Expeditions with MCUP

The Battle with Data

Realities of Bringing Artificial Intelligence to the Battlefield

Lieutenant Colonel Ryan Whitty, USMC

https://doi.org/10.36304/ExpwMCUP.2022.09

Abstract: When describing the future battlefield, many military practitioners speculate on the impact of artificial intelligence (AI), and others even demand its use. This article offers a basic understanding of the technology they wish to employ. First and foremost, quality, labeled, and organized data feeds an AI system. Also, many AI constructs prove fragile when exposed to too little or tainted data and risk becoming predictable to an adversary. Applying a sound application model, which accounts for human interaction with AI implementation, may help ensure that military engagements do not become purely data-driven.

Keywords: artificial intelligence, machine learning, data, signature management, future battlefield

---

LtCol Ryan Whitty is a U.S. Marine Corps officer currently serving as a Commandant of the Marine Corps Fellow at the MIT Lincoln Laboratory. He has 19 years of experience including operational tours and deployments with I Marine Expeditionary Force (I MEF), the 11th Marine Expeditionary Unit, and III MEF. Whitty attended the U.S. Naval Academy, the Naval Postgraduate School, Marine Corps University, and the Joint Forces Staff College. He holds a bachelor's degree in systems engineering and master's degrees in electrical engineering and military studies. The views expressed in this article are solely those of the author. They do not necessarily reflect the opinions of Marine Corps University, the U.S. Marine Corps, the Department of the Navy, or the U.S. government.

Introduction

When describing the future battlefield, many military practitioners speculate on the impact of artificial intelligence (AI), and others even demand its use. They often do this because AI already drives parts of the economy and steers prototype self-driving cars. However, many of those who are unindoctrinated in the fundamentals of this technology fail to realize that these commercial examples represent a choice by engineers to use AI because it offers the right solution to a specific problem. These examples often use machine learning (ML), which performs a specific data characterization task extremely well but does not represent an artificial general intelligence.[1] ML algorithms can produce impressive results but prove highly dependent on the training data they receive, which leaves them fragile to attack. Nonetheless, military planners and leaders face current decisions on where to implement and how to invest in AI and ML, the consequences of which risk losing potential battlefield advantages in semiautonomy and speed of decision making. In fact, section 226 of the National Defense Authorization Act for Fiscal Year 2022 requires a review of "the potential applications of artificial intelligence and digital technology to the platforms, processes, and operations of the Department of Defense," making an understanding of AI and ML applications all the more urgent.[2] In addition, promising investments and research through such organizations as the Joint Artificial Intelligence Center, the U.S. Naval Research Laboratory, and the U.S. Air Force–Massachusetts Institute of Technology AI Accelerator, to name a few, may soon deliver capabilities that commanders must decide how to use. A basic understanding of ML, its dependency on data, and its vulnerabilities reveals fundamental risks and limitations of potential military

applications. These serve to inform a basic model by which a military planner can decide where and how to leverage an AI-based system. This model, along with ensuring data hygiene and tracking the AI systems training currency, can help ensure that military applications do not fall into a trap of letting AI make inaccurate or the wrong kinds of decisions.[3]

Background (ML Basics)

Current implementations of ML train a computer system to do one task very well. For example, current ML systems can read road signs. That said, however, an ML algorithm for image processing that is applied to other data modalities will rarely yield commensurate results. For example, a system designed for another purpose would not be able to use an image processing algorithm to identify human speech with the same confidence. Even when presented with the same mode of data, such as images, a ML algorithm that performed well at identifying one type of image, such as road signs, may not perform well when trained to identify another image, such as trees. According to Kai-Fu Lee, a leader in both U.S. and Chinese AI innovation, implementing a ML system "requires massive amounts of relevant data, a strong algorithm, a narrow domain, and a concrete goal. If you're short any one of these, things fall apart."[4] For this reason, applications outside of image, audio, finance, and signal processing often remain aspirational or topics of research and development. The internet has provided the means to collect the millions of data points needed to develop many of the first robust AI systems.[5] However, this data then required organization and labeling, often by humans, so that an ML system could train. To continue with the example above, a human likely looked at, identified, and labeled

every image of a stop sign that an ML system initially learned from. Ground-truth labeled data on which to train AI systems for military applications often remains scarce. The thousands—if not millions—of already labeled examples of an image of a bird or stop sign reflect a scale of time and resources not necessarily available to label the unique sensor data needed to train military AI systems. Absent finding already labeled data that can translate to the military application, one must implement a system for gathering and labeling new training data. For these reasons, the quality and availability of data present time and resource obstacles to any military application.

When provided with quality data, one can train an ML program under two basic models: supervised and unsupervised. Supervised training requires examples with ground-truth labels (i.e., the correct answer). After classifying new data, the ML model can compare its decision to the correct solution. As it receives more and more training data, the ML algorithm then seeks to minimize the error between its decisions and the ground-truth label. As more data and examples feed through the ML, the accuracy of its response may increase, but too many similar examples can overfit the model, making it misclassify new data on implementation.[6] In unsupervised learning, the ML model does not have an example of correct classification; rather, it groups data into similar categories. Again, the training data set influences how precise these groupings become in their similarity. A supervised learning algorithm may prove more useful when presented with labeled data and seeking a specific classification ability, while unsupervised learning may prove initially more useful to a new data set. It is important to note that mistakes are inherent to both approaches.

Both supervised and unsupervised learning apply to military applications. For imagery analysis, one ML system could potentially identify a particular type of aircraft or version of a vehicle, while a different ML algorithm may only lump all vehicles and aircraft into separate groups. An ML system could also quickly identify a change in an observed environment. This identified change could trigger other sensors or ML systems for a closer look. Current military examples of AI implementation exist. The Air Force recently stated that it has implemented AI to support targeting.[7] Chinese military thinkers view AI and intelligent weapons as potentially decisive technology in future warfare.[8] According to Paul Stockton, China and Russia, with the assistance of AI, "convey microtargeted [information operations] on a massive scale."[9] Intelligence, surveillance, and reconnaissance (ISR); information operations; and imagery analysis provide an example of ready military ML applications.

Beyond building large amounts of organized and digestible data for training, the ML system must reside in a place where it has access to new data and the computing capacity to execute the ML algorithm. As it turns out, graphics processing units, the same thing driving a modern gaming computer, can provide the processing horsepower needed. However, the storage space and bandwidth needed to move and process ML data can make pushing ML systems toward tactical applications challenging. Acknowledging these limitations means that one may soon push applications such as navigation and sensor-to-shooter fires forward to the tactical edge.[10] Other achievable applications, such as complex aircraft maintenance, require significant data collection, development, and testing, followed by a lighter trained implementation at the tactical edge.[11]

Potentially, one could overcome such obstacles by training the ML system on a supercomputer and implementing the trained algorithm forward, possibly aided by dedicated cloud architectures. Perhaps mesh networks and distributed computing approaches will help to overcome this challenge. However, even to implement these solutions would require a massive overhaul of current tactical data communications. The ability to access and process data will determine where AI applications take place on the battlefield. Without the ability to push and pull real-time data at the tactical edge, ML battlefield implementations will remain limited.

Vulnerabilities (Good and Bad Data)

Beyond requiring a large quantity of data to train an ML system, the data must also be of good quality. *Quality* means that the data is taken from a diverse set of sources under varied circumstances and then labeled and presented in a manner that the ML system can ingest. More importantly, the ML implementation must ensure the integrity of the data and the accuracy of the labeling. The old adage "garbage in, garbage out" rings true for ML, and false positives or other undesirable results can proliferate in the absence of high-quality curated training data. Several research examples have demonstrated this against ML image processing algorithms. Simply adding noise or slight changes in hue and brightness to pixels in a photo can force a misclassification, even though the image appears the same to the human eye.[12] Likewise, altering an object physically by placing stickers on a stop sign has forced an ML system to misclassify the sign.[13] (A human, on the other hand, would still perceive the stop sign and ignore the stickers.) Similarly, Google has demonstrated that a simple "patch" placed in the

corner of an image prevents proper ML classification.[14] While these examples focus on image applications of ML, they illustrate the fragility of ML systems in training and the importance of data integrity used in any ML system. These problems are tolerable for many commercial applications, but they take on added salience in a military context.

This dependency on data introduces two main approaches to subvert an ML system. An attacker can either poison the data before the system learns or present pathological data to a trained system.[15] Pixel manipulation and image patching could provide a means to attack during the model training phase (i.e., preoperation). Physically placing stickers on a stop sign or a poster in a room could serve to attack a trained ML system in operations. The military practitioner can easily hypothesize tampering with data before and after an adversary's ML system trains on it. These possibilities range from conducting a cyberattack on a database to applying simple and consistent means of camouflage to ensure that a signature in training does not match the one used in operations. In addition, deviations from normal ways of employing weapons systems could potentially influence an ML system's classification results. Any military application of ML must protect against such attacks by emphasizing the importance of ensuring that data remains unmanipulated, that it comes from reliable sources, and that it is properly labeled. All the while, a military AI system must have a sustainment plan for updating trained models with new relevant data sources as they become available. Such maintenance may be hard to achieve in the compressed timeframe of battlefield operations.

To protect against such possibilities and to counter adversary applications of ML, data management becomes paramount. First, the U.S.

Department of Defense (DOD) must develop a means to track friendly data exposures in a labeled and organized fashion. This means tracking every time a friendly capability may reasonably have exposure to adversary ISR or is leaked unwittingly through espionage, hacking, industry, or the press. It is equally important for friendly forces to maintain a picture of what data they can expect an adversary to have. The adversary may have access to quality surveillance data or even an ML model that the DOD has implemented. With this dataset and the associated analysis, one could potentially wargame and account for possible adversary ML-based capabilities.[16] After a technical wargame, AI experts could arrive at recommendations for camouflage, deceptions, or even data attacks. Alternatively, with the knowledge of what ML techniques an adversary uses, coupled with this friendly data, one could estimate friendly vulnerabilities or reveal predictability in ML-based decisions. Consequently, understanding adversary AI and ML algorithms should be a priority. This method of data tracking should help identify what items friendly forces should keep concealed until their use on the battlefield and guide their effective application.

However, identification of critical data alone will not be enough. Military Services must also provide the means for training that allows for technical signature management. Having more ready access to secure facilities and radio frequency (RF)-shielded hangars or training only during appropriate light and cloud-cover conditions are all concepts that units must implement. One cannot erase old data collected by an adversary, but as the DOD embarks on new capabilities, U.S. forces can manage their data exposure or change signatures moving forward. Simply changing the shade of paint on an aircraft or slightly varying an RF form factor could mitigate

past signature exposures. Such capabilities as expanded Joint Strike Fighter deployments, high-mobility artillery rocket systems, or new tactical formations such as the Marine littoral regiment should apply all of these principles as part of their fielding and training plans. In summary, the principles to apply include ensuring the integrity of ML training data, protecting and varying friendly signature data, and ensuring the confidentiality of friendly ML implementations.

Where to Implement AI (Risk)

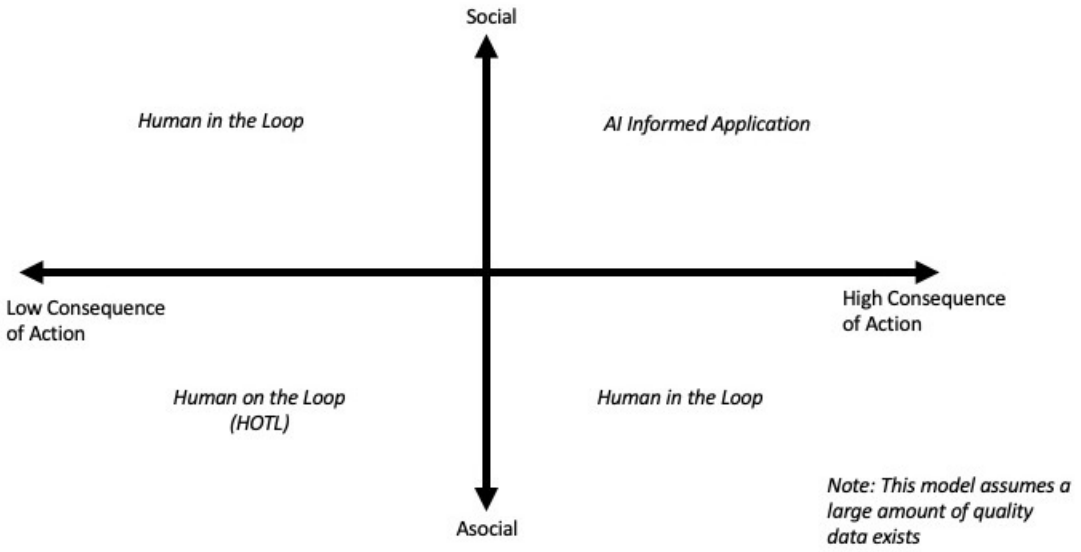
Risk will also drive where the military should implement AI and how much autonomy the system should have. The Massachusetts Institute of Technology (MIT) Lincoln Laboratory presented one model to describe the AI "domain of impact," which attempted to define existing commercial and military applications of AI. In this model, a low consequence of action application included using AI for a robot vacuum, while a high consequence of action application placed lives at risk, such as medical diagnostics.[17] This model placed AI investments in categories across different levels of available data and consequences of action. For example, based on this model, initially investing in AI to analyze ISR imagery has a lower consequence of action than using AI to directly engage targets. The same data can inform both ISR detection and weapons engagement, but clearly the consequence of action increases with engagement.

Another model relies more on the human element when trying to explain what AI may supplant in the future. Kai-Fu Lee has analyzed the potential human roles that AI could replace. He approaches AI applications from a different perspective, by reviewing if the potential use of AI replaces

a social or asocial role, while still recognizing the requirement for quality data. As an example from the medical field, a psychiatrist's role is highly social while that of a radiologist is asocial. Lee contends that AI investment can most readily supplant asocial roles when provided lots of labeled and digestible data.[18] This point of view agrees with what economists have already predicted concerning automation in the workplace replacing "routine manual and cognitive skills," which do not require great creative thinking or personal interactions.[19] For military purposes, how social a role AI plays relates to the human elements of warfare such as troop morale, political ends, and the will to resist. Using only a social-based model for a military application means that one may consider replacing certain human roles in prioritizing targets and fires with AI. However, the consequence of action for fires would insist that a human remain in the loop.

Combining both the social considerations and consequence of action in a two-axis analysis provides an example of an initial assessment tool of whether a military activity would benefit from an AI application and how much human oversight should remain (figure 1).

Figure 1. Model for assessing a military application of AI as derived from MIT Lincoln Laboratory and Kai-Fu Lee research



Source: courtesy of the author.

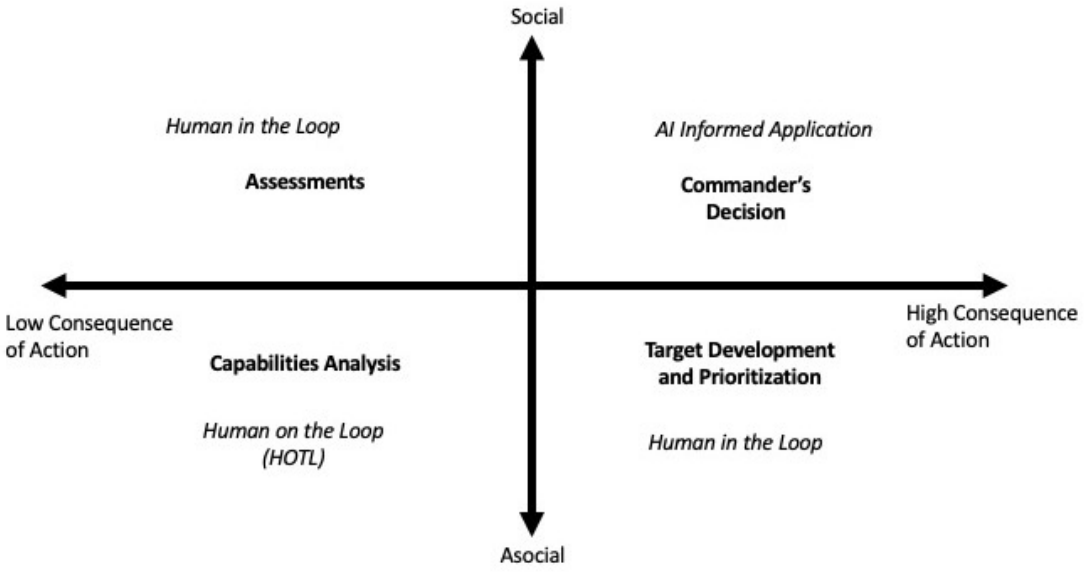
As both Lee and the MIT Lincoln Laboratory emphasize the importance of data, step one of a combined model requires the curation of a large amount of quality, labeled data. The existence of the necessary data and a complete understanding of the existing system only implies that AI may work. Other important technical considerations abound, including but not limited to computational tractability, the precision of optimality descriptions, and the suitability of a learned model rather than a prescriptive, or programmed, solution. Beyond technical considerations, other criteria inform if one should use AI. Applying a two-axis analysis using the consequence of action and the social aspect assesses how one should use AI rather than if they can. A complete assessment of an AI application requires reviewing the availability of data and technical suitability of AI followed by assessing the social and consequence of action elements.

The hybrid model illustrated in figure 1 suggests that the best places for initial military AI application lie with asocial applications with a lower consequence of action. This same lower-left quadrant also offers the opportunity for initial implementations of a human-on-the-loop (HOTL) model. This means that AI would provide battle management options in compliance with the rules of engagement (the legal framework), with the possibility of human vetoing to ensure that an AI recommendation meets ethical requirements.[20] However, absent human intervention, the HOTL AI system will execute the actions. HOTL stands in contrast to a human-in-the-loop model, in which an operator would provide active input to the AI decision process. The more an application lands in the highly social or high consequence of action area, the more a human must remain in the loop before a decision takes place. Finally, one would avoid direct applications of AI in the upper right quadrant of this model, where a highly social activity with a high consequence of action takes place.

Figure 2 illustrates a sample application of this method to common Joint targeting activities. Target development and prioritization, which is largely technical and asocial in nature, results in a high consequence of action, requiring a human in the loop. Conversely, battle damage assessments have a lower consequence of action but a high human-social role when determining what effects the fires had on the adversary. Weapons pairing and capabilities analysis on its own rests in the asocial area, with a relatively lower consequence of action only requiring a human on the loop. A commander's decision has high social aspects and consequence of action and should remain only AI-informed. This simple application serves as an illustration and provides a relative assessment

between the examples. While this model provides an example of how one may explore where to apply AI, many others exist as well. Using any such model, a complete analysis of technical suitability of AI for any application along with the quality of data must take place first.

Figure 2. Sample application of the AI assessment model to common Joint targeting activities



Source: courtesy of the author.

Using AI (Trust)

This AI application model alone does not answer the important question of trust in the AI system. By its very nature, AI produces a result without the user (or even the designer) knowing completely why the AI made a decision. This "black box" leaves significant ethical and trust gaps. For this reason, technical research into providing risk controls and representing uncertainty pushes forward.[21] While the model in figure 1 points toward a good place to

apply AI with the appropriate human interactions, it does not mean that the military practitioner will trust the AI results. The first place to gain trust comes from ensuring data hygiene and integrity, as has been previously discussed. Beyond this, one must have confidence in the performance of the system. [22] A military AI implementation would not always be a static system. As an adversary adjusts its equipment or new sensors come online, the AI system would constantly need exposure to expanded and current data to ensure its classification decisions remain accurate and happen for the correct reason.

Much like military pilots and other specialists must remain qualified to employ their weapon systems, an AI implementation would require a sustained training and evaluation plan. A similar AI reoccurring validation in the performance would involve retraining with updated datasets. This revalidation in performance must take place for any AI system, as current models used do not achieve general intelligence. It is much easier to retrain a human operator to account for a new or novel data input, while an AI algorithm may completely fail to work when introduced to new or novel data. If an AI system were to classify an image as friend or foe, for example, the human in or on the loop would want to know that the AI system uses a current and rigorously tested model. Additionally, anyone retraining an AI system will also wish to compare the current performance to past performance metrics to know if the system has improved or not. Degradation in performance could indicate degradation or even compromise in the data and a need to retrain the system before employing it. Much like a trained military technician needs to keep current

qualifications on a weapons platform, an AI system's currency and proficiency at a task should remain tracked.

These and other challenges led to the establishment of five DOD principles of artificial intelligence ethics: responsible, equitable, traceable, reliable, and governable.[23] More recently, the National Security Commission on Artificial Intelligence called for the National Institute of Standards and Technology to set measures and "tools for qualified confidence in AI."[24] Implementing an AI system in an appropriate area, properly managing data, and ensuring current AI training can all serve to build trust in a military AI system.

Conclusion

As the AI research community begins to deliver capability to the DOD, the uninitiated military practitioner will correctly seek to understand the warfighting implications of AI and ML. However, some military leaders who are in a hurry to implement AI technologies as an end in itself risk failing to understand the basic principles of AI technology. AI serves as a potential means to solve a problem but not always the best one. First and foremost, quality, labeled, and organized data feeds an AI system. In their initial development, AI/ML combat-related applications will likely rely on data from image and signal processing. Furthermore, the current ML constructs prove fragile when exposed to too little or tainted data. Even if provided a sound implementation of AI, a military application now risks becoming predictable to an adversary who observes the AI's use or acquires the same technology. As the DOD embarks on new capabilities and emphasizes signature management in military operations, the management of data will prove

paramount in assessing the employment of AI systems. Applying a sound application model, which accounts for human interaction with AI implementation, will help ensure that military engagements do not become purely data-driven. The use of body counts during the Vietnam War illustrates what can happen when data alone drives military thinking, and implementing AI without attention to human and social considerations must be avoided.[25] The same understanding of data will also inform U.S. AI systems' risk to attack and the nation's ability to counter adversary AI/ML capabilities. By applying AI in the proper areas, ensuring data hygiene, adhering to ethic principles, and tracking systems training currency, the military practitioner can trust AI while mitigating new attack vectors. Applying AI in the wrong ways will open easy attack vectors to an adversary, fail to recognize the human elements of warfare, lead to valuable wasted resources, produce a predictable response, and ultimately fail to create the desired battlefield advantage.

---

[1] *Artificial general intelligence* refers to a yet unachieved form of AI that can learn similar to how a human learns.

[2] National Defense Authorization Act for Fiscal Year 2022, Pub. L. no. 117-81, 135 Stat. 1541 (2021).

[3] *Data hygiene* refers to ensuring that data is up to date, accurate, free of duplicates, and properly formatted.

[4] Kai-Fu Lee, *AI Superpowers: China, Silicon Valley, and the New World Order* (New York: Houghton Mifflin Harcourt, 2018), 10.

[5] Katy Warr, *Strengthening Deep Neural Networks: Making AI Less Susceptible to Adversarial Trickery* (Sebastopol, CA: O'Reilly, 2019).

[6] "Overfitting," IBM [International Business Machines Corporation], accessed 2 June 2022.

[7] David Hambling, "Artificial Intelligence Is Now Part of U.S. Air Force's 'Kill Chain'," *Forbes*, 28 October 2021.

[8] Elsa B. Kania, *"AI Weapons" in China's Military Innovation* (Washington, DC: Brookings, 2020).

[9] Paul Stockton, *Defeating Coercive Information Operations in Future Crises: National Security Perspective* (Laurel, MD: Johns Hopkins Applied Physics Laboratory, 2021), viii.

[10] Alexia Schulz and Pierre Trepagnier, *AI at the Tactical Edge* (Lexington, MA: MIT Lincoln Laboratory, 2021).

[11] "The JCF and the Combatant Commands: A Symbiotic Relationship," Joint Artificial Intelligence Center, 3 June 2020.
[12] Warr, *Strengthening Deep Neural Networks*, 9.
[13] Kevin Eykholt et al., "Robust Physical-World Attacks on Deep Learning Visual Classification" (paper presentation, 2018 Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 19 June 2018).
[14] Warr, *Strengthening Deep Neural Networks*.
[15] Apostolos P. Fournaris, Aris S. Lalos, and Dimitrios Serpanos, "Generative Adversarial Networks in AI-Enabled Safety-Critical Systems: Friend or Foe?," *Computer* 52, no. 9 (September 2019): 78–81, https://doi.org/10.1109/MC.2019.2924546.
[16] Hussein Abbass et al., "Computational Red Teaming: Past, Present and Future," *IEEE Computational Intelligence Magazine* 6, no. 1 (February 2011): 30–42, https://doi.org/10.1109/MCI.2010.939578; and Bryce G. Hoffman, *Red Teaming: How Your Business Can Conquer the Competition by Challenging Everything* (New York: Crown Business, 2017).
[17] Dave Martinez et al., *Artificial Intelligence: Short History, Present Developments, and Future Outlook: Final Report* (Lexington, MA: MIT Lincoln Laboratory, 2019), 10.
[18] Lee, *AI Superpowers*, 155–56.
[19] Georgios Petropoulos, "The Impact of Artificial Intelligence on Employment," Bruegel, 31 July 2018.
[20] BGen Jean-Michel Verney, FR AF (Ret), Col Thomas Vinçotte, FR AF (Ret), and Laurent le Quement, "Human-on-the-Loop," in *Joint Air & Space Power Conference 2021 Read Ahead* (Kalkar, Germany: Joint Air Power Competence Centre, 2021).
[21] Stephen Bates et al., "Distribution-Free, Risk-Controlling Prediction Sets," *Journal of the ACM* 68, no. 6 (December 2021): 1–34, https://doi.org/10.1145/3478535.
[22] Connor McLemore and Charles Clark, "The Devil You Know: Trust in Military Applications of Artificial Intelligence," *War on the Rocks*, 23 September 2019.
[23] C. Todd Lopez, "DOD Adopts 5 Principles of Artificial Intelligence Ethics," Department of Defense, 25 February 2020.
[24] *National Security Commission on Artificial Intelligence: Final Report* (Washington, DC: National Security Commission on Artificial Intelligence, 2021), 137.
[25] Maj Ken Hampshire, USMCR, "Every Marine a Data Scientist?," U.S. Naval Institute *Proceedings* 145, no. 6 (June 2019).