How we define the needs and objectives of AI and humans within human–machine teams.

Absent fundamental changes in how we access, train and grow people in leadership positions, and how we reform the institutions and teams within which they operate, we risk getting the trust balance in the human–machine relationship wrong and will fail to harness AI's full transformative potential.

113. David Epstein, *Range: How Generalists Triumph in a Specialized World* (London: Macmillan, 2019).

About the Authors

Christina Balis is Global Campaign Director of Training and Mission Rehearsal at QinetiQ. Her 20 years' experience across both sides of the Atlantic encompasses consulting, industry and public policy settings, with particular focus on defence, global security and transatlantic relations. She has been a fellow in the Europe Programme of the Center for Strategic and International Studies in Washington, DC, vice president for strategy and corporate development at Serco Inc., and principal and head of European Operations of Avascent in Paris. She holds a MA and a PhD in International Relations from the Johns Hopkins School of Advanced International Studies in Washington, DC and Bologna, Italy, and business degrees from the UK and Germany.

Paul O'Neill is the Director of Military Sciences at RUSI. With over 30 years' experience in strategy and human resources, his research interests cover national security strategy and organisational aspects of defence and security, particularly organisational design, human resources, professional military education and decision-making. He is a CBE, Companion of the Chartered Institute of Personnel and Development, Visiting Professor at the University of Winchester and member of the UK Reserve Forces External Scrutiny Team.

Annex

The following individuals kindly spoke to the authors in the development of this paper:

- Professor Patrick Baker, Head of Science Air Information Experimentation, Rapid Capabilities Office, RAF.
- General Sir Richard Barrons, Co-Chairman and Co-Founder, Universal Defence & Security Solutions; former Commander, Joint Forces Command, MoD.
- Robert Bassett-Cross, CEO, Adarga.
- Lieutenant Colonel Al Brown, SO1 Land Systems, Dstl; Visiting Fellow, University of Oxford.
- Stephen Cassidy, Chief Researcher – Future Organisation, BT.
- Air Commodore Blythe Crawford, Commandant Air and Space Warfare Centre, RAF.
- Brigadier Stefan Crossfield, Head Information Exploitation and Chief Data Officer, British Army.
- Major General John Collyer, Director Information, British Army.
- Dr Dionysios Demetis, Associate Professor (Senior Lecturer), Hull University Business School (UK); Visiting Professor, Texas A&M University (US).
- Martin Howard, Senior Leadership Team, Cyber, NATO Policy and Strategy, Universal Defence & Security Solutions.
- Professor Björn Johansson, Research Director, Swedish Defence Research Agency; Co-Chair NATO SAS-143.
- Dr Angeliki Kerasidou, Associate Professor in Bioethics, University of Oxford.
- Dr Oliver Lewis, Co-Founder, Rebellion Defence.
- Professor Michael Mainelli, Executive Chairman, Z/Yen Group; Alderman, City of London.
- Major General Jim Morris, Commander, Standing Joint Force Headquarters, UK Strategic Command.
- John Ridge, Director Strategy and Enterprise Services, UK Strategic Command.
- Paul Rimmer, Visiting Professor, Department of War Studies, King's College London; former Deputy Chief of Defence Intelligence, MoD.
- Professor Eugene Sadler-Smith, Professor of Organisational Behaviour, Surrey Business School, University of Surrey.
- Rob Solly, Director of Research Partnerships, Improbable.
- Air Commodore Steve Thornber (Ret.), former Defence Intelligence, MoD.
- Colonel François Villiaumey (Ret.), former Deputy Director, Ecole de Guerre, France.
- Guy Williams, Head UK Defence, Palantir Technologies.