Machine Intelligence and Trust: the Implications of AI for Joint Operations

Michael Mayer
Norwegian Defence Research Establishment
Instituttveien 20
2027 Kjeller
NORWAY

michael-john.mayer@ffi.no

Key words: Artificial intelligence, machine intelligence, trust, human autonomy teaming, joint operations

ABSTRACT

Advances in the field of artificial intelligence continue to expand the range of potential military applications for this group of technologies. This paper explores the crucial role of trust in human-machine teaming for joint operations and the potential implications of relying on AI to supplement human cognition. Trusting machine intelligence will be a crucial component in future operations if AI is relied upon to accurately process sensor data, operate autonomous systems and platforms, or provide advantageous decision support through proposed operational concepts such as decision-centric warfare that envision a central command and control role for machine intelligence. Given these technical and doctrinal developments, the concept of trust becomes highly relevant for the use of machine intelligence in military operations at the tactical and operational levels and correctly calibrated trust levels are fundamental for safe and effective operations. After a brief review of recent advances in machine intelligence and an exploration of the concept of trust, the paper outlines current and potential applications for AI on the battlefield, and challenges stemming from either insufficient or unjustifiably high levels of trust.

1.0 INTRODUCTION

Throughout history, technology has expanded the domains of armed conflict, the tempo of tactical engagements, the geographic breadth of the battlefield, and the means by which commanders communicate with their forces. Technological innovations – both military and civilian – have altered how militaries fight and how states plan and conduct those conflicts. In the 21st century, few advances have so far garnered as much attention as the group of technologies known collectively as artificial intelligence (AI). AI is poised to usher in a new era in which machine intelligence and autonomy are generating distinctly new concepts for the planning and execution of military operations. Algorithmic warfare may lead to something unique: systems that augment or even displace human decision-making processes, and at speeds that may exceed the cognitive capacity of human planners.

The integration of emerging technologies raises any number of fundamental organisational and ethical issues that deserve attention. Using qualitative social science methodology, this paper will focus on one important aspect of human-autonomy teaming (HAT): encouraging the appropriate levels of trust in machine intelligence. A vast body of academic literature exists that focuses on trust in automation or robotics, but less work is available regarding specific military applications. What challenges and opportunities for trust calibration when AI is operationally deployed in joint operations? After a brief review of AI and an overview of the likely applications of machine intelligence on the battlefield, the paper explores the concept of trust and trust calibration before analysing the pitfalls and potential solutions for encouraging appropriate levels of trust.

2.0 ADVANCES IN AI

For decades, humans have been fascinated with the possibility of imbuing 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" [1, p. 13]. Two broad approaches to AI emerged during the earliest days of the digital age. A top-down *expert system* approach used complex preprogramed 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 uses a bottom-up *machine learning* approach that modelled the way humans learn by detecting patterns within data. Neural networks are a form of machine learning modelled after the human brain that are able to identify complex patterns by using multiple (and therefore "deep") layers of artificial neurons, are fundamental to the technique known as "deep learning" [2]. Through its ability to find relationships within data sets, such techniques are also termed "connectionist" [3].

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 an ability to separate the learning from the data set upon which it trains and therefore can be applied to other problems. Whereas rules-based algorithms can perform exceedingly well at narrowly defined tasks, deep learning approaches are able to rapidly find patterns and in effect teach themselves for applications for which "brute force" expert-system computational approaches are ineffective [4]. A number of recent AI advances demonstrate an ability to mimic creativity, generating effective approaches to problem solving that can appear counterintuitive to humans [5].

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 to human cognition, 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 fall short when attempting inductive reasoning where it must make general observations about a data set or an environment. Massive training sets of data are still necessary for most machine learning, even though new approaches (including generative adversarial networks (GAN) and "less than one-shot" or LO-shot learning) requiring very small datasets are emerging [6]. Image recognition algorithms are easily confused, and cannot immediately or intuitively understand 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 [7].

While AI technologies continue to make significant progress in becoming more adaptable, anything approaching human-like artificial general intelligence remains elusive [8]–[10]. 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 normalization of its very substantial progress. As 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" [11, p. 12]. Although symbolic AI and the various forms of machine learning have comprised the bulk of recent progress in the field, perhaps with the exception of attempts to fuse both approaches, the future remains uncertain [3], [12]. Some speculate that the progress resulting from machine learning techniques may plateau, while others remain optimistic [9], [13]. Related technological advances, such as computer chip design in the short term and quantum computing in the long term, may influence the pace of further progress [14], [15].

3.0 ARTIFICIAL INTELLIGENCE IN JOINT OPERATIONS

For many military applications, however, narrow uses of AI are more than adequate. 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 afforded by AI are part of a dramatically different future operating environment envisioned by analysts such as Christian Brose and former defence officials such as Robert Work [16], [17]. If these predictions regarding the effects of artificial intelligence come to fruition, they will have wide-ranging implications for the planning and implementation of joint operations. Existing and near-future applications can be divided into three categories: data integration and analysis, autonomous systems, and decision support software.

3.1 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. In a military context, wireless communication and radars can leverage machine-learning algorithms for optimal use of the electromagnetic spectrum [18]. For unmanned or remotely piloted aircraft, onboard algorithms allow sensors to independently conduct preliminary data analysis and thereby reduce bandwidth requirements. Algorithms are already useful for analysing sensor data across a range of systems and platforms.

In addition to these integrated applications, the conscious and active use of AI for data analysis extends to intelligence, surveillance, and reconnaissance (ISR) efforts. 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 that can enhance situational awareness [19]. 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 [20]. Some 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 [21].

3.2 Autonomous systems

A second category of AI applications encompasses a range of autonomous systems. Autonomy is a term that defies precise and concise definitions. 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 that could determine its own course of action, deliberate decisions not restrained by pre-programmed responses, an ability to learn and compile "experience", and no longer completely predictable in its actions [22]. Paul Scharre and Michael Horowitz described three dimensions of autonomy in a 2015 paper: the *human-machine command and control relationship* simplified by determining whether a human is "in", "on" or "out of" the decision-making loop; the *complexity* and abilities of the machine or system; and the *type of function being automated* [23].

Within the context of AI, it is worth noting that the distinction between automated and autonomous becomes blurred as 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 [22]. Autonomy therefore describes a sliding scale of independent machine functionality along a number of variables, including level of human-machine interaction, an ability to independently sense and adapt to changing contexts, and decision-making abilities to accomplish some set of predetermined goals and continuously learn and improve from those decisions.

A broader definition of autonomy might include 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 [23], [24]. 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, airborne 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 transport cargo or perform refuelling duties. Concepts known colloquially as "loyal wingman" programs seek to develop uninhabited aircraft that can operate alongside piloted craft, thereby offering networked sensors, additional munitions and expanded tactical options. Autonomous ships will soon give maritime commanders a similar capability at sea, and ground-based systems are also currently under development [25], [26]. Advancements in size, weight and power characteristics for data processing, novel manufacturing processes and AI appear likely to enable large numbers of small, unmanned systems that can be controlled and coordinated in swarm formations using artificial intelligence [27]. With lower unit costs compared with piloted aircraft and uploadable navigation, battle management, and targeting software, autonomous systems are poised to dramatically increase the number of platforms on the battlefield [28], [29].

3.3 Decision support and decision-centric warfare

Military commanders currently rely on machine intelligence in their decisionmaking 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 assists situational awareness and provides options for warfighters. Future decisionmaking aids may bring about further developments. 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, the now-common term emerged during the nascent information technology age of the 1960s to distinguish the authority and responsibility of *command* from the processes and framework that create the necessary conditions for the commander to exert *control* over the implementation and execution of operations [30]. Although observation of tactical engagements by higher-level commanders and political leaders has become more commonplace, it may be that the operational level may be the most appropriate for having humans "on the loop" if autonomous systems are deployed. Even with fleets of self-synchronizing autonomous surface vessels or aerial systems, the need to coordinate the broader operational effort will remain human-centric. If that is the case, however, operational planning and coordination may need assistance from AI 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 coordinate a network of disaggregated forces. The concept proposes a hybrid C2 configuration that utilizes human command and machine control, whereby commanders choose tasks in need of completion from a set of recommended courses of action (COA) and most advantageous manned and unmanned force components available from the AI-enabled decision support system [31, p. 35]. Concepts integrating AI and autonomous systems in this fashion are a logical – albeit ambitious – progression given the perceived advantages of rapid machine-based decision-making, particularly if a connected battlespace allows for data fusion among a disparate but linked network. The sheer volume of available information is such that machine intelligence will be required to understand and act upon that data in an advantageous manner.

4.0 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. An expert group report under the rubric of the Nato 2030 initiative recommended that the 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"[32]. Official statements [33] 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 [34]. As machine intelligence becomes more capable of increasingly complex cognitive functions and an ability to operate independently, humans will need to view AI and autonomous systems as partners just as much as tools. With any partnership, trust is a crucial to effective cooperation.

4.1 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" [35, p. 53]. 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" [35, p. 53]. 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" [35, p. 51]. 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. As Keng Siau and Weiyu Wang [36, p. 47] noted, trust is dynamic and is typically built gradually via two-way interaction, but can also be strongly affected by initial impressions. 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* [35, p. 59].

4.2 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 be termed *process* information that reveal 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 [35, p. 59].

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 can eventually be replaced by the more durable qualities of predictability and dependability [37, p. 411]. In a comprehensive 2015 survey of scholarly articles on trust and automation, Kevin Hoff and Masooda Bashir [37] 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.

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 [37, p. 413]. For Nato, 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 given the alliance's multinational character [35, p. 57]. Factors such as attitudes towards power and authority or balancing between individual or collective interests can 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 [35, p. 58].

Along with the initial impact from dispositional trust, situational trust is the model's second component with 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 [37, p. 418].

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 pre-existing knowledge of automation, whether it comes via previous experience from other automated systems or based on the reputation of the automation in question. Their expectations regarding automation and knowledge regarding its performance 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 [37, p. 421]. 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. After the initial levels of trust garnered from prior experience or the design features baked into the system, the operator continually and dynamically gauges trust based on factors such as reliability, predictability, system utility, and when and how errors occur, including how the operator is alerted to them [37, p. 424].

4.3 Trust calibration and misalignment

Significant effort has been devoted to creating trust between humans and automated systems, but experience has demonstrated that excessive trust can also be problematic. Among the most common tendencies of 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, where operators fail to respond when automation malfunctions or make incorrect decisions to follow automated recommendations [38]. One study suggests 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 [27].

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 argued in a detailed 2014 article about the crash that the crew, so accustomed to relying on automated flying aids, were unable to comprehend what was actually happening to the aircraft when a faulty airspeed indicator led to a string of faulty decisions

and an ultimate failure to make the proper adjustments. Langewiesche's succinctly summarized 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" [39].

Rather than focusing ways to increase human trust of automated systems, developers often strive to elicit appropriate or calibrated levels of trust that correlate to the system's capabilities. 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 [35, p. 55]. 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 overcorrect their trust levels and lower their expectations to a level below the capabilities of the system – thereby transitioning directly from overtrust to distrust [40, p. 16].

4.4 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 in the paper distinguished between rigid, pre-determined, and predictable automated tasks versus unrestrained, dynamic, and unpredictable autonomy. One recent survey article on human autonomy teaming by Thomas O'Neill et al. 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" [41, p. 4].

In practice, therefore, this distinction is more graduated and perhaps better understood as a sliding scale with automated functionality on one end and autonomous functionality on the other. Even this type of graduated approach has only limited utility due the fact that technological progress and human expectations naturally will consider autonomous functionality to be automated as 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, such as an autonomous AI-empowered cyber defence that acts upon threats in an unpredictable and unscripted fashion, but the network defences are considered automated.

In a thought-provoking article on trusting 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" [42]. Similarly, 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" [41, p. 4]. 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 particular 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.

5.0 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 joint 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" [37, p. 407]. For the Alliance, this trust has an additional interoperability dimension that further complicates trust calibration. Most existing weapons systems employed across Nato are of a similar character and art, despite dissimilar characteristics and manufacturers. With the introduction of autonomous agents with which member states have established a certain comfort level, that trust may not necessarily be transferable to personnel from different cultural backgrounds and attitudes toward machine intelligence. Even within each state's military forces, however, issues of trust calibration will likely vary according the tasks performed machine intelligence across the three categories mentioned above: data integration and analysis, autonomous weapons systems, and decision support.

5.1 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 decisionmaking that result. 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.

Perhaps the most immediate and obvious concern with this type of application 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 would 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 [43]. Of course, this too would be subject to the same weaknesses as the decision process itself, but could add a useful corrective to human tendencies toward automation bias.

5.2 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 entails other unique and challenging aspects of human autonomous teaming. These systems represent a truer embodiment of autonomous agents with 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 [44, p. 2.11].

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 foundations for trust such as situational awareness and understanding [40, p. 21]. This transparency can enhance learned trust as operators become more proficient and experienced with autonomous agents. A number of transparency models are possible, including communication 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 [45].

Transparency is one potential design feature for enhancing human-autonomy teaming. Any number of other engineering details relating to the interface can be influential, but may also be challenging strike a 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 [46]. Attributing autonomous agents with human characteristics is a natural psychological phenomenon that can enhance cooperation, but such anthropomorphising can have negative effects, including creating unfortunate emotional attachments to explosive ordnance disposal robots or encouraging overtrust in autonomous agents due to human-like speech patterns [35], [47].

Dispositional trust may be most influential during the initial interactions between humans and physical autonomous agents, and there is evidence that service members are sceptical to autonomous weapons [48]. However, achieving proper trust calibration over time may be most dependent on situational and learned trust. The human judgement to either rely on machine intelligence in high-risk situations or leave the critical tasks to other humans even if that choice is suboptimal may ultimately be a highly personal one. As with human teaming, such decisions are often based in previous experience from similar situations, suggesting that comprehensive training exercises with autonomous agents will 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 [49], [50]. Roff and Danks submit that the context in which training occurs might also be consequential, emphasising the variations between a low-risk environment such as basic training and more advanced exercises that simulated battlefield environments. Additional, they suggest leveraging the transitive property of trust by creating an autonomous agent “liason 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 [42].

5.3 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 some warn that deployment of autonomous systems on the battlefield may bring about adaptation at the operational level as well [16], [31], [51]. Greater numbers of autonomous platforms operating independently – along with tactical decisionmaking 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 20th century commander will not survive in that environment” without the assistance of machine intelligence that manages that data [52]. This use of machine intelligence incorporates the benefits and risks of trust discussed in the previous two sections and adds yet 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 coordinated 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 joint operations planners, the element of trust becomes an additional factor for evaluating the readiness and efficacy of combat units. Joint operations will likely be more complex with the influx of autonomous agents even without potential concepts such as decision-centric algorithmic warfare.

Of the incremental technological developments ranging from increased autonomy in sensor data analysis, a shift from automated to autonomous operation for certain platforms, or greater numbers of autonomous units on the battlefield, decision-centric warfare concepts that incorporate AI directly into command and control structures may be the most dramatic. The existing awareness of the potential strategic implications of tactical decisionmaking 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 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.

Trust is a phenomenon occurring in situations of uncertainty and risk, two aspects of operational planning and control that machine intelligence can potentially mitigate with both 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" [53, p. 48]. 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 in itself involves some sort of machine-based "trust". 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 [54]. For commanders managing autonomous agents as the human "on the loop", providing situational understanding has been shown to be more effective than simply providing options from which an operator can choose [38].

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 given 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 [55].

6.0 CONCLUSIONS

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 to the more automated processes on the sliding scale from automation to autonomy. Given the likely functions of machine intelligence in joint operations, much of this research remains highly relevant – particularly 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 and promise of artificial intelligence.

7.0 REFERENCES

[1] N. J. Nilsson, *The Quest for Artificial Intelligence: A History of Ideas and Achievements*. Cambridge: Cambridge University Press, 2009. doi: 10.1017/CBO9780511819346.

[2] G. Lewis-Kraus, "The Great AI Awakening," *New York Times*, New York, Dec. 14, 2014. [Online]. Available: https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html

[3] R. Toews, "To Understand The Future of AI, Study Its Past," *Forbes*, Nov. 17, 2019. https://www.forbes.com/sites/robtoews/2019/11/17/to-understand-the-future-of-ai-study-its-past/ (accessed Aug. 11, 2021).

[4] I. Sample, "'It's able to create knowledge itself': Google unveils AI that learns on its own," *the Guardian*, Oct. 18, 2017. http://www.theguardian.com/science/2017/oct/18/its-able-to-create-knowledge-itself-google-unveils-ai-learns-all-on-its-own (accessed Aug. 10, 2021).

[5] C. Baraniuk, "How Google's balloons surprised their creator - BBC Future," *BBC News*, Feb. 24, 2021. Accessed: Jun. 18, 2021. [Online]. Available: https://www.bbc.com/future/article/20210222-how-googles-hot-air-balloon-surprised-its-creators?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter_axiosfutureofwork&stream=future

[6] K. Hao, "A radical new technique lets AI learn with practically no data," *MIT Technology Review*, Oct. 16, 2020. Accessed: Sep. 09, 2021. [Online]. Available: https://www.technologyreview.com/2020/10/16/1010566/ai-machine-learning-with-tiny-data/

[7] D. Heaven, "Why deep-learning AIs are so easy to fool," *Nature*, vol. 574, no. 7777, pp. 163–166, Oct. 2019, doi: 10.1038/d41586-019-03013-5.

[8] P. Allen and M. Greaves, "Paul Allen: The Singularity Isn't Near," *MIT Technology Review*, Oct. 12, 2011. https://www.technologyreview.com/2011/10/12/190773/paul-allen-the-singularity-isnt-near/ (accessed Aug. 10, 2021).

[9] B. Dickson, "Is DeepMind's new reinforcement learning system a step toward general AI?," *VentureBeat*, Aug. 03, 2021. https://venturebeat.com/2021/08/03/is-deepminds-new-reinforcement-learning-system-a-step-toward-general-ai/ (accessed Aug. 06, 2021).

[10] J. Schrittwieser *et al.*, "Mastering Atari, Go, chess and shogi by planning with a learned model," *Nature*, vol. 588, no. 7839, pp. 604–609, Dec. 2020, doi: 10.1038/s41586-020-03051-4.

[11] Stanford University, "Artificial Intelligence and Life in 2030," Stanford University, Stanford, 2016. [Online]. Available: https://ai100.stanford.edu/sites/g/files/sbiybj9861/f/ai100report10032016fnl_singles.pdf

[12] M. Cummings, "Rethinking the maturity of artificial intelligence in safety-critical settings," *AI Mag.*, vol. 42, no. 1, pp. 6–15, 2021.

[13] J. Somers, "Is AI Riding a One-Trick Pony?," *MIT Technology Review*, Sep. 29, 2017. https://www.technologyreview.com/2017/09/29/67852/is-ai-riding-a-one-trick-pony/ (accessed Aug. 10, 2021).

[14] K. Roy, A. Jaiswal, and P. Panda, "Towards spike-based machine intelligence with neuromorphic

computing," *Nature*, vol. 575, no. 7784, pp. 607–617, Nov. 2019, doi: 10.1038/s41586-019-1677-2.

[15] T. 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*, Seoul Republic of Korea, Jun. 2020, pp. 456–461. doi: 10.1145/3387940.3391469.

[16] C. Brose, "The new revolution in military affairs," *Foreign Aff.*, vol. 98, no. 3, pp. 122–134, 2019.

[17] C. Pellerin, "Deputy Secretary: Third Offset Strategy Bolsters America's Military Deterrence > U.S. Department of Defense > Defense Department News," *Department of Defense News*, Arlington VA, Oct. 31, 2016. Accessed: Aug. 10, 2021. [Online]. Available: https://www.defense.gov/Explore/News/Article/Article/991434/deputy-secretary-third-offset-strategy-bolsters-americas-military-deterrence/

[18] European Defence Agency, "Stronger communication & radar systems with help of AI." https://eda.europa.eu/news-and-events/news/2020/08/31/stronger-communication-radar-systems-with-help-of-ai (accessed Sep. 10, 2021).

[19] C. Pellerin, "Project Maven to Deploy Computer Algorithms to War Zone by Year's End," *U.S. Department of Defense*, Jul. 21, 2017. https://www.defense.gov/Explore/News/Article/Article/1254719/project-maven-to-deploy-computer-algorithms-to-war-zone-by-years-end/ (accessed Aug. 13, 2021).

[20] S. Seckel, "AI algorithm trained to predict what ISIL forces will do in different situations," Sep. 21, 2015. https://phys.org/news/2015-09-ai-algorithm-isil-situations.html (accessed Aug. 13, 2021).

[21] B. Tingley, "The Pentagon Is Experimenting With Using Artificial Intelligence To 'See Days In Advance,'" *The Drive*, Jul. 30, 2021. https://www.thedrive.com/the-war-zone/41771/the-pentagon-is-experimenting-with-using-artificial-intelligence-to-see-days-in-advance (accessed Aug. 02, 2021).

[22] A. Heider and M. B. Catarrasi, "Future Unmanned System Technologies - Legal and Ethical Implications of Increasing Automation," Joint Air Power Competence Centre (JAPCC), Kalkar Germany, 2016.

[23] M. C. Horowitz and P. Scharre, "An Introduction to Autonomy in Weapon Systems," Center for New American Security (CNAS), Washington, D.C., 2015.

[24] E. Tyugu and D. Branch, "Artificial intelligence in cyber defense." 2011 3rd International Conference on Cyber Conflict, 2011.

[25] A. Chuter, "British shell out seed funding for 'loyal wingman' combat drone," *Defense News*, Jan. 25, 2021. https://www.defensenews.com/global/europe/2021/01/25/british-shell-out-seed-funding-for-loyal-wingman-combat-drone/ (accessed Jun. 21, 2021).

[26] S. Trimble and L. Hudson, "U.S. Air Force Plots Fleet Insertion Path For 'Loyal Wingman,'" *Aviation Week*, Mar. 06, 2020.

[27] M. L. Cummings, "Human Supervisory Control of Swarming Networks." 2nd annual swarming: autonomous intelligent networked systems conference (pp. 1-9)., Jun. 2004.

[28] P. Scharre, "Robotics on the Battlefield Part II: The coming swarm," Center for New American

Security (CNAS), Washington DC, Oct. 2014.

[29] A. Illanchinski, *AI, Robots, and Swarms*. Alexandria, VA: CNA, 2017. [Online]. Available: https://www.cna.org/cna_files/pdf/DRM-2017-U-014796-Final.pdf

[30] D. R. Pigeau and C. McCann, "Re-conceptualising command and control," *Can. Mil. J.*, vol. Spring 2002, pp. 53–64, 2002.

[31] B. Clark, D. Patt, and H. Schramm, "Mosaic Warfare - Exploiting artificial intelligence and autonomous systems to implement decision-centric operations," Center for Strategic and Budgetary Assessments, Alexandria, VA, 2020.

[32] NATO, "NATO 2030: United for a New Era," Nov. 2020. [Online]. Available: https://www.nato.int/nato2030/independent-group/

[33] T. M. Cronk, "Hicks Announces New Artificial Intelligence Initiative," *U.S. Department of Defense*, Jun. 22, 2021. https://www.defense.gov/Explore/News/Article/Article/2667212/hicks-announces-new-artificial-intelligence-initiative/ (accessed Aug. 19, 2021).

[34] J. Johnson and L. Selby, "Department of the Navy Science and Technology Strategy for Intelligent Autonomous Systems," Jul. 2021. [Online]. Available: https://www.nationaldefensemagazine.org/-/media/sites/magazine/2021-dist-a-don-st-strategy-for-intelligent-autonomous-systems-2-jul-2021.ashx

[35] J. D. Lee and K. A. See, "Trust in Automation: Designing for Appropriate Reliance," *Hum. Factors*, vol. 46, no. 1, pp. 50–80, Mar. 2004, doi: 10.1518/hfes.46.1.50_30392.

[36] K. Siau and W. Wang, "Building Trust in Artificial Intelligence, Machine Learning, and Robotics," *Cut. Bus. Technol. J.*, vol. 31, no. 2, pp. 47–53, Mar. 2018.

[37] K. A. Hoff and M. Bashir, "Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust," *Hum. Factors*, vol. 57, no. 3, pp. 407–434, May 2015, doi: 10.1177/0018720814547570.

[38] K. L. Mosier, U. Fischer, B. K. Burian, and J. A. Kochan, "Autonomous, Context-Sensitive, Task Management Systems and Decision Support Tools I: Human-Autonomy Teaming Fundamentals and State of the Art," 2017, doi: 10.13140/RG.2.2.25859.60966.

[39] W. Langewiesche, "The Human Factor," *Vanity Fair*, Oct. 2014. https://archive.vanityfair.com/article/2014/10/the-human-factor (accessed Sep. 03, 2021).

[40] M. Konaev, T. Huang, and H. Chahal, "Trusted partners: human-machine teaming and the future of military AI," Center for Security and Emerging Technology, Washington DC, Feb. 2021.

[41] T. O'Neill, N. McNeese, A. Barron, and B. Schelble, "Human–Autonomy Teaming: A Review and Analysis of the Empirical Literature," *Hum. Factors J. Hum. Factors Ergon. Soc.*, p. 001872082096086, Oct. 2020, doi: 10.1177/0018720820960865.

[42] H. M. Roff and D. Danks, "'Trust but Verify': The Difficulty of Trusting Autonomous Weapons Systems," *J. Mil. Ethics*, vol. 17, no. 1, pp. 2–20, Jan. 2018, doi: 10.1080/15027570.2018.1481907.

[43] M. Aitken, N. Ahmed, D. Lawrence, B. Argrow, and E. Frew, "Assurances and Machine Self-Confidence for Enhanced Trust in Autonomous Systems." RSS 2016 Workshop on Social Trust in

Autonomous Systems. qav.comlab.ox.ac.uk., 2016.

[44] Research and Technology Organization, *Human-Autonomy Teaming: Supporting Dynamically Adjustable Collaboration.* Neuilly-sur-Seine: NATO, Research and Technology Organisation, 2020. Accessed: Jul. 29, 2021. [Online]. Available: https://bit.ly/2ZH0aRW

[45] J. B. Lyons, M. A. Clark, A. R. Wagner, and M. J. Schuelke, "Certifiable Trust in Autonomous Systems: Making the Intractable Tangible," *AI Mag.*, vol. 38, no. 3, pp. 37–49, Oct. 2017, doi: 10.1609/aimag.v38i3.2717.

[46] "Army research enables conversational AI between Soldiers, robots," *www.army.mil*. https://www.army.mil/article/237580/army_research_enables_conversational_ai_between_soldiers_robots (accessed Jun. 18, 2021).

[47] M. L. Cappuccio, J. C. Galliott, and E. B. Sandoval, "Saving Private Robot: Risks and Advantages of Anthropomorphism in Agent-Soldier Teams," *Int. J. Soc. Robot.*, Feb. 2021, doi: 10.1007/s12369-021-00755-z.

[48] J. Galliott, "Risks and Benefits of Autonomous Weapon Systems: Perceptions among Future Australian Defence Force Officers," *J. Indo-Pac. Aff.*, vol. Winter 2020, pp. 17–34, 2020.

[49] S. Freedberg, "Learn By Losing: Give AI To OPFOR First - Breaking Defense Breaking Defense - Defense industry news, analysis and commentary," *Breaking Defense*, Nov. 16, 2020. Accessed: Jun. 18, 2021. [Online]. Available: https://breakingdefense.com/2020/11/learn-by-losing-give-ai-to-opfor-first/

[50] S. Freedberg, "eBullet Brings Richer Realism To Army Training; No More Laser Tag," *Breaking Defense*, Nov. 30, 2020. Accessed: Jun. 21, 2021. [Online]. Available: https://breakingdefense.com/2020/11/ebullet-brings-richer-realism-to-army-training-no-more-laser-tag/

[51] J. Allen and A. Husain, "On Hyperwar," *Proceedings*, Jul. 2017, Accessed: Sep. 07, 2021. [Online]. Available: https://www.usni.org/magazines/proceedings/2017/july/hyperwar

[52] "'A 20th Century Commander Will Not Survive': Why The Military Needs AI - Breaking Defense Breaking Defense - Defense industry news, analysis and commentary." https://breakingdefense.com/2021/01/a-20th-century-commander-will-not-survive-why-the-military-needs-ai/ (accessed Jun. 18, 2021).

[53] C. Crosby, "Operationalizing Artificial Intelligence for Algorithmic Warfare," *Mil. Rev.*, p. 10, 2020.

[54] J. Y. C. Chen, S. G. Lakhmani, K. Stowers, A. R. Selkowitz, J. L. Wright, and M. Barnes, "Situation awareness-based agent transparency and human-autonomy teaming effectiveness," *Theor. Issues Ergon. Sci.*, vol. 19, no. 3, pp. 259–282, May 2018, doi: 10.1080/1463922X.2017.1315750.

[55] M. A. Thomas, "Time for a Counter-AI Strategy," *Strateg. Stud. Q.*, vol. Spring 2020, pp. 3–8, 2020.