CHAPTER 5

Artificial Intelligence in Military Decision-Making

Avoiding Ethical and Strategic Perils with an Option-Generator Model

Shannon E. French and Lisa N. Lindsay

There is little doubt that artificial intelligence (AI) will become widely deployed in military decision-making. There is still time, however, to determine how to do it right, in both the ethical and practical sense. There are clear perils to avoid when integrating AI into military decision-making, but also some potentially promising opportunities. Above all, human decision-making should not be ceded to machines, for a number of reasons that we will present. However, there are ways in which AI could help humans make better decisions in military contexts. We will argue that the most ethical and valuable role for AI in military decision-making is as an option generator.

1 Automation Bias

The growing push toward integrating AI into military decision-making is fueled in part by the widely held position that AI technology will decrease errors in military decision-making and lead to fewer overall deaths during warfare. The errors in question include both practical and moral failures. Practical failures can result from inaccurate or incomplete information, faulty analysis, inadequate training, and a host of other factors, while moral failures tend to be produced by a toxic combination of psychological stressors, character flaws, poor leadership, and a lack of discipline. Defending the use of robots and other automated systems in the military, Ron Arkin has essentially argued that humans are too emotionally vulnerable to be trusted to do the right thing in combat conditions. Citing surveys in which military personnel admit to unethical views about the importance (or lack thereof) of obeying the laws of war, Arkin asserts that humans are too often overcome by intense feelings such as rage and fear (or terror) that effectively hijack their brains and can lead even to the perpetration of war crimes. In his book *Governing Lethal Behavior in Autonomous Robots*, Arkin avers that, '… it seems unrealistic to expect normal human beings by their very nature to adhere to the Laws of Warfare when

 | DOI:10.1163/9789004507951_007

confronted with the horror of the battlefield, even when trained.'[1] He believes robots can do better.

Others lean less on the claim that automated systems would be more ethical than human troops and focus instead on the idea that technology can simply make faster decisions than humans, and that that speed in itself creates a strategic advantage for the side that deploys it. This of course depends on whether it is empirically true that speed of decision-making actually does produce an advantage. Boyd's well-known OODA loop concept of military decision-making encourages the belief that rapid decisiveness wins engagements. However, Boyd's critics have pointed out that the OODA loop model is only applicable in certain specific tactical settings (such as air-to-air combat), and does not translate well to, for example, the urban combat domain where decision-making happens simultaneously across many levels of command and a bad decision made quickly is *not* always better than a delayed decision.[2] A more nuanced version of the argument that technology-derived decision-making in combat may be superior to human decision-making suggests not that mere speed is the decisive factor, but that programmed systems are capable of making complex calculations that take into consideration more data or aspects of a problem than most human minds could manage to weigh at any one time. This is an idea to which we will return later.

Intentionally replacing human decision-making with AI decision-making is ill advised from both an ethical and a strategic perspective, and we will say more about that shortly. However, there is a serious and often unappreciated risk of *unintentionally* overriding human decision-making with AI decision-making that needs to be confronted first. Widespread use of AI – as well as other automated technology – in military settings can proliferate a phenomenon called automation bias. Automation bias occurs when humans are in an automated environment that orients them to be mostly observers rather than agents.[3] In this environment, humans exhibit an increased trust in automated systems, and often seek neither to confirm nor deny the validity of an evaluation or decision made by a computer or other automated system. They simply accept the automated system's judgment as final (and superior).

The presence of AI systems creates an environment where military members are likely to trust judgments of AI over their own and those of their human

1 Ron Arkin, *Governing Lethal Behavior in Autonomous Robots* (Taylor & Francis Group 2009) 36.

2 See Jim Storr, 'A Critique of Effects-Based Thinking,' (2005) *The RUSI journal*, Volume 150, Number 6, (December 2005), 32.

3 Linda J. Skitka, Kathleen L. Mosier, Mark Burdick, 'Does automation bias decision-making?' (1999) 51 *International Journal of Human-Computer Studies* 991.

peers. In the late 1990s, Linda Skitka did a study on automation bias using a flight simulator. Her study showed that in a test environment with an automated computer aid specifically stated *not* to be 100% accurate, participants would often still trust the computer over the other instruments in the cockpit that *were* stated to be 100% accurate.[4] This is worrisome because the participants seemed blinded by their own assumptions or notions about the accuracy and reliability of the automated aid, despite being specifically told that it was fallible (and less reliable than the other systems).

Alongside this, the study reached the conclusion that people in an automated environment were less vigilant about checking the accuracy of their systems and indicators than those in a non-automated environment. Participants in an environment with an automated flight aid only noticed 59% of problems unannounced by the automated aid, whereas participants without any automated aid noticed 97% of the same problems.[5] The type of error committed by those in the aided environment is one of omission, where a human not alerted to a problem by an automated system will not notice the problem, nor check to make sure that no problems in fact exist. The high level of success in noticing unannounced problems by those participants without an automated aid can be attributed to their vigilance in checking dials and indicators, and processing that information to determine the existence, type, and severity of a problem.

The other type of automation bias error is that of commission, in which a human acts according to the prescriptions of an automated system, even when other non-automated systems are indicating something different or contradictory to the automated system. An example of this type of error can be seen in a small 1992 study done in the NASA Ames Advanced Concepts Flight Simulator.[6] This study, which Skitka references in her own work on automation bias, included one scenario in which an 'auto-sensed checklist' suggested that the flight crew shut down the #1 engine due to fire damage. However, 'traditional engine parameters indicated that the #2 engine was actually more severely damaged.'[7] Three-quarters of the crews in the auto-sensed checklist scenario shut down the #1 engine, while only one quarter of participants using a paper checklist did the same.

This shows how an automated aid can quickly diminish the vigilance and diligence of people in verifying information given to them. Any crew in the

4 Skitka, Mosier and Burdick (n 3) 991.

5 Skitka, Mosier and Burdick (n 3) 991.

6 Kathleen Mosier, Everett Palmer & Asaf Degani, 'Electronic checklists: Implications for decision making' (1992) *Proceedings of the Human Factors Society 36th Annual Meeting* 7.

7 Skitka, Mosier and Burdick (n 3) 991.

auto-sensing scenario could have chosen to look at the analog dials and indicators in the cabin to determine if in fact the #1 engine was severely damaged. However, rather than carefully reading the analog dials and other indicators in the cockpit to form a judgment based on information and experience, the crews often opted instead to follow directions from a computer they had placed a high level of trust in. The analog devices could communicate the same information to an experienced pilot as a computer, but it takes more mental work on the behalf of the pilot to conclude that information from the indicators given to them.

When discussing this study, Skitka says, '[a]nalysis of the crews' audiotapes also indicated that crews in the automated condition tended to discuss much less information before coming to a decision to shut down the engine, suggesting that automated cues short circuited a full information search.'[8] This short-circuiting is one of the most dangerous aspects of automation bias. The presence of automated systems makes it less likely that those working with them will be vigilant in routinely checking nonautomated systems, nor will they put much effort into using other available tools to verify a decision reached by an automated system. These lapses are symptoms of the increased trust humans have in computerized or automated systems. This trust itself can also have perilous repercussions, both for the human *in situ*, and those who are affected by his or her actions – or inactions.

In July 1988, the crew of the USS *Vincennes* fell prey to the effects of automation bias with grave consequences. The ships' Aegis radar system (which was at the time set to a semi-automatic mode wherein humans worked with the system to decide what to fire upon and when) misidentified an Iranian passenger jet as an F-14 Iranian fighter jet. Despite other data indicating that the plane was not a fighter jet – including the plane broadcasting civilian radar and radio signals – no one in the command crew of eighteen disagreed with the computer's classification. They authorized the system to shoot, and only afterwards realized their horrific mistake. All 290 passengers and crew onboard the civilian plane were killed.

Peter Singer recounts this event in *Wired for War*, as well as a similar incident during the 2003 invasion of Iraq in which U.S. Patriot missiles shot down a pair of allied planes that the system misidentified as Iraqi rockets. The soldiers in this event had 'veto power,' over the system, but unfortunately they were 'unwilling to use [it] against the quicker (and what they viewed as better) judgment of a computer.'[9] Elke Schwarz describes this kind of environment,

8 Skitka, Mosier and Burdick (n 3) 991.

9 Peter Singer, *Wired for War* (The Penguin Press 2009) 125.

saying, '[s]et against a background where the instrument is characterized as inherently wise, the technology gives an air of dispassionate professionalism and a sense of moral certainty to the messy business of war.'[10]

Now, over twenty years after the Skitka study highlighted the dangerous existence of automation bias, humans have increased the use of automated systems in all areas of life, both civilian and military alike and, if anything, are more trusting than ever before of automated guidance. Artificial intelligence, as a technology that is little understood by the general public and often sold as 'better' than human ability, puts us at even higher risk for automation bias. Despite documented failures, people focus on reports that seem to suggest that artificial minds are superior to organic ones, including news of computers beating humans at strategy games like chess and Go. While television advertisements find humor in people following incorrect Google maps directions straight into a lake, the reality of overreliance on automation is much less amusing. In a military context, automation bias can have life or death consequences. Just as infantry check the proper functioning of their weapons, those using AI systems in their military roles are obligated to make sure their tools – both automated and not – are working correctly, too.

It is also vital to remember that no matter how advanced or capable of 'machine learning' they are, AI and other automated systems were originally programmed by humans, and we are fallible, biased creatures. Rather than freeing us from our natural weaknesses, such systems unfortunately have the potential to establish them more firmly – to 'bake them in,' as it were. There are already many recorded cases of algorithms that were designed with the goal of providing superhuman objectivity, but instead merely amplified human prejudices and replicated character flaws. One example that shows this effect clearly was the attempt in 2016 by the company Beauty.AI to program a computer to judge an international beauty contest. The results turned out to be dramatically biased in favor of lighter-skinned contestants. Analysis revealed that the programmers who had provided the original images to the system for it to 'learn' what beauty was had relied almost exclusively on photos of young white women:

> Beauty.AI – which was created by a 'deep learning' group called Youth Laboratories and supported by Microsoft – relied on large datasets of photos to build an algorithm that assessed beauty. While there are a

10 Elke Schwarz, 'Technology and moral vacuums in just war theorising' (2018) *Journal of International Political Theory* 1.

> number of reasons why the algorithm favored white people, the main problem was that the data the project used to establish standards of attractiveness did not include enough minorities, said Alex Zhavoronkov, Beauty.AI's chief science officer.[11]

In another famous case, Microsoft created an AI 'chatbot' which it named Tay and allowed it to 'learn' how to communicate from interactions with real humans on Twitter. The results were deeply disturbing, as Tay soon started sending out hate-filled racist, anti-Semitic, misogynist, and otherwise extremely offensive tweets.[12] And as journalist Sam Levin notes, '[A]fter Facebook eliminated human editors who had curated "trending" news stories …, the algorithm immediately promoted fake and vulgar stories on news feeds.'[13] It, too, copied the worst aspects of the content it encountered.

This is a problem commonly referred to as 'garbage in, garbage out,' but it cannot be solved just by keeping AI systems away from the more pernicious elements on social media. Bias can be more difficult to keep out of AI programming than one might at first imagine. As legal scholar Jerry Kang has exhaustively documented, human bias is extremely difficult to avoid and is quite often unrecognized or unconscious. As one illustration, Kang cites the following study:

> Shooter Bias. Social cognitionist Joshua Correll created a video game that placed photographs of a White or Black individual holding either a gun or other object (wallet, soda can, or cell phone) into diverse photographic backgrounds. Participants were instructed to decide as quickly as possible whether to shoot the target. Severe time pressure designed into the game forced errors. Consistent with earlier findings, participants were more likely to mistake a Black target as armed when he in fact was unarmed (false alarms); conversely, they were more likely to mistake a White target as unarmed when he in fact was armed (misses). Even more striking is that Black participants showed similar amounts of 'shooter bias' as Whites.[14]

---

11 Sam Levin, 'A beauty contest was judged by AI and the robots didn't like dark skin,' *The Guardian*, (8 September 2016).

12 James Vincent, 'Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day,' *The Verge* (24 March 2016).

13 Levin (n 11).

14 Jerry Kang, 'Trojan Horses of Race,' *UCLA Journal of Scholarly Perspectives* (1 January 2007) 3, 43.

Thus, as long as humans program AI, bias and other human vices will come along for the ride, and sometimes be amplified. While this does not make it certain that AI will make any worse decisions than people would, it is a warning that AI systems should never be assumed to be more accurate or objective than a person would be. They should certainly not be trusted to be anything like infallible. In another recent example of a high-profile AI failure, IBM marketed its Watson supercomputer to doctors as a source of medical guidance for treating cancer patients (as 'Watson for Oncology'). However, the results were strongly criticized by real doctors as worse than useless:

> One example in the documents is the case of a 65-year-old man diagnosed with lung cancer, who also seemed to have severe bleeding. Watson reportedly suggested the man be administered both chemotherapy and the drug 'Bevacizumab.' But the drug can lead to 'severe or fatal hemorrhage,' according to a warning on the medication, and therefore shouldn't be given to people with severe bleeding. …[15]

Interestingly, when the negative study results came back to IBM, they blamed Watson's failures on the way that he was 'trained' by doctors at Memorial Sloan Kettering (MSK) Cancer Center, using synthetic data:

> According to the report, the documents blame the training provided by IBM engineers and on doctors at MSK, which partnered with IBM in 2012 to train Watson to 'think' more like a doctor. The documents state that – instead of feeding real patient data into the software – the doctors were reportedly feeding Watson hypothetical patients data, or 'synthetic' case data. This would mean it's possible that when other hospitals used the MSK-trained Watson for Oncology, doctors were receiving treatment recommendations guided by MSK doctors' treatment preferences, instead of an AI interpretation of actual patient data.[16]

Whatever the source of Watson's errors, this case highlights the danger of organizations that manage risk on the level of life and death giving in to the temptation to rush to adopt AI systems that offer guidance that sounds authoritative but may in fact be at least as likely to be mistaken as human judgment

15 Jennings Brown, 'IBM Watson reportedly recommended cancer treatments that were "unsafe" and "incorrect," ' *Gizmodo* (25 July 2018).

16 Brown (n 15).

(if not more so). All that does is introduce another possible point of failure into the system.

2 Ethical Deskilling of the Military

Another potential way in which the integration of artificial intelligence into a military decision-making role could prove to be harmful if not handled correctly is if it leads to the 'moral deskilling' of the military.[17] Deskilling occurs when the opportunity for a skill to be practiced is diminished or eliminated, leading to the decreased ability of a human to perform that skill well. Both practical and ethical skills are put at risk with the introduction of artificial intelligence or other automated aids in military decision-making. For our purposes here, we will focus on ethical deskilling, though many of the same arguments and conclusions will apply to practical deskilling, as well.

In many militaries around the world, the expected professional military ethic is communicated and taught through virtue ethics. There is a code promoting particular virtues that is inculcated into all new recruits and (ideally, if not always successfully) reinforced until it becomes core to the identity of every person who serves. This particular brand of thinking is popular in the United States, where virtue ethics are the backbone of the professional military ethic in all military branches. Virtue ethics differs from other styles of ethical thinking, such as duty-bound deontology or greatest-good-for-the-greatest-number utilitarianism, in that it calls upon each agent to act ethically in everything they do as a matter of habit, as a way of embodying certain key virtues (e.g. courage, commitment, integrity, honor, loyalty, discipline, etc.). By this approach, troops (and especially those in leadership roles) are encouraged not simply to reference a rule or perform a calculation to make ethically charged decisions, but instead to act according to the military virtues instilled within themselves. Virtues are expected to be an intrinsic part of each action and decision in which a member of the military engages. More than guiding principles, virtues are meaty and deep character traits that shape a person and the way they move through the world. Acting ethically should be not second- but first-nature to a truly virtuous person.

17 This phrasing of the concept was first brought to our attention by Shannon Vallor, 'The Future of Military Virtue: Autonomous Systems and the Moral Deskilling of the Military,' (2013) *2013 5th International Conference on Cyber Conflict* (CYCON *2013*) Tallinn, 1. For more in-depth analysis, see Vallor's excellent *Technology and the Virtues*, 2016.

In his *Nicomachean Ethics*, Aristotle says, '… moral virtue comes about as a result of habit.'[18] Habituation of ethical behavior is essential to both properly forming and maintaining good character. As Shannon Vallor notes, Aristotle '… also reminds us that virtue is *more* than just skill or know-how; it is a state in which that know-how is reliably put into action when called for, and is done with the appropriate moral concern for what is good.'[19] This state cannot be attained, or maintained, in an environment where humans are denied the opportunity to practice their virtues. The passive observer role that artificial intelligence systems can easily push humans into is one such environment.

Use of autonomous systems in any facet of military work performed by a person takes away the opportunity for military personnel to practice their virtues in that role. Just like muscles in the body, virtues must be exercised in order to maintain their strength and effectiveness. In the specific context of military decision-making, handing this role – in whole or in part – to autonomous systems will reduce the ability of military personnel to practice all of their virtues and skills honed for that purpose. This could very well lead to a decreased ability of troops to respond to any ethically fraught issues that arise when the AI is not available.

Ethical deskilling of the military is something to be avoided for several reasons. First, the longer that autonomous systems are used in decision-making capacities and humans are excluded from that role, the more likely it becomes that the humans will feel less sure of themselves when it comes to questioning the autonomous system and challenging its authority. As discussed previously, automation bias makes people much less likely to question automated systems or to try to verify their conclusions. The people who work around autonomous decision-making systems will grow used to not having to make any decisions themselves, and thus will be much less prepared to do so than people who are consistently and actively practicing their virtuous habits and decision-making skills.

There is an additional concern. Due to what some scholars refer to as the 'black box' nature of how AI makes decisions, currently it is almost if not totally impossible to determine what reasons or justifications an AI has for making a particular decision.[20] This makes it impossible for most humans (who are not

18 Aristotle, *Nicomachean Ethics* Book II, 1 (trans. W. D. Ross) <http://classics.mit.edu/Aristotle/nicomachaen.2.ii.html> accessed 12 May 2020.

19 Vallor (n 17) 1, emphasis original.

20 However, there are some efforts underway to find out what's going on inside the 'black box,' see https://futurism.com/third-wave-ai-darpa/ 'DARPA is funding research into AI that can explain what it's "thinking."' <https://futurism.com/third-wave-ai-darpa/> accessed 12 Mai 2020.

programmers and the designers of algorithms) to understand an AI system's decision-making process and possibly learn something from it. We do not want a result where soldiers are neither practicing their own ethical decision-making skills, nor are they able to learn from the new entities that are making the decisions. This essentially leaves soldiers in a mere observer-caretaker role in which they watch AI systems make decisions and perhaps are also responsible for ensuring their proper functioning.

This can be problematic when a soldier who has not had to make a decision in a long time is working with an artificial intelligence that may be malfunctioning and making unethical decisions. How able will the soldier be to see if an AI is creeping towards the edges of legally and ethically permissible action? A lack of practice in having to carefully weigh objectives, costs, and benefits of an action within legal and ethical restraints makes it less likely that a soldier will see when an AI is currently or is about to act unethically. For an analogy, consider how the United States Naval Academy struggled with the decision to keep or remove courses in celestial navigation in the era of GPS navigation systems. In the end, after taking the course out their requirements for midshipmen, they ended up restoring it, because they received feedback from the Fleet that celestial navigation was both a good back-up skill in case of a failure of electronic navigation (or tampering with it by an enemy) and a way for sailors to recognize when their GPS systems might be malfunctioning:

> The Navy and other branches of the U.S. military are becoming increasingly concerned, in part, that they may be overly reliant on GPS. … In a big war, the GPS satellites could be shot down. Or, more likely, their signal could be jammed or hacked. … [Rear Admiral Michael] White, who heads the Navy's training, says there is also a desire to get back to basics. Over the past decade, electronic navigation systems on ships have become easier to use, so less training is required. He says the Navy is bringing back celestial navigation to make sure its officers understand the fundamentals. 'You know, I would equate it to blindly following the navigation system in your car: If you don't have an understanding of north/south/east/west, or perhaps where you're going, it takes you to places you didn't intend to go,' he says. In fact, there has been at least one incident in the past decade when a Navy ship ran aground partly because of problems with the electronic navigation system, investigators say.[21]

21 Geoff Brumfiel, 'U.S. Navy Brings Back Navigation By The Stars For Officers' NPR, Science, Morning Edition (22 February 2016) <https://text.npr.org/s.php?sId=467210492> accessed 12 May 2020.

Ethical decision-making – including when and how to show moral courage – will also become an area of vulnerability if it is not taught and practiced.

The most detrimental peril of ethical deskilling of the military is that it takes something valuable away from military personnel. Virtues and the habituation of good character stay with a military member long after they have left the service. It seems safe to say that most people do not join the military to become a worse person – they join to become a better person and a better member of society for having served their country. Denying them the opportunity to cultivate virtues and character properly and fully is depriving them of one of the aspects of military culture that is intrinsically beneficial to military members. While the military teaches many valuable skills – such as survival, navigation, and marksmanship – habituated ethical virtues and the ability to make ethical decisions under pressure can provide someone with benefits in all areas of their life, military and civilian. In order to maintain an ethical military, all its members must be able to properly habituate and inculcate virtues as a benefit to themselves, as well as the institution of the military, and society as a whole. While in reality this goal remains aspirational (and the many challenges of professional military education (PME) are beyond the scope of this discussion), it certainly should not be rendered nearly impossible by the overuse of AI systems.

3 Ceding Strategic Advantage

There is a reliable pattern in human history, in which the development of a new military technology by any group or nation sparks an 'arms race' among competitors to at least catch up with or ideally leap ahead of all others to deploy the latest tool in combat. It is therefore unsurprising that there are some panicked voices in NATO or among its allies insisting that the determination of governments in, for example, China and Russia, to focus resources on the speedy advancement and utilization of AI in military applications represents a serious threat that can only be answered by wholeheartedly diving into an AI arms race. So we see statements like these in a memo from Deputy Secretary of Defense Patrick Shanahan, released in June 2018:

> This effort is a Department priority. Speed and security are of the essence. I expect all offices and personnel to provide all reasonable support necessary to make rapid enterprise-wide AI adoption a reality.[22]

22 Official memorandum from Deputy Secretary of Defense Patrick Shanahan (27 June 2018).

These are the closing words of the memo, emphasizing speedy and extensive AI adoption for the Department of Defense. However, it should be noted that earlier in the same memo, Shanahan cautions that, 'we must pursue AI applications with boldness and alacrity while ensuring strong commitment to military ethics and AI safety.'[23]

The lessons of history do not definitively support the idea that a 'be first, or be last' approach is strategically sound. There are four key points to remember: (1) being one of the first to adopt a new military technology does not guarantee immediate advantage or ultimate supremacy in the wielding of that technology, (2) even if technological superiority is achieved by the earliest adopters, being the technologically superior side in an asymmetric conflict is absolutely no assurance of victory, (3) introducing any new technology introduces new potential points of failure, and (4) taking time to develop a new technology more carefully and deliberately can allow you anticipate potential weakness and both harden your own systems against them and identify how to exploit the flaws in your opponents' systems.

Being the first out of the gate to field a new weapon has produced a mixed bag of outcomes. Gunpowder was invented by China in the 9th century AD, but was not really used effectively until the 13th century when Islamic troops in Egypt were armed with small cannon and other gunpowder-based projectile weapons.[24] Certainly, the English army benefited tremendously from the edge that the use of the longbow gave them against the French at Agincourt. However, the first submarine, used by the Colonials against the British in the American Revolution, was a dismal failure. On the other hand, it could be argued that the development of the tank helped break the stalemate in World War I. The Germans flew the first jet fighters in World War II and were generally innovative and early adopters, but, due to many factors, they ultimately lost the war. Meanwhile, both radar and the atomic bomb were new technologies vital to the victory of the Allies. The Soviet Union had many technological firsts, and yet they lost the Cold War, in part because they were outspent and driven to extremes (such as the Caspian Sea Monster) in their attempts to outpace the West.

Similarly, while it seems counter-intuitive to doubt that superior technology will yield victories, this has simply not been consistently the case in asymmetric conflicts. In ancient times, the technologically inferior Gauls defeated the Romans (and the Celts gave them a run for their money). The ill-equipped

23 Shanahan (n 22).

24 Michael Marshall, 'Timeline: Weapons Technology,' *New Scientist* (7 July 2009).

Colonials beat back the British army (although arguably they might not have been able to do so without the support of the French fleet). And more recently, the United States found itself unable to dominate the conflict in Vietnam, despite having objectively superior weapons and technology across the board. It is simply not the case that low-tech always loses against high-tech. Small modern drones can be knocked out of the sky by a simple hand-thrown spear or a trained falcon. Asymmetric advantage is a red herring. The seemingly technologically inferior side in an asymmetric conflict quite frequently wins, against the odds.

This is not to say that developing new technology is a waste of time. It can be an essential component to victory, particularly in more evenly matched conflicts. The caution here is only against the dangerous assumption that it will always provide a decisive edge. To put this in more positive terms, there is no need for a country like the United States to be made nervous or intimidated by news of the aggressive pursuit of AI by its rivals. The correct response is to focus on research, not to rush anything into the field. Researching new military technology is necessary, if only to be prepared with countermeasures against whatever opponents might design. The approach should not be to mirror the enemy's moves, but to carefully identify all the potential advantages and flaws of this new technology: not to do it first, but to do it better. Such research should simultaneously focus on effective countermeasures to exploit the vulnerabilities in the new technology. If some forms of AI are a bad idea, the smart move is not to waste resources copying those forms, but to design tools to defeat them. In this way, the hastiness of others becomes more an opportunity than a threat.

All programmed and automated systems introduce two potential points of failure: predictability and hackability. Human troops cannot easily be reprogrammed to turn against their own side. It is possible to do so, but it will never be as straightforward as changing a few lines of code. At the same time, humans can be astonishingly adaptable and creative. Artificial systems may have success against people within the strict confinement of a rule-based game like chess or Go. In the real world, rigid rules are replaced by ever-changing circumstances and seemingly irrational or unpredictable but often surprisingly effective human responses to extreme peril. Perhaps advances in quantum computing will ultimately yield artificial intelligence that is as flexible as the human mind and capable of leaps of intuition and imagination. Until that time, however, talented human strategists, with that elusive quality Clausewitz referred to as the *coup d'oeil*, will prevail. It is incumbent on military leaders not to allow themselves to be carried away by the siren song of the latest breakthroughs in emerging technology and hand over actual military

decision-making to AI. This would be foolishly ceding a strategic advantage for the fashion of the season.

4 Moral and Legal Responsibility

While new technologies are being integrated into both civilian and military society, often ethical and legal standards regarding their use lag behind. Artificial intelligence is a tool that many fields, from banking to defense, have wanted to get their hands on and use as quickly, and as much, as possible. There have been some early successes that have added to the excitement, such as the use of AI to stitch together and extrapolate from archeological data to give more detailed images of the past than were ever possible and to decipher ancient manuscripts.[25] New technology that is appraised as better than current tools is often rapidly implemented with little thought to the human, social, or political repercussions of its longer-term or more extensive use. This has been particularly true with AI, which seems to have endless capabilities and possible applications. It is unfortunately both common and dangerous that questions of legality and morality surrounding the use of new technology are typically left to be debated only after something problematic has occurred. This overall trend needs to be fought against, and many are already doing so, with vigor, as seen in the robust scholarship around the legal or ethical permissibility of using autonomous weapons systems and so-called 'killer robots.'[26] For all emerging tech, but especially any with potentially lethal consequences, conversations about the possible consequences of the use of emerging technology need to happen early in the R&D phase, so that appropriate safeguards can be built into the design (rather than having to be retrofitted in response to documented harms). As Damon Horowitz says, '[w]e want the people building the technology thinking about what we should be doing with the technology.'[27]

The 'should' is important here, and points to the inherently normative nature of creating and introducing, let alone widely implementing, any new technology. Deeply considering how a new technology could be used or

25 *News Network Archaeology*, 'Using AI to Uncover Mysteries of the Voynich Manuscript' (26 January 2018).

26 See: Bradley Jay Strawser (ed), 'Killing by remote control: The ethics of an unmanned military' (2013); Patrick Lin, Ryan Jenkins, and Keith Abney (eds), 'Robot Ethics 2.0, From Autonomous Cars to Artificial Intelligence' (2017); Noel Sharkey, 'Saying 'No!' to Lethal Autonomous Targeting,' (2010) *Journal of Military Ethics*, Volume 9 Issue 4, 369.

27 Damon Horowitz, 'We need a "moral operating system," ' TEDxSiliconValley talk (2011).

abused, particularly outside its original scope or purpose, helps to anticipate problems and, where possible, design them out of existence. Even when it is impossible to block a potentially harmful use (or misuse) of a product, having forewarning of the issue at least allows for clear communication (advance warning is far preferable to surprise) and the development of countermeasures in advance. This is important not only when the concern is about new technology getting into the 'wrong hands,' but also even when it will be in the right hands. The U.S. military has a troubling history of implementing systems without ample time for them to be carefully studied and tested from a safety perspective, let alone from a legal or ethical standpoint. The Bradley fighting vehicle and the osprey are just two well-known examples of flawed systems rushed into use. There are also the tragedies of service personnel who were sent into irradiated areas before the effects of nuclear weapons were understood or the combat troops who were given unreliable, jam-prone M-16s in Vietnam: 'from Gettysburg to Hamburger Hill to the streets of Baghdad, the American penchant for arming troops with lousy rifles has been responsible for a staggering number of unnecessary deaths.'[28]

It is not enough, however, to take steps to anticipate what might go wrong with a new system. It is also important to consider in advance how accountability can be determined when something does go wrong. Is it possible to determine what should be attributable to the new tool and what is the fault of the operator? If AI is given too great a role in military decision-making, such determinations of legal and moral responsibility may become quite murky. In the case of the *USS Vincennes* described earlier, did the crew members decided to shoot down the aircraft the Aegis system identified as an enemy fighter? Or did they allow the system to act from its own findings? In that event, the Aegis system required a human to approve its findings and authorize it to fire, so attribution of responsibility for the catastrophe can reasonably be tied to the crewmembers who gave the go-ahead to the system. Still, the influence of the system on the crew's decision was significant. It is not difficult to imagine even more nuanced circumstances in which an AI system assumed to operate exclusively within algorithmically determined targeting parameters in line with the relevant laws of war and rules of engagement offers only one targeting option to troops, and innocents die as a result. Those seeking to assign blame could choose to take aim at the manufacturer, the code writers, the system technician on the ground, or a myriad of other possibilities. In the end, no one might truly be held accountable. The less human oversight and the greater

28 Robert H. Scales, 'Gun Trouble,' *The Atlantic* (January/February 2015).

decision-making influence AI systems have, the greater the problem of legal and ethical responsibility grows.

Patrick Lin has argued that, 'as robots become more autonomous, a case could be made to treat robots as culpable legal agents.'[29] He goes on to examine how autonomous robots may fit into a category he calls 'quasi-agents,' which typically includes children and the mentally disabled in today's legal understanding. Legal quasi-agents are not expected to have the same faculties of judgment as a full agent, and thus have 'diminished responsibility' for their actions.[30] Any consideration of how to treat artificial intelligence in the legal realm will likely follow from determinations about the legal status of robots generally. However, it is not necessarily a good idea for robots, or AI, to be given any legal status. What societal benefits would be reaped from being able to legally prosecute a robot or automated system? Surely humans would not feel safer after a robot has been punished (whatever that may entail), but only when the cause for a robot's harmful, unethical, and/or illegal actions is discovered and corrected. Keeping automated systems, and AI in particular, out of any decision-making role eliminates the need to consider them in any legally culpable sense. If these systems cannot be said to be legally responsible for errors, then it is the humans in charge who will be culpable. This puts the onus on the humans to closely observe automated systems as they work, and emphasizes the need for humans not to take the determinations of an automated system as the last word on any matter.

Not permitting AI to make decisions in a military context will help guard against automation bias, the dangers of which we explored earlier, and will keep humans in an ethically-charged role wherein they can be held to appropriate legal and moral responsibility for their actions. Elke Schwarz saliently describes what happens when humans are not the decision-makers:

> [there is a] moral vacuum that technologies of ethical decision-making create in their quest to 'secure' moral risk. A moral vacuum opens when certain parameters of harm are no one's responsibility; when the decision that harm is permissible has been determined through technological means.[31]

29 Patrick Lin, George Bekey, and Keith Abney, 'Autonomous Military Robotics: Risk, Ethics, and Design,' Report for the Department of the Navy, Office of Naval Research (2008).

30 Lin, Bekey and Abney (n 29).

31 Schwarz (n 10).

5 The Promise of AI as an Option-Generating Advisor

Artificial intelligence is not an inherently unethical creation. While it can cause myriad different problems if permitted to make its own decisions or to overly influence the decisions of human operators, this does not mean it cannot be used well in more appropriate manners and contexts. AI can be a positive supplement to military operations if it is employed as a tool to help humans, rather than as a replacement for or authority over them. There is a specific role in which military decision-making that AI could be immensely beneficial – that of an option generator. In this role, an automated system could even be, to borrow language from the responsibilities of military chaplains, 'an ethical advisor to command.'[32]

One of the well-documented strengths of AI is that it can process large amounts of information faster than humans.[33] This capability of AI makes it well suited to assist with time-sensitive and data-heavy tasks, both of which are easily found in military contexts. We have explained the perils of having AI systems control military decision-making, but AI could enhance human decision-making. By flagging logical fallacies, challenging assumptions, exposing blind spots in reasoning, and by processing information quickly it may provide a number of potential courses of action. This could help guard against common impediments to good decision-making by humans under stressful conditions.

There are many factors that can derail effective and ethical decision-making in high-pressure situations. For example, Kevin Mullaney and Mitt Regan have done in depth analysis of one single minute of the 19 November 2005 incident in Haditha, Iraq, in which U.S. Marines ultimately killed 24 unarmed civilians.[34] Among many other valuable insights, they have determined that the marines in question, due to their emotional state following the death of one of their own from an IED, failed to detect various visual cues that should have indicated essential information such as that the civilian car involved in the incident was

32 OPNAVINST 1730. (See (5) 'The chaplain shall serve as the principal advisor to the commander on all matters related to religious ministry and shall advise on ethical and moral matters and issues pertaining to the command.') And as Rev. Dr. Nikki Coleman has noted, there are generally not enough chaplains to go around to perform this role for all commands.

33 Lin, Bekey and Abney (n 29).

34 Prof. Mitt Regan (Georgetown Law and the US Naval Academy) and CDR Kevin Mullaney, Ph.D. (US Naval Academy), 'One Minute in Haditha: Morality and Non-Conscious Decision-Making,' presented at the North American ISME (International Society for Military Ethics) conference, Case Western Reserve University, Cleveland, Ohio (25 January 2018).

riding too high to have been filled with weapons and explosive equipment. An AI system could conceivably be designed to make a rapid scan of the surroundings and pass that kind of information on to human decision-makers in a neutral way that would not predetermine what action should follow from that information (thus avoiding the risk of undermining human authority through automation bias). In other words, the AI system might simply register something like, 'Apparent civilian car detected. Not carrying a heavy load.'

Of even greater value might be a well-timed, calm reminder by the automated system of the rules of engagement: 'Likely civilians detected; current rules of engagement prohibit firing unless fired upon; no incoming rounds detected.' Rather than being a spur to lethal action, as in the *USS Vincennes* case, the automated system could serve as a backstop against impulsive behavior. In conducting variations of his famous experiments about obedience and human responses to authority, Stanley Milgram discovered that the negative influence of a corrupt authority could be overcome by the introduction the competing influence of a rebellious peer or secondary authority.[35] By just asking questions (e.g. 'Are you sure we should be doing this?'), they were able to 'break the spell' of the original authority and awaken the conscience of the experiment subjects. Without giving direct contradictory orders, an AI advisor could provide a valuable second opinion or voice of reason to remind troops of their true mission and core values.

In cases where time permits those in command to weigh different options, AI systems could present the pros and cons of different options – again, aiding decisions, not making them. This could be used in operational settings and to help military leaders think through non-combat-related ethical issues and perhaps avoid the reckless violations and waves of corruption that have at times seemed to sweep through and decimate the service (e.g. the 'Fat Leonard' scandal in the U.S. Navy). It is clear that some commanders could use an AI voice of reason: their own automated 'Jiminy Cricket.'

Such systems must be programmed never to provide only one potential course of action. The moment AI offers just one option, it becomes the voice of authority, and all of the previously detailed perils to human autonomy and accountability return. While AI could be programmed to know and remind members of the military about such guidance as the Uniform Code of Military Justice (UCMJ), the Law of Armed Conflict (LOAC) or International Humanitarian Law (IHL), and particular Rules of Engagement (ROEs), the

35 Stanley Milgram, *Obedience to Authority: An Experimental View* (New York: Harper Perennial Modern Thought 2009) 107 and 118.

interpretation of these laws and principles must be left up to the humans themselves. Beyond the concerns already raised, the door must always be left open for human agents to decide that something is unethical, even if it is technically legal and permitted. Human moral agents have the capacity to recognize gray areas and nuance, to feel empathy, and sometimes to know instinctively when something 'just isn't right.' No AI system should undermine that ability. What we commonly think of as instinct or conscience is more than likely the result of the lived experience of navigating a complex world and the well-engrained, habituated virtues acquired through that experience.

Without interfering with instinct or blocking the nudge of conscience (and in some cases even by being the source of the latter), artificial intelligence can present multiple viable options in moments of decision that could help reduce the negative impact of the psychological and physiological effects of combat stress, such as tunnel vision, becoming trapped in a false dilemma, and experiencing paralyzing sensory overload. For example, tunnel vision can impede troops' ability to think 'big picture' and consider effects beyond their current engagement, possibly leading to actions that may solve an immediate problem, but create other, greater problems within the mission. Using AI as an option-generator could be programmed to work within and support strategic, operational, and tactical objectives. While programming would have to account for how to balance among these objectives, they all would be considered in the option-generation process to create viable choices for troops or leaders to consider.

Similarly, artificial intelligence as an option-generator could help prevent soldiers from falling into the trap of false dilemmas: believing themselves to be in binary, 'either/or' scenarios, when in fact other options for action exist. Implementing AI as an option-generator in the military decision-making context will guard against this phenomenon by consistently presenting a range of appropriate options. AI as an option-generator could alleviate some of the pressure from the amount of data that troops must gather and process when attempting to develop solutions to a pressing problem. This would address the issue of sensory overload.

The combination of time-sensitivity and data-leadenness in military decision-making make artificial intelligence attractive as a means to produce appropriate options. This is not to say that troops and military leaders are not adept at or capable of formulating solutions to pressing mission problems, only that they might benefit from using AI to generate options quickly that balance legal, ethical, and practical restraints, especially in chaotic conditions or within layered domains. Situating artificial intelligence as an option-generator

and not as a decision-maker keeps it in the realm of tools to be used by human service members, not as a replacement for them.

6 Conclusion

Much of the language around AI references the position of a human in or out of a decision-making 'loop.' Yet even in situations where a human is 'somewhere in the loop,' supposedly checking and observing the system to make sure it is working properly, artificial intelligence should never be allowed to make decisions or to offer only one option to the people it advises. Humans are far too trusting of computers and other advanced systems. While technology is often implemented to reduce human error, AI systems can manifest new kinds of errors for humans to make that are just as deadly as the old ones. Artificial intelligence can make mistakes on its own, as well, and the people who work around it may not be vigilant enough to catch them in time.

This does not mean that there is no opportunity for the military to benefit from the use of AI, but rather that it should be designed and deployed as an option-generator for decision-making humans, not an artificial authority figure. AI should not tell service members what to do (let alone do it for them), but should instead provide them more information and perspective than they might be able to access themselves. As an option-generator, taking myriad factors into consideration, AI could act as a bulwark against the natural psychological pressures that can drive humans to make poor or unethical decisions in the stress of combat or other high-stakes situations. Automated systems could even serve as a sort of artificial conscience, reminding military leaders of the ramifications and possible consequences of their decisions.

War is an enterprise with profound human costs. As such, it should be conducted primarily by humans. As autonomous agents, humans can bear the moral responsibility for their actions and be held accountable for them. Technology cannot be used as a shortcut to perform roles using less effort or expenditures than the moral weight of those roles demands. There is too much at stake, including resisting the moral and practical deskilling of the military, avoiding the hidden bias that can be deeply embedded in automated and machine-learning systems, and maintaining strategic advantage.

Any new technology that is implemented in military decision-making can have serious and far-reaching consequences, and ought to be used only as an aid or tool for human operators – never as a replacement for them. There are ways that AI could assist humans in making ethical decisions in the complex context of modern warfare, by presenting information quickly in a digestible

manner and offering a range of options to the human decision-maker. Yet AI systems remain fallible and their advice should never be privileged over human judgment, instinct, and experience. War is not chess or a game of Go, and any military that sublimates human decision-makers to AI systems will lose more than its soul.

References

Aristotle, Nicomachean Ethics Book II, 1 (trans. W. D. Ross) <http://classics.mit.edu/Aristotle/nicomachaen.2.ii.html> accessed 12 May 2020.

Arkin, Ron, Governing Lethal Behavior in Autonomous Robots (Taylor & Francis Group 2009).

Brown, Jennings, 'IBM Watson reportedly recommended cancer treatments that were "unsafe" and "incorrect," ' *Gizmodo* (25 July 2018).

Brumfiel, Geoff, 'U.S. Navy Brings Back Navigation By The Stars For Officers' NPR, Science, Morning Edition (22 February 2016) <https://text.npr.org/s.php?sId=467210492> accessed 12 May 2020.

Horowitz, Damon, 'We need a "moral operating system," ' TEDxSiliconValley talk (2011).

Kang, Jerry, 'Trojan Horses of Race,' *UCLA Journal of Scholarly Perspectives* (1 January 2007).

Levin, Sam, 'A beauty contest was judged by AI and the robots didn't like dark skin,' *The Guardian*, (8 September 2016).

Lin, Patrick, George Bekey, and Keith Abney, 'Autonomous Military Robotics: Risk, Ethics, and Design,' Report for the Department of the Navy, Office of Naval Research (2008).

Lin, Patrick, Ryan Jenkins, and Keith Abney (eds), *Robot Ethics 2.0, From Autonomous Cars to Artificial Intelligence* (Oxford University Press 2017).

Marshall, Michael, 'Timeline: Weapons Technology,' *New Scientist* (7 July 2009).

Milgram, Stanley, *Obedience to Authority: An Experimental View* (New York: Harper Perennial Modern Thought 2009).

Mosier, Kathleen, Everett Palmer & Asaf Degani, 'Electronic checklists: Implications for decision making' (1992) Proceedings of the Human Factors Society 36th Annual Meeting.

News Network Archaeology, 'Using AI to Uncover Mysteries of the Voynich Manuscript' (26 January 2018).

Official memorandum from Deputy Secretary of Defense Patrick Shanahan (27 June 2018).

Regan, Mitt and Kevin Mullaney, 'One Minute in Haditha: Morality and Non-Conscious Decision-Making,' presented at the North American ISME (International Society for

Military Ethics) conference, Case Western Reserve University, Cleveland, Ohio (25 January 2018).

Scales, Robert H., 'Gun Trouble,' *The Atlantic* (January/February 2015).

Schwarz, Elke, 'Technology and moral vacuums in just war theorising' (2018) *Journal of International Political Theory*.

Sharkey, Noel, 'Saying 'No!' to Lethal Autonomous Targeting,' (2010) *Journal of Military Ethics*, Volume 9 Issue 4.

Singer, Peter, *Wired for War* (The Penguin Press 2009).

Skitka, Linda J., Kathleen L. Mosier, Mark Burdick, 'Does automation bias decision-making?' (1999) 51 *International Journal of Human-Computer Studies*.

Storr, Jim, 'A Critique of Effects-Based Thinking,' (2005) *The RUSI journal*, Volume 150, Number 6, (December 2005).

Strawser, Bradley Jay (ed), *Killing by remote control: The ethics of an unmanned military* (Oxford University Press 2013).

Vallor, Shannon, 'The Future of Military Virtue: Autonomous Systems and the Moral Deskilling of the Military,' (2013) 2013 5th International Conference on Cyber Conflict (CYCON 2013) Tallinn.

Vallor, Shannon, *Technology and the Virtues* (Oxford University Press 2016)

Vincent, James, 'Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day,' *The Verge* (24 March 2016).