# Coding Just War Theory:
# Artificial Intelligence in Warfare

***by Dana Gingrich***

Leveraging Artificial Intelligence (AI) to improve decision making is not a new concept. The proliferation of AI, however, makes this period in history a strategic inflection point. Tomorrow's wars will be fought and won using AI. The United States military must understand this technology and utilize AI to retain the advantage. Today's leaders will be responsible for integrating AI within the current battlefield and must therefore consider the impacts of employing AI within the principles of Just War Theory.

The U.S. military is at the threshold of another frontier to determine how to synchronize multiple domains—including space, cyber, information, and others—to avoid decisive conflict or to defeat potential threats. AI has demonstrated the ability to develop superior strategies never considered by humans; AI should inform our strategic options. There are ethical implications, though, on how AI determines optimum strategies. Humans currently define the variables and parameters for the machine and the machine optimizes based on a given criteria. The computer's objectivity could eventually counter the accepted principles of *jus en bello*, principles for waging war, which would require strategists to constrain the power of AI.

Our adversaries are constantly looking for opportunities to expose and exploit our critical vulnerabilities. What if our adversaries are willing to code different rules? Our adversary's AI could develop strategies that put the United States in an untenable position. In such a situation, if the U.S. does not have the ability to fight and win, then we must consider additional ethical implications for *jus ad bellum,* principles for going to war. The U.S. must find a new balance between upholding Just War Theory and leveraging the full power of artificial intelligence to fight and win in this new era of warfare.

**Major Dana Gingrich currently serves in the 75th Ranger Regiment, U.S. Army Special Operations Command. He earned an M.B.A. from Stanford Graduate School of Business and recently graduated as the Gen. George C. Marshall Distinguished Honor Graduate from the U.S. Army Command and General Staff Officers Course.**

## The technology is here. What is to be done about it?

In 2015, Google's DeepMind Artificial Intelligence Lab sought to develop a computer program to learn the Chinese board game Go, more than 2,500 years old and widely considered the world's most complex strategic board game. Within two years, the DeepMind computer program, AlphaGo Zero, defeated the Go world grandmaster two games in a row, evidencing the power of advancing technology.[1] The competition yielded two amazing outcomes. First, the computer taught itself. After researchers coded the Go rules into the program, AlphaGo Zero determined the optimal strategy by playing itself over 4.9 million times, learning from its experience. Second, and most importantly, AlphaGo employed techniques that no human player had ever considered.[2]

**While recognizing AI's superior strategy, we must consider the inherent limitations of defining all necessary parameters for a wartime scenario.**

To achieve this overwhelming feat of technology, DeepMind employs a deep neural network AI algorithm that allows the system to learn unconstrained. Traditional computer analytics rely on large data sets of previous human experiences that introduce a bias into the AI system. By only coding the Go rules into AlphaGo Zero, the computer scientists allowed the computer to learn in an unconstrained and unbiased environment. DeepMind has replicated the power of this technology with a new program, AlphaZero. Scientists programmed AlphaZero with the rules for Chess, Shogi, and Go. The computer produced similar results in each strategy game, innovative strategies never previously considered and undeniable defeats over a human grandmaster. AlphaZero produced these results in "9 hours for Chess, 12 hours for Shogi, and 13 days for Go."[3]

While recognizing AI's superior strategy, we must consider the inherent limitations of defining all necessary parameters for a wartime scenario. AlphaGo and AlphaZero developed strategies within a structured game based on programmed optimization criteria. War is not that precise. The military defines end-state with respect to friendly forces, enemy forces, terrain, and civilians. The computer programmer would have to input all the cross-domain capabilities available to strategists, the "rules" of the game, and a weighted value for each of the end states. AI would generate vastly different strategic approaches depending on the application of the resources and the prioritization of the end states. This does not mean that the different approaches are irrelevant, but it does create an inherent risk to blindly accepting AI's solution.

## What does an AI enhanced military strategy look like?

In 1982, Stanford University computer science professor, Douglas Lenat, decided to participate in Traveller, a well-known naval war game competition in Silicon Valley, California. The war game provided a budget for each team and more than 140 factors to consider when creating its fleet. The teams could allocate funds for whatever attributes (armament, protection, maneuverability) they deemed crucial for their fleet to survive and win.[4]

Lenat had no previous military experience and was competing against highly respected strategists from across the political-military spectrum. He did, however, understand the power and capabilities of AI. Lenat developed a computer program, Eurisko, to consider the value of all factors when determining the optimal strategy. Eurisko ran thousands of simulations learning the strengths and weaknesses of different fleets. It then calculated the optimal

size, speed, and lethality of the naval fleet given the constrained budget. Lenat entered the competition with a 96-ship fleet of small, defensible vessels. The average competitor, with expertise in military strategy, had 20 well-armed, technologically advanced warships. Eurisko recognized this and fought a battle of attrition, losing 50 ships but also destroying all but one of the competitor's.[5]

Lenat competed with Eurisko again the following year. The Traveller commission wanted to prevent the computer from winning decisively, so they did not release the game parameters until the week before the competition. The updated rules placed more importance on maneuverability than protection. In that week, Eurisko developed a new strategy that resulted in scuttling its own damaged ships to maintain tempo and speed. Eurisko, again, achieved victory.[6]

The computer objectively determines the optimal way to achieve the end-state given the current means, as demonstrated by Eurisko. Although the computer developed the superior strategy two years in a row, it never considered the second and third order effects of its "winning" strategies. In the second scenario, what happens to sailor morale when the commander begins scuttling friendly ships? What happens to national resolve? The computer's strategy might win the tactical scenario and lose the war.

## What are the ethical implications for the United States?

The United States military accepts Just War Theory as foundational to professional military ethics. Two tenets of Just War Theory are *Jus ad Bellum* and *Jus en Bello,* the principles for going to war and the principles for waging war respectively. Analysts generally consider eight components of *jus ad bellum*, discussed later in this paper, when studying a nation's decision to go to war. Analysts then consider two components of *jus en bello*, discrimination and proportionality, when studying conduct in war.[7] How can the military strategist blend the power of AI with the accepted principles of Just War Theory?

Computer scientists developed AlphaZero and Eurisko to "conduct war" in their respective strategy games. Although computer scientists used different AI techniques to develop Eurisko and AlphaZero, the systems acted with considerable similarity. The AI systems considered the rules in the form of parameters, the variables to be acted upon, and the end result in the form of optimization criteria. Both systems found the most efficient and effective strategy to achieve that end-state and were successful. Military leaders must leverage this power to develop innovative solutions to multi-domain operations. Today's leaders must consider how to integrate space-based systems, offensive and defensive cyber, and joint service capabilities to deter potential adversaries or to defeat threats.

**The value of AI technology for the military is not only the capability to rapidly assist decision-making but also to develop unbiased approaches to future challenges.**

The value of AI technology for the military is not only the capability to rapidly assist decision-making but also to develop unbiased approaches to future challenges. Both AlphaZero and Eurisko learned in an unconstrained environment, meaning they taught themselves based on simple parameters. The output for Eurisko was a winning strategy, but also one that scuttled friendly ships without stopping to save the lost crew. For AI to be a viable tool in the U.S. military, the strategy must be scrutinized under Just War Theory.

Since these AI systems would be developed to "conduct war," they must consider proportionality and discrimination.

Proportionality in Just War Theory means military actions should not cause unnecessary destruction. Considering only proportionality, an unconstrained AI solution could cripple the adversary's critical infrastructure and economy with a precisely launched cyber-attack, winning the war without firing a shot. This would potentially optimize for minimizing physical destruction but would not consider the impact on non-combatants. Discrimination, as the second tenet, means military actions should target combatants. Optimizing for discrimination could result in a drawn-out conflict focused solely on combatants that results in numerous military casualties. The subjectivity of these two tenets creates a significant challenge for an objective computer. The current military leader must simultaneously consider tenets of *jus en bello* and utilize judgment to weigh the costs in both unnecessary destruction and non-combatant casualties.

**The United States faces a major challenge to codify Just War Theory in AI.**

Although objective in decision-making, AI could still consider both proportionality and discrimination as parameters to be addressed in recommending a strategy. As demonstrated with AlphaZero and Eurisko, AI has superior intuition across a range of strategic scenarios against leading human experts. Its ability to rapidly consider thousands of possible outcomes in making decisions puts AI's human opponent in an almost unwinnable situation. This gives a nation with such technology a clear advantage in warfare.

Moving forward, the international community will have to consider this power asymmetry when incorporating AI into wartime scenarios and would most likely seek to establish an international norm that all potential belligerents should follow. Proponents of constraining AI are currently petitioning the United Nations Security Council to consider international standards for AI's application in warfare.[8] Universal standards for AI's consideration of non-combatants would support Just War Theory and *jus en bello* principles, and those standards would prevent military leaders from simply toggling the two tenets to determine an outcome when considering *jus en bello*.

The United States faces a major challenge to codify Just War Theory in AI. Because AI has the power to put its opponents in an untenable situation, our adversaries may not be willing to program the same rules. Does this possibility add more weight to a solution from the international community?

## Can international norms constrain AI's application in modern warfare?

> *So in war, the way is to avoid what is strong and to strike at what is weak. – Sun Tzu*[9]

Predicting a solution's potential breach is no reason to fail to implement a solution in the first place. International treaties and norms have been established across the range of science and technology to guide individual behavior for the collective good. Establishing an international norm that forces AI algorithms to consider proportionality and discrimination would preserve just conduct of war, but it would also constrain the power of AI to determine a range of innovative solutions to fight and win. The United States must consider the latter part when assessing a potential belligerent's willingness to constrain its AI systems.

Two recent violations of international norms should give the United States military pause when considering constraining the power of AI. In 2007, the Chinese military shot down Tiangong-1, a weather satellite, at 500 miles above the Earth, which drew sharp criticism for the violation of international norms established

in the Outer Space Treaty of 1967.[10] The incident spread thousands of debris particles which prevented the further use of that orbit and demonstrated a Chinese military capability that threatened U.S. satellites.

In 2018, Chinese researcher, He Jiankui, published a paper claiming to have successfully edited the human genome of two embryos later born as twin girls. Although He Jiankui was attempting to alter a gene that would prevent these girls from ever contracting HIV, his human gene-editing experiment gravely violated scientific norms.[11] At the Second International Summit for Human Genome Editing, scientists from across the globe rebuked He's violation as "a failure to meet ethical standards for protecting the welfare of research subjects."[12] There is, however, an encouraging aspect to these two violations of international norms: in both instances, the Chinese people strongly opposed the violations. In 2007, the Chinese government received backlash from its citizens for violating international space norms, and in 2018, a group of 122 Chinese researchers signed an open letter criticizing their colleague's actions.[13] The question remains would this internal criticism exist or exert enough force to constrain China's use of AI against an existential threat in warfare?

Sun Tzu, an ancient Chinese military strategist, stated that "in war the victorious strategist only seeks battle after the victory has already been won."[14] AI has demonstrated superior intuition and decision-making against human experts. Humans study the innovative strategies developed by AI, but still cannot sustain that level of thinking to achieve victory. For every human action, AI has a superior counteraction. This has ethical implications for a nation's decision to go to war against an adversary with superior AI capabilities.

One foundational tenet for *jus ad bellum*, the law for going to war, is that the nation has a reasonable hope of success. The criterion for this tenet suggests that "if defeat is inevitable, then avenues other than war should be pursued."[15] The military's role in the United States, though, is to provide civilian leaders with a military capable of projecting national power in the interests of the country. Military leaders must reconcile this tenet of *jus ad bellum* with incorporating AI into warfare. One possible solution to leverage the full power of AI and to uphold Just War Theory is to keep the human in the loop.

**Human-in-the-loop systems incorporate the power of AI with human judgment.**

Human-in-the-loop systems incorporate the power of AI with human judgment.[16] This combined effort must start when assessing *jus ad bellum*. Civilian and military leaders should come together to determine the strategic and military objectives for the computer to optimize against. AI could prescribe a range of actions across the Diplomatic, Informational, Military, and Economic elements of national power that provides leaders comprehensive strategies that achieve the end-state. If the decision is made to go to war, military leaders should leverage unconstrained AI systems to inform operational approaches, especially in an era of multi-domain operations. This assists commanders and staffs in understanding how they synchronize the range of multi-domain capabilities to mass effects while also considering the tenets of *jus en bello*. AI develops approaches that human commanders simply fail to consider. With human-in-the-loop decision-making, the commander then utilizes his or her experience, intuition, and judgment to determine how best to consider proportionality and discrimination when executing AI's operational approach.

## Conclusion

This generation of military leaders will incorporate Artificial Intelligence into warfare. Unconstrained AI systems have developed

superior approaches to strategy games that human experts had yet to consider. With less than a week to learn their respective strategy games, AlphaZero and Eurisko markedly defeated world experts. The computer, though, demonstrated these capabilities in games with clearly defined rules and objectives. Warfare does not have clearly defined rules. Just War Theory strives to provide universal tenets that leaders must consider when deciding to go to war and when conducting war. Incorporating AI into warfare also means considering the principles of *jus en bello*, proportionality and discrimination, understanding that considering additional objectives, such as limiting non-combatant casualties or minimizing unnecessary destruction, would constrain the possible solutions that AI would determine due to the computer's objectivity.

AI's impartiality creates a dilemma for the military leader incorporating AI into warfare. If potential adversaries are willing to allow unconstrained AI systems to inform decision-making, then constraining AI could put the U.S. military in an unwinnable situation. This violates accepted tenets of *jus ad bellum*, the principles for going to war. The U.S. military leader, however, cannot allow unconstrained AI to make decisions on the battlefield without considering *jus en bello*. Therefore, the U.S. military leader must retain the decision-making authority. Military leaders must allow unconstrained AI to inform strategic and operational approaches in multi-domain operations while utilizing their experience, intuition, and judgment when considering proportionality and discrimination. There are times where a commander must prioritize proportionality over discrimination, resulting in more civilian casualties, but this decision is made by a human who contemplates the second and third order effects of the decision made. When first introducing AI into warfare, human-in-the-loop decision-making allows U.S. military leaders to leverage the power of artificial intelligence while upholding the principles of the Military Ethic through the tenets of Just War Theory. ***IAJ***

## NOTES

1 "AlphaGo at The Future of Go Summit, 23-27 May 2017," DeepMind Technologies, accessed October 10, 2018, https://deepmind.com/research/alphago/alphago-china.

2 David Silver, "Mastering the Game of Go without Human Knowledge," Nature 550 (19 October 2017): 354–359.

3 David Silver et al., "AlphaZero: Shedding new light on the grand games of chess, shogi, and Go," DeepMind Technologies, accessed February 15, 2019, https://deepmind.com/blog/alphazero-shedding-new-light-grand-games-chess-shogi-and-go.

4 George Johnson, "Eurisko, The Computer with a Mind of its Own," *The Alicia Patterson Foundation* (1984), accessed October 10, 2018, http://aliciapatterson.org/stories/eurisko-computer-mind-its-own.

5 Ibid.

6 Malcolm Gladwell, "How David Beats Goliath," *The New Yorker* (May 11, 2009), accessed October 15, 2018, https://www.newyorker.com/magazine/2009/05/11/how-david-beats-goliath.

7 Bill Rhodes, *An Introduction to Military Ethics*, (Santa Barbara, CA: Praeger Security International, 2009), 24-25.

8 Janosch Delcker, "France, Germany under fire for failing to back 'killer robots' ban," *Politico*, last modified April 1, 2018, accessed March 13, 2019, https://www.politico.eu/article/artificial-intelligence-killer-robots-france-germany-under-fire-for-failing-to-back-robots-ban.

9 Sun Tzu, *The Art of War*, (Filiquarian Publishing, 2006), 34.

10 William J. Broad and David E. Sanger, "Flexing Muscle, China Destroys Satellite in Test," *The New York Times*, January 19, 2007, accessed March 10, 2019, https://www.nytimes.com/2007/01/19/world/asia/19china.html.

11 "Red moon rising," *The Economist*, January 12-18, 2019, 9.

12 Rob Stein, "Science Summit Denounces Gene-Edited Babies Claim, But Rejects Moratorium," *NPR*, November 29, 2018, Accessed March 4, 2019, https://www.npr.org/sections/health-shots/2018/11/29/671657301/international-science-summit-denounces-gene-edited-babies-but-rejects-moratorium.

13 "Red moon rising," 69.

14 Sun Tzu, 18.

15 Rhodes, 25.

16 Richard Waters, "Artificial Intelligence: when humans coexist with robots," *Financial Times*, October 8, 2018, accessed March 15, 2019, https://www.ft.com/content/bcd81a88-cadb-11e8-b276-b9069bde0956.