Defense & Security Analysis

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/cdan20

Trusting machine intelligence: artificial intelligence and human-autonomy teaming in military operations

Michael Mayer

DEFENSE & SECURITY ANALYSIS
https://doi.org/10.1080/14751798.2023.2264070

Routledge
Taylor & Francis Group

OPEN ACCESS
Check for updates

Trusting machine intelligence: artificial intelligence and human-autonomy teaming in military operations

Michael Mayer

Norwegian Defence Research Establishment (FFI), Kjeller, Norway

ABSTRACT

Continuous advances in artificial intelligence has enabled higher levels of autonomy in military systems. As the role of machine-intelligence expands, effective co-operation between humans and autonomous systems will become an increasingly relevant aspect of future military operations. Successful human-autonomy teaming (HAT) requires establishing appropriate levels of trust in machine-intelligence, which can vary according to the context in which HAT occurs. The expansive body of literature on trust and automation, combined with newer contributions focused on autonomy in military systems, forms the basis of this study. Various aspects of trust within three general categories of machine intelligence applications are examined. These include data integration and analysis, autonomous systems in all domains, and decision-support applications. The issues related to appropriately calibrating trust levels varies within each category, as do the consequences of poorly aligned trust and potential mitigation measures.

KEYWORDS

Artificial intelligence; autonomous platforms; decision-centric warfare; the future battlefield; trust

Introduction

Throughout history, technology has played a key role in the evolution of armed conflict. New technologies and platforms have expanded the number of warfighting domains to include the seas, the skies, outer-space and Cyber-space.[1] Technology has increased the tempo of tactical engagements, the geographic breadth of the battlefield, the means by which commanders communicate with their forces, and the ways in which states plan and conduct armed conflicts.[2] In the twenty-first century, a group of technologies known collectively as artificial intelligence (AI) appears poised to usher in a new era in which machine-intelligence and autonomy will enable distinctly new concepts and procedures for the planning and execution of military operations.[3] The growing availability of large quantities of data has encouraged an insatiable appetite for information that requires expedient and dispassionate analysis, a role for which AI is uniquely suited. The fusion of AI decision-making, improved sensors, and agile robotics will enable new systems capable of independently performing all phases of the observe-

CONTACT Michael Mayer Michael-John.Mayer@ffi.no

orient-decide-act (OODA) decision-making loop. The influx of autonomous systems on the battlefield and potential for algorithm-based warfare will gradually augment, or even displace, human decision-making processes within some or all of these steps, and at speeds that may exceed the cognitive capacity of human planners.

Currently, algorithms contribute to a range of military systems, from communication equipment to sensors to air defence systems. In many respects, a modern military force is already reliant on certain forms of AI. At its most capable, however, future applications of machine-intelligence promise something new – a non-human collaborative partner able to make proactive "decisions" within the context of shifting circumstances on the battlefield. The inherent advantages of this capability can only be realised if humans are comfortable relying on AI – not just as a tool, but as a member of the team. This article will therefore focus on one specific aspect of human-autonomy teaming (HAT): establishing appropriate levels of trust in machine intelligence. A vast body of academic literature exists that focuses on trust in automation, or robotics, for commercial applications, but academic research specifically addressing military applications is less plentiful. In particular, this essay explores how autonomous systems leveraging AI are used in dissimilar military contexts and how these varying contexts influence trust.

The basic argument presented here is threefold. First, AI technologies are being developed for military use within three broad categories that cut across the tactical, operational, and strategic levels of warfare: algorithmic solutions for data integration and analysis, autonomous systems utilising machine intelligence, and decision-support software that augments human decision-making. Second, I argue that the proliferation of AI in military operations will necessarily lead to more interaction between humans and intelligent machines. Operations will increasingly rely on safe and effective human-machine teaming, which in turn relies on humans constantly evaluating and granting appropriate levels of trust – known as trust calibration – to intelligent machines.[4] Issues relating to proper trust calibration can vary within each of the three categories, as will the implications of trust misalignment. Third, ensuring optimal human-machine teaming with AI will, therefore, depend on identifying potential trust issues within each of these categories and devising appropriate technical or doctrinal adaptations to address them. After a brief review of AI and an overview of the likely applications of machine-intelligence on the battlefield, I explore the concept of trust and trust calibration before analysing the pitfalls and possibilities for encouraging appropriate levels of trust in each of the three categories.

Advances in artificial intelligence

For decades, humans have been fascinated with the possibility of infusing machines with some form of artificial intelligence, defined by Nils Nilsson as "that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment."[5] Two broad approaches to AI emerged during the earliest days of the digital age. A top-down expert system approach used complex pre-programmed rules and logical reasoning to analyse a particular data set. For well-defined environments with predictable rules – applications such as analysing laboratory results or playing chess – the performance of expert systems or "symbolic" AI (based on symbolic logic) depended largely on processing speeds and the quality of the algorithms. The other broad category used a bottom-up machine learning approach modelled after the way humans learn by detecting patterns within data. Neural networks are a form of machine-learning inspired by the human brain that can identify complex patterns using multiple (and therefore "deep") layers of artificial neurons, a technique that is fundamental to the approach known as "deep learning."[6] Through an ability to find relationships within data sets, such methods are also termed "connectionist."[7]

The differences between top-down, rule-based symbolic systems; and bottom-up machine learning connectionist techniques, are substantial, particularly regarding the potential range and flexibility of their applications. Deep learning approaches are notable due to their ability to separate the learning from the data set upon which they train, and the software therefore can be applied to other similar problems. Whereas rules-based algorithms perform exceedingly well at narrowly defined tasks, deep learning approaches are able to find rapidly patterns and; in effect, teach themselves applications for which expert-system computational approaches are less effective.[8] A number of recent AI advances demonstrate an ability to mimic creativity, or generate novel approaches to problem-solving, that can appear counter-intuitive to humans. Examples include high altitude balloons with AI navigation controls that discovered optimal and unexpected techniques for utilising wind patterns, or using AI to develop more effective designs for machine-parts using additive manufacturing.[9]

In general, however, AIs remain narrow or "brittle" in the sense that they function well for particular applications, but remain inflexible when used for others. Compared with humans, machine-intelligence is far superior when applying rules of logic to a data set given that machine computational speeds far exceed the human brain, but falls short when attempting inductive reasoning where it must make general observations about a data set, or an environment. Large amounts of training data are still necessary for most machine learning, even though new approaches including self-supervised learning, techniques for generating simulated data, such as the use of generative adversarial networks (GAN), and "less than one-shot" or LO-shot learning, requiring very small datasets are emerging.[10] Image recognition algorithms can become easily confused, and cannot immediately, or intuitively, interpret situational context as well as humans. This brittleness extends to other problems such as games. Whereas AI often exhibits superhuman capabilities in video games, they often cannot transfer that expertise to a new game with similar rules or playing mechanics.[11]

While AI technologies continue to make significant progress in becoming more adaptable, anything approaching human-like artificial general intelligence remains elusive.[12] This is partly due to our surprisingly limited grasp of the biology and chemistry involved when humans process information – what is generally referred to as human cognition. Neuroscientists still do not fully understand how the brain functions, which limits efforts to model digital processes on biological ones. AI techniques such as deep-learning have therefore enjoyed a symbiotic and mutually beneficial relationship with cognitive neuroscience.[13] Evaluating the near-term future of AI is further complicated by the incremental progress of the technology. The hype surrounding AI – fuelled in no small part by the success of deep-learning approaches – has led to both unrealistic expectations surrounding the future of the technology and a normalisation of its very substantial progress. Some have termed this the "AI effect." One report noted, "AI brings a new technology into the common fold, people become accustomed to this technology, it stops being considered AI, and newer technology emerges."[14] Some speculate that the progress resulting from machine-learning techniques may plateau, while others remain optimistic.[15] Some see the potential in attempts to merge symbolic AI approaches with the various forms of machine-learning.[16] The near-term future therefore remains uncertain. Related technological advances, including computer chip design in the short-term, and quantum computing in the long-term, may also influence the pace of further progress.[17]

Artificial intelligence in military operations

For many military applications, however, narrow uses of AI are more than adequate. Many algorithmic solutions already in use by militaries around the globe can be considered "artificial intelligence" and there is no shortage of proposed uses for AI. The possible military capabilities enabled by AI are part of a dramatically different future operating environment envisaged by analysts such as Christian Brose, and former defence officials such as Robert Work.[18] If these predictions regarding the effects of artificial intelligence come to fruition, they will have wide-ranging implications for the planning and implementation of operations. Existing and near-future applications can be divided into three categories: data integration and analysis, autonomous systems, and decision-support software. As with most typologies, the categories do not have completely clean edges and some applications cut across several of the labels. Notably, however, the potential consequences of leveraging AI in a military context – and thus the risks of poorly calibrated trust – increase from data analysis to autonomous systems and ultimately decision-support applications. Furthermore, the integration of autonomous systems in military force structures suggests a cumulative effect, as well. Trusting AI to process sensor data is a necessary step for allowing autonomous systems to operate alongside human personnel, and future AI-supported decision-making at the operational level will require an additional layer of trust resting atop the trust relationships infused in the human-autonomy teaming among military units fielding autonomous systems.

Data integration and analysis

The use of AI in the operation of various capabilities and platforms may oftentimes go unnoticed for the average user simply due to its integrated role in system architectures. Examples of this include civilian satellite navigation, internet search engines, or online translation tools. Military applications include wireless communication, or radars, that leverage machine-learning algorithms for optimal use of the electromagnetic spectrum.[19] For unmanned or remotely piloted aircraft, onboard algorithms allow sensors independently to conduct preliminary data analysis and thereby reduce bandwidth requirements. Algorithms are already useful for analysing sensor data across a range of systems and platforms.[20]

In addition to these integrated applications, the conscious and active use of AI for data analysis extends to intelligence, surveillance, and reconnaissance (ISR) efforts. As James Johnson notes, machine-learning algorithms "could significantly improve existing machine vision and other signal processing applications, identify patterns from large data-sets of signals and imagery, and enhance autonomy and sensor fusion applications."[21] The US Air Force created the Algorithmic Cross Functional Team in 2017 to apply AI to image analysis in its efforts to identify and track targets, and establish patterns of life to enhance situational awareness.[22] In cyberspace, pattern-recognition algorithms can similarly determine a network's normal operating pattern to enable easier identification of deviances that may signal the presence of an intruder. The use of AI for open-source intelligence (OSINT) analysis can identify individuals or even make rough near-term predictions about insurgent activity.[23] Experimental AI applications such as the Global Information Dominance Experiments (GIDE) sift through massive amounts of multisource data for patterns and trends to make predictions about a range of future events.[24]

Autonomous systems

A second category of AI applications comprises a broad range of autonomous systems. Autonomy is a term that remains challenging to define precisely or concisely. A 2016 report by the Joint Air Power Competency Centre (JAPCC) distinguished automation – which involves machines preforming predictable, bounded pre-defined tasks set by humans – from a fully autonomous system. The authors characterised an autonomous system as one that could determine its own course of action, deliberate decisions not restrained by pre-programmed responses, have an ability to learn and compile "experience," and therefore no longer be completely predictable in its actions.[25] Paul Scharre and Michael Horowitz described three dimensions of autonomy in a 2015 paper: (a) the human-machine command and control relationship, simplified by determining whether a human is "in," "on", or "out of" the decision-making loop; (b) the complexity and abilities of the machine or system; and, (c) the type of function being automated.[26]

Within the context of AI, it is worth noting that the distinction between automated and autonomous systems becomes blurred. Machine-intelligence is highly relevant for a number of automated functions that enable autonomous systems, including system operations and self-diagnostics, autopilots, combat software and target tracking/identification, and self-guided weaponry.[27] Autonomy therefore describes a sliding scale of independent machine functionality along a number of variables, including the level of human-machine interaction, an ability independently to sense and adapt to changing contexts, decision-making abilities to accomplish some set of pre-determined goals, and the ability continuously to learn and improve from those decisions.

The less stringent definition of autonomy might encompass current military assets ranging from air and missile defence systems, counter-rocket or artillery systems, active protection systems for ground vehicles, loitering munitions, advanced cruise missiles, and Cyber capabilities.[28] While autonomous systems are currently deployed in most warfighting domains, the next generation of autonomy will leverage AI to enable even greater independence from human direction. Currently under development are space, maritime, airbourne, and ground-based platforms and systems that, as the JAPCC report outlined, represent a qualitative evolution, from a tool at the disposal of a tactical commander, to a partner with which humans will have to interact and cooperate.

Autonomous aircraft will soon perform logistical tasks such as transporting cargo, or refuelling duties. New operational concepts known colloquially as "loyal wingman" programmes envisage larger unmanned platforms that operate alongside piloted craft, thereby offering more options for networking sensors, or additional munitions, and thus greater tactical flexibility. Autonomous ships will give maritime commanders a similar capability at sea, and ground-based systems are currently under development as well.[29] New manufacturing processes will reduce production costs along with reduced size, weight, and power requirements for scalable AI software. This will likely enable the deployment of large numbers of small unmanned systems, which will be controlled and co-ordinated in swarm formations with battle management and targeting software that can be quickly uploaded and updated to "retrain" effectively the system with a few keystrokes.[30] Autonomous systems are therefore poised to increase the overall number of platforms on the battlefield.

Decision support and decision-centric warfare

Military commanders already rely on machine-intelligence in their decision-making processes, ranging from algorithmically derived collateral damage estimates, to targeting solutions for air and missile defence systems. For a range of systems, computer-generated data analysis enhances situational awareness and provides options for war-fighters. Future decision-making aids may bring about further developments. Compared with current time-consuming operational planning paradigms, Steven Davis observes that "AI can lead decision-makers towards optimal solutions when presented with many that are merely suitable, feasible, or complete."[31] The introduction of large numbers of autonomous weapon systems using AI decision-making software may influence the operational level of war, particularly command and control (C2) aspects of military operations.

Appropriately enough, this now-common term emerged during the nascent information technology age of the 1960s to distinguish the authority and responsibility of command from the processes creating the necessary conditions for the commander to exert control over the implementation and execution of operations.[32] Although it has become commonplace for higher-level commanders and political leaders to observe particular tactical engagements, the operational level may nevertheless be the most appropriate for having humans "on the loop" if autonomous systems were deployed. Even with fleets of self-synchronizing autonomous surface vessels or aerial systems, the need to co-ordinate the broader operational effort will remain human-centric. If that is the case, however, operational planning and co-ordination may need AI assistance simply to maintain an advantageous and effective battle rhythm.

This is the motivation behind the so-called "decision-centric" concept of warfighting. One such concept developed by the Defense Advanced Research Projects Agency (DARPA), known as Mosaic Warfare, utilises AI to co-ordinate a network of disaggregated forces. The concept proposes a hybrid C2 configuration with human command and machine control. Commanders choose tasks in need of completion from a set of AI-generated courses of action (COA) based on its overview of available manned and unmanned force components.[33] Another approach outlined by Davis utilises a maritime decision-support architecture with a middle layer of AI based on operational functions such as protection, sustainment, or fires.[34] Concepts integrating AI and autonomous systems are a logical – albeit ambitious – progression given the perceived advantages of rapid machine-based decision-making. This is particularly applicable when a connected battlespace allows for data-fusion amongst a disparate, but linked, network. The sheer volume of available information may lead to a dependence on machine-intelligence simply because machine cognition will be needed to understand and act upon those data in an advantageous and timely manner.

Trust and machine intelligence

The anticipated role of machine-intelligence in all areas of military operations – from sensor data to weapons systems to operational decision support – suggests a growing reliance on AI. For example, an expert group report under the rubric of a North Atlantic Treaty Organisation (NATO) initiative recommended that the military alliance "should encourage the incorporation of AI into strategic and operational planning. It should exploit the power of AI-driven technologies to enhance scenario planning exercises and long-term preparedness."[35] Official statements and publications such as the US Navy's recently-released policy on intelligent autonomous systems emphasises trust as an important component of reliance, and includes questions such as how and when humans should trust machines.[36] As machine intelligence becomes more capable of increasingly complex cognitive functions and hones its ability to operate independently, humans will need to view AI and autonomous systems as partners just as much as tools. Similar to any partnership, trust is a crucial to effective human-machine cooperation.

Defining trust

Trust is one of many concepts that initially appears intuitive, but becomes more complex upon further inspection. Not surprisingly, multiple definitions and conceptualisations of trust have emerged over the past decades. After reviewing some of the various attempts to define the term, the authors of one influential article concluded that, "these definitions highlight some important inconsistencies regarding whether trust is a belief, attitude, intention, or behaviour. These distinctions are of great theoretical importance."[37] One popular definition from Mayer et al. (1995) contends that trust is the "willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that party."[38] A more recent and simplified definition of trust is "the attitude that an agent will help achieve an individual's goals in a situation characterised by uncertainty and vulnerability."[39] The presence of vulnerability and therefore risk is a significant component of trust since it attaches a potential cost for misplaced trust.

Although the building blocks of human-machine teaming are distinct from human interpersonal relationships, many of the fundamentals are comparable. Keng Siau and Weiyu Wang note that trust is dynamic, and typically is built gradually via two-way interaction, but can also be strongly affected by initial impressions.[40] Some scholars have posited that generating trust occurs initially through the predictability of future behaviour, which is then repeatedly confirmed through consistent behaviour that establishes dependability, and finally evolves into a broad judgement of reliability akin to faith.[41]

Trust in automation

Three similar elements influence trust in automation. The past and current performance of the automation, along with information about what the system actually does, parallels predictability. Details about the automation's design, and whether it will achieve the goals set by the operator, can reveal relevant and useful information about how the system operates, thereby eliciting the same dynamics as dependability. Finally, the purpose, or rationale, behind the automation; and whether its use aligns with the designer's intent, has an abstract quality of transference (trust the designer's intent, therefore trust the automation) similar to faith.[42]

For many scholars, it is at this point that human interpersonal relationships and human trust in machines begin to differ. Whereas people are usually sceptical of strangers and trust builds gradually as described above, humans often have initial, faith-based expectations that machines will work perfectly. This initial trust quickly erodes when errors arise, but faith eventually can be replaced by the more durable qualities of predictability and dependability.[43] In a comprehensive 2015 survey of scholarly articles on trust and automation, Kevin Hoff and Masooda Bashir developed a three-part trust model that takes this initial trust in machines (dispositional trust) as its starting point and adds context (situational trust) and experience (learned trust) to the mix.[44]

They posit that dispositional trust of automation is the most stable of the three and most influenced by culture, age, gender, and personality traits. Most of these variables have a demonstrative impact but with few clear tendencies.[45] The role of culture – which can be defined as a "set of social norms and expectations that reflect shared education and life experience" – represents a particularly salient factor.[46] Professional background or technical training constitutes one cultural difference that influences how individuals approach automation. Attitudes towards power and authority, or even views regarding the balance between individual or collective interests, can also play a role. One study of trust in e-commerce services among customers in Iceland, Finland, and Sweden revealed significant differences regarding dispositional trust, with customers in Finland harbouring the greatest scepticism and those in Iceland exhibiting the highest levels of trust.[47]

Along with the initial impact from dispositional trust, situational trust is the model's second component having a substantial role in developing trust in automation. Contextual factors may include external variabilities such as system complexity, operator workload that affects automation monitoring, environmental factors that influence the risks and benefits of automation, or organisational structures. Relevant situational trust factors considered "internal" to the human operator might include self-confidence, subject-matter expertise in the domain being automated, the operator's ability to focus (affected by stress, sleep, boredom, internal motivation), or even a positive mood – which has been linked to higher levels of initial trust in automation.[48]

The third and final component of the model is learned trust, which encompasses a broad set of variables relevant to trust in automation. An operator often has some preexisting knowledge of automation, whether it comes via previous experience from other automated systems, or based on the reputation of the automation in question. Expectations regarding automation and second-hand knowledge regarding its performance can influence trust even before an operator interacts with the system. The initial interaction is influenced first by the automation's design features: its appearance, ease of use, modes of communication, and transparency.[49] Design choices relating to the human-machine interface, such as display layout or types of voice commands, can play a significant role in eliciting trust. Once the initial levels of trust are established through prior experience or design features of the system itself, operators continually and dynamically gauge their level of trust. This may rely on factors such as reliability, predictability, system utility, and when and how errors occur – including how the operator is alerted to them.[50]

Trust calibration and misalignment

Significant effort has been devoted to creating trust between humans and automated systems, but past experience has demonstrated that excessive trust can also be problematic. Amongst the most common tendencies associated with automation "overtrust," or misuse, include complacency and automation bias. Operators overseeing mostly reliable automated systems tend to become complacent, and therefore less vigilant in their monitoring routines and assume – not surprisingly – that systems are functioning normally. A related issue is automation bias, whereby a human operator fails to respond to automation malfunctions, or makes incorrect decisions to follow automated recommendations.[51] One study concluded that pilots using a computer-generated recommendation system for de-icing procedures outperformed those without the aid, as long as the computer provided correct advice, but performed more poorly when the advice was incorrect. In another study, operators responsible for in-flight retargeting of Tomahawk cruise missiles appeared to more acceptant of automated recommendations as the level of automation increased, suggesting the existence of automation bias.[52]

Automation bias appears to have contributed to a number of commercial aircraft disasters, included the loss of Air France flight 447 in 2009. Veteran journalist William Langewiesche, in a detailed 2014 article about the crash based on the cockpit crew's conversations recovered from the aircraft's flight recorder, focused on automation as a contributing factor. Langewiesche argued that the pilots were so accustomed to relying on automated flying aids that misleading information from a faulty airspeed indicator created uncertainty and an inability to comprehend what was actually happening to the aircraft. This led to a string of faulty decisions and repeated failures to make the proper – and in retrospect relatively simple – adjustments that might have avoided the tragedy. His succinctly summarised thesis was that "automation has made it more and more unlikely that ordinary airline pilots will ever have to face a raw crisis in flight – but also more and more unlikely that they will be able to cope with such a crisis if one arises."[53]

Rather than focusing on ways to increase human trust of automated systems, developers often strive to elicit calibrated levels of trust that correlate to the system's capabilities. Trust calibration simply describes a process by which human interactions with machine-automation, or machine intelligence, strive to achieve an ideal state in which the human places an appropriate amount of trust in machine intelligence based on its strengths and weaknesses. With properly calibrated trust levels as a target, overtrust can be understood as trust that exceeds the capabilities of the system, whereas distrust describes the opposite situation in which the operator trusts the system less than its capabilities might dictate.[54] Achieving the proper trust alignment sounds simple enough, but often can be complicated by normal human responses. As noted above, operators usually have high performance expectations when using systems, particularly those with machine-intelligence. When errors occur, human operators tend to over-correct their trust levels and lower their expectations to a level below the capabilities of the system – thereby transitioning directly from overtrust to distrust.[55]

Automated versus autonomous systems

Most of the research into human-machine teaming over the past decades has focused on automated systems. A fundamental question for which there are few clear answers is the extent to which automated systems differ from autonomous systems. The distinction mentioned earlier differentiated between rigid, pre-determined, and predictable automated tasks, versus unrestrained, dynamic, and unpredictable autonomy. One recent survey article on human autonomy teaming noted, "the division between the two is a matter of degree and the differences are a moving target … .at what point automation might be better described as autonomy is an open question."[56]

In practice, therefore, this distinction is more graduated and perhaps better understood as a continuum with automated functionality at one end, and autonomous functionality at the other. Even this type of graduated approach has only limited utility. We humans tend gradually to regard autonomous functionality as something more akin to automation once we become more comfortable with its performance and reliability. To add further nuance, it may even be the case that autonomous systems could have an automated function. An autonomous AI-enabled Cyber-defence may act independently to handle threats in an unpredictable and unscripted fashion, but the network defences themselves might considered automated.

In a thought-provoking article dealing with trust in autonomous weapons systems, Heather Roff and David Danks question a similar binary attitude categorising autonomous systems either as a tool "where reliability and predictability of behaviour is sufficient to 'trust' the system," or "a moral agent with values and preferences, in which case the threshold for 'trust' would be significantly higher."[57] Similarly, Thomas O'Neill et al. introduces the concept of computer-based "autonomous agents," as "distinct entities that represent unique roles on the team that would otherwise have to be filled by a human."[58] While acknowledging Roff and Danks' discomfort with the binary concept of moral agent versus tool, the distinction nevertheless has some value in conceptualising the differences between trusting automation and trusting autonomy. Rather than simply performing pre-defined actions for a narrow set of circumstances, the autonomous agent relies to a greater degree on something akin to judgement. Trusting this judgement combines the dispositional and situational trust related to the performance of automated systems with an increased focus on process and purpose, which entails a deeper understanding of the agent's values and preferences.

AI and trust calibration on the battlefield

The potential for machine-intelligence to provide new capabilities and enhance the performance of existing ones can be a significant factor for military operations, as long as the human operators have properly calibrated levels of trust in the systems being operated. As Hoff and Bashir observed, "just as it does in interpersonal relationships, trust plays a leading role in determining the willingness of humans to rely on automated systems in situations characterised by uncertainty."[59] The effective incorporation of autonomy, therefore, depends in part on individual traits, cultural backgrounds, and attitudes towards machine intelligence. In alliance or coalition operations, interoperability issues can arise if some member states have a well-established and well-calibrated relationship with human-machine teaming, whilst others do not. Trust calibration may not necessarily be transferable to personnel across different cultural backgrounds. Even within each state's military forces, however, issues of trust calibration will likely vary according to the tasks performed by machine-intelligence across the three categories mentioned above: data integration and analysis, autonomous weapons systems, and decision-support.

Trust calibration for AI data integration and analysis

For many military applications, the role performed by machine-intelligence has already been so fully integrated in the system architecture that it may not even be noticeable. Applications can include automated language translation tools, AI-steered frequency selection for communications equipment, the integration of sensor data to create a holistic view of the battlefield for platform operators, or an intelligent digital entity monitoring computer networks for signs of intrusion. For these types of functions, the AI is making "choices" and influencing the human operator's understanding of the situation, which in turn has an effect on cognition and the human decision-making that results. This use of machine intelligence fits more comfortably in the definition of an automated system. Issues of trust calibration are therefore more familiar and more thoroughly studied.

An immediate and obvious concern is the high level of dispositional, or initial, trust most operators are likely to grant these types of systems, perhaps even unaware the extent to which the AI is shaping the information environment. Proper trust calibration for military applications could involve human-machine interface-design features that both elicit trust, but provide adequate levels of transparency, particularly regarding the robustness of the data upon which the machine intelligence bases its conclusions. One study suggested that autonomous agents should have an ability to evaluate its own self-confidence, including uncertainties in its own knowledge base as well as uncertainties about its own state of operation and uncertainties about its reasoning processes.[60] Of course, this too would be subject to the same weaknesses as the decision-making process itself, but could add a useful corrective to human tendencies toward automation bias.

Another challenge for human operators who depend on machine-intelligence for situational awareness is temporal in nature. During a future conflict, the time available to make decisions may be severely truncated, incentivising a reliance on machine-cognition. When forced to act quickly, humans may choose not to analyse the self-confidence levels of an autonomous agent, or critically evaluate AI-generated data when making time-critical decisions. In certain domains, other sensors might provide a useful secondary source for verification, whereas humans in other situations may be completely reliant on information provided by machine intelligence. AI-enabled tools in the Cyber-domain represent a complex hybrid between data analysis and autonomous systems, as machine intelligence monitors networks to protect against potential intrusions, and enables offensive Cyber-weapons to analyse and circumvent an opponent's smart defences.[61]

Independent verification of such tools to ensure adequate situational awareness may be challenging, particularly given operational time constraints.

Trust calibration for autonomous systems

Interactions with autonomous systems in the physical world – whether it be a ground-based "packbot" system, an unmanned refuelling drone, an autonomous surface vessel, or an autonomous weapon system, involve the same issues as the algorithmic entities discussed above, but entail other unique and challenging aspects of human autonomous teaming. These systems represent a truer embodiment of autonomous agents filling a defined role within a team, and are often discussed in terms of human agent interaction (HAI). Therefore, the characteristics of successful interpersonal teaming have greater relevance, including strong communication, shared mental models regarding intentions and motivations, and an ability to act predictably and collaboratively.[62]

One study conducted under the auspices of the US Defence Department's Autonomy Research Pilot Initiative examined interactions between a military unit and its autonomous "packbot" squad member, finding that displaying data about the robot's intent and logic strengthened some of the basic foundational building blocks for trust, such as situational awareness and understanding.[63] This transparency can enhance learned trust as operators become more proficient and experienced with autonomous agents. A number of transparency models are possible, including communicating the agent's intentions and goal structures, or its understanding of the tasks, an analytical model that focuses on the agent's inner workings and algorithms, communicating the agent's understanding of the external environment, or a teamwork model that emphasises the division of labour within the team.[64]

Transparency is one potential design feature for enhancing human-autonomy teaming. Engineering details relating to the machine-interface can be influential to striking the proper balance between eliciting trust and encouraging over-trust. Natural language processing and synthetic speech has made significant strides, enabling conversational communication between humans and robots that improves transparency and trust.[65] Attributing human characteristics to autonomous agents is a natural psychological phenomenon that can enhance co-operation, but anthropomorphising can have negative effects including unfortunate emotional attachments to explosive ordnance disposal robots, or encouraging overtrust in autonomous agents due to human-like speech patterns.[66]

Dispositional trust may be most influential during the initial interactions between humans and physical autonomous agents. For example, one Australian study concluded that service members harbour deep-seated scepticism of autonomous weapon systems.[67] However, achieving proper trust calibration over time may depend primarily on situational and learned trust. The human judgement to rely on machine intelligence in high-risk situations; or leave the critical tasks to other humans even if that choice is sub-optimal, may ultimately be a highly personal one. As with human teaming, such decisions are often based on previous experience from similar situations, which suggests that comprehensive training exercises with autonomous agents can be an important component in trust calibration. Training with autonomous systems has been touted as a logical step to encourage trust in human autonomy teaming, with the added benefit of providing additional AI training data.[68] Roff and Danks caution that the context in which training occurs might also be consequential, and emphasise the variations between a low-risk environment such as basic training and more advanced exercises that simulated battlefield environments. Additionally, they suggest leveraging the transitive property of trust by creating an autonomous agent "liaison officer" within each unit that works more closely with the system to understand its logic, motivations, and processes. Trust calibration for the remaining members of the unit might then be more easily conveyed through the liaison officer, although this approach has its limitations as well.[69]

Trust calibration for operational decision support systems

The issues relating to effective human autonomy teaming discussed above will have an immediate impact at the sub-tactical and tactical levels, but deployment of autonomous systems on the battlefield may bring about adaptation at the operational level as well.[70] Greater numbers of autonomous platforms operating independently – along with tactical decision-making occurring at machine speeds – will pose challenges for human cognition and may become a limiting factor in disrupting an adversary's decision loops. Considering the threats an adversary can pose in multiple domains and the amount of information required to respond adequately and promptly, one US military leader concluded that "a twentieth century commander will not survive in that environment" without the assistance of machine intelligence to manage the data.[71] The use of machine-intelligence at the operational level is likely cumulative, incorporating the benefits and risks of trust discussed in the previous two sections and adding another layer of complexity.

Leveraging machine intelligence for decision-support at the operational level has clear parallels with data analysis at the tactical level, particularly the susceptibility to automation bias and tendencies to overlook the sometimes-subtle decision-making effects of AI. Furthermore, the potential addition of co-ordinated groups – perhaps even swarms – of autonomous weapons or platforms introduces new challenges to existing C2 procedures such as joint targeting that may themselves require automation in a potentially more fast-paced and dynamic environment. For operations planners, the element of human-machine trust becomes an additional factor for evaluating the readiness and efficacy of combat units. A decision-centric warfare concept that incorporates AI directly into command-and-control structures may be the most dramatic application of autonomy. An appreciation of the potential strategic implications stemming from missteps in tactical decision-making has become even more poignant with the advent of continuous news coverage and social media. An important part of human autonomy teaming in the military sphere involves the consideration of the autonomous agent's ability to act with an awareness of the conflict's strategic and political context, as well as within the legal framework of the international laws of armed conflict. This consideration becomes greatly amplified at the operational level, as AI-assisted information flows and autonomous control over groups of autonomous platforms combine with the consequences of autonomous agent actions at the tactical level. This is particularly concerning given the potential conflict escalation dynamics associated with machine-intelligence and autonomous systems due to compressed decision-making times, confirmation bias, and the clinical cost–benefit rewards system of machine learning.[72]

Trust is a phenomenon occurring in situations of uncertainty and risk. These are two aspects of operational planning and control that machine-intelligence can potentially mitigate with fewer personnel in harm's way and improved information processing leading to enhanced situational awareness. As noted in a recent article, AI for algorithmic warfare must remain flexible and reduce operational complexity, including an ability to "independently compose and adjudicate courses of action."[73] Trusting machine intelligence to act as the moral agent "in the loop" for planning and approving specific COAs involves an adequate level of comfort in allowing the autonomous agent to evaluate tactical decisions appropriately, which of itself involves some sort of machine-based "trust." As Davis argues, an AI-based decision-support architecture at the operational level "would act as a trusted agent, condensing the amount of information for which the commander was responsible."[74] Existing research suggests that operators overseeing, or managing, autonomous agents should be given as much situational data as possible, particularly since some studies suggest that situational awareness degrades as the number of autonomous agents increases.[75] For commanders managing autonomous agents as the human "on the loop," enhancing situational understanding has been shown to be more effective than simply providing options from which an operator can choose.[76]

Another issue that could emerge relating to trust and machine intelligence is the somewhat paradoxical nature of trust and tactical advantage. Existing research suggests that predictable behaviour within similar circumstances engenders trust, but this predictability can be a vulnerability on the battlefield if an adversary has similar data analysis tools and can predict algorithmic patterns. After only a few instances of observing the algorithmic tactics and behaviours of autonomous agents, their actions might be anticipated and thereby countered. To be sure, adaptations can be incorporated into the behaviour of the agents to refrain from repeating identical manoeuvres during aerial combat, for example, but this lack of predictability will make human-autonomy trust more challenging, however advantageous it may be in a tactical sense. The potential for adversarial interference with one's own training data, or algorithms, will also remain a concern and a justified reason for scepticism.[77]

Such tactical and operational level issues can create a number of strategic dilemmas. Time constraints may force human operators to relinquish important aspects of the decision-making process to autonomous agents in order to gain an advantage over an adversary. In virtual experiments conducted by the US Army, humans tended to micromanage their drone swarms and were consistently defeated by AI-controlled units due to slower human cognition.[78] Such a dynamic is likely to encourage operators to forego a sober evaluation of an autonomous agent's trustworthiness for the sake of expediency. The gradual reduction of human cognition in decision-making is likely to have unpredictable strategic implications relating to deterrence and escalation control at all levels, particularly when trust issues are involved. As James Johnson has argued, "uncertainty created by AI threats to strategic stability could be either the result of an adversary's exaggerated faith in its effectiveness or (and perhaps more concerning) the false belief that a particular AI capability is operationally effective when it is not."[79]

Conclusion

It seems clear that military applications for autonomous systems will continue to expand as the incentives for policy-makers appear to outweigh the risks. Machine intelligence represents a unique set of technologies that goes beyond incremental improvements to range, speed, or accuracy of weapons systems. Autonomous agents in each of the three categories can potentially alter the battlefield in unpredictable ways, which presents both danger and opportunity to military leaders looking for the most optimal means of integrating AI technology into existing and planned force structures. Given that warfare remains a human-centred endeavour, the importance of human autonomous teaming is likely to become an increasingly important aspect of military operations.

Research-based knowledge on aspects of trust in human-autonomy teaming is wide-ranging and comprehensive, but much of the empirical data naturally relates primarily towards the more automated processes side of the sliding scale from automation to autonomy. Given the anticipated functions of machine intelligence in a military context, much of this research nevertheless remains highly relevant – particularly regarding aspects of cultural differences related to dispositional trust or common phenomena such as automation bias. The challenges to proper trust calibration vary according to the type and category of application, and eliciting sufficient human trust in physical autonomous systems may be more challenging than integrated machine-learning software for ISR data analysis. Ultimately, it remains crucial that appropriate and calibrated levels of trust are achieved to best harness the potential of artificial intelligence.

Notes

1. See Eric Wesley and Robert Simpson, *Expanding the Battlefield: An Important Fundamental of Multi-Domain Operations* (Arlington, VA: Association of the United States Army, April 2021).
2. Martin Van Creveld, *Technology and War: From 2000 BC to the Present* (New York: Macmillan, 1991); *Max Boot, War Made New: Technology, Warfare and the Course of History – 1500 to Today* (New York: Gotham Books, 2006).
3. Steven I. Davis, 'Artificial Intelligence and the Operational Level of War', *Defense and Security Analysis* 38, no. 1 (2022).
4. Margarita Konaev, Tim Huang, and Husanjot Chahal, *Trusted Partners: Human-Machine Teaming and the Future of Military AI* (Washington, DC: Center for Security and Emerging Technology, February 2021), 11.
5. Nils J. Nilsson, *The Quest for Artificial Intelligence: A History of Ideas and Achievements* (Cambridge: Cambridge University Press, 2009).
6. Gideon Lewis-Kraus, 'The Great AI Awakening', *New York Times*, December 14, 2014.
7. Rob Toews, 'To Understand the Future of AI, Study its Past', *Forbes*, November 17, 2019.
8. Ian Sample, '"It's Able to Create Knowledge Itself": Google Unveils AI That Learns on Its Own', *The Guardian*, October 18, 2017.
9. Chris Baraniuk, 'How Google's Balloons Surprised Their Creator – BBC Future', *BBC News*, February 24, 2021; Miles O'Brian, 'How 3D Printing is Spurring Revolutionary Advances in Manufacturing and Design', *PBS Newshour*, March 21, 2018.
10. Khari Johnson, 'Nvidia Researchers Devise Method for Training GANs with Less Data', *VentureBeat* (blog), 7 December 2020.
11. Douglas Heaven, 'Why Deep-Learning AIs Are so Easy to Fool', *Nature* 574, no. 7777 (9 October 2019): 163–6.
12. Paul Allen and Mark Greaves, 'Paul Allen: The Singularity Isn't Near', *MIT Technology Review*, October 12, 2011; Ben Dickson, 'Is DeepMind's New Reinforcement Learning System a Step Toward General AI?', *VentureBeat* (blog), August 3, 2021; Julian Schrittwieser et al., 'Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model', *Nature* 588, no. 7839 (December 2020): 604–9.

13. Michael L. Littman et al., *Gathering Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence (AI100) 2021 Study Panel Report. Stanford University* (Stanford, CA: Stanford University, September 2021).
14. Peter Stone et al., *Artificial Intelligence and Life in 2030: The One Hundred Year Study on Artificial Intelligence* (Stanford: Stanford University, 2016).
15. Dickson, 'Is DeepMind's New Reinforcement Learning System a Step Toward General AI?'; James Somers, 'Is AI Riding a One-Trick Pony?', *MIT Technology Review*, September 29, 2017.
16. Toews, 'To Understand The Future of AI, Study Its Past'; Missy Cummings, 'Rethinking the Maturity of Artificial Intelligence in Safety-Critical Settings', *AI Magazine* 42, no. 1 (2021): 6–15.
17. Kaushik Roy, Akhilesh Jaiswal, and Priyadarshini Panda, 'Towards Spike-Based Machine Intelligence with Neuromorphic Computing', *Nature* 575, no. 7784 (November 2019): 607–17; Thomas Gabor et al., 'The Holy Grail of Quantum Artificial Intelligence: Major Challenges in Accelerating the Machine Learning Pipeline', in *Proceedings of the IEEE/ACM 42nd International Conference on Software Engineering Workshops* (ICSE '20: 42nd International Conference on Software Engineering, Seoul Republic of Korea: ACM, 2020), 456–61.
18. Christian Brose, 'The New Revolution in Military Affairs', *Foreign Affairs* 98, no. 3 (2019): 122–34; Cheryl Pellerin, 'Deputy Secretary: Third Offset Strategy Bolsters America's Military Deterrence', *Department of Defense News*, October 31, 2016.
19. European Defence Agency, 'Stronger Communication & Radar Systems with Help of AI', *EDA Website*, August 31, 2020.
20. Zachary Davis, 'Artificial Intelligence on the Battlefield: Implications for Deterrence and Surprise', *Prism* 8, no. 2 (2019): 114–31.
21. James Johnson, 'Deterrence in the Age of Artificial Intelligence and Autonomy: A Paradigm Shift in Nuclear Deterrence Theory and Practice?', *Defense and Security Analysis* 36, no. 4 (2019): 422–48.
22. Cheryl Pellerin, 'Project Maven to Deploy Computer Algorithms to War Zone by Year's End', *U.S. Department of Defense*, July 21, 2017.
23. Scott Seckel, 'AI Algorithm Trained to Predict What ISIL Forces Will Do in Different Situations', *Pys.org* website, September 21, 2015.
24. Brett Tingley, 'The Pentagon Is Experimenting With Using Artificial Intelligence To See Days In Advance', *The Drive*, July 30, 2021.
25. Andre Heider and Maria Beatrice Catarrasi, 'Future Unmanned System Technologies – Legal and Ethical Implications of Increasing Automation' (Kalkar Germany: Joint Air Power Competence Centre (JAPCC), 2016).
26. Michael C. Horowitz, Lauren Kahn, and Casey Mahoney, 'The Future of Military Applications of Artificial Intelligence: A Role for Confidence-Building Measures?', *Orbis* 64, no. 4 (1 January 2020): 528–43.
27. Heider and Catarrasi, 'Future Unmanned System Technologies – Legal and Ethical Implications of Increasing Automation'.
28. Michael C. Horowitz and Paul Scharre, *An Introduction to Autonomy in Weapon Systems* (Washington, DC: Center for New American Security, 2015); Enn Tyugu and D Branch, 'Artificial Intelligence in Cyber Defense', *3rd International Conference on Cyber Conflict* (2011).
29. Andrew Chuter, 'British Shell out Seed Funding for "Loyal Wingman" Combat Drone', *Defense News*, January 25, 2021; Steve Trimble and Lee Hudson, 'U.S. Air Force Plots Fleet Insertion Path For "Loyal Wingman"', *Aviation Week*, March 6, 2020.
30. Paul Scharre, *Robotics on the Battlefield Part II: The Coming Swarm* (Washington, DC: Center for New American Security, October 2014); Andrew Illanchinski, *AI, Robots, and Swarms* (Alexandria, VA: CNA, 2017); M. L. Cummings, 'Human Supervisory Control of Swarming Networks', *2nd Annual Swarming and Autonomous Intelligent Networked Systems Conference*, June 2004.

31. Davis, 'Artificial Intelligence and the Operational Level of War'.
32. Dr Ross Pigeau and Carol McCann, 'Re-Conceptualising Command and Control', *Canadian Military Journal* (Spring 2002): 53–64.
33. Bryan Clark, Dan Patt, and Harrison Schramm, *Mosaic Warfare – Exploiting Artificial Intelligence and Autonomous Systems to Implement Decision-Centric Operations* (Alexandria, VA: Center for Strategic and Budgetary Assessments, 2020), 35.
34. Davis, 'Artificial Intelligence and the Operational Level of War'.
35. North Atlantic Treaty Organisation, 'NATO 2030: United for a New Era', November 26, 2020.
36. Terri Moon Cronk, 'Hicks Announces New Artificial Intelligence Initiative', *U.S. Department of Defense*, June 22, 2021; Joan Johnson and Lorin Selby, 'Department of the Navy Science and Technology Strategy for Intelligent Autonomous Systems', July 2, 2021.
37. John D. Lee and Katrina A. See, 'Trust in Automation: Designing for Appropriate Reliance', *Human Factors* 46, no. 1 (March 2004): 50–80.
38. Lee and See, 'Trust in Automation', 53.
39. Ibid., 51.
40. Keng Siau and Weiyu Wang, 'Building Trust in Artificial Intelligence, Machine Learning, and Robotics', *Cutter Business Technology Journal* 31, no. 2 (March 2018): 47–53.
41. Lee and See, 'Trust in Automation', 59.
42. Ibid.
43. Kevin Anthony Hoff and Masooda Bashir, 'Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust', *Human Factors* 57, no. 3 (1 May 2015): 407–34, https://doi.org/10.1177/0018720814547570.
44. Hoff and Bashir, 'Trust in Automation'.
45. Ibid., 413.
46. Lee and See, 'Trust in Automation', 57.
47. Ibid., 58.
48. Hoff and Bashir, 'Trust in Automation', 418.
49. Ibid., 421.
50. Ibid., 424.
51. Kathleen L. Mosier, Ute Fischer, Barbara K. Burian, and Janeen A. Kochan, *Autonomous, Context-Sensitive, Task Management Systems and Decision Support Tools I: Human-Autonomy Teaming Fundamentals and State of the Art* (Moffett Field, CA: National Aeronautics and Space Administration, 2017).
52. Cummings, 'Human Supervisory Control of Swarming Networks'.
53. William Langewiesche, 'The Human Factor', *Vanity Fair*, October 2014.
54. Lee and See, 'Trust in Automation', 55.
55. Konaev, Huang, and Chahal, *Trusted Partners*, 16.
56. Bridget O'Neal, 'Zeolite CNT 3DS Printing Materials Uxsed in Load Bearing Structures', *3D Print.com website*, March 11, 2020.
57. Heather M. Roff and David Danks, '"Trust but Verify": The Difficulty of Trusting Autonomous Weapons Systems', *Journal of Military Ethics* 17, no. 1 (January 2018): 2–20.
58. Thomas O'Neill, Nathan J. McNeese, Amy Barron and Beu Schelble, 'Human–Autonomy Teaming: A Review and Analysis of the Empirical Literature', *Human Factors: The Journal of the Human Factors and Ergonomics Society* 64, no. 5 (2022): 904–38.
59. Hoff and Bashir, 'Trust in Automation', 407.
60. Matthew Aitken, Nisar Ahmed, Dale Lawrence, Brian Argrow, and Eric Frew, 'Assurances and Machine Self-Confidence for Enhanced Trust in Autonomous Systems', paper presented at the RSS Workshop on Social Trust in Autonomous Systems (2016).
61. Christopher Whyte, 'Problems of Poison: New Paradigms and "Agreed" Competition in the Era of AI-Enabled Cyber Operations', paper presented at the 12th International Conference on Cyber Conflict (2020).

62. Research and Technology Organization, *Human-Autonomy Teaming: Supporting Dynamically Adjustable Collaboration* (Neuilly-sur-Seine: NATO, Research and Technology Organisation, 2020), 2.11.
63. Konaev, Huang, and Chahal, *Trusted Partners*, 21.
64. Joseph B. Lyons, Matthew Clark, Alan Wagner and Matthew Schuelke, 'Certifiable Trust in Autonomous Systems: Making the Intractable Tangible', *AI Magazine* 38, no. 3 (2 October 2017): 37–49.
65. 'Army Research Enables Conversational AI between Soldiers, Robots', www.army.mil (accessed June 18, 2021).
66. Lee and See, 'Trust in Automation'; Massimiliano L. Cappuccio, Jai C. Galliott, and Eduardo B. Sandoval, 'Saving Private Robot: Risks and Advantages of Anthropomorphism in Agent-Soldier Teams', *International Journal of Social Robotics*, February 12, 2021.
67. Jai Galliott, 'Risks and Benefits of Autonomous Weapon Systems: Perceptions among Future Australian Defence Force Officers', *Journal of Indo-Pacific Affairs* (Winter 2020): 17–34.
68. Sydney Freedberg, 'Learn By Losing: Give AI To OPFOR First', *Breaking Defense*, November 16, 2020; Sydney Freedberg, 'EBullet Brings Richer Realism To Army Training; No More Laser Tag', *Breaking Defense*, November 30, 2020.
69. Roff and Danks, 'Trust but Verify'.
70. Brose, 'The New Revolution'; Clark, Patt, and Schramm, 'Mosaic Warfare'; John Allen and Amir Husain, 'On Hyperwar', *Proceedings*, July 2017.
71. '"A 20th Century Commander Will Not Survive": Why the Military Needs AI', *Breaking Defense*, January 12, 2021.
72. Johnson, 'Deterrence in the Age of Artificial Intelligence and Autonomy', 435.
73. Courtney Crosby, 'Operationalizing Artificial Intelligence for Algorithmic Warfare', *Military Review* (July–August 2020): 42–51.
74. Davis, 'Artificial Intelligence and the Operational Level of War', 16.
75. Jessie Y. C. Chen et al., 'Situation Awareness-Based Agent Transparency and Human-Autonomy Teaming Effectiveness', *Theoretical Issues in Ergonomics Science* 19, no. 3 (4 May 2018): 259–82.
76. Mosier et al., 'Autonomous, Context-Sensitive, Task Management Systems and Decision Support Tools I'.
77. M. A. Thomas, 'Time for a Counter-AI Strategy', *Strategic Studies Quarterly* (Spring 2020): 3–8; Erik Lin-Greenberg, 'Allies and Artificial Intelligence: Obstacles to Operations and Decision-Making', *Texas National Security Review* 3, no. 2 (Spring 2020): 57–76.
78. Sydney Freedberg, 'Let Your Robots off the Leash – or Lose: AI experts', *Breaking Defense*, November 19, 2020.
79. James Johnson, 'Artificial Intelligence & Future Warfare: Implications for International Security', *Defense and Security Analysis* 35, no. 2: 147–69.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributor

Dr. Michael Mayer is a senior advisor at the Norwegian Defence Research Establishment (FFI), focusing primarily on emerging technologies and their potential effects on future military operations.