THE LANDPOWER ESSAY SERIES

LPE 19-1 / JANUARY 2019 • PUBLISHED BY THE INSTITUTE OF LAND WARFARE

Training the Machines: Incorporating AI into Land Combat Systems

by Lieutenant Colonel Stephan Pikner, U.S. Army

Introduction

Recent developments in artificial intelligence (AI), or machine learning, have the potential to revolutionize how humans interact with technology. Rather than merely responding to direct inputs in predefined ways, systems that can sort through vast amounts of data and refine their network structures are rapidly improving their ability to predict and categorize. These advances are driven primarily by the massive quantity of information that can be captured and correlated. The more data that is integrated, the more precise the model becomes—whether it is driving patterns on city roads, overhead imagery of farm fields or medical scans of disease-prone organs. AI also has broad military applications, ranging from cybersecurity to aviation maintenance diagnostics to higher echelon military intelligence analysis.[1] Its potential in military applications, however, is sharply limited by the rarity of war. Lacking real-world data to train on and unable to draw from historical cases, AI-enabled land combat systems* may be severely limited in their effectiveness, especially in the critical opening phases of war.

This is not a call to abandon research into military applications of AI. Instead, Army leaders must recognize the limitations of machine learning in particular contexts, especially in situations in which adaptation and the recognition of causal links are critically important. Unquestioning acceptance of AI and the wholesale dismissal of it are both flawed approaches. Effectively integrating these systems into Army combat formations requires a general understanding of how they predict and classify, how they are trained and how they can best complement the innate strengths and adaptability of our nation's warfighters.

What Drives AI Predictions and Classifications?

The power of AI lies in its ability to pull from increasingly massive amounts of data collected from a range of sources and discern correlations between variables to predict behavior. As the amount of data grows, so does the power of these predictions. One example, if controversial: By integrating closed-circuit television image data, mobile phone eavesdropping and retinal scans at police checkpoints, the Chinese regime has built the powerful and deeply intrusive Integrated Joint Operations Platform system to monitor the ethnic minority Uighur population in Xinjiang Province. The data predicts people who demonstrate patterns of behaviors deemed threatening to the regime, which then arrests those people and places them into re-education camps.[2] More widely, technology firms are routinely mating their increasingly fine-grained inventory of human characteristics—captured through the integration of data generated by mobile phone applications, publicly accessible financial and demographic characteristics, and consumption patterns—with deliberate variation of product characteristics. This integration allows internet retailers and social media platforms to use machine learning techniques to sharpen and personalize their business models in ways that were previously impossible.

* Land combat systems refers to tactical-level weapons platforms, such as armored vehicles, helicopters, air defense or indirect fire systems, whose effectiveness may be improved with artificially intelligent navigation, targeting, self-defense and maintenance diagnostic capabilities.

AI also excels at classifying objects. By fusing input from an array of sensors and comparing it with a library of known characteristics, AI models can rapidly and accurately categorize vast amounts of data. Ambiguous cases may be sent to humans for adjudication, and these decisions are used to further improve the computer model. Most cybersecurity defenses rely on machine learning classifier systems to detect and counter malicious software. These systems are trained on data samples that simulate cyberthreats, and through this training, the systems refine their models to more accurately classify, for example, spam and legitimate email traffic.[3]

Both of these AI abilities—predictions based on hidden correlations in vast swaths of data and the categorization of objects—are driven by Bayesian updating. Bayesian processes center on weighing prior beliefs against the strength of new information to generate a posterior estimate. This balance of existing beliefs about the likelihood of an event and empirical evidence is critical since evidence is rarely certain. For example, a positive medical test does not necessarily mean that the patient has that disease—in fact, given the rarity of most illnesses and the general bias of tests to generate false positives over far more dangerous false negatives, a single positive result only marginally increases the probability that the patient is ill. The presence of additional risk factors can refine the prior belief about the probability of having the disease, and subsequent, independent tests can further update the likelihood that the patient is sick. The final determination of that patient's health is then used to refine the larger model's initial assumptions about the prevalence of the disease in the wider population.[4]

The refinement of this model of prior beliefs and new information is the learning process that drives AI. Each iteration of classification or prediction sharpens the prior beliefs that underpin the model. The model must begin somewhere, though. In Bayesian terms, this can be a set of "flat priors," or neutral initial beliefs. Increasing experience iteratively strengthens the prior beliefs of the model, enabling it to more accurately incorporate specific information into a prediction about the true classification of an object. For example, autonomous vehicles are trained on real-world streets to get experience driving through the sometimes-ambiguous environment of actual traffic. By comparing the prior, assumed behavior of an object (such as a bicycle) with its observed behavior, the model refines its prior beliefs about bicycles in general.

Advantages of Humans over AI in Making Connections

Unlike bicycles on the road, though, war is rare. The amount of data needed to train artificially intelligent land combat systems is simply nonexistent, and without this context, such systems will not live up to their potential in the critical, opening phases of conflict. This is true, to an extent, of Soldiers as well—training in peacetime is only a rough approximation of combat. Humans have three advantages in this regard, though.

First, they can buy into the training scenario. For example, an infantry squad at the National Training Center (NTC) in Fort Irwin, California, knows that the "building" they are clearing is really a stack of shipping containers, but they accept that artifice and can train effectively on the relevant tasks.

Second, humans draw on cues in the operational environment and through historical study to update their prior beliefs. A patrol in a typically bustling, peaceful village market will intuitively increase its security posture if all the shops are inexplicably closed one day, even without specific prior experience of being ambushed in that setting. This heightened awareness may not be from a particular training scenario or real-life experience but could come from the squad leader having read about similar circumstances in a war in the same region a hundred years ago. The human experience of war throughout history has been captured in countless stories and books, but such knowledge cannot easily be translated into specific training data for an artificially intelligent combat system.

Third, and most critically, humans can reason causally, rather than solely through correlations. This is particularly important in complex environments with vast amounts of extraneous data or in unique ones where there isn't enough preexisting information to form meaningful prior beliefs. For example, an autonomous vehicle trained in a controlled environment without public alcohol consumption may classify a person holding a vodka bottle and staggering across the road as a normal pedestrian rather than a person who is inebriated. While humans can quickly connect the causal logic between a half-empty bottle of alcohol and inebriation, machines that learn solely through correlation may be overwhelmed by every other possible measurable variable and miss the causal link between the bottle and the behavior.[5]

The inability of AI to reason causally can result in fundamentally mistaken findings. In a recent case, a machine-learning-based decision support system used to explore mortality rates among pneumonia patients found that having asthma increased the chances of survival. This counterintuitive correlation missed the causal mechanism behind the lower mortality. The protocol for treating asthmatic pneumonia patients in the hospitals studied was to admit them directly into the intensive care unit (ICU), where they received a much higher level of care. The ICU was the causal factor in their increased survival, not having asthma.[6]

Complications for AI in Land Combat Systems

While the asthma mistake was quickly discovered by doctors who reviewed the decision support system's output and were able to trace the relatively simple process by which it reached this odd and mistaken conclusion, the complications for artificially intelligent land combat systems making such mistakes are threefold.

First, while pneumonia patients enter hospitals at a relatively constant rate, and medical staff are generally familiar with their circumstances, the relative rarity of war and variations between conflicts make finding such obvious machine-learning mistakes much harder. These mistaken correlations may then be added into the larger, mostly opaque model. Even "human in the loop" controls that are intended to act as a brake or safety on AI may fail, as the people entrusted to veto the machine's decision can be as fallible or overwhelmed as the system itself.[7]

Second, in more complex neural networks—systems that form hidden layers of links between the observable input and output layers—the "reasoning" steps are unobservable. The network is designed to evolve on its own as it learns, creating new links and adjusting probabilities based on updated beliefs about the accuracy of new information and the environment itself. With a limited set of inputs and an objective, attainable "right answer" (e.g., that metallic red octagon along the road is, in fact, a stop sign), this is less of a problem, as the machine reconfigures its neural network in the case of a missed classification as part of its training.[8]

Though this process of deep learning has resulted in incredible progress in classification accuracy and speed, the driving logic is still correlational. Furthermore, it is opaque; programmers do not know why machines generate certain outputs, making the diagnosis of odd results—such as the lower mortality rate among asthmatics with pneumonia—nearly impossible.[9] This weakness is exacerbated, again, by the rarity and variation of land combat. An AI-enabled land combat system based on an opaque neural network trained either in a peacetime simulation or in another, now irrelevant operational context not only will generate poor outputs but also will be coy as to how those conclusions were reached. There is a real risk that such a system overlearns from irrelevant training and fails to adapt rapidly enough.

Multiple classifier systems, which incorporate and weigh several classifier models, may help overcome this weakness. In such a system several classifiers, all observing the same object, each independently assesses it using their own model. Their varied results are integrated by a fuser, which weighs the reliability of the various models to return a decision.[10] In the context of AI-enabled land combat systems, adjustments of the fuser to account for changes in the operational environment or evolving enemy tactics, techniques and procedures may increase the reliability of the entire system. The inputs from previously effective classifiers, trained in another context, may be less reliable in a unique context, and preemptive rebalancing of the weights allotted to each independent classifier may result in a more capable overall system. This process is akin to one that routinely happens on military staffs: Faced with a unique problem, leaders will adjust whose recommendation they rely on. The opinions of those with deep but narrow experience in a particular operational context will be less valuable to the commander in a completely new operational context. Multiple classifier systems function in a broadly similar way, allowing for a degree of control over the opaque neural networks that make up each independent classifier.

Third, in addition to the scarcity of combat training data and the opacity of deep learning neural networks, AI-enabled land combat systems must deal with the challenge of an adaptive adversary that does not want its forces to be readily classified as such. In contrast to street signs or pneumonia mortality, where the true class of the object is ultimately knowable and is compared with the prediction to iteratively refine the model, enemies hide. This ambiguity limits how strongly machines can learn real-life lessons.

Training an artificially intelligent system in a controlled environment such as the NTC, where the machine's classifications of opposing force targets can be updated with surety, may result in overly strong prior beliefs of

enemy characteristics. The more ambiguous evidence about the enemy's signatures and locations in real-life combat may struggle to overturn these strongly formed prior beliefs. Systemic cases of false negatives, in which both the human trainer and the AI classification system fail to correctly identify a real-life threat, may create large blind spots in the ability of Army systems to find adversary forces.

In a sense, this is not new—militaries have been using deception and camouflage for millennia. Adversary exploitation of weaknesses in classification by friendly deep neural networks is a likely next step and may result in AI systems being spoofed by manipulated inputs.[11] Again, this is not itself novel—chaff was developed to counter radar, for example. What is central is that warfighters recognize this potential and understand the limitations and remedies, all while not dismissing outright the valuable potential of artificially intelligent systems to Soldier effectiveness.

Conclusion

War in the land domain is complex and ambiguous, for intelligent machines as much as for humans. Breakthroughs in AI in peacetime applications and other warfighting domains may not easily and reliably translate to ground combat, and an overestimation of the power of machine learning may lead to, at best, disappointment. To effectively integrate these powerful but limited tools, leaders must understand their innate logic and inherent constraints and train both their Soldiers and machines to complement each other's strengths. This process of adaptation, training and refinement will be most critical in the earliest phases of combat in which AI systems are still largely basing their decisions on neural networks and prior Bayesian beliefs formed on, essentially, the last war.

This is not solely a problem for acquisition officers, engineers or computer scientists. Instead, it is fundamentally a task for combat leaders employing these systems as powerful complements for Soldiers in their units. However, the problem is not insurmountable, and by complementing the potential of AI with the skills, adaptability and creativity of our Soldiers, the Army could dramatically improve its lethality, survivability and effectiveness on the battlefields of tomorrow.

★ ★ ★

Lieutenant Colonel Stephan Pikner is an Army Strategist currently studying at Georgetown University as part of the Advanced Strategic Planning and Policy Program.

Endnotes

1 James Mingus and David Dilly, "On Warfare and Watson: Invest Now to Win with AI," *ARMY*, 21 August 2017, https://www.ausa.org/articles/warfare-and-watson-invest-artificial-intelligence.

2 "Inside Xinjiang: Apartheid with Chinese Characteristics," *The Economist*, 2 June 2018.

3 Patrick McDaniel, Nicolas Papernot and Z. Berkay Celik, "Machine Learning in Adversarial Settings," *IEEE Security & Privacy* 14, no. 3 (2016): 68–72.

4 Judea Pearl and Dana Mackenzie, *The Book of Why: The New Science of Cause and Effect* (New York: Basic Books, 2018), 104–107.

5 Pearl and Mackenzie, *The Book of Why*, 31.

6 Federico Cabitza, Raffaele Rasoini and Gian Franco Gensini, "Unintended Consequences of Machine Learning in Medicine," *Journal of the American Medical Association* 318, no. 6 (2017): 517–518.

7 Sydney Freedberg, Jr., "Why a 'Human in the Loop' Can't Control AI: Richard Danzig," *Breaking Defense*, 1 June 2018, https://breakingdefense.com/2018/06/why-a-human-in-the-loop-cant-control-ai-richard-danzig.

8 McDaniel, Papernot and Celik, "Machine Learning in Adversarial Settings," 70–71.

9 Pearl and Mackenzie, *The Book of Why*, 360.

10 Michał Woźniak, Manuel Graña and Emilio Corchado, "A Survey of Multiple Classifier Systems as Hybrid Systems," *Information Fusion* 16 (2014): 3–17.

11 Nicolas Papernot et al., "Practical Black-Box Attacks against Machine Learning," *Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security* (2017): 506–519.