HARVARD LAW SCHOOL
NATIONAL SECURITY JOURNAL

ONLINE ARTICLE

The Fetishization of "The Human" in the Critique of Autonomous Weapons*

Kevin Jon Heller†

* This article is the last in a symposium on Kevin Jon Heller's "The Concept of "the Human" in the Critique of Autonomous Weapons," published in this journal in 2023. It responds to prior articles in the symposium which can be found in the Harvard National Security Journal Online at https://harvardnsj.org/onlineedition.

† Professor of International Law & Security, Department of Political Science, University of Copenhagen (Centre for Military Studies); Special Advisor to the Prosecutor of the International Criminal Court on War Crimes.

Table of Contents

INTRODUCTION

At the beginning of their response to my article, Elke Schwarz and Neil Renic say that "[w]e know and like Kevin."[1] Bo does not say that she likes me in her response, but I'm confident that she does, because we've known each other for quite some time and have always got along well. Regardless, I like all three of them, and I am deeply grateful for their willingness to reflect so deeply on my arguments. Although I doubt that we will ever see eye to eye on the potential of autonomous weapon systems (AWS), both responses help clarify precisely where the disagreements between us lie.

In what follows, I address five topics: (1) technological optimism; (2) the human/machine comparison; (3) the nature of international humanitarian law; (4) "humanity" in war; and (5) human/machine teaming.

I. TECHNOLOGICAL OPTIMISM

Both responses accuse me of being too techno-optimist when I argue that "combat using machines will eventually be more ethical and more humane than combat with human soldiers."[2] Bo focuses on the present, noting "the claimed superior accuracy of AWS is a dubious proposition, as shown by current debates around accuracy rates as well as by the reality of AI-enabled targeting in current conflicts."[3] Schwarz & Renic focus on the future, claiming that the promise of AWS technology "is often wildly overstated and overhyped," because "[s]ystems that employ AI for the full kill-chain are likely to be marred by incomplete, low-quality, incorrect, or discrepant data" that "will lead to highly brittle systems and biased, harmful outcomes."[4]

These points are well taken. We do indeed need to be realistic not only about what autonomous weapons can accomplish now, but also – and more importantly – how AWS technology will develop over time. Given the punctuated equilibrium of technological development, it is impossible to know with any certainty what AWS will be capable of next year, much less in 10 or 20. Moreover, Schwarz & Renic are absolutely right to be skeptical of techno-optimist predictions concerning AWS, given that such predictions are almost invariably "pushed forward by private actors with a direct financial interest in normalizing the technology and lowering barriers for use."[5] States cannot and must not outsource to capitalist corporations their obligation under Art. 36 of the First Additional Protocol to determine whether the use of a particular autonomous weapon "would, in some or all circumstances, be prohibited by this Protocol or by any other rule of international law."[6]

---

[1] Elke Schwarz & Neil Renic, *On the Pitfalls of Technophilic Reason: A Commentary on Kevin Jon Heller's "The Concept of 'the Human' in the Critique of Autonomous Weapons,"* HARV. NAT'L SEC. J. ONLINE 1 (2024).
[2] Kevin Jon Heller, *The Concept of "the Human" in the Critique of Autonomous Weapons*, 15 HARV. NAT'L SEC. J. 1, 4 (2023).
[3] Marta Bo, *Countering the "Humans vs. AWS" Narrative and the Inevitable Accountability Gaps for Mistakes in Targeting: A Reply to Kevin Jon Heller*, HARV. NAT'L SEC. J. ONLINE 6 (2024).
[4] Schwarz & Renic, *supra* note 1, at 7.
[5] *Id.* at 8.
[6] Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts, art. 36, June 8, 1977, 1125 U.N.T.S. 3.

II. COMPARING HUMANS AND MACHINES

That said, stubborn techno-pessimism is no more intellectually defensible than blind techno-optimism. AWS technology may never live up to the claims of its corporate cheerleaders, but the question is not *whether* it will improve – no one seriously questions that it will – but *how much* it will improve.

Moreover, and this is the key point, my argument does not depend on autonomous weapons achieving some fixed, predetermined, and potentially impossible degree of compliance with international humanitarian law (IHL). As the article makes clear, the issue of IHL compliance is strictly relational: using an AWS is permissible in a particular combat situation only when the autonomous weapon is capable of complying with IHL at least as well as human soldiers. In some combat situations, such as close-up fighting in dense urban areas, that may be decades from now – and perhaps never. But in other situations, such as combat in the air and at sea, AWS may well outperform human soldiers in the near future. Do we know precisely when? Of course not. But that is not a problem for my argument, because the burden will always be on the state that wants to use an autonomous weapon to show that the machine/human comparison is satisfied.

My argument, in short, is consequentialist. And because it is consequentialist, it is empirical – neither *inherently* techno-optimist nor *inherently* techno-pessimist. Is the same true of my critics' arguments? I don't know the answer that question, because the responses do not make clear whether the authors reject the very idea of comparing humans and machines or simply don't believe that autonomous weapons will ever, in any combat situation, come out on top of that comparison.

Bo's response suggests that she leans toward the former position, because she claims that "at a more conceptual level, comparing the ability of human combatants to comply with IHL norms with the ability of AWS to do so is problematic."[7] Schwarz & Renic, by contrast, appear more ambivalent. Sometimes they seem to reject the human/machine comparison *tout court*, such as when they when they claim that such a "side-by-side comparison is not productive towards furthering the pressing moral and legal debate on autonomy in weapon systems."[8] Other times, however, they seem to believe that AWS will never be able to comply with IHL as well as human soldiers, such as when they insist that my claim to the contrary "compared, as it must be, to autonomous violence as it *actually* manifests, rather than to an imagined system of perfect rationality."[9]

Insofar as either Bo or Schwarz & Renic simply reject comparing humans and machines, they are subject to same critique I level at scholars such as Leveringhaus and Geiss[10]: namely, that it is ethically problematic to reject AWS *even if* they would cause fewer civilian casualties and less unnecessary combatant suffering than human soldiers. By contrast, if they are simply claiming that AWS will never be able to comply with IHL as well as human soldiers, their position is not problematic at all – because presumably they would, like "soft" deontologist Robert Sparrow,[11] change their mind if it turned out that, despite their techno-pessimism, they were wrong.

---

[7] Bo, *supra* note 3, at 9.
[8] Schwarz & Renic, *supra* note 1, at 6.
[9] *Id*. at 7.
[10] Heller, *supra* note 2, at 18.
[11] *Id*.

As I have noted elsewhere,[12] to answer that (empirical) question we must (1) assess the capabilities of AWS and human soldiers as they stand now; (2) compare those capabilities in specific combat situations; and (3) forecast the best we can how the capabilities of AWS and human soldiers will improve over time. Schwarz & Renic may well be right that I am too techno-optimist when I conduct the analysis – particularly with regard to (3). But neither they nor Bo provide reason to believe that a techno-pessimist analysis, much less an anthropo-optimist one, is more compelling.

Start first with the human side of the equation. Neither response challenges the idea that cognitive and social biases, negative emotions, and physiological limitations profoundly distort human decision-making, particularly in dangerous and uncertain situations like combat. Instead, Schwarz & Renic claim that my "excessively technologized approach pathologizes the human agent in war as not merely flawed, but irredeemably so."[13] That criticism implies they believe human soldiers can in fact be redeemed – or "debiased," to use the cognitive-psychology term – but they neither present evidence that such debiasing is possible nor provide an explanation of why my assertion to the contrary is incorrect. Instead, Schwarz & Renic casually dismiss my critique of human decision-making as "press[ing] human capabilities into an analytical framework fit for … [a robotic system] and analyz[ing] warfare solely as an engineering problem."[14] Yet it is not I who is doing that; I am simply reporting the research of dozens of cognitive psychologists who study how humans make decisions. If that research is not normative enough for Schwarz & Renic's tastes, their problem is with cognitive psychology as a discipline, not with me.[15]

For her part, Bo questions whether it possible to compare the two "[w]ithout clear benchmarks for human decision-making," noting that comparison of "AI errors with human errors in targeting" is difficult if "statistics on the latter are lacking."[16] This is a legitimate concern, because it is obviously impossible to experimentally replicate combat situations with complete verisimilitude. (Psychologists are not permitted to kill their test subjects.) That said, cognitive psychologists are fully aware of this problem and have still managed to study decision-making in a manner that indicates how likely human soldiers are to make mistakes in actual combat; indeed, my article cites a wide variety of such studies,[17] none of which Bo provides any reason to question. We also have extensive statistics concerning how accurate actual soldiers are when they fire their weapons. Bo dismisses "using accuracy rates" because of their "limited usefulness,"[18] but she never explains why that is the case. Minimizing civilian

---

[12] *See generally* Kevin Jon Heller & Lena Trabucco, *Beyond the Ban: Comparing the Ability of Autonomous Weapon Systems and Human Soldiers to Comply with International Humanitarian Law*, 46 FLETCHER F. WORLD AFFS. 15 (2022)

[13] Schwarz & Renic, *supra* note 1, at 6.

[14] *Id*.

[15] Indeed, the statement Schwarz & Renic dismiss as leaving them unsure whether I am discussing "human capabilities, or those of a robotic system," *id*. – concerning how techniques for debiasing bad statistical reasoning have not been shown to work in complex situations like combat – is a quote from a book, not my mechanistic interpretation of the book's findings. *See* Heller, *supra* note 2, at 48.

[16] Bo, *supra* note 3, at 9.

[17] *See, e.g*., Heller, *supra* note 2, at 35 (stereotyping), 38 (anchoring bias), 42 (sleep deprivation), 43 (cognitive overload), 44 (stress).

[18] Bo, *supra* note 3, at 9.

harm in combat requires more than the ability to distinguish between combatants and civilians. It also requires the ability to shoot combatants without hitting civilians instead.[19]

On the machine side of the equation, the responses provide a number of reasons to be skeptical of my argument that autonomous weapons promise to make combat more humane. Many of those reasons are unpersuasive, such as Schwarz & Renic's insistence that "[s]ystem outcomes are inherently unpredictable, and the probabilistic nature of AI reasoning implicitly recognizes error and accident as a feature, not a bug, of the system."[20] That is true, but I am not sure what follows from it. All non-autonomous weapons have error rates, yet militaries still use them. Human soldiers make errors, too. Perhaps the errors of non-autonomous weapons and human soldiers are bugs, not features. Does that matter to a dead civilian? If the unpredictable errors of AWS kill fewer civilians than the predictable errors of human soldiers using non-autonomous weapons that themselves predictably fail, what is the moral argument against using AWS?

To be sure, Schwarz & Renic and Bo reject the idea that autonomous weapons will eventually make fewer errors than human soldiers. Both emphasize, for example, that there will almost certainly be significant problems with the data used to train AWS, making accurate targeting unlikely. Schwarz & Renic thus write that "[a]utonomous systems tend to be built and tested on rather limited samples of data. Sometimes it is synthetic data, and sometimes inappropriate data."[21] Bo, meanwhile, cites a variety of research indicating that "algorithmic bias" is almost inevitable for "AI technologies" because – ironically enough – humans are responsible for "annotating/labelling/classifying data samples, feature selection, modelling, model evaluation and post-processing after training."[22] Even worse, Bo insists – and Schwarz & Renic agree[23] – that the gap between the workshop and the battlefield means that such data problems cannot be adequately resolved through testing. As she writes:

> [A] difficulty with testing is that any "accuracy rate" is developed on the basis of a sample of data, which will not necessarily give an indication of how the same model may function in new circumstances in the future. This is particularly problematic given the complex, unpredictable and dynamic nature of armed conflict in the first place.[24]

Although these are serious criticisms that cannot easily be dismissed, two responses are warranted. The first will come as no surprise to readers of my article: namely, that the criticisms apply equally to humans. As the article discusses *ad nauseum*, human decision-making is riven with biases, both cognitive and social. Those biases are not precisely the same as the biases

---

[19] Moreover, although not a legal requirement, it is always better to disable a combatant instead of killing him. Shooting a combatant in the arm or leg instead of the head or chest requires significant accuracy on the part of the attacker.

[20] Schwarz & Renic, *supra* note 1, at 7.

[21] *Id*. at 7; *see also* Bo, *supra* note 3, at 9 (noting that "it is essential to consider the issue of algorithmic bias. This issue shows how humans influence AI technologies and their output, in some cases ultimately resulting in misidentification of targets.").

[22] Bo, *supra* note 3, at 9 (quoting Ingvild Bode, *Falling Under the Radar: the Problem of Algorithmic Bias and Military Applications of AI*, HUMANITARIAN L. & POL. BLOG (March 14, 2024), https://blogs.icrc.org/law-and-policy/2024/03/14/falling-under-the-radar-the-problem-of-algorithmic-bias-and-military-applications-of-ai/).

[23] *See* Schwarz & Renic, *supra* note 1, at 7 ("Autonomous systems tend to be built and tested on rather limited samples of data… problematic enough, before we even consider the messy complexities of the battlefield.").

[24] Bo, *supra* note 3, at 8.

that affect AWS,[25] but human and machine biases can each lead to violations of IHL, which is what matters. Moreover, testing can no more compensate for the biases of human soldiers than it can for the biases of machinic ones: "given the complex, unpredictable and dynamic nature of armed conflict," it is simply impossible to train soldiers to perform in all the circumstances they may face in combat.[26] So even successful training "will not necessarily give an indication of how the same model [or, in this case, human soldier] may function in new circumstances in the future."[27]

My response is not intended to be flippant. The point is an important one: objecting to the use of AWS on data and testing grounds makes sense only if the data and testing problems with human soldiers are less serious. Not only do Schwarz & Renic and Bo provide no evidence that is the case, common sense suggests otherwise.

To begin with, consider training. There are two basic limits on a military's ability to train combatants (human or machine) to comply with IHL on the battlefield: (1) the number of situations the combatants can be exposed to and (2) the verisimilitude of the training exercises. Autonomous weapons can clearly be exposed to more combat situations during training than human soldiers, because unlike human soldiers AWS do not need to eat, sleep, or visit their families. And it is far easier to make combat situations approximate real-world conditions for AWS than it is for human soldiers, because the inhumanity of AWS obviate the need to recreate the situational factors that are most responsible for bad human decision-making: noise, heat, fatigue, anger, and – above all else – fear. Indeed, the impossibility of exposing human soldiers to a realistic fear of death during training is enough *by itself* to doubt whether human soldiers can ever be trained as well as autonomous weapons to avoid errors.

These considerations are specific to combat training. Critiques of autonomous weapons must also take into account that, in general, some of their limitations are – to quote Schwarz & Renic – "likely to be mitigated as more data becomes available, hardware becomes more sophisticated, and technology generally advances."[28] That is not simply speculation: scientists have successfully debiased AI in a number of non-military contexts.[29]

To be sure, it remains "an open question"[30] (to quote Schwarz & Renic again) whether *all* of the limits of AWS will eventually be overcome, or whether even *enough* limits will be overcome for AWS to be lawfully used in combat. It is nevertheless anything but blind techno-optimism to believe that militaries are far more likely to find ways to improve their autonomous weapons than they are to find ways to improve their human soldiers. Again: human decision-making is difficult to debias in the best of situations and almost impossible to debias in dangerous and uncertain situations such as combat. Schwarz & Renic might not *like* the idea that human soldiers are irredeemably flawed, but that doesn't mean they are redeemable. When predicting whether autonomous weapons will ever comply with IHL better than human soldiers, therefore, techno-optimism seems a safer bet than techno-pessimism.

---

[25] Though they are not that different. Are racial biases not often driven by incomplete or erroneous information?
[26] *Contra id.*
[27] *Contra id.*
[28] Schwarz & Renic, *supra* note 1, at 8.
[29] For a useful overview, see generally Emilio Ferrara, *Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies*, 6 SCI 1 (2024).
[30] Schwarz & Renic, *supra* note 1, at 8.

III. THE NATURE OF IHL

To be fair to Schwarz & Renic and to Bo, their critique of autonomous weapons goes beyond simply downplaying human flaws or being pessimistic about AWS technology. Their basic claim is more fundamental: namely, that IHL compliance requires the kind of judgment that only humans possess. Schwarz & Renic thus assert that "[c]ritics are right to doubt whether autonomous weapons can align with IHL frameworks, designed as these frameworks are with human capacities and limitations in mind,"[31] while Bo insists that the very comparison of humans and IHL is flawed because "compliance with the principle of distinction and proportionality requires more than object recognition and classification."[32]

This is a very common critique of autonomous weapons, one that my article addresses at length. The problem is that Schwarz & Renic and Bo, like the critics I discuss in the article, never explain precisely *why* IHL compliance always requires human judgment. They simply assert that it does.

Consider, for example, Bo's claim that "despite progress in AI, many states recognize that difficulty remains in training a system to correctly recognize civilians no longer directly participating in hostilities, wounded, or surrendering."[33] I acknowledge in the article that there will be DPH situations in which human judgment is necessary; indeed, I provide an example of one (the two boys playing with toy guns).[34] But I also explain at length why, in terms of IHL compliance, recognizing surrender *can* in fact be reduced to "object recognition and classification."[35] Bo never explains why I am wrong about that – nor does she explain why human judgment is necessary in the one hypothetical example she offers, mistaking a pickup for a military tank.[36] How is distinguishing a pickup from a tank a uniquely human endeavor that requires more than object recognition and classification? Do humans not make the distinction based on their knowledge of what tanks look like and what trucks look like? The cognitive task is the same regardless of whether the soldier is a machine or a human – and humans aren't very good at it, as the misidentification statistics from Afghanistan and Iraq that I cite in the article indicate.[37]

Instead of providing a hypothetical situation in which human judgment is supposedly necessary for IHL compliance, Schwarz & Renic simply condemn the very effort to unpack the cognitive requirements of IHL, particularly the principle of distinction, on the ground that doing so "once again stacks the odds in favour of machine logics."[38] But I am not doing that. *IHL is*. IHL is predicated on a series of binaries: combatant/civilian; military objective/civilian object; fighting/surrendering; etc. Those binaries, in turn, are generally predicated on visible behavior, not by intention: soldiers in uniform or combatants wearing a fixed and distinctive sign are targetable even if they have no intention of fighting; a tank can be attacked even if it is being used to transport the wounded (subject to proportionality); a soldier who wants to surrender is not doing so until he raises his hands or waves a white flag. And that emphasis on the visible is not an accident. On the contrary, *it is precisely IHL's "machine logic"* – the fact that its central

[31] *Id.* at 3.
[32] Bo, *supra* note 3, at 9.
[33] *Id*. at 8.
[34] Heller, *supra* note 2, at 25–26.
[35] *Id*. at 21–23.
[36] Bo, *supra* note 3, at 10.
[37] Heller, *supra* note 2, at 59.
[38] Schwarz & Renic, *supra* note 1, at 6.

prohibitions can generally be reduced to simple, easy to apply rules that do not require complicated judgments about intention – *that enables soldiers to comply with it*.

To be sure, the machine logic of IHL does not guarantee compliance. No matter how clear the rules, human soldiers will sometimes misinterpret or misapply them. But at the very least clarity facilitates compliance. I hate to imagine how often human soldiers would violate IHL if its rules did, in fact, take the form that Schwarz & Renic and Bo seem to believe they do – as constantly requiring complicated assessments of subjective mental states (emotions, intent, etc). As I discuss in the article, and as amply demonstrated by cognitive-psychological literature, humans are terrible at identifying the mental states of other people.[39]

Once we understand that IHL is based on a machine logic most of the time, the attractiveness of employing logic-based machines in warfare becomes evident. Humans aren't even very good at "simple" cognitive tasks like object recognition and classification. Machines may not be better at those tasks now. Schwarz & Renic's discussion of the current state-of-the art in AI is an important reminder to take corporate claims with a large grain of salt. But it is at least *possible* (not certain) that in *many* situations (not all) autonomous weapons will eventually be able to distinguish combatants from civilians, military objectives from civilian objects, and surrendering soldiers from non-surrendering ones *better than* (not perfectly) human soldiers. If that prediction turns out to be correct, what is the moral objection to using AWS?

IV. "HUMANITY" IN WAR

My argument, again, is unabashedly consequentialist. Until humans stop going to war, I support them doing whatever they can to ensure that war produces as few civilian casualties and as little unnecessary combatant suffering as possible. If autonomous weapons can one day help do that by complying with IHL better than human soldiers, I am in favor of developing and using them. If they cannot, I will happily become a card-carrying member of the Campaign to Stop Killer Robots.

My consequentialism assumes, of course, that ensuring compliance with IHL is the best way to minimize civilian casualties and unnecessary combatant suffering during war. That assumption is not self-evident: there is no question that IHL authorizes as much violence as it restrains.[40] Short of renouncing war, however, IHL functions much like Churchill's democracy: the worst way to restrain violence, except for all the others.

Schwarz & Renic have a very different understanding of the relationship between war and violence. Specifically, they seem to believe that best way to restrain violence during war is to ensure that war is fought only by human soldiers, because only humans can understand what it means to kill. Consider the following quotes:

> It matters that these machines are not human; that they cannot understand human contexts; are unable to draw on plural experiences of their own; are not sensitive to vulnerability; and lack a conception of the value of life, or indeed, the horrors of a violent death.[41]

---

[39] Heller, *supra* note 2, at 33–36.
[40] *See generally* SAMUEL MOYN, HUMANE: HOW THE UNITED STATES ABANDONED PEACE AND REINVENTED WAR (2021).
[41] Schwarz & Renic, *supra* note 1, at 3.

> Is pre-programming a system, the specific effects of which cannot be known with exactitude… really the same as making and affecting the in-the-moment decision to kill? Is the human agency and judgement we consider to be morally relevant for taking another human life present in sufficient quantities in both instances?[42]

Once again, the logic of the argument is left unstated. *Why* will allowing machines that "lack a conception of the value of life, or indeed, the horrors of a violent death" to kill lead to greater violence during war? *How* does having "human agency… present in sufficient quantities" for each "in-the-moment decision to kill" ensure that war does not descend into barbarism?

Much like the deontologists I criticise in my article, Schwarz & Renic never directly answer these questions. Instead, they simply assume that human war is humane war, at least in comparison to war fought by machines.

It is possible, of course, to fill in the blanks in their argument. Perhaps Schwarz & Renic believe that war fought by human soldiers will, in fact, be more IHL-compliant, and thus more humane, than war fought by human soldiers. There are tendrils of that argument in their response, such as their thorough-going techno-pessimism. But they never make the argument explicitly, almost certainly because it would commit them to supporting AWS if it turns out (someday) that machines *can* comply with IHL better than humans.

It seems more likely, then, that Schwarz & Renic's argument is a version of the compassion argument made by Asaro, Leveringhaus, Geiss, and others: a human must always be involved in "making and affecting the in-the-moment decision to kill,"[43] because only humans can choose *not* to kill. If so, I stand behind the position I take in the article, which is that the unintended consequences of choosing not to kill when killing is lawful make not killing more likely to dehumanize war than to humanize it[44] – especially when we factor in all the biases and negative emotions that inevitably accompany "positive" emotions like compassion.

Schwarz & Renic and I are equally committed to restraining the violence of war. Our disagreement is about means, not ends: whereas they see humans as the final bulwark against unrestrained violence, I see them as precisely the cause of war's barbarity. My techno-optimism may indeed be, as Schwarz & Renic claim, "an article of faith."[45] But I would humbly suggest that their anthropo-optimism is no less of one.

V. HUMAN/MACHINE TEAMING

The final criticism I want to respond to, from Bo, is that comparing autonomous weapons to human soldiers in terms of IHL compliance is an oversimplification:

> [C]urrent and foreseen developments in the uses of military AI run counter to this narrative. Military AI use cannot be thought of as a single automated weapons system. Rather, human-machine teaming is the go-to approach to the integration of military AI undertaken by many states

[42] *Id*. at 3–4.
[43] *Id.* at 4.
[44] Heller, *supra* note 2, at 61–62.
[45] Schwarz & Renic, *supra* note 1, at 8.

> …. The real and pertinent question is thus not whether machines are "better" than humans but rather how human-machine teaming is currently and is likely to be configured along with the consequences of these configurations.[46]

This criticism can be overstated. Some states – particularly less powerful ones that develop autonomous weapons on a shoestring budget or buy them "off the rack" – will use AWS in the manner contemplated by my article: as a direct replacement for human soldiers. Ukraine is an example, as the *New York Times* recently reported.[47] Bo is nevertheless right to insist that we examine how AWS will function when they are working alongside humans instead of replacing them.

Two of Bo's specific fears about human/machine teaming call for comment. The first is that instead of humans controlling AI-equipped systems, AI-equipped systems will control humans. "The use of algorithmic DSS within complex environments," she suggests, "can… hamper users' autonomy by shaping their choices."[48] That is no doubt true, and the example Bo provides – Israel's reliance on AI-enabled systems such as Lavender and Gospel to generate targets that ostensibly can be lawfully attacked – creates significant cause for concern. The term "hamper," which Bo borrows from Taylor Kate Woodcock, is nevertheless a loaded one. If algorithmic DSS helps human soldiers identify attackable targets *more accurately* than they would without algorithmic DSS, thus reducing civilian harm, "enable" would be more accurate than "hamper." Would such algorithmic DSS not be desirable?

It is an open question, of course, whether AI-enabled targeting systems are more likely to enable or hamper human autonomy. My point – once again – is that this is an empirical question, one that cannot be answered either categorically or *a priori*. Perhaps, on balance, algorithmic DSS will hamper more than it enables. But perhaps not.

The second of Bo's fears is that "the speed and scale of AI-enabled target production and nomination … might affect compliance with IHL."[49] The primary culprit, according to the *Opinio Juris* blog post Bo cites,[50] is automation bias: the tendency of humans to put too much faith in the recommendations of machines. This is also an entirely reasonable fear, because a vast amount of research supports the existence – and power – of automation bias. That bias is a problem in the context of algorithmic DSS, however, only insofar as the AI that is part of a human/machine team is making *inaccurate* recommendations. If it is making *accurate* ones, the human member of the team is likely to be the problem, because of a related human/machine bias: undertrust, where humans disregard the recommendations of machines even when they have no reason to do so.[51] In other words, when it comes to IHL compliance, we want humans in the loop if we cannot trust autonomous weapons, and we want humans out of the loop if we

---

[46] Bo, *supra* note 3, at 4–5.
[47] Paul Mozur & Adam Satariano, *A.I. Begins Ushering in an Era of Killer Robots*, N.Y. TIMES (July 2, 2024), https://www.nytimes.com/2024/07/02/technology/ukraine-war-ai-weapons.html ("Until recently, a human would have piloted the quadcopter. No longer. Instead, after the drone locked onto its target – Mr. Babenko – it flew itself, guided by software that used the machine's camera to track him.").
[48] Bo, *supra* note 3, at 7 (quoting Taylor Kate Woodcock, *Human/Machine(-Learning) Interactions, Human Agency and the International Humanitarian Law Proportionality Standard*, 38 GLOB. SOC'Y 100, 112 (2024)).
[49] *Id*. at 7.
[50] Marta Bo & Jessica Dorsey, *The 'Need' for Speed – The Cost of Unregulated AI Decision-Support Systems to Civilians*, OPINIO JURIS (Apr. 4, 2024), https://opiniojuris.org/2024/04/04/symposium-on-military-ai-and-the-law-of-armed-conflict-the-need-for-speed-the-cost-of-unregulated-ai-decision-support-systems-to-civilians/.
[51] *See* Heller, *supra* note 2, at 51.

can.

CONCLUSION

Schwarz & Renic end their response to my article by saying that "humanity is an endless disappointment in war, but we suspect that we will miss it when it is gone, especially if it is cleared away to make room for flawed technologies that fall short of their promise and deaden our moral imagination."[52] That sentence encapsulates the difference between critics of AWS like Schwarz, Renic, and Bo and critics of the human like me. They start from the premise that there is something ineffable about humanity that ensures war fought by human soldiers will always be more humane than war fought by machines. I do not start from an equivalent premise about machines. On the contrary: my belief that autonomous weapons will one day be able to comply with IHL better than human soldiers is not a premise but a *prediction*. This prediction could turn out to be wrong, whether because militaries discover ways to make human soldiers perform better in combat or – and this is more likely – AI technology never improves to the point where we would permit machines to "decide" to take human life. But at least my position is falsifiable; Schwarz and Renic and Bo can't say the same. For them, humans are inherently superior to machines, now and forever, and suggesting otherwise is simply to hinder "our efforts to restrain the worst excesses and impulses of war."[53] What those efforts are, and why those efforts are more likely to succeed than making use of an intelligence not subject to human "excesses and impulses," they never tell us.

Such faith in humanity is no doubt comforting for those who fear machines. But it is also curious, because humans have proven time and again that they have the ability to kill each other with no help from artificial intelligence at all.

[52] Schwarz & Renic, *supra* note 1, at 10.
[53] *Id.* at 9.