GC REAIM
GLOBAL COMMISSION ON RESPONSIBLE AI IN THE MILITARY DOMAIN

TRENDS
ترينـدز للبحوث والاستشـارات
TRENDS RESEARCH & ADVISORY

Foreign, Commonwealth & Development Office

GC REAIM Expert Policy Note Series

Context is Everything: Policy Implications of the Military AI Responsibility Contextualization Framework

Herwin Meerveld and Roy Lindelauf

May 2025

POWERED BY The Hague Centre for Strategic Studies

GC REAIM Expert Policy Note Series

Context is Everything: Policy Implications of the Military AI Responsibility Contextualization Framework

Authors: Herwin Meerveld and Roy Lindelauf

May 2025

Cover photo: unsplash

The Global Commission on Responsible Artificial Intelligence in the Military Domain (GC REAIM) is an initiative of the Government of the Netherlands that was launched during the 2023 REAIM Summit on Responsible Artificial Intelligence in the Military Domain in The Hague. Upon request of the Dutch Ministry of Foreign Affairs, the Hague Centre for Strategic Studies acts as the Secretariat of the Commission.

The GC REAIM Expert Policy Note Series was funded by the Foreign, Commonwealth and Development Office (FCDO) of the United Kingdom. GC REAIM Experts maintained full discretion over the topics covered by the Policy Notes. The contents of the GC REAIM Expert Policy Note series do not represent the views of the Global Commission as a whole. The Policy Notes are intended to highlight some key issues around the governance of AI in the military domain and provide policy recommendations. This Policy Note was released during the Commission Meeting in Abu Dhabi, hosted and graciously supported by Trends Research and Advisory.

HCSS
Lange Voorhout 1
2514 EA The Hague

Follow us on social media:
@hcssnl

The Hague Centre for Strategic Studies
Email: info@hcss.nl
Website: www.hcss.nl

1. Introduction

NATO's Principles of Responsible Use (PRUs) of AI provide a general framework for navigating ethical challenges of AI use in military operations, ensuring alignment with ethical and legal standards across different operational settings. However, the practical implementation of these principles varies significantly depending on the context in which AI is deployed. The debate about Responsible AI in the Military Domain (REAIM) should therefore be guided by the following adage: context is everything.

For instance, the domains of sea and space demonstrate how distinct operational environments shape the implementation of the responsibility principle of transparency. In the sea domain, autonomous underwater vehicles operate in highly variable environments influenced by water currents, salinity, and noise from marine life or vessels.[1] Threats and objects of interest are often obscured, requiring complex signal processing and probabilistic classification.[2] In contrast, the space domain involves satellites operating in a vacuum under predictable orbital mechanics, though uncertainties such as debris movement and sensor limitations exist.[3] Consequently, transparency in the sea domain must address noisy, uncertain data and provide tools for interpreting probabilistic outputs, while in the space domain, it emphasizes the traceability of deterministic calculations. In short: the *nature of the physical environment* matters with respect to the ethical implementation of AI in a military operation.

Time sensitivity further differentiates the military domains. In underwater autonomous operations, decisions, such as threat identification or mine avoidance, often need to be made in real time with minimal operator intervention, necessitating transparency mechanisms that prioritise real-time explainability and operator trust.[4] In space, where collision avoidance decisions typically unfold over hours or days, transparency can involve detailed pre- and post-mission audits, enabling rigorous operator reviews of AI decision-making processes.[5] Finally, the *consequences of errors* also vary significantly. In the sea domain, misclassifying a mine or failing to detect a hostile vessel could result in loss of assets, human lives, or conflict escalation. Transparency mechanisms in this context must therefore include confidence levels and fail-safe mechanisms to mitigate risks in high-stakes scenarios. By contrast, in the space domain, errors in collision avoidance can lead to satellite loss, debris creation, and long-term implications for the

[1] Arif Wibisono et al., 'A Survey on Unmanned Underwater Vehicles: Challenges, Enabling Technologies, and Future Research Directions', *Sensors* 23, no. 17 (January 2023): 7321, https://doi.org/10.3390/s23177321.
[2] Erin M. Fischell and Henrik Schmidt, 'Classification of Underwater Targets from Autonomous Underwater Vehicle Sampled Bistatic Acoustic Scattered Fields', *The Journal of the Acoustical Society of America* 138, no. 6 (21 December 2015): 3773–84, https://doi.org/10.1121/1.4938017.
[3] Ya-zhong Luo and Zhen Yang, 'A Review of Uncertainty Propagation in Orbital Mechanics', *Progress in Aerospace Sciences* 89 (1 February 2017): 23–39, https://doi.org/10.1016/j.paerosci.2016.12.002.
[4] Dan Yu et al., 'Enhancing Autonomous Underwater Vehicle Decision Making through Intelligent Task Planning and Behavior Tree Optimization', *Journal of Marine Science and Engineering* 12, no. 5 (May 2024): 791, https://doi.org/10.3390/jmse12050791.
[5] P. Ravi et al., 'AI for Satellite Collision Avoidance — Go/No Go Decision-Making', vol. 2852, 2023, 6043, https://ui.adsabs.harvard.edu/abs/2023LPICo2852.6043R.

orbital environment, requiring transparency efforts that focus on model validation and reliability to prevent cascading failures involving multiple stakeholders. These examples underscore the importance of tailoring the principle of transparency to the unique requirements of each context. The sea domain benefits from mission-specific explainability tools that support real-time operations, while the space domain demands interoperability and accountability frameworks suited to a broader array of stakeholders. Understanding these distinctions ensures that transparency efforts are both effective and meaningful, addressing the operational, technical, and ethical challenges unique to each context.

Other PRUs are also context dependent and therefore require similar considerations. Moreover, defining a specific military context involves multiple dimensions beyond just the operational domain. In an earlier paper, we developed the Military AI Responsibility Contextualization (MARC) framework to address this complexity.[6] This framework helps differentiate the operationalization of PRUs across various dimensions, ensuring that both ethical and technical aspects are appropriately developed within specific contexts. After briefly introducing this framework and illustrating it with use-cases we provide several policy implications of the proposed framework for the responsible development of AI in the military domain.

[6] Meerveld, H.W., et al. "Operationalising Responsible AI in the Military Domain: A Context-Specific Assessment" In Ethics and Information Technology, *under review.*

2. The MARC framework

The Military AI Responsibility Contextualization (MARC) framework provides a non-deterministic approach to operationalize overarching principles of responsible AI use in the military domain.[7] It defines three key dimensions to characterize the contexts of military operations and is presented schematically in figure 1.

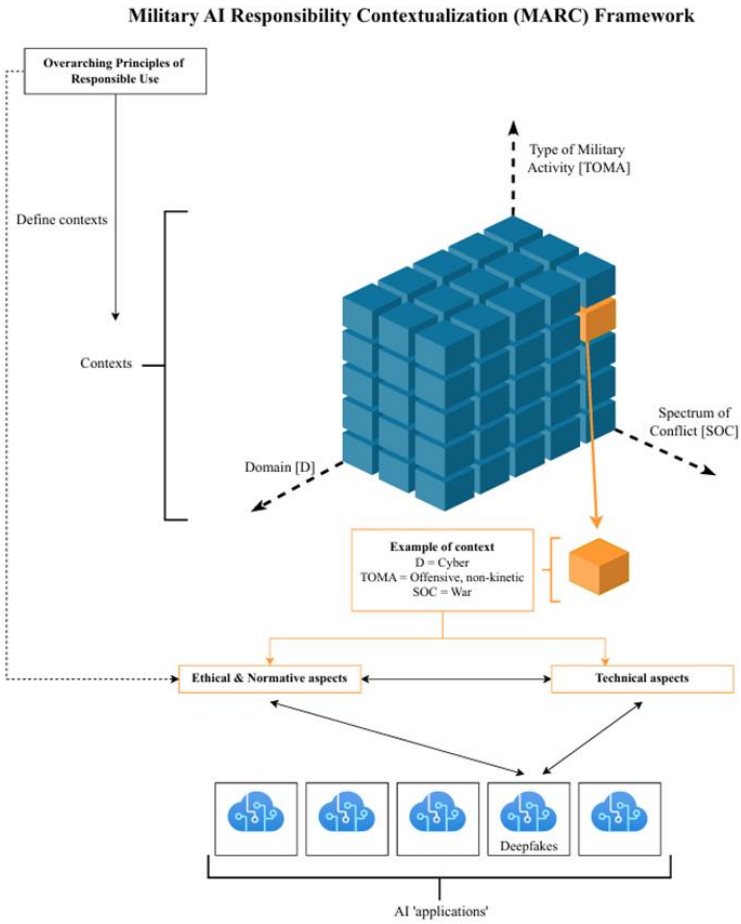

Figure 1: *Military AI Responsibility Contextualization framework, distinguishing military contexts across three dimensions: military domain, spectrum of conflict and type of military activity. Source: Meerveld et al., "Operationalising Responsible AI in the Military Domain: A Context-Specific Assessment" In Ethics and Information Technology, under review.*

The key dimensions employed by the MARC framework are the domain of operation, the type of operation, and the spectrum of conflict. Each of these dimensions contributes to a comprehensive understanding of the environment, objectives, and constraints of a given mission. By combining the five military domains — land, sea, air, space, and cyberspace — with five types of operations — offensive kinetic, offensive non-kinetic,

[7] Introduced in: Meerveld, H.W., et al. "Operationalising Responsible AI in the Military Domain: A Context-Specific Assessment" In Ethics and Information Technology, *under review.*

defensive kinetic, defensive non-kinetic, and service and support operations — and three phases across the spectrum of conflict — ranging from peace, grey zone, to high-intensity warfare — we obtain a framework that encompasses 75 distinct contexts. This framework not only aids in facilitating AI-enabled military decision-making but also reveals the critical challenge of operationalizing high-level principles of responsible use of AI in a manner that aligns with the demands of each unique context. While the MARC framework offers a structured way to contextualize military AI operations, we recognize that many real-world missions straddle both physical and digital domains or fall within ambiguous zones of conflict. The boundaries between kinetic and non-kinetic, offensive, and defensive actions are not always clear-cut, especially in emerging forms of hybrid warfare. Nevertheless, we believe the MARC framework remains valuable, as it enables nuanced contextualization; when multiple domains are involved, relevant contexts can be combined into analytical subsets to reflect the complexity of the operational environment.

The principles of responsible use — such as lawfulness, responsibility and accountability, explainability and traceability, and bias mitigation — require different interpretations and applications depending on the combination of domain, operation type, and phase of conflict. In other words, context is everything. Existing frameworks often treat these principles as universally applicable, overlooking the nuanced requirements imposed by varying operational conditions. This disconnect poses a serious challenge for both military operators and data scientists, as the lack of clear, actionable guidance on responsible AI use leads to inconsistencies in its application. In other words, current approaches for the operationalization of PRUs often overlook the varied requirements and perspectives of various user groups.[8] For NATO and allied forces, this fragmentation not only risks undermining mission success - by reducing trust in AI-driven systems - but also threatens the alliance's moral authority, as ethical lapses can erode legitimacy and public confidence.

[8] Kristian González Barman, Nathan Wood, and Pawel Pawlowski, 'Beyond Transparency and Explainability: On the Need for Adequate and Contextualized User Guidelines for LLM Use', *Ethics and Information Technology* 26, no. 3 (17 July 2024): 47, https://doi.org/10.1007/s10676-024-09778-2.

3. Illustrative use cases

To illustrate the practical relevance of the MARC framework, we sketch several diverse use cases that span different domains, operational types, and spectra on the conflict spectrum.

3.1 Cybersecurity and autonomous agents

As AI technology rapidly advances, enabling autonomous and intelligent operations at the tactical edge, future military operations are increasingly expected to integrate traditional warfare with decentralized, technological advanced capabilities closer to the tactical edge.[9] In high-intensity combat, where decisions may involve life-and-death consequences, ethical and legal standards demand full human control over weapon systems. However, cyber operations often preclude direct human oversight of every action. As a result, accountability in this domain shifts towards strategic oversight: ensuring that autonomous cyber agents (ACAs) are designed to act in predictable, explainable ways, and that their behaviour remains traceable.

ACAs are software agents operating in digital environments with the capacity to sense, decide, and act autonomously in pursuit of defined cybersecurity goals. For example, agents may be tasked with detecting anomalous activity in network traffic, isolating compromised nodes, or responding to distributed denial-of-service (DDoS) attacks in real time. Some existing or prototypical systems include Google's Chronicle Autonomic Security Operations platform, which automates threat detection and response using machine learning, or the U.S. Defense Advanced Research Projects Agency's (DARPA) Cyber Grand Challenge (CGC) prototypes, which demonstrated fully autonomous systems capable of identifying and patching software vulnerabilities without human intervention.[10]

Consider for instance the use of ACAs in the context of a defensive non-kinetic type of operation, in the grey zone on the spectrum of conflict (cyberspace domain). In operations below the threshold of warfare the requirements for responsible use differ dramatically. This context might involve countering disinformation campaigns or securing critical infrastructure without direct physical engagement. The principle of proportionality becomes especially salient, as cyber responses must be calibrated to avoid escalating tensions or violating sovereignty. For instance, an ACA defending a smart electrical grid must detect and neutralize intrusions without unintentionally affecting civilian services or crossing legal boundaries. In such politically sensitive environments, transparency and collaboration with civilian agencies are essential to maintain legitimacy, while mechanisms of accountability must ensure that ACA actions

[9] Adib Bin Rashid et al., 'Artificial Intelligence in the Military: An Overview of the Capabilities, Applications, and Challenges', *Int. J. Intell. Syst.* 2023 (1 January 2023), https://doi.org/10.1155/2023/8676366.
[10] David Brumley, 'The Cyber Grand Challenge and the Future of Cyber-Autonomy', *;;Login:*, 2018, https://www.usenix.org/publications/login/summer2018/brumley.

remain consistent with international legal and ethical norms.[11] Understanding the behaviour of these agents—what triggers their actions, how they make decisions, and where their boundaries lie—is crucial to maintaining operational trust and responsible use. This underscores the need for explainability-by-design, robust testing protocols, and oversight architectures that balance autonomy with accountability.

3.2 Cognitive warfare

The rise of deep fake technology and AI-generated content has necessitated the development of more effective detection methods. However, challenges persist, including rapid progress in generative techniques, limited high-quality datasets, and the need for more efficient and robust detection systems.[12] Deep fakes can amplify disinformation through social media platforms and digital ecosystems, exploiting the rapid spread of misinformation to create confusion and disrupt decision-making during crises. This impact is especially acute in environments where quick, accurate responses are critical to maintaining operational integrity and public trust. Both for military forces and civilian first-responders, distinguishing authentic communications from fabricated content can be decisive in ensuring effective command and control during crises. Due to cognitive biases, it is difficult to fully correct perceptions that arise due to misinformation, especially if the targeted population has been exposed to them repeatedly.[13] The human brain interprets ease of processing as a signal of truthfulness[14] and it is particularly susceptible to disinformation when analytic thinking is not employed[15], for example when emotional response is elicited (e.g., fear or anger). To counter these threats, military and governmental organizations are prioritizing the development of AI-based early detection and verification systems.[16] These tools aim to identify synthetic media in real-time, enabling the rapid response necessary to mitigate the disruptive effects of deep fakes. However, the evolving sophistication of deep fake generation poses a constant challenge and constitutes a continuous arms race, demanding ongoing innovation in detection technologies.[17] With respect to governmental use, AI systems must not only be effective but also explainable and

[11] Tim Krause et al., 'Cybersecurity in Power Grids: Challenges and Opportunities', *Sensors* 21, no. 18 (January 2021): 6225, https://doi.org/10.3390/s21186225.
[12] Yuxiang Zhang et al., 'Deepfake Detection System for the ADD Challenge Track 3.2 Based on Score Fusion', in *Proceedings of the 1st International Workshop on Deepfake Detection for Audio Multimedia*, DDAM '22 (New York, NY, USA: Association for Computing Machinery, 2022), 43–52, https://doi.org/10.1145/3552466.3556528.
[13] Ullrich K. H. Ecker, Stephan Lewandowsky, and David T. W. Tang, 'Explicit Warnings Reduce but Do Not Eliminate the Continued Influence of Misinformation', *Memory & Cognition* 38, no. 8 (1 December 2010): 1087–1100, https://doi.org/10.3758/MC.38.8.1087.
[14] Lisa K. Fazio et al., 'Knowledge Does Not Protect against Illusory Truth.', *Journal of Experimental Psychology: General* 144, no. 5 (October 2015): 993–1002, https://doi.org/10.1037/xge0000098.
[15] Gordon Pennycook and David G. Rand, 'Lazy, Not Biased: Susceptibility to Partisan Fake News Is Better Explained by Lack of Reasoning than by Motivated Reasoning', *Cognition* 188 (July 2019): 39–50, https://doi.org/10.1016/j.cognition.2018.06.011.
[16] Yisroel Mirsky and Wenke Lee, 'The Creation and Detection of Deepfakes: A Survey', *ACM Comput. Surv.* 54, no. 1 (2 January 2021): 7:1–7:41, https://doi.org/10.1145/3425780.
[17] Maria Pawelec, 'Decent Deepfakes? Professional Deepfake Developers' Ethical Considerations and Their Governance Potential', *AI and Ethics*, 25 September 2024, https://doi.org/10.1007/s43681-024-00542-2.

transparent to ensure accountability and compliance in their deployment. Ethical considerations are thus paramount in addressing the use of deep fakes in cognitive warfare. While countering disinformation is critical, measures must be implemented in ways that respect human rights and international norms. Systems designed to detect and combat deep fakes must balance the need for operational effectiveness with principles of fairness, privacy, and accountability. Across all use cases, the NATO principles, human rights and just war norms contingent on the context are relevant, but their application is shaped by the operational context.

3.3 Drone swarms for increased situational awareness

Situational awareness (SA)—the ability to perceive, understand, and predict conditions in an environment—is critical for effective decision-making in time-sensitive scenarios. Over the past three decades, military operations have become increasingly complex, characterized by conflicts involving state and non-state actors and blurred boundaries between warfare, organized crime, and human rights violations. These complexities necessitate tailored, context-sensitive solutions rather than one-size-fits-all approaches. Moreover, global urbanization trends have shifted the theatre of war to densely populated cities, as demonstrated by conflicts in Aleppo (Syria), Mogadishu (Somalia), and various Ukrainian cities, including Donetsk and Mariupol. These urban settings demand technological solutions capable of navigating dense, dynamic environments while minimizing collateral damage and addressing ethical considerations, such as the impact of military actions on civilian populations.

Despite advances in sensor technologies, current systems still face critical limitations in reliably collecting and analysing data within complex urban and natural environments. These constraints can delay critical insights, such as detecting enemy positions, improvised explosive devices (IEDs), or assessing casualties, leading to mission failures and increased risk to personnel. For troop-contributing nations, low situational awareness undermines their ability to accurately assess threats, compromising both force protection and political backing for multinational security and humanitarian efforts.

Similarly, in civilian emergency scenarios—such as natural disasters or large-scale accidents—poor situational awareness diminishes survival chances for victims and increases risks for first responders. Recent developments in unmanned aerial systems (UAS) have highlighted the potential for drone *swarms*—groups of small, interconnected drones operating collaboratively—to enhance situational awareness.[18] Compared to legacy reconnaissance or combat platforms like the Northrop Grumman Global Hawk or the General Atomics MQ-9 Reaper, which are costly and infrastructure-dependent, drone swarms present a more agile, scalable, and cost-effective alternative, particularly suited to contested and infrastructure-poor environments. Military organizations are increasingly prioritizing "small, smart, and cheap" solutions, with swarm technology

[18] Somaiyeh MahmoudZadeh et al., 'Holistic Review of UAV-Centric Situational Awareness: Applications, Limitations, and Algorithmic Challenges', *Robotics* 13, no. 8 (August 2024): 117, https://doi.org/10.3390/robotics13080117.

emerging as a key focus due to its expected operational advantages in contested environments.[19] Yet the implementation of Human-Swarm Teaming (HST) remains fraught with challenges.

First, swarm systems place unduly high cognitive demands on human operators, requiring supervisory control, trust in automation, and distributed decision-making under conditions of uncertainty.[20] Removing humans from the decision loop risks skill degradation and raises ethical and legal concerns. Hence, ensuring meaningful human control necessitates flexible and adaptive autonomy, a concept that yet is underexplored in swarm-control design. Second, drone (and for that matter swarm) operations rely heavily on GNSS-based navigation, which can be spoofed, jammed, or rendered ineffective in so-called urban-canyons.[21] This emphasizes the need for resilient alternative navigation techniques and energy-efficient intra-swarm communication architectures – for instance neuromorphic systems – that can maintain functionality in GNSS-denied environments. Third, decentralized drone swarms face challenges in aggregating noisy, incomplete, and inconsistent spatial-temporal data that is often characteristic of military and first-responder environments.[22] Novel AI techniques for data fusion and game-theoretical and other algorithms for swarm coordination are thus essential for effective operations. Fourth, the deployment of HST systems brings critical questions of human agency, accountability for errors, and compliance with ethical and legal norms to the forefront. Fifth, successful integration of these technologies demands not only new operational doctrines but also comprehensive training programs that efficiently integrate HST systems into end user environments, including active engagement and cooperation with public institutions and stakeholders. These multifaceted challenges underline the need for a structured approach to operationalizing responsible AI in HST contexts, an issue taken up in Section 5.

[19] Isuru Munasinghe, Asanka Perera, and Ravinesh C. Deo, 'A Comprehensive Review of UAV-UGV Collaboration: Advancements and Challenges', *Journal of Sensor and Actuator Networks* 13, no. 6 (December 2024): 81, https://doi.org/10.3390/jsan13060081.

[20] Joseph P. Distefano, Souma Chowdhury, and Ehsan Esfahani, 'Exploring Human-Swarm Interaction Dynamics in Cyber-Physical Systems: A Physiological Approach', *Journal of Integrated Design and Process Science* 27, no. 3–4 (1 November 2023): 200–210, https://doi.org/10.1177/10920617241292155.

[21] Shlomi Hacohen et al., 'Improved GNSS Localization and Byzantine Detection in UAV Swarms', *Sensors* 20, no. 24 (January 2020): 7239, https://doi.org/10.3390/s20247239.

[22] Salvatore Rosario Bassolillo, Egidio D'Amato, and Immacolata Notaro, 'A Consensus-Driven Distributed Moving Horizon Estimation Approach for Target Detection Within Unmanned Aerial Vehicle Formations in Rescue Operations', *Drones* 9, no. 2 (February 2025): 127, https://doi.org/10.3390/drones9020127.

4. Policy implications

To operationalize responsible military AI across diverse contexts and for different use-cases (as sketched above), concrete policy measures are needed that go beyond abstract principles. The following subsections sketch two initiatives, learning from past AI incidents and fostering ethical stakeholder engagement.

4.1 Learning from AI incidents.

An AI Incident Database (AIID) addresses the growing need for a systematic approach to documenting failures in artificial intelligence systems, like how aviation and cybersecurity industries track incidents.[23] Without a centralized database, AI practitioners repeatedly make the same mistakes, leading to recurring failures in safety-critical areas such as law enforcement, healthcare, and autonomous systems. An AIID collects real-world AI failures from diverse sources, allowing stakeholders—including engineers, risk officers, and policymakers—to identify trends, mitigate risks, and develop more robust AI applications. By providing searchable, structured data on AI incidents, the database fosters transparency and accountability in AI deployment. For military applications, such an incident database aids in ensuring reliable and ethical AI integration in defense systems. AI-driven military technologies, including autonomous weapons, surveillance systems, and decision-support algorithms, require rigorous oversight to prevent unintended consequences. An AIID can enable defence organizations to analyse past failures, anticipate risks, and implement best practices before deploying AI in critical missions. Regarding the MARC framework, an AIID can aid in refining the ethical and technical aspects related to specific contexts by learning from previous AI-related incidents and their respective context.

An AIID's structured approach to incident reporting and MARC validation, as discussed in this paper, highlights the importance of taxonomy-based classification of AI incidents to identify patterns, mitigate risks, and improve AI governance. For military applications, incorporating a standardized incident taxonomy—such as the CSET AI Harm Taxonomy or the Goals, Methods, and Failures (GMF) Taxonomy—would enable defence organizations to analyse past AI failures systematically, improving risk assessment and response mechanisms, and analyse MARC contexts more deeply.[24] By integrating mandatory AI incident reporting protocols, like those outlined in the European Union AI Act, military AI systems can be monitored for both actual and potential harm, ensuring greater accountability. Additionally, fostering cross-sector collaboration with academia, industry, and international defence bodies would enhance the database's effectiveness in capturing incidents across diverse operational environments, and thus feeding more accurate information into MARC context analyses.

---

[23] Sean McGregor, 'Preventing Repeated Real World AI Failures by Cataloging Incidents: The AI Incident Database', *Proceedings of the AAAI Conference on Artificial Intelligence* 35, no. 17 (18 May 2021): 15458–63, https://doi.org/10.1609/aaai.v35i17.17817.

[24] Kevin Paeth et al., 'Lessons for Editors of AI Incidents from the AI Incident Database', *Proceedings of the AAAI Conference on Artificial Intelligence* 39, no. 28 (11 April 2025): 28946–53, https://doi.org/10.1609/aaai.v39i28.35163.

To implement an AIID effectively, military AI governance should focus on timely AI incident monitoring, ensuring that potential failures—such as unintended targeting errors or autonomous system malfunctions—are identified and addressed before deployment. A tiered classification system should differentiate between low-impact AI failures (e.g., minor sensor misreads) and high-risk AI incidents (e.g., lethal autonomous weapon misfires), enabling proportional responses. Finally, integrating AIID-based simulations and risk assessments into MARC analyses and military training programs would prepare personnel to recognize and mitigate AI-related failures early, ensuring that AI-driven military operations adhere to safety, legal, and ethical standards.

4.2 Stakeholder engagement through guidance ethics.

To continuously evaluate the ethical, operational, and technical aspects of specific (MARC) contexts in which military AI can be applied, it is recommended to conduct guidance ethics approaches.[25] Central to the guidance ethics approach is the participatory engagement of stakeholders to identify and deliberate on the societal implications of technology, ensuring that ethical considerations are embedded within the operational context. By facilitating dialogues among military operators, planners, technologists, ethicists, lawyers and policymakers, the guidance ethics approach enables the co-creation of actionable strategies that align technological capabilities with ethical principles tailored to specific mission parameters. This approach emphasizes the interplay between technological, human, and environmental factors, offering a means to ground ethical considerations in the operational realities of military engagements. By systematically applying this methodology to the distinct contexts defined by the combination of military domains, operation types, and spectra of conflict, it is possible to generate actionable, context-specific ethical guidelines. This section elaborates on the application of the guidance ethics methodology to military contexts and sketches its utility in resolving the challenges posed by context-specific requirements.

The process begins with a comprehensive case analysis aimed at understanding the specific dimensions of the military context (as provided by the MARC framework) in question. This involves defining the operational parameters by situating the mission within the military domain, the type of military activity, and the spectrum of conflict. Each of these factors introduces unique variables that must be considered to ensure ethical principles are appropriately operationalized. Following the analytical phase, the methodology prioritizes structured dialogue among a diverse group of stakeholders. This dialogue serves as the centrepiece of the guidance ethics approach, bringing together military planners, ethicists, technologists, policymakers, and other relevant actors to collaboratively explore the ethical dimensions of the operation. The inclusion of varied perspectives ensures that the dialogue captures the complexities of the context while fostering a shared understanding of the challenges. For instance, in a space domain, peace operations, and peaceful competition scenarios, stakeholders might

[25] Verbeek Peter-Paul and Tijink D, 'Guidance Ethics Approach: An Ethical Dialogue about Technology with Perspective on Actions', 2020, https://ecp.nl/wp-content/uploads/2020/11/Guidance-ethics-approach.pdf.

discuss the implications of deploying satellites that enhance intelligence capabilities while simultaneously preserving international norms and preventing the weaponization of space. This collaborative engagement is critical for identifying key ethical values—such as sovereignty, accountability, and proportionality—that should guide decision-making.

The outcomes of these discussions form the basis for generating ethical and technical aspects, and actionable courses of action (COAs) tailored to the specific context. These possible courses of action, along with their considerations, should be entered into the online MARC repository and database for the specific context(s) under consideration. One of the strengths of such a guidance ethics methodology is its iterative nature, which allows for continuous refinement of ethical guidelines for each operational context. The application of guidance ethics to military contexts not only ensures that ethical principles are deeply rooted in operational realities but also enhances the legitimacy and effectiveness of military operations. By addressing the specific ethical requirements of each context, this methodology mitigates the risks associated with a one-size-fits-all approach to ethical decision-making. Furthermore, the participatory and transparent nature of the process fosters trust among stakeholders, both within the military and in the broader international community.

In conclusion, a guidance ethics methodology provides a robust approach for analysing and addressing the context-specific ethical challenges of military operations as addressed by the MARC framework. By combining detailed case analysis, inclusive dialogue, and the generation of actionable options for specific contexts, this approach offers a practical means of aligning military's principles of responsible AI use with the diverse demands of contemporary conflict. As military engagements continue to evolve in complexity, the systematic application of this methodology will be essential for ensuring that ethical AI considerations remain at the forefront of operational planning and execution. Resolving this issue requires a systematic approach to tailoring the principles of responsible use to specific contexts. Such an approach must account for the operational priorities, objectives, and ethical considerations unique to each combination of domain, operation type, and conflict spectrum. By addressing these contextual requirements, militaries can ensure that its actions remain not only operationally effective but also ethically defensible, reinforcing legitimacy in an increasingly complex and interconnected battlespace.

Bibliography

Barman, Kristian González, Nathan Wood, and Pawel Pawlowski. 'Beyond Transparency and Explainability: On the Need for Adequate and Contextualized User Guidelines for LLM Use'. Ethics and Information Technology 26, no. 3 (17 July 2024): 47. https://doi.org/10.1007/s10676-024-09778-2.

Bassolillo, Salvatore Rosario, Egidio D'Amato, and Immacolata Notaro. 'A Consensus-Driven Distributed Moving Horizon Estimation Approach for Target Detection Within Unmanned Aerial Vehicle Formations in Rescue Operations'. Drones 9, no. 2 (February 2025): 127. https://doi.org/10.3390/drones9020127.

Brumley, David. 'The Cyber Grand Challenge and the Future of Cyber-Autonomy'. ;;Login:, 2018. https://www.usenix.org/publications/login/summer2018/brumley.

Distefano, Joseph P., Souma Chowdhury, and Ehsan Esfahani. 'Exploring Human-Swarm Interaction Dynamics in Cyber-Physical Systems: A Physiological Approach'. Journal of Integrated Design and Process Science 27, no. 3–4 (1 November 2023): 200–210. https://doi.org/10.1177/10920617241292155.

Ecker, Ullrich K. H., Stephan Lewandowsky, and David T. W. Tang. 'Explicit Warnings Reduce but Do Not Eliminate the Continued Influence of Misinformation'. Memory & Cognition 38, no. 8 (1 December 2010): 1087–1100. https://doi.org/10.3758/MC.38.8.1087.

Fazio, Lisa K., Nadia M. Brashier, B. Keith Payne, and Elizabeth J. Marsh. 'Knowledge Does Not Protect against Illusory Truth.' Journal of Experimental Psychology: General 144, no. 5 (October 2015): 993–1002. https://doi.org/10.1037/xge0000098.

Fischell, Erin M., and Henrik Schmidt. 'Classification of Underwater Targets from Autonomous Underwater Vehicle Sampled Bistatic Acoustic Scattered Fields'. The Journal of the Acoustical Society of America 138, no. 6 (21 December 2015): 3773–84. https://doi.org/10.1121/1.4938017.

Hacohen, Shlomi, Oded Medina, Tal Grinshpoun, and Nir Shvalb. 'Improved GNSS Localization and Byzantine Detection in UAV Swarms'. Sensors 20, no. 24 (January 2020): 7239. https://doi.org/10.3390/s20247239.

Krause, Tim, Raphael Ernst, Benedikt Klaer, Immanuel Hacker, and Martin Henze. 'Cybersecurity in Power Grids: Challenges and Opportunities'. Sensors 21, no. 18 (January 2021): 6225. https://doi.org/10.3390/s21186225.

Luo, Ya-zhong, and Zhen Yang. 'A Review of Uncertainty Propagation in Orbital Mechanics'. Progress in Aerospace Sciences 89 (1 February 2017): 23–39. https://doi.org/10.1016/j.paerosci.2016.12.002.

MahmoudZadeh, Somaiyeh, Amirmehdi Yazdani, Yashar Kalantari, Bekir Ciftler, Fathi Aidarus, and Mhd Omar Al Kadri. 'Holistic Review of UAV-Centric Situational Awareness: Applications, Limitations, and Algorithmic Challenges'. Robotics 13, no. 8 (August 2024): 117. https://doi.org/10.3390/robotics13080117.

McGregor, Sean. 'Preventing Repeated Real World AI Failures by Cataloging Incidents: The AI Incident Database'. Proceedings of the AAAI Conference on Artificial

Intelligence 35, no. 17 (18 May 2021): 15458–63. https://doi.org/10.1609/aaai.v35i17.17817.

Mirsky, Yisroel, and Wenke Lee. 'The Creation and Detection of Deepfakes: A Survey'. ACM Comput. Surv. 54, no. 1 (2 January 2021): 7:1–7:41. https://doi.org/10.1145/3425780.

Munasinghe, Isuru, Asanka Perera, and Ravinesh C. Deo. 'A Comprehensive Review of UAV-UGV Collaboration: Advancements and Challenges'. Journal of Sensor and Actuator Networks 13, no. 6 (December 2024): 81. https://doi.org/10.3390/jsan13060081.

Paeth, Kevin, Daniel Atherton, Nikiforos Pittaras, Heather Frase, and Sean McGregor. 'Lessons for Editors of AI Incidents from the AI Incident Database'. Proceedings of the AAAI Conference on Artificial Intelligence 39, no. 28 (11 April 2025): 28946–53. https://doi.org/10.1609/aaai.v39i28.35163.

Pawelec, Maria. 'Decent Deepfakes? Professional Deepfake Developers' Ethical Considerations and Their Governance Potential'. AI and Ethics, 25 September 2024. https://doi.org/10.1007/s43681-024-00542-2.

Pennycook, Gordon, and David G. Rand. 'Lazy, Not Biased: Susceptibility to Partisan Fake News Is Better Explained by Lack of Reasoning than by Motivated Reasoning'. Cognition 188 (July 2019): 39–50. https://doi.org/10.1016/j.cognition.2018.06.011.

Peter-Paul, Verbeek, and Tijink D. 'Guidance Ethics Approach: An Ethical Dialogue about Technology with Perspective on Actions', 2020. https://ecp.nl/wp-content/uploads/2020/11/Guidance-ethics-approach.pdf.

Rashid, Adib Bin, Ashfakul Karim Kausik, Ahamed Al Hassan Sunny, Mehedy Hassan Bappy, and Yu-an Tan. 'Artificial Intelligence in the Military: An Overview of the Capabilities, Applications, and Challenges'. Int. J. Intell. Syst. 2023 (1 January 2023). https://doi.org/10.1155/2023/8676366.

Ravi, P., A. Zollo, R. Kahle, and H. Fiedler. 'AI for Satellite Collision Avoidance — Go/No Go Decision-Making', 2852:6043, 2023. https://ui.adsabs.harvard.edu/abs/2023LPICo2852.6043R.

Wibisono, Arif, Md Jalil Piran, Hyoung-Kyu Song, and Byung Moo Lee. 'A Survey on Unmanned Underwater Vehicles: Challenges, Enabling Technologies, and Future Research Directions'. Sensors 23, no. 17 (January 2023): 7321. https://doi.org/10.3390/s23177321.

Yu, Dan, Hongjian Wang, Xu Cao, Zhao Wang, Jingfei Ren, and Kai Zhang. 'Enhancing Autonomous Underwater Vehicle Decision Making through Intelligent Task Planning and Behavior Tree Optimization'. Journal of Marine Science and Engineering 12, no. 5 (May 2024): 791. https://doi.org/10.3390/jmse12050791.

Zhang, Yuxiang, Jingze Lu, Xingming Wang, Zhuo Li, Runqiu Xiao, Wenchao Wang, Ming Li, and Pengyuan Zhang. 'Deepfake Detection System for the ADD Challenge Track 3.2 Based on Score Fusion'. In Proceedings of the 1st International Workshop on Deepfake Detection for Audio Multimedia, 43–52. DDAM '22. New York, NY, USA: Association for Computing Machinery, 2022. https://doi.org/10.1145/3552466.3556528.

About the Authors

Herwin Meerveld

Herwin Meerveld is Lieutenant Colonel in the Royal Netherlands Air Force. Currently, he is coordinator of the Data Science Centre of Excellence of the Dutch Ministry of Defence. He is an external PhD candidate at Tilburg University. His research focusses on the role of data science in military strategic decision-making. He recently demonstrated that Large Language Models can make wargaming more effective and accessible for strategic decision-making.

Roy Lindelauf

Roy Lindelauf is professor of Data Science at the Netherlands Defense Academy and leads the Data Science Center of Excellence at the Netherlands Ministry of Defence. In addition, he holds an endowed chair at Tilburg University as professor of data science, safety, security.

HCSS
Lange Voorhout 1
2514 EA The Hague

Follow us on social media:
@hcssnl

The Hague Centre for Strategic Studies