An International Journal in Philosophy, Religion, Politics, and the Arts

Existenz

Volume 17, No 1, Spring 2022 ISSN 1932-1066

Artificial Intelligence and The Just War Proportionality Principle

Joseph O. Chapa
United States Department of the Air Force
chapajoe@gmail.com

Abstract: In this essay, I evaluate the relationship between uncertainty imposed by modern applications of artificial intelligence and the *jus in bello* just war proportionality principle. I look first at the structure of the proportionality principle and argue that, whenever possible, military institutions have moral obligations to improve a commander's ability to make accurate predictions about the goods and harms that will result from a contemplated military action. I then address the uncertainty and unforeseen consequences that can result from the use of modern, AI-enabled systems. I argue that, in the face of this potential uncertainty, and based on the previous claim, military institutions have an obligation to reduce the uncertainty that can result from AI-enabled systems in the military context. Finally, I argue that there are two broad means of reducing that uncertainty. The first one is to improve the algorithm's performance; and the second one is to increase commanders' training and education in artificial intelligence technology and operational details so that they can more capably recognize and predict system flaws and failures.

Keywords: Proportionality; just war; *jus in bello*; artificial intelligence; machine learning; uncertainty; evidence-relative; fact-relative; deep learning.

The literature on the ethics of using artificial intelligence for military operations has been dominated by discussions of lethal autonomous weapons systems.[1] To be sure, lethal autonomous weapons systems raise novel ethical questions—about responsibility and accountability and questions concerning whether taking a person's life via lethal autonomous system violates that person's dignity. These questions notwithstanding, the application of artificial intelligence-enabled technology to the military context is much broader than the autonomy of lethal weapons.

One underexplored area in ethical debates regarding the military use of artificial intelligence systems is the relationship between the uncertainty imposed by artificial intelligence applications and the principle of proportionality under the just war tradition—specifically, the proportionality that applies to combat operations or *jus in bello* (as opposed to the proportionality principle that governs the decision to wage war or *jus ad bellum*). In this essay, I will address primarily the moral implications of employing artificial-intelligence-enabled tools within the provisions of the proportionality principle. I look first at the structure of the proportionality principle and argue that, whenever possible, military institutions have moral obligations to improve a commander's ability to make accurate predictions about the goods and harms that will result from a contemplated military action. I then address

[1] The views expressed are those of the author and do not necessarily reflect those of the United States Air Force, the Department of Defense, or the United States Government.

the uncertainty and unforeseen consequences that can result from the use of modern, AI-enabled systems. I argue that, in the face of this potential uncertainty, and based on the previous claim, military institutions have an obligation to reduce the uncertainty that can result from AI-enabled systems in the military context. Finally, I argue that there are two broad means of reducing that uncertainty. The first is to improve the algorithm's performance; and the second is to increase the commander's training and education in artificial intelligence technology so that they can more capably recognize and predict system flaws and failures.

The Proportionality Principle

Scholars disagree as to what, precisely, is the object of the proportionality principle. For instance, some approaches to the just war proportionality principle mirror the international legal proportionality principle. On this view which is supported by Art. 51 of the Geneva Convention, a military action is proportionate if and only if the concrete and direct military advantage anticipated outweighs the loss of civilian lives and harm to civilian property.[2] Others have defined proportionality more broadly to capture all the moral implications of a military action. For instance, Jeff McMahan argues that a military action is proportionate if and only if the weighted moral good to be achieved is greater than the weighted moral costs. McMahan distinguishes between proportionality in the narrow sense and in the wide sense,[3] in both instances, he is concerned with the comparison between morally weighted harms and the good to be achieved. My purpose here is not to adjudicate between competing conceptions of proportionality. Instead, my intent is to show that whatever definition of just war proportionality one adopts, one must weigh moral costs and moral benefits.

Whatever set of metrics one adopts as relevant to the proportionality calculus, there are at least two senses in which one can conduct the proportionality calculus: either with respect to the evidence available to the decision-maker at the time of the decision; or with respect to the outcomes that did, in fact, result from the decision. In Derek Parfit's terms, these are, respectively, evidence-relative and fact-relative determinations.[4] And so, one might ask, is the just war tradition's *jus in bello* proportionality principle applicable in the evidence-relative sense or in the fact-relative sense? The answer is as unsurprising as it is unsatisfactory: it depends on what function the proportionality principle is understood to perform. If the aim is to use the proportionality principle to evaluate the character or blameworthiness of the agent, then one ought to focus on the evidence-relative sense. However, if the aim is to use the proportionality principle to capture the just and unjust distributions of goods and harms that resulted from the action under discussion, then one ought to focus on the fact-relative sense. This apparent ambiguity in whether it is more important to evaluate actions in the evidence-relative or fact-relative sense is not unique to proportionality nor to just war theory. In fact, this dichotomy between evidence-relative and fact-relative evaluations can pertain to any decision that has moral consequences.

There are long-standing criticisms of approaches to military ethics that focus on fact-relative justifications. Suppose a soldier fights on the side of a war that is objectively unjust but suppose that based on all the evidence available to the soldier, he justifiably believes that he fights for the just side. According to Parfit's distinction, the harm the soldier causes is morally permissible in the evidence-relative sense; but morally impermissible in the fact-relative sense. Those in what can be broadly called the revisionist just war theory camp would argue that the soldier's moral liability to defensive harming is grounded in the fact that he poses an unjust threat in the fact-relative sense—he threatens to harm enemy soldiers who have done nothing to give up their right not to be killed. In contrast, those in the traditionalist camp argue that combatants on both sides of the conflict, by engaging in military acts of war, have given up their rights not to be killed independent of the justice or injustice of their cause. Moreover, critics of the revisionist view argue that a theory of

[2] Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts (Protocol I), of 8 June 1977, Art. 51/5/b, p. 266, https://ihl-databases.icrc.org/assets/treaties/470-AP-I-EN.pdf.

[3] Jeff McMahan, *Killing in War*, New York, NY: Oxford University Press 2009, pp. 20-1.

[4] Parfit provides a third category, namely, belief-relative considerations, but this category is not relevant to the present discussion. Derek Parfit, *On What Matters: Volume 1*, ed. Samuel Scheffler, Oxford, UK: Oxford University Press 2011, pp. 150-1.

the just war grounded in fact-relative determinations cannot be "action-guiding."[5] In other words, if what is proportionate is defined in part by evidence that is unavailable to the military commander, then the claim that "military commanders ought not to commit military actions that fail the proportionality test" cannot provide moral guidance for commanders' actions. The moral principle is not, *ex hypothesi*, action-guiding. According to this critique, the normative use of "ought" loses all its moral force. Such a critique can admit that moral principles that focus only on fact-relative considerations might have *ex post facto* explanatory power when one attempts to catalogue the actual distribution of goods and harms, but they cannot help agents to choose the right action *ex ante* because each agent lacks access to the universal set of relevant real-world facts of any given case.

Just war approaches tend to treat fact-relative and evidence-relative determinations of proportionality as mutually exclusive. As I show in a later section, however, once one takes into account the commander's epistemic position, evidence-relative moral judgments can be brought into closer alignment with fact-relative determinations—and this is an important moral project for military organizations.

Proportionality and Uncertainty: Historical Examples

Although artificial intelligence does raise novel concerns about uncertainty and AI technology, to one extent or another, proportionality has always had to account for uncertainty. Anecdotal evidence is easy enough to find. For instance, William Hitchcock argues that on the first day of The Battle of the Bulge approximately 3,000 French civilians had been killed.[6] Whether or not the Allied battle was proportionate is dependent upon whether the good (or the "concrete and direct military advantage") to be achieved was sufficient to outweigh that moral cost. Some might agree that, all things considered, for the Allies this cost was worth paying to prevent German forces from splitting the Allied Forces in two and, ultimately, for the Allies to bring the war in Europe to an end. But General Dwight Eisenhower sent General George Patton's Third Army to reinforce the 101st Airborne Division at Bastogne, not knowing exactly how many civilians would have been killed. Even if reasonable people agree that the Allied Forces' participation in the Battle of The Bulge was *ex post facto* proportionate, Eisenhower could not have known with certainty *ex ante* that the moral benefits would have outweighed the moral costs for he could not have known with any certainty how many civilians would be killed. Eisenhower acted as a military commander under conditions of uncertainty and yet managed to submit to the proportionality principle.

Anecdotal historical cases such as The Battle of The Bulge cannot provide a systematic account of the role of uncertainty in the proportionality calculus for the analysis of each case relies upon counterfactuals and suppositions. There are historical touch points, though, that can provide, if not a systematic account of uncertainty, then at least a perceptible trend in uncertainty and proportionality over time. The history of aerial bombardment is one such example.

The standard measurement for precision in air-to-surface weapons is called "circular error probable" (CEP), which is a circle of some radius that captures fifty percent of the bombs on a given weapons employment.[7] If a particular weapon has a CEP of 100 feet, that means that, on a given weapons employment, fifty percent of the weapons released will fall in a circle with a 100-foot radius. As sensor and guidance system technologies have improved over time, the CEP for the U.S. military's air-to-ground weapons has decreased.

For instance, during the Second World War, U.S. Army Air Forces bombing runs generally resulted in a CEP of roughly 1,200 feet; though that number was reduced to 1,000 feet with intentional training and disciplined flight profiles.[8] During that conflict, U.S.

[5] Henry Shue, "Keeping Exceptions Exceptional in War: Could Any Revisionist Theory Guide Action?" in *Walzer and War: Reading Just and Unjust Wars Today*, eds. Graham Parsons and Mark A. Wilson, Cham, CH: Palgrave Macmillan 2020, pp. 189-214.

[6] William I. Hitchcock, *The Bitter Road to Freedom: A New History of the Liberation of Europe*, New York, NY: Free Press 2008, p. 3.

[7] Technically, a circle in which fifty percent of munitions fall is called "CEP 50." Analysts might also refer to a larger circle in which 90 percent of munitions fall as "CEP 90." For clarity it needs to be stated, I refer to "CEP" and "CEP 50" synonymously throughout.

[8] John T. Correll, "The Emergence of Smart Bombs," *Air Force Magazine* (1 March 2010), https://www.airandspaceforces.com/article/0310bombs/. [Henceforth cited as ESB]

forces had to launch dozens of aircraft, each carrying as many as forty bombs, to strike a single target. For instance, B-17s, B-24s, and B-29s could carry eight, sixteen, and forty 500lb bombs, respectively.[9] As Richard Hallion has put it,

> For example, in the summer of 1944, 47 B-29s raided the Yawata steel works from bases in China; only one plane actually hit the target area, and with only one of its bombs. This single 500 lb general purpose bomb... represented one quarter of one percent of the 376 bombs dropped over Yawata on that mission...It took 108 B-17 bombers, crewed by 1,080 airmen, dropping 648 bombs to guarantee a 96 percent chance of getting just two hits inside a 400 x 500 feet German power-generation plant.[10]

Developments in guidance systems decreased the CEP in each successive decade. In the Korean War, U.S. Air Force crews began employing free fall bombs with a CEP of 750 feet, and with additional training, reduced the CEP to 375 feet. Referring to declassified information, Barry Watts reports that the advent of laser-guided munitions during the war in Southeast Asia brought the CEP down to 25 feet,[11] while dive bombing with free-fall bombs still yielded CEPs as high as 500 feet. Advances in aircraft avionics systems produced drastic reductions in CEP even for free-fall bombs during the 1991 Gulf War. In that conflict, medium to high altitude bombing—even with unguided, free-fall bombs—resulted in a CEP of just 30 feet when measured "in peacetime low altitude training."[12] And precision-guided munitions came to maturity in that conflict. In response to a report by a leading senior military official, the Department of Defense noted that "the CEP for laser guided munitions are measured in feet, not hundreds of feet."[13] As technology allowed for increased precision, military commanders became increasingly reliant upon these precision-guided munitions. In the war in Southeast Asia, just one percent of air-to-ground munitions were precision-guided. By the 1991 Gulf War, that rate had increased to eight percent. By the post-9/11 wars in Iraq and Afghanistan, roughly seventy percent of U.S. air-to-ground munitions were precision-guided (*ESB*). There are, of course, outliers here and there throughout this eighty-years history, but the general trend has been a decrease in uncertainty about air-to-surface munition performance over time.

Even as high-tech militaries reduce CEP and thereby increase precision, these trends do not remove uncertainty from the proportionality calculus. Indeed, since precision is, as a matter of convention, measured with reference to where just half of the bombs will fall, even this measure of precision implies an element of imprecision. On this point Maja Zehfuss writes:

> If a weapon is said to have a CEP of 10 meters, then every other time it is fired in a test the weapon will land within a 10-meter radius of the designated target. In the other 50 percent of cases, it will land *somewhere else*, more than 10 meters away from the target.[14]

Even now, as militaries such as the US rely almost exclusively on precision-guided munitions, a vanishingly small CEP does not imply certainty. In a *Washington Post* Opinion piece, Eliot Cohen cautions that "smart weapons periodically go stupid,"[15] performing in ways that surprise operators and manufacturers alike.

The history of air-to-ground munitions, even if it has trended toward increased precision, has

[9] Benjamin Brimelow, "Here Are the Bombers the US Has Used to Dominate Skies All over the World for over 80 Years," *Business Insider* (10 September 2020), https://www.businessinsider.com/air-force-bombers-b17-b24-b29-b52-b1-b2-b21-2020-9.

[10] Richard P. Hallion, "Precision Guided Munitions and the New Era of Warfare," *Air Power Studies Centre Working Papers* 53 (1995), https://man.fas.org/dod-101/sys/smart/docs/paper53.htm.

[11] Barry D. Watts, *Six Decades of Guided Munitions and Battle Networks: Progress and Prospects*, Washington, DC: Center for Strategic and Budgetary Assessments 2007, p. 9.

[12] James A. Winnefeld, Preston Niblack, and Dana J. Johnson, *A League of Airmen: U.S. Air Power in the Gulf War*, Santa Monica, CA: RAND 1994, p. 127.

[13] Kwai-Cheung Chan, *Operation Desert Storm: Evaluation of the Air Campaign. Report to the Ranking Minority Member, Committee on Commerce, House of Representatives*, Washington, DC: United States General Accounting Office 1997, p. 132.

[14] Maja Zehfuss, "Targeting: Precision and the Production of Ethics," *European Journal of International Relations* 17/3 (September 2011), 543-566, here p. 548.

[15] Eliot A. Cohen, "Five Myths About Cruise Missiles," *The Washington Post*, 12 September 2013.

always included some degree of uncertainty. During the Second World War, commanders could know with certainty that most of their forces' bombs would not hit their intended targets. But, when those bombs missed their targets, commanders could not predict with any certainty to what degree they would miss or in which direction. Bombs that failed to hit their targets would cause unintended harm, but commanders could not know with certainty in advance exactly how much harm or to whom. Even in later decades in which CEP has been reduced to mere feet, when commanders and operators make a proportionality determination, they must do so, not with perfect omniscience about what the weapon will necessarily do, but instead with a reasonable expectation of what the weapon should do, knowing full well that it might do something else.

None of this historical analysis is done to suggest that proportionality has been inapplicable or that the *jus in bello* principles of just war have failed to apply to airpower. Quite to the contrary, the *jus in bello* proportionality principle has been sufficient to govern aerial bombardment, despite technologically driven uncertainties in outcomes. Commanders have had to determine, not whether a considered military action is proportionate under ideal conditions, but whether the military action would be proportionate under non-ideal, real-world conditions. The same is true in the age of artificial intelligence enhanced weapons. Proportionality is sufficient to govern military operations, despite technologically driven uncertainties in outcomes.

Proportionality and Uncertainty: Theoretical Support

As indicated by the brief overview of the history of CEP above, uncertainty is endemic to the proportionality calculus. The proportionality principle requires commanders to weigh the anticipated goods against the anticipated harms a considered course of action will cause (or allow). No person has perfect knowledge of the future and, as a result, no one can guarantee the accuracy of predictions about anticipated goods and harms that might result from a contemplated action. This uncertainty is, perhaps, exacerbated under conditions of war. Carl von Clausewitz remarks realistically regarding uncertainty in war:

> three quarters of the factors on which action in war is based are wrapped in a fog of greater or lesser uncertainty.[16]

Given such fog, military commanders might operate from an even less complete epistemic position than decision makers in other contexts.

There is a trend, though, in recent scholarship on *jus in bello* proportionality that engages in questions of proportionality only from a position of omniscience and not from the epistemically limited position of commanders in the real world. Adil Haque, a critic of this trend, has made the relationship between uncertainty and proportionality explicit. He writes,

> The proportionality rule requires attacking forces to predict the harm that an attack may inflict on civilians, typically in the form of collateral damage estimates (CDEs). On my view, attacking forces must also predict the harm that their attacks will prevent in current or future military operations. Such predictions may seem impractical but in fact are ubiquitous in warfare and reflect a basic skill of responsible command.[17]

Likewise, Patrick Tomlin has made a case for considering subjective proportionality in addition to considering proportionality considerations in the fact-relative sense. Tomlin argues that the philosophical literature on proportionality often assumes that agents will have omniscience as it pertains to the outcomes of their actions or failures to act. He writes:

> Perfect knowledge is assumed about the attack and the defensive options available to the defensive agent. This is unproblematic insofar as we are interested in what Thomas Hurka calls "objective proportionality," where we weigh the actual harm caused against the actual good it achieves. But few, if any, defensive agents (be they private individuals, states, or soldiers) will ever face a violent situation in which they possess perfect knowledge of x, y, and z. Therefore, how we should make proportionality calculations under conditions of empirical uncertainty is an important moral and practical question in the ethics of war and self-defense.[18]

There is a tension here: the proportionality principle that requires commanders to conceive of actual (fact-

[16] Carl von Clausewitz, *On War*, transl. Michael E. Howard and Peter Paret, Princeton, NJ: Princeton University Press 1989, p. 101.

[17] Adil Ahmad Haque, *Law and Morality at War*, Oxford, UK: Oxford University Press 2017, pp. 195-6.

[18] Patrick Tomlin, "Subjective Proportionality," *Ethics* 129/2 (January 2019), 254-283, here pp. 254-5.

relative) harms that will result in the real world, but it requires them to do so from a limited (evidence-relative) epistemic position.

It is here that one can find an important interrelationship between fact-relative and evidence-relative considerations under the proportionality principle. A closer look at a real-world historical case can demonstrate this fact: In 1998, the United States' intelligence community was aware that Osama bin Laden's al Qaeda organization posed a threat. Al Qaeda had already taken responsibility for the August 1998 bombings of United States embassies in Kenya and Tanzania. According to the 9/11 Commission Report, President Clinton authorized cruise missile strikes on two targets believed to be central to al Qaeda's operations. The first was a supposed meeting of al Qaeda leadership in Khost, Afghanistan. The second was a pharmaceutical facility in Sudan suspected of developing nerve gas for bin Laden.[19]

The missiles struck their intended targets; however, bin Laden was not in Khost, but was in Kandahar, and there is considerable debate as to whether the pharmaceutical facility was, in fact, developing chemical weapons. Suppose for the sake of argument that if the intelligence on these two targets had been accurate then the cruise missile strikes would have been proportionate. In such a case, the harm caused would include harm to civilian bystanders who were not liable to be killed as well as harm to bin Laden and other members of al Qaeda, at least some of whom were liable to be killed. The Clinton Administration considered a third target, a bin Laden-owned tannery. The argument for striking the target was that it would hurt bin Laden financially. Again, according to the 9/11 Commission Report, President Clinton himself removed the tannery from the target list

> because he saw little point in killing uninvolved people without doing significant harm to Bin Laden. [*CR* 117]

This decision appears to have been grounded in a consideration of proportionality. The good to be achieved is measured in the prevention of harm al Qaeda would cause in the future. This means that if in 1998 bin Laden had been targeted successfully this might have prevented the terrorist attacks on the United States in 2001 and, ultimately, prevented the United States' war in Afghanistan.[20] Surely, in this case, the weighted good to be achieved is greater than the weighted harm caused.

If the function (or one function) of the proportionality principle is to reduce weighted moral harm, then it matters a great deal whether the military decision-maker accurately predicts the goods and harms that might result from these decisions. Suppose a military commander evaluates two possible courses of action, A and B. Based on the available evidence, the commander believes that both A and B will achieve some morally justified objective and he also believes that A will achieve the objective at the cost of ten noncombatant lives and that B will achieve the objective at the cost of a hundred noncombatant lives. Suppose further the commander justifiably believes that the good to be achieved is significant enough to justify the loss of ten noncombatants, but not significant enough to justify the loss of a hundred. Thus, the commander chooses to execute course of action A on the grounds that A is proportionate while B is not.

However, assume that the commander is mistaken for reasons that could not have been anticipated from the available evidence *ex ante*, and it is in fact A that will cause a hundred noncombatant deaths and B that will cause only ten. Course of action A is morally permissible in the evidence relative sense, but morally impermissible in the fact-relative sense. And *vice versa*, course of action B is morally permissible in the evidence-relative sense, and morally impermissible in the fact-relative sense. If fact-relative and evidence-relative moral judgments are independent of one another, then the argument must end here. There is nothing more to do but to acknowledge the distinction and lament the tragedy resulting from the commander's limited epistemic position.

[19] Thomas H. Kean, Lee H. Hamilton, et al., *The 9/11 Commission Report: Final Report of the National Commission on Terrorist Attacks Upon the United States*, New York, NY: W. W. Norton & Company 2004, pp. 116-7. [Henceforth cited as *CR*]

[20] This is a helpful hypothetical in that it allows the anticipated benefits significantly to exceed the costs. Though I am confident that killing bin Laden in 1998 would have altered the course of the relationship between the United States and al Qaeda, I am not confident that killing bin Laden would have dissuaded al Qaeda from pursuing large-scale terrorist attacks—on the scale of the 9/11 attacks—against the United States.

In the case of *jus in bello* proportionality, evidence-relative and fact-relative moral judgements are linked in that affecting the evidence available to the decision maker can change the resulting facts. In general, making better predictions about the goods and harms that will result from an action can ultimately reduce the weighted moral harm that results. If one begins from an assumption that military commanders intend to submit to the proportionality principle, if military commanders better anticipate the goods and harms that result from their potential decisions, they will make choices that cause less morally weighted harm. Provided that there is little that commanders can do to improve their epistemic positions, then perhaps this claim is uninteresting; but, as I argue below, especially with reference to artificial-intelligence-enabled systems, there are at least some contexts in which a commander's epistemic position might be improved.

Uncertainty and Artificial Intelligence

There are different uses of the term, "artificial intelligence," and the degree to which artificial intelligence imposes uncertainty depends upon one's definition of it. I am focused on a specific sub-discipline within artificial intelligence, namely, deep learning, which began to receive a considerable increase in attention in the second decade of this century. To justify my emphasis on deep learning, a brief review of it will be of some value.

Artificial intelligence ethics arose as a field of study in response to developments in deep learning. Unlike in the previous generations of software that had been called "artificial intelligence," in deep learning neither end users nor developers can predict exactly what the system output will be. This is because deep learning is not deterministic like traditional software, but instead, it is statistical. A deep neural network is trained on a set of training data. The model learns to make predictions based on that training data. Then, when the model is exposed to real-world data, it relies upon the patterns it learned to recognize in the training data to make predictions about the real-world data. The benefit of these systems is that they can identify patterns in the data that humans are likely to miss. The downside is that, because the model might identify as important patterns that are different from those humans would identify, the model can produce surprising results. Sometimes, these surprising results can have ethical implications. A deep-learning milestone was set in 2012 in the context of an object classification competition where a deep neural network called AlexNet outperformed previous deep neural networks:

> We trained a large, deep convolutional neural network to classify the 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into the 1000 different classes. On the test data, we achieved top-1 and top-5 error rates of 37.5% and 17.0% which is considerably better than the previous state-of-the-art...We also entered a variant of this model in the ILSVRC-2012 competition and achieved a winning top-5 test error rate of 15.3%, compared to 26.2% achieved by the second-best entry.[21]

These capabilities set in motion the deep learning revolution that necessitated the establishment of a designated field of ethics to address the ethical implications of unintended and unforeseen consequences that can result from deep learning systems. Since the data output is directly impacted by whatever data is available to the model, unintended consequences can result in not just in outputs that are factually wrong, but depending on the context, also in outputs that are morally significant. The most well-known cases in industry and academia were cases in which the deep neural network produced outcomes that were biased on the basis of gender, race, ethnicity, age, health, social status, wealth, and so on, even if none of the developers intended bias on these grounds.

The same technological phenomenon can occur in the military context, albeit with different implications. Just as surprising factual errors that result from deep learning systems in the civilian context can result in morally significant outcomes, surprising factual errors in the combat context can also result in morally significant outcomes. For instance, suppose a commander employs a deep learning tool that helps the commander to understand the operational environment. If so, the proportionality calculus the commander conducts will be based on, *inter alia*, the commander's interpretation of the AI-enabled system output. If the system generates outputs that are factually wrong, those outputs can affect the

[21] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, *Imagenet Classification with Deep Convolutional Neural Networks*, https://papers.nips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.

commander's expectation of goods and harms that will result from the military action, and thereby to decrease the accuracy of the commander's proportionality calculus. The commander's actions, ultimately, may end up being morally permissible in the evidence-relative sense, but morally impermissible in the fact-relative sense.

If there is nothing anyone can do to improve a commander's epistemic position and thereby to change the commander's evidence-relative moral judgment, there would be nothing more to say. But in the case of uncertainty from deep learning systems, military institutions do have the ability to reduce uncertainty and thereby to bring the commander's proportionality calculus into closer alignment with the actual outcomes. Merely by observing the snapshot in time at which a commander chooses a course of action, the factors that contribute to whether a decision is evidence-relative permissible seem fixed—there is little that the commander can do at that moment to improve the commander's epistemic position. Only if one takes a broader view, one can ask what the military institution could have done prior to that moment of making a battlefield decision, in order to ensure that the commander's *ex ante* predictions about goods and harms are as close as possible to what will be the *ex post facto* evaluation of the goods and harms that result from the chosen course of action.

There are at least two ways in which military institutions can improve upon a commander's epistemic position with respect to deep learning systems, and thereby ultimately to enable the commander to conduct the proportionality calculus more accurately. Both of these methods fall under the category that Tim Rudner and Helen Toner call "AI assurance"; they argue that

> to ensure the safety of a machine learning system, human operators must understand why the system behaves the way it does, and whether its behavior will adhere to the system designer's expectations.[22]

AI assurance is not merely about system performance, but also about the operator's understanding of system performance. To use the preceding discussion of precision-guided munitions and CEP as an analogy, it is not enough merely for commanders to endeavor to get munitions to hit their intended targets. Commanders must also understand the conditions under which munitions may miss their targets and where those munitions might fall. In the case of deep learning, it is incumbent upon commanders to know the conditions under which those models may produce factually errant outputs and how to recognize those factually errant outputs.

The first means by which military institutions can improve a commander's epistemic position with respect to system outputs is to make deep learning systems more explainable—that is, to design deep learning algorithms in such a way that the procedures by which they move from input to output are perceptible and understandable to users. Much has been written on this topic, so I will set it to one side here.

The second way for military institutions to affect a commander's epistemic position is through deliberate education and training in AI. As commanders increasingly understand how the underlying machine learning or deep learning computer science techniques produce system outputs, they will be better able to identify and predict poor system performance. This is not merely a claim that operators should practice with and become proficient with their tools. Instead, this is a more nuanced claim, that practicing with a tool enabled by deep learning neural networks is insufficient, precisely because proficiency in using the tool does not require knowledge of its inner workings. And yet, the ability to predict and identify poor system performance does require knowledge of its inner workings. The same is not true for traditional weapons systems. Commanders do not need to understand modern optics to employ laser-guided missiles; nor to understand thermodynamics to employ thermobaric warheads. But they do need to understand—at least to some degree—how deep neural networks work if they are to improve their ability to identify operational conditions that will negatively affect system performance and, ultimately, to make accurate predictions about the goods and harms that are likely to result from a contemplated military action.

This alignment between a commander's expectation of system performance and actual system performance will make it more likely that courses of action that are proportionate in the evidence-relative sense are also proportionate in the fact-relative

[22] Tim G. J. Rudner and Helen Toner, *Key Concepts in AI Safety: An Overview*, Washington, DC: Center for Security and Emerging Technology at Georgetown University, March 2021, p. 4, https://cset.georgetown.edu/publication/key-concepts-in-ai-safety-an-overview/.

sense. This alignment will enable commanders to choose courses of action that will achieve greater morally weighted good while causing less morally weighted harm. AI education in the military can put commanders in a better position to achieve the practical goal of the proportionality principle which is to reduce harm to noncombatants.

Conclusion

I have argued, first, that the just war *jus in bello* proportionality principle, on any interpretation, consists in a weighing of goods and harms. This is true in at least two senses. In the evidence-relative sense, commanders have an obligation to choose only courses of action that are proportionate—those in which the anticipated morally weighted goods exceed the anticipated morally weighed harms. In the fact-relative sense, there will be a specific distribution of goods and harms that does, in fact, result from the military action, independent in the commander's expectations. An action that is morally permissible in the fact-relative sense might be morally impermissible in the evidence-relative sense. In such cases, military commanders might cause more harm than good in the real world, even though they are morally excused for doing so.

If this relationship between proportionality in the fact-relative sense and in the evidence-relative sense applies, then military institutions have an obligation, insofar as it is feasible, to improve commanders' epistemic situation—to bring into closer alignment the evidence-relative proportionality calculus and the fact-relative proportionality calculus. This approach to proportionality is especially relevant in cases in which commanders employ deep-learning-enabled systems. This is so because these systems can impose an additional layer of uncertainty, over and above the fog of war conditions that already pervade combat operations.

I further argued that in the case of deep-learning-enabled systems, bringing anticipated goods and harms into alignment with actual goods and harms relies upon, among other things, enabling commanders adequately to determine the operational output accuracy of these system. Clearly, the processes through which deep learning-enabled systems arrive at their outputs must be accessible to the comprehension of their operators, and therefore, military institutions ought to provide commanders with AI education to improve this comprehension.

Ultimately, I have argued that, under the principle of proportionality, broadly understood, military institutions have moral obligations to teach their commanders, and those who will become commanders, how artificial intelligence systems work, and doing so to such a degree of competence that military commanders will know when, and perhaps more importantly, know when not, to rely upon deep-learning-enabled systems.