Ethical and Social Perspectives on New Military Technologies

Royal Military Academy, Brussels

26-27 February 2026

ABSTRACT BOOK

Table of Contents

KEYNOTES

Lode Dewaegheneire – Autonomous Weapons Systems: "From Terminator to Responsible by Design", the Evolution of Diplomacy on AWS

Abstract: Since the first discussions on Autonomous Weapon Systems at the United Nations in 2014, technological advancements have accelerated at an extraordinary pace, leaving international deliberations and initiatives struggling to keep up. The broader societal debate surrounding Artificial Intelligence, along with emerging military applications of AI, has compelled stakeholders to reassess the scope, methodologies, and ambitions of international negotiations and regulatory efforts. Lode Dewaegheneire will guide participants through the evolution of global dialogues on the prohibition and limitation of autonomous weapon systems, highlighting how rapid technological progress has shaped both the dynamics and the substantive content of these discussions.

Biography: After an operational career as a helicopter pilot at the Belgian Air Force, Lode Dewaegheneire has been active in disarmament for two decades. He was military advisor for humanitarian disarmament at the Defence staff (ACOS STRAT) and Senior Verification Coordination Officer at the Organisation for the Prohibition of Chemical Weapons.

Lode is currently an independent expert in disarmament affairs. He is closely involved with the *Campaign to Stop Killer Robots* and a member of the *Forum of the Arms Trade*. His key areas of expertise are humanitarian disarmament, with a focus on autonomous weapons systems, and transparency measures.

Lode holds a master's degree from the Royal Military Academy, Brussels and a master's degree in international relations from the University of Liege.

Sofie van de Maarel – Moral injury and new moral dilemmas with technology

Abstract: Moral dilemmas are inherently part of military operations, but what if a moral dilemma leaves lasting damage? The term 'moral injury' has been in the spotlight for a number of years to describe the inner conflict that can arise after incidents that clash with one's moral compass. In this keynote, Dr. Sofie van der Maarel (Netherlands Defense Academy) discusses this concept in greater depth. She also addresses the new kinds of moral dilemmas that can arise with technological innovations. Experimenting with drones, AI and robotics introduces

additional challenges for soldiers. How can these be understood and navigated in the context of heightened readiness in Europe?

Biography: Sofie van der Maarel is Assistant Professor of Military Ethics and Leadership at the Netherlands Defence Academy and affiliated to Radboud University Nijmegen. She is an anthropologist in the field of international relations, specialized in military innovation, future warfare, and soldiers' lived experiences and moral dilemmas.

Elke Schwarz & Nathan G. Wood – Discussion: The Challenge of Military AI: Ethics, Governance, and Strategic Realities

Abstract: How is artificial intelligence reshaping military practice, and what should we be most concerned about? This moderated discussion brings together Dr. Elke Schwarz and Dr. Nathan Wood to explore the ethical and practical challenges of AI integration in defense contexts. Beginning with their primary concerns about current AI deployment in military settings, this open discussion will aim to move through three key areas. First, our speakers will reflect on the human element in AI-enabled military systems, including questions of intuitive understanding, experiential knowledge, and the deeply human dimensions of responsible use that technical solutions alone may not address. Second, our speakers will analyse the relationship between armed forces and technology providers, from established defense contractors to Silicon Valley startups, with attention to how trust functions in military procurement and the often-overlooked realities of defense acquisitions. Finally, our speakers will address urgent contemporary challenges, from the security risks of large language models to European defense priorities in an era of grey-zone conflict and infrastructure vulnerability.

Biographies: ***Elke Schwarz***, Professor of Political Theory at Queen Mary University of London, holds a PhD from the London School of Economics and Political Science (LSE), an MA in Conflict Studies from the War Studies Department at King's College London (KCL) and a Bachelor in Business Studies from Belmont University (USA). Prior to coming to QMUL, she held academic positions at UCL, Anglia Ruskin University and Leicester University. Her work focuses on the nexus of ethics, technology and politics / warfare with a specific emphasis on new and emerging military technologies, including military Artificial Intelligence (AI), autonomous weapon systems, drones and robots.

She is the author of Death Machines: The Ethics of Violent Technologies (Manchester University Press) and her work on military AI and autonomous weapon systems has been widely published in a range of international publications. She is Vice-Chair of the International Committee of Robot Arms Control (ICRAC), an Associate of the Imperial War Museum (IWM) and an RSA Fellow. Elke is also a 2022/23 Centre for Apocalyptic and Post-Apocalyptic Studies (CAPAS) Fellow and 2024 Leverhulme Research Fellow with a project on the politics of apocalyptic AI.

She works with a number of civil society organisations and serves on several editorial boards for publications and book series on politics and technology.

Nathan Wood is a Junior Research Group Leader at the German Aerospace Center (DLR) in a project focused on ethical and legal aspects of autonomous and AI-enabled systems in the military. His research focuses on the ethics and laws of war, especially as these relate to emerging technologies, artificial intelligence, autonomous weapon systems, outer space warfare, and other aspects of future conflict. He has works published or forthcoming in Ethics and Information Technology, AI and Ethics, Philosophical Studies, War on the Rocks, The Journal of Military Ethics, and numerous other journals, and is a contributor for the upcoming Oxford Handbook of Remote Warfare. In 2024, Wood was a Research Fellow at CETE-P, working on a project about military ethics funded by the Czech Science Foundation.

INVITED SPEAKERS

James Gillard – Autonomous Weapons and the Problem of Unknown Unknowns

Abstract: There is growing philosophical debate about the permissibility of deploying autonomous weapons systems (AWS). Influentially, Simpson and Muller (2016) argue that autonomous weapons may be used if their error rate is no worse than human performance, ie. within a certain risk-based tolerance. Meanwhile DoD Directive 3000.09 insists that autonomous weapons only be deployed after testing establishes that they will "function as anticipated in realistic operational environments…" (2012, 4.a.(1).(a)) In this paper, however, I argue that all risk-based approaches to autonomous weapons face a problem: it is widely accepted in machine learning literature that pre-deployment testing cannot, in practice, provide reliable probability estimates for the rate of AWS failure.

Autonomous weapon systems are expected to rely on Convolutional Neural Networks (ConvNets) and Vision Transformers (ViTs) for target recognition. However, both are inherently vulnerable to "shortcut learning" and "spurious correlations." Instead of classifying targets based on relevant features (insignia, weapon orientation, etc.), they are prone to rely on irrelevant cues which coincidentally correlated with correct outputs in training (e.g., lighting conditions, facial features) (Geirhos et al. 2020; Scheirer et al. 2013).

That such systems are prone to spurious correlations leads to what is often called the "unknown-unknowns problem." Pre-deployment testing may suggest that systems have a given failure rate but, given the opacity of such systems, it does not reveal the specific conditions under which they are liable to fail. When systems are later deployed in more varied environments, and in the presence of adversaries, evidence from pre-deployment testing is highly likely to be structurally insufficient to ground any reliable probability estimate of their real-world failure. For instance, Kalra and Paddock (2016) show that proving the safety of driverless cars (within a 95% confidence) would take a fleet of 100 cars driving 24/7 for about 12.5 years. When comparing the relative predictability of civilian driving behaviour to the unpredictability of theatres of war, the problem of unknown-unknowns becomes particularly salient.

The standard suggested solutions to this problem are unpersuasive. Most AI interpretability techniques, for instance, are fundamentally post-hoc rather than predictive. Meanwhile, industry standard assurance frameworks (NIST AI RMF 1.0, AMLAS, STPA) remain constrained by the difficulty of predicting the conditions under which AWS will err. I conclude that given the practical impossibility of calculating the error rate of AWS, risk-based approaches based on the notion of engineering tolerance are insufficient to justify fully autonomous AWS deployment in all but the narrowest of contexts. High levels of human supervision (eg. "semi-autonomous" weapons) and highly context-specific AWS are perhaps the way to go.

Biography: James Gillard, University of Texas at Austin

Bibi Imre-Millei – Ethics and Canadian Drone Operators

Abstract: Abstract During Canada's involvement in Afghanistan, Canadian Armed Forces (CAF) members used a variety of unmanned/uncrewed aerial systems (UAS) or drones in a variety of ways. The equipment was often leased from companies or allied states, but despite challenges, small communities of drone operators developed during this time. After Afghanistan, drone use all but disappeared until drones started reappearing in all branches in the late 2010s and early 2020s. However, the implementation of this technology has had numerous challenges, one of which is the way that approaches to ethics have been implemented. The CAF is currently using unarmed drones, with plans underway to integrate armed drones.

This presentation focusses on 55 interviews with Canadian army and air force members who use or have used UAS in some capacity as part of the CAF, in conjunction with a review of the types of ethical education these drone operators receive. Using this data, I argue that the current version of the Defence Ethics Program (DEP), as well as the courses drone operators take as part of their training do not prepare them for the ethical challenges they might face which are specific to the type of technology they use in both conflict scenarios, or on exercises.

In the army in particular, drone operators tend to be lower ranked non-commissioned members and non-commissioned officers. Participants noted that the role requires a high level of critical thinking and self-reflection, and that the authoritative and rank structured training of the army (especially their basic training and organizational socialization) did not prepare them for the more flexible, critical role required of drone operators. During the interviews themselves, ethical self-reflection, as well as abstract and hypothetical thinking were also difficult tasks for participants.

In the air force on the other hand, the psychological impacts of drone use on the drone operator are often conflated with the potential ethical issues in both developing policy and interview settings. While these two issues are related, ethical issues are much broader than this specific concern. During interviews, participants often had trouble untangling what ethical problems might arise when it comes to distance warfare and had trouble distinguishing between crewed and uncrewed aircraft when it came to conversations about ethics.

Biography: Bibi Imre-Millei is PhD Student in the Department of Political Science at Lund University in Sweden. Her research focuses on military identity and technology in Canada and Sweden. Bibi is also a fellow at the Centre for International and Defence Policy at Queen's University in Canada.

Lina Vidauskytė – Technological Courage and the Ethics of Remote Warfare

Abstract: In the era of rapidly emerging military technologies, the notion of courage becomes both an ethical and symbolic challenge. A telling case was the 2013 U.S. Department of Defense initiative to establish the Distinguished Warfare Medal for drone operators and cyber warriors. Because it ranked above the Bronze Star and Purple Heart—decorations historically tied to bodily risk and sacrifice—the medal provoked strong opposition. Within months, it was revoked and, in 2014, replaced by the modest "R" device attached to existing awards to denote remote achievements. This episode reveals deep uncertainty over whether remote action can be equated with classical martial courage.

Traditionally, military courage has been inseparable from physical exposure and the possibility of death. Remote and augmented warfare, however, introduces a new form: technological courage. Here, the critical test lies not in bodily endurance but in sustaining psychological pressure, exercising discipline at a distance, and maintaining precision in technologically mediated combat. The symbolic politics of awards reflects this shift. The recent creation of medals for U.S. border patrol agents (Mexican Defense Border medal)—operating in technologized but non-combat environments—indicates that martial virtue is expanding into technopolitical spaces of security and surveillance, where the meaning of bravery is increasingly detached from the battlefield.

From a phenomenological perspective, the transformation of martial courage unfolds in three distinct experiential shifts.

- First, courage is traditionally rooted in bodily exposure—the direct vulnerability to harm and death in combat.

- Second, in technologically mediated warfare, courage becomes a hybrid embodiment: through FPV drones and AR systems, the soldier's perception and agency are extended into the battlefield, creating a fusion of human intention and machine interface.

- Third, courage evolves into a lived experience where risk is no longer confined to physical injury but emerges in psychological and moral dimensions—such as decision-making under remote pressure, ethical ambiguity, and emotional detachment.

These shifts redefine the horizon of martial virtue, where the human and the technological are interwoven not only functionally, but existentially.

Contemporary warfare thus reshapes not only tactics and strategy but also the very meaning of martial virtue. By tracing the interplay between awards, symbolic recognition, and immersive technologies, this paper shows that the transformation of courage is already underway, redefining ethical frameworks and destabilizing traditional hierarchies of military honor. Yet how

this evolving notion of courage will reshape future concepts of military ethics remains an open and pressing question.

Biography: Dr. Lina Vidauskytė, Professor of Ethics at the Lithuanian Military Academy, specializes in the history of warfare, philosophy of technology, and phenomenology. Her recent work explores honor, ethical challenges of the enhanced soldier, and FPV drone operator phenomenology.

Patrik Baard – Virtue ethics and Autonomous Weapon Systems: What, if any, role for martial virtues in AWS?

Abstract: Virtues have traditionally had a central place in military ethics. Yet, despite the rise of discussions on ethical aspects of autonomous weapon systems (AWS), few have concerned the relevance of virtue ethics, opting instead for establishing deontologically-styled principles.

The presentation is premised on the following assumptions. First, that there are substantial normative costs of omitting virtues in military contexts (Sparrow 2013: 105). Omitting martial virtues per se leads to insufficient ethical guidance in war and we should be cautious about omitting them and what we replace them with. Second, there are ethical risks of AWS. Closely related is how a central concept such as responsibility is upset by systems that can inter alia autonomously choose targets, creating a responsibility gap (Taddeo 2024). I will consider how virtue ethics relate to that gap. AWS ought to be deployed only when meeting strict and justified ethical demands. Third, the focus on deontologically-styled principles reinforces a "programmatic view" on ethics, a view which virtue ethics can abate. Thus, if guidance of ethical risks of AWS is needed, and there are normative costs of omitting virtues in military operations that principles cannot replace, then martial virtues should be included and considered in military operations concerning AWS.

Yet, while that conclusion reinforces long-standing arguments, it fails to provide guidance to the case of AWS. I will suggest that virtue ethics can provide guidance on issues such as meaningful ethical control. I will also critically consider whether AWS can facilitate ethical behavior (Riesen 2022) and, possibly, virtues. In a forward-looking manner, the presentation draws on recent discussions in general AI ethics on virtuous AI (Stenseke 2024), virtuous relations to AI (Coeckelbergh 2010), and ascribing fictional agency (Mallory 2023).

References

Coeckelbergh, M., 2010. 'Robot rights? Towards a social-relational justification of moral consideration', *Ethics and Information Technology* 12, 209–221

Mallory, F., 2023. 'Fictionalism about chatbots', *Ergo: An open access journal of philosophy*. DOI: 10.3998/ergo.4668

Riesen, E., 2022. 'The moral case for the development and use of autonomous weapon systems', *Journal of Military Ethics* 21, 132–150

Sparrow, R., 2013. 'War without virtue?', in Strawser, B. J. (ed.), *Killing by remote control*, OUP, 84–105

Stenseke, J., 2024. 'Artificial virtuous agents in a multi-agent tragedy of the commons', *AI & Society* 39, 855–872
Taddeo, M., 2024. *The ethics of artificial intelligence in defence*, OUP

Biography: Patrik Baard is a researcher at the Swedish Defence Research Agency. He holds a PhD in philosophy from the Royal Institute of Technology, Stockholm, Sweden. He has worked in various areas of applied philosophy, including ethics, political philosophy, and decision-making under uncertainty.

Gabrielle Joni Verreault – Civilians, Drones, and Doctrine: The Ethics of DIY Warfare in Ukraine

Abstract: In the digital age, the character of warfare is undergoing a profound shift. Beyond the use of autonomous systems or advanced military-grade AI, a quieter transformation is underway: the blurring of lines between combatants and civilians, facilitated by widely available commercial technologies. This paper examines the ethical and legal implications of civilians repurposing commercial drones for military use, with a focus on the ongoing Ukrainian conflict.

Since Russia's annexation of Crimea in 2014 and especially following the full-scale invasion of Ukraine in 2022, thousands of Ukrainian civilians have taken up non-traditional roles in the war effort. Motivated by patriotism, necessity, or solidarity, civilians have engaged as cyber-operators, drone engineers, and information warriors. One of the most striking developments is the widespread grassroots innovation surrounding commercial drones, which are being modified for reconnaissance, targeting, and even kinetic operations. These practices have not only contributed to Ukraine's tactical resilience but also shaped new forms of civilian-led military doctrine.

However, this trend raises complex ethical dilemmas. Civilian participants are knowingly engaging in dual-use repurposing, often infringing on intellectual property rights. Companies, in turn, are compelled to address the unintended military applications of their products. These developments challenge established legal and moral frameworks: Do such civilian engineers constitute legitimate military targets? What moral principles justify their engagement, and what ethical responsibilities do they assume? How should states and international institutions respond to this new class of "irregular" actors?

Drawing on fieldwork conducted in Ukraine between November 2022 and November 2024, along with a review of emerging literature on civilian participation in modern conflict, this presentation will explore the motivations, ethical self-conceptions, and moral reasoning of these civilian actors. It will also examine the implications for Just War Theory, particularly the principles of discrimination and proportionality, as well as the evolving nature of military virtue and responsibility.

The findings aim to inform the development of the first draft of ethical guidelines for civilian engagement in technologically mediated conflicts. By addressing the doctrinal, legal, and moral gray zones that these actors inhabit, this paper contributes to applied military ethics in an era where the boundaries of warfare are increasingly crowd-sourced, decentralized, and digitized.

Biography: Gabrielle Joni Verreault, Université de Montréal

Maciek Zając – Setting Ethical Standards for Autonomous Weapons Use: the Case of the Russian Invasion of Ukraine

Abstract: The scholarly discourse on the legality and ethicality of using autonomous weapons is currently moving away from unsustainable absolutist claims about their universal and insurmountable infeasibility and into more practically oriented matter of setting correct standards and procedures for using them in compliance with the International Humanitarian Law. Practical problems include establishing proper testing & verification procedures, opaqueness & predictability, susceptibility to adversarial manipulation, the scope of engagement to be considered individual attacks, creation of surrender procedures, and, last but not least, the matter of the actual standards of legal and ethical performance to be set for these weapons vis-à-vis particular obligations.

The Russo-Ukrainian War provides early examples of use and case studies of emerging practices. It also provides possibly the clearest example in recent memory of both a Walzerian supreme emergency and of a manifestly unjust war conducted in a manifestly unjust manner by a mostly volunteer force with manifestly mercenary motivations. The paper aims to investigate the intersection of general issues related to AWS implementation with these conditions. Do these conditions provide justification for relaxing performance standards for Ukrainian AWS if this generated considerable military advantage by allowing their use in otherwise impermissible environments? Which standards and rules should remain unnegotiable even in conditions of supreme emergency? The paper utilizes Walzer's distinction between rules rooted in deep morality and rules rooted in war convention to explore novel – or perhaps not so novel – dilemmas of surrender to AWS, including enemy wounded in collateral harm calculations, or the enemy infantry "playing dead".

Divergent answers to seemingly similar dilemmas showcase the need for proceeding on the case-by-case basis with careful attention to the spirit and purpose, not only to the letter of the law. Balancing the demands of humanity and effective defense against a genocidal threat while simultaneously struggling to apply a fundamentally novel technology is not easy – but I argue that it can be done, and that grim realities of contemporary warfare may ultimately and with proper effort lead to the emergence of a significantly more humane style of fighting.

Biography: Maciek Zając received his PhD from the University of Warsaw for a thesis arguing autonomous weapons could be ethically used if carefully regulated. He has since published in *Philosophy & Technology*, *Ethics & Information Technology*, *Journal of Military Ethics* and other journals on that topic and other matters related to war ethics. His current postdoc project, funded by the Polish National Science Centre, focuses on ethical issues arising from Russian aggression against Ukraine.

Kevin P. Schieman & Grace M. Ryan – Agency Based Framework for Operationalizing Autonomous Weapons Systems

Abstract: United States policy mandates that autonomous weapons systems "be designed to allow commanders and operators to exercise appropriate levels of human judgment over the use of force" (Hicks, 2023). While significant research has focused on developing explainable, governable, and traceable autonomous systems, the critical question of what constitutes "appropriate human judgment" in operational contexts remains underexplored. This paper addresses this gap by proposing human agency as the central concept in effectively operationalizing machine autonomy in military settings. Drawing on the standard model of agency (Himma, 2008) we analyze how choice and understanding function as necessary conditions for exercising human judgment, specifically focusing on cases in which lethal autonomous systems could be employed. Through a series of case studies examining human-machine teaming in simulated combat environments, we identify specific cognitive and operational factors that enable or inhibit effective human judgment. Our findings contribute to the operationalization of new technologies by providing:(1) a framework for discussing human-machine interaction designs based on agency-enhancing criteria; (2) evidence-based recommendations for training protocols that develop the specific competencies required for appropriate judgment; and (3) integration guidelines that preserve meaningful human control while leveraging autonomous capabilities, thus enhancing a user's efforts. These contributions address the pressing need for practical approaches to implementing autonomous systems that satisfy both ethical requirements and operational effectiveness in complex military environments.

Biography: Kevin P. Schieman & Grace M. Ryan, Dept. of Law and Philosophy, US Military Academy

James Johnson – Can AI behave ethically during military crises? A framework for human-centric moral reasoning in high-stakes AI decision support

Abstract: This paper examines the ethical implications of integrating artificial intelligence-powered decision-support systems (AI-DSS) into strategic decision-making. As AI-DSS enhances battlefield awareness and accelerates operational tempo, it also risks undermining human moral agency through automation bias, anthropomorphism, and over-reliance on machine outputs. Focusing on the socio-technical and psychological dimensions of human-AI interaction, the study explores how these systems may reshape ethical deliberation, responsibility, and judgment in high-stakes environments. To address these challenges, the paper proposes a hybrid ethical framework that integrates elements of virtue ethics, deontological constraints, and consequentialist reasoning. This approach aims to support—not supplant—human decision-making by embedding ethical considerations into the design and deployment of AI-DSS. Key contributions of the paper include adapting ethical exemplarism to military applications, articulating normative design criteria for AI-DSS, and developing policy-relevant safeguards, including mandatory override mechanisms, trust calibration systems, and context-specific training. The findings suggest that AI-DSS should be understood not as a passive tool but as an active participant in an evolving decision ecology, raising urgent questions for strategy, policy, and the governance of autonomous warfare.

Biography: Dr. James Johnson is a Senior Lecturer and Director of Strategic Studies in the Department of Politics and International Relations at the University of Aberdeen. He is the author of *The AI Commander: Centaur Teaming, Command, and Ethical Dilemmas, AI and the Bomb: Nuclear Strategy and Risk in the Digital Age*, and *Artificial Intelligence and the Future of Warfare: USA, China & Strategic Stability*.

Daniel Møller Ølgaard – Military imaginaries of AI: Exploring how military professionals navigate socio-technological change

Abstract: Artificial Intelligence (AI) is generally assumed to have a pivotal role to play in future warfare. In the last decade or so, a veritable choir of political leaders, military decision-makers and industry representatives have publically imagined how AI will revolutionize military affairs, and mused about the seemingly endless possibilities for victory created by this. At the same time, a global community has evolved around the challenge of developing and integrating AI into military organizations and practices in a responsible and legal manner. In light of these discussions and visions, this presentation explores the crucial voices of a group of actors that have so far received little attention: The commanders and soldiers tasked with integrating new AI technologies into military practices. Based on field observations and interviews with participants in a pilot study on the integration of AI tools into software-based operational

planning practices at the Danish Army's Officers' School, it focuses on how army officers' make sense of, navigate and anticipate the possibilities and challenges associated with military AI in mundane, everyday settings, far from government offices, corporate boardrooms and military headquarters. Exploring and understanding these seemingly mundane visions, as I will show, is a necessary but underappreciated precondition for the responsible development and use of military AI.

Biography: Daniel Møller Ølgaard, Royal Danish Defence College

Christian Enemark – Anthrobots, Victimhood, and Ethical Human-Machine Warfare

Abstract: The governments of some technologically advanced states envisage that their military personnel will soon operate in combat teams alongside robots that incorporate AI technology. This innovative practice is expected to afford advantages including increased war-fighting mass, better and faster decision-making, and the transfer of physical risks from warriors onto robots. It could also raise a novel ethical question: as a matter of 'human-machine teaming' in future warfare, how should human warriors interact with AI-enabled robots? Recent research on human-robot interaction (HRI) suggests that designing and/or assigning humanlike characteristics could usefully increase humans' trust in those robots. However, another lesson from civilian settings is that robot anthropomorphisation might sometimes cause dangerous confusion and/or harm. It is therefore important now to consider how we should think ethically about military 'anthrobots'. When and why is it right or wrong to anthropomorphize such robots, and what designs and behaviours would count as 'responsible anthropomorphism'?

Arguably, an answer to these question should be pursued by reference to three different visions of robots within human-machine teams: 'assistants' for warriors; 'users' of force; and 'victims' in war. This presentation focuses on the third vision, and it assesses the proposition that humans are morally required to treat humanlike robots humanely. It begins by discussing some examples of military personnel seemingly affording moral consideration to a robot. In the context of war, one of the ethical risks associated with such behaviour is that military anthrobots might be overprotected (as a matter of *jus in bello* proportionality) in comparison to any humans who are present in the same warzone. The presentation then addresses the broader questions of whether or why humans should regard some robots as moral patients. After acknowledging objections to the idea of artificial moral patiency, attention turns to the problem that anthrobots might perpetuate harmful human prejudices and vicious dispositions. Here, virtue ethics is relevant, and there are lessons that might be drawn from research into the ethics of human encounters with anthropomorphic sexbots. The tentative conclusion is that, when military anthrobots are envisioned as 'victims' in war, such anthropomorphism should be permitted; but only if it reinforces human warriors' humane respect for humanity (human beings) in general.

Biography: Christian Enemark is Professor of International Relations at the University of Southampton in the United Kingdom. As a researcher and educator, he explores emergent issues of military ethics, arms control, and global health security.

Nicholas Daniel Johnston – Automating Devil's Advocacy: The Ethics of Adversarial AI in the Military

Abstract: In the last few years, intelligence agencies have begun developing intelligent decision support systems (IDSS) which, rather than make recommendations, challenge analysts. In 2023, the Intelligence Advanced Research Projects Activity (IARPA) launched REASON (Rapid Explanation, Analysis, and Sourcing Online) a program designed to participate in analysis by pointing out overlooked evidence and gaps in reasoning. Academics, entrepreneurs, and intelligence experts pursue similar avenues, using large language models (LLM) to play the role of a critic (ARC, 2025; Chiang et al., 2024; Roberts, 2025; Zhao et al., 2025). These adversarial roles include red-teaming, contrarian structured analytic techniques (SATs) like devil's advocacy, and generation of alternative hypotheses.

This paper argues four points, two conceptualize adversarial AI in relation to IDSS, and two identify ethical opportunities and concerns. First, we argue that adversarial applications of AI can be seen as a response to dominant concerns in AI ethics. They appear to respond to worries about over-reliance (Dzindolet et al., 2003), complacency (Parasuraman & Manzey, 2010) automation bias (Cummings, 2017), and sycophancy (Kwik, 2025). Likewise, adversarial applications of AI promise to slow down and promote reflection rather than increase time-pressure (Rieger & Manzey, 2022). Second, these applications differ significantly from most IDSS, warranting their own ethical treatment. At one level, the implicit function of adversarial AI in military intelligence and decision making appears like design for what scholars call "positive friction" intended to stimulate deliberation (Chen & Schmidt, 2024; Inan et al., 2025) or slow down decisions with "speed bumps" (MIT Sloan Review). More than mere friction though, adversarial AI can function as agents themselves in technosocial arrangements like multi-agent collaborations (Zhao et al., 2025).

Third, a key ethical opportunity which adversarial AI offers is the promotion of moral autonomy in terms of reason responsiveness. Positive friction can be used to help users consciously review their reasons, such that when they do act, it is with greater awareness of why. Fourth, we caution that more than merely augmenting analysis or decision making, adversarial AI's attraction is in part its ability to alleviate demands on labor which contrarian techniques require. This sort of social epistemic scaffolding can shape the distribution of labor, cultivation of expertise, and discursive culture of an organization, potentially to its detriment by deskilling.

We conclude that adversarial AI's ethical benefits are realized if it is designed to promote reason responsiveness without diminishing the organizational place of dissent.

References

Chen, Z., & Schmidt, R. (2024, June). Exploring a behavioral model of "positive friction" in human-AI interaction. In International Conference on Human-Computer Interaction (pp. 3-22). Cham: Springer Nature Switzerland.

Cummings, M. L. (2017). Automation bias in intelligent time critical decision support systems. In Decision making in aviation (pp. 289-294). Routledge.

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. International journal of human-computer studies, 58(6), 697-718.

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. Human factors, 52(3), 381-410.

Inan, M., Sicilia, A., Dey, S., Dongre, V., Srinivasan, T., Thomason, J., ... & Alikhani, M. (2025). Better slow than sorry: Introducing positive friction for reliable dialogue systems. arXiv preprint arXiv:2501.17348.

Kwik, J. (2025). Digital Yes-Men: How to Deal With Sycophantic Military AI?. Global Policy, 16(3), 467-473.

Rieger, T., & Manzey, D. (2022). Human performance consequences of automated decision aids: The impact of time pressure. Human factors, 64(4), 617-634.

Stackpole, B. (2024) To help improve the accuracy of Generative AI, add speed bumps, MIT Sloan. Available at: https://mitsloan.mit.edu/ideas-made-to-matter/to-help-improve-accuracy-generative-ai-add-speed-bumps

Zhao, X., Zhao, S., Yu, H., Zhang, L., & Li, Q. (2025). AgentCDM: Enhancing Multi-Agent Collaborative Decision-Making via ACH-Inspired Structured Reasoning. arXiv preprint arXiv:2508.11995.

Biography: Nicholas Johnston is a PhD candidate in philosophy at TU Delft in the Netherlands, working on the ethics of military technologies, particularly those involved in the intelligence cycle. His research focuses on the technosocial effects of emerging technologies on the division of labor and expertise. Additional work investigates the ethical attitudes of scientists and engineers towards military research.