On the Road to Designing Responsible AI Systems in Military Cyber Operations

Clara Maathuis
Open University of the Netherlands, Heerlen, The Netherlands
clara.maathuis@ou.nl

Abstract: Military cyber operations are increasingly integrating or relying to a specific degree on AI-based systems in one or more moments of their phases where stakeholders are involved. Although the planning and execution of such operations are complex and well-thought processes that take place in silence and with high velocity, their implications and consequences could be experienced not only by their targeted entities, but also by other collateral friendly, non-friendly, or neutral ones. This calls for a broader military-technical and socio-ethical approach when building, conducting, and assessing military Cyber Operations to make sure that the aspects and factors considered and the choices and decisions made in these phases are fair, transparent, and accountable for the stakeholders involved in these processes and the ones impacted by their actions and largely, the society. This resonates with facts currently tackled in the area of Responsible AI, an upcoming critical research area in the AI field that is scarcely present in the ongoing discourses, research, and applications in the military cyber domain. On this matter, this research aims to define and analyse Responsible AI in the context of cyber military operations with the intention of further bringing important aspects to both academic and practitioner communities involved in building and/or conducting such operations. It does that by considering a transdisciplinary approach and concrete examples captured in different phases of their life cycle. Accordingly, a definition is advanced, the components and entities involved in building responsible intelligent systems are analysed, and further challenges, solutions, and future research lines are discussed. Hence, this would allow the agents involved to understand what should be done, what they are allowed to do, and further propose and build corresponding strategies, programs, and solutions e.g., education, modelling and simulation for properly tackling, building, and applying responsible intelligent systems in the military cyber domain.

Keywords: cyber operations, cyber weapons, military operations, targeting, artificial intelligence, responsible AI.

1. Introduction

"Human technology starts with an honest appraisal of human nature. We need to do the uncomfortable thing of looking more closely at ourselves." (Tristan Harris)

Old conflicts continue in different forms and through new battles. These battles are conducted not only on conventional battlefields, but also in the information environment i.e., cyberspace. Therein, different actors i.e., state, non-state, and hybrid (Maathuis, Pieters & Van Den Berg, 2021) build skills/force for achieving goals through effective strategies. Over 100 states can launch military Cyber Operations against adversaries (Maathuis, Pieters & Van Den Berg, 2018) while having well-prepared cyber commandos and units (Smeets, 2018). In this process, intelligent technologies are used at increasing rate and scale for building intelligent systems to conducting military Cyber Operations (Brantly, 2016). Since AI is a disruptive technology containing a set of multiple technologies, one could say that is aligned with Thomas Edison's perspective on electricity: 'it is a field of fields…it holds the secrets which will reorganize the life of the world' (Schmidt et al, 2021). Thus, AI changes the world (NATO, 2021) and relationships between humans and machines, diffuses rapidly and broadly (Schmidt et al, 2021), and does these inside and through its natural environment i.e., cyberspace (Hartmann & Giles, 2020) no matter if adaptation to world's problems can be difficult since human intelligence processes are complex.

AI applications for military Cyber Operations are reconfiguring the action of an intelligent-cyber weapon if the state of an exploited vulnerability is changed by dynamically finding and exploiting another vulnerability, adapting weapon's action for limiting/avoiding unintended effects on collateral actors (Cox et al, 2019), or conducting proportionality assessment (Martellini & Trapp, 2020). However, such complex activities require vast-amounts of data, high computing power, up-to-date intelligence, advanced process knowledge, and compliance to the applicable legal-ethical frameworks. Ultimately, the ones responsible for targeting decisions are military Commanders, meaning that if they knew or should have known that the weapon used would produce massive collateral damage on civilian side, they should be responsible (Hallao et al, 2017). Additionally, AI systems are software-based, thus vulnerable to attack vectors (Reding & Eaton, 2020) through exploiting e.g., unknown software vulnerabilities, unproper communication defence, or failure of critical processes.

While the existing body of knowledge contains a rich plethora of studies relevant for grasping ethical aspects in military Cyber Operations, to the best of our knowledge, concrete definitions and assessments of challenges and corresponding solutions lack. This is the knowledge gap that this research tackles through transdisciplinary research in military Cyber Operations, military operations, AI ethics, and RAI fields where extensive literature review on scientific resources (e.g., IEEE publications/standards) and governmental resources (e.g., NATO and EU-Commission) was conducted focusing on the concepts, methods, and techniques relevant for building and conducting military Cyber Operations. Hence, following research objectives are addressed:

1. To propose a definition for RAI when building and conducting military Cyber Operations.
2. To propose an analytical model that captures the entities involved in these processes.
3. To structure and analyse challenges and recommendations for integrating RAI systems in these processes.

The remainder of this article is structured as follows. Section 2 presents related studies that consider diverse aspects for designing RAI systems when building and conducting military Cyber Operations. Section 3 proposes a definition for RAI and an analytical model in military Cyber Operations. Section 4 discusses challenges encountered when building and using RAI systems and presents recommendations applicable when using them in military Cyber Operations. Section 5 presents concluding remarks and future research ideas.

2. Related research

Research and practitioner communities formulate relevant questions and seek to build intelligent systems with a good purpose while being aware of their possible negative impact which should be prevented or eliminated when signals of its presence are detected. Hereof, Zhu et al (2022) stress that building and conducting AI-based military operations raises concerns on ethical risks associated, thus critical from a humanitarian standpoint. Additionally, the authors mention benefits like increasing accuracy and precision for decision-making, intelligence and targeting activities: facts of major importance in military Cyber Operations. Furthermore, Hartmann & Giles (2020) emphasize that due to increased data availability, computing power, and publicly available tools, cyber offenders can use successfully intelligent techniques that reach large audiences and produce significant harm e.g., deepfakes and artificial humanoid disinformation campaigns. Thusly, Hallaq et al (2017) envisage that future cyber strategies rely on AI while considering corresponding ethical issues and legal questions. These points call for diving into relevant aspects from the ongoing research and practitioner perspectives in the military ethics, AI Ethics, and RAI fields.

Dobos (2020) argues for understanding relevant aspects like power, conflict dynamics, moral conditioning and damage in war context. Furthermore, Finney & Mayfield (2018) point the importance of properly expressing self-awareness and an ethical code of behaviour e.g., the fiduciary duty of military officers when conducting military operations. Moreover, Kaurin (2016) analyse warriors' meaning in contemporary warfare which encapsulates warriors' personal identity, demands on them, and experience: aspects relevant when capturing and embedding human values when building RAI in military Cyber Operations. Petrozzino & Shapiro (2020) recommend the following actions for achieving ethical AI systems: i) creating ethical principles that drive organizational policies for supporting ethical analysis and open dialogue, ii) creating training and awareness on role-based AI ethics for leaders, policymakers, developers, and users, and iii) establishing diverse multidisciplinary AI teams to analyse ethical aspects from multi-stakeholder perspective. Canca (2020) considers that ethical principles are formed regarding autonomy, beneficence for avoiding harm and doing good, plus justice. Since responsibility has multiple meanings, Cheng, Varshney & Liu (2021) address it broader through social responsibility of AI i.e., human-value driven process where values like fairness, transparency, accountability, responsibility, safety, privacy and security, and inclusiveness are the principles, while designing socially responsible AI algorithms is the means. Peters et al (2020) propose two conceptual frameworks for integrating ethical analysis in engineering practices: the first considers integrating wellbeing support and ethical impact analysis in each engineering phase, and the second argues for wellbeing supportive design while reflecting and structuring ethical analysis. For managing AI ethical aspects through educating AI systems, Baker-Brunnbauer (2021) scrutinizes that the systems could be implicit ethical being forced preventing unethical results, explicit ethical by explicitly pointing the actions allowed/not allowed, and full ethical by benefiting free will and intention while having consciousness. Brundage et al (2020) consider institutional, software, and hardware mechanisms for building trustworthy AI systems: institutional for shaping or clarifying the incentives of people involved, software for embedding or enhancing interpretability, privacy-preserving aspects of AI systems, and hardware for securing hardware systems and processes.

IEEE developed the IEEE 7000 – 2021 standards for tackling ethical concerns during system design like the IEEE P7001 on Transparency of Autonomous Systems for developing autonomous systems able to assess own actions and understand decisions made, and IEEE P7002 on Data Privacy Process for managing privacy issues for systems collecting personal data (IEEE P7000, 2021). As Cyber Operations are software-based activities, relevant principles, guidelines, and methodologies could be proposed following such standards. Accordingly, EU aims to turn Europe into a hub for trustworthy AI as the Commissioner Thierry Breton said: "AI is a means, not an end…Today's proposals aim to strengthen Europe's position as a global hub of excellence in AI from the lab to the market" (EU Commission, 2021a). Hence, the European Commission came forward with useful programs and strategies like AI strategy, Coordinated Plan on AI, and Data Governance Act (EU Commission, 2021b). Particularly, the European Commission established seven key requirements for assuring trustworthy AI: human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, societal and environmental well-being, and accountability (EU Commission, 2019). Moreover, NATO (2020a)'s Deputy Secretary General Mircea Geoana argues that 'there are considerable benefits of setting up a transatlantic digital community operating on AI and emerging and disruptive technologies, where NATO can play a key role as a facilitator for innovation and exchange'. NATO stresses that a dynamic adoption of new technologies like AI and their responsible governance are fundamentally important (NATO, 2020b). From the same angle, the U.S. DoD campaigns for the adoption of AI ethical principles in (non-)combat functions for upholding legal, ethical, and policy commitments in this domain. Accordingly, DoD 'will exercise appropriate levels of judgement and care, while remaining responsible' for building and using AI capabilities, plus equitable, traceable, reliable, and governable (U.S. DoD, 2020).

The above discussed studies contribute to defining RAI for military Cyber Operations and to identifying and analysing challenges and recommendations embedding academic and practitioner perspectives.

3. Definition

Since the beginning of AI, people were interested formulating questions of not only technical nature, but also ethical trying to propose answers to aspects like its capability to emulate or surpass human intelligence, design choices, and the meaning, scale, and severity of its (mis)use (Russell & Norvig, 2021). AI is changing 'the face and pace of warfare' and could be used responsibly as force multiplier to support military decision-making processes through accuracy, precision, speed, and easier integration in other battlefields (Meritalk, 2021) e.g. target localization with network/communication information and access point or even broader through a common operating picture, automatically detecting target's vulnerabilities and building corresponding exploits for efficient engagement, and collateral damage prevention on civilian infrastructure (Slayer, 2020), or using intelligent decision making support system for proportionality assessment and targeting decisions (Maathuis, Pieters & Van Den Berg, 2021). However, until now responsibility was indirectly tackled in military Cyber Operations through notions like 'attack', 'target', and 'proportionality' mainly through legal lenses. This is the underlying motivation of this article as responsibility does not only imply considering, interpreting, and integrating principles and norms, but also socio-ethical values when building military Cyber Operations while taking precautionary measures for preventing, containing, limiting, and avoiding unintended effects (Agarwal & Mishra, 2021). Correspondingly, the underlying questions would be: How to build responsible AI-based systems and solutions in respect to principles, norms, and values when developing and conducting military Cyber Operations? And, as Dignum (2019) suggests: Who or what should be responsible for AI-based systems' decisions and actions? Can an AI-based system be accountable for its actions? To find answers for such critical questions, a proper definition for RAI in this domain is required while considering specific characteristics of cyberspace e.g., being able to directly influence and impact other battlefields/domains (Brantly, 2016). Hence, Dignum (2019) calls for a human-centred approach focused on human well-being and alignment with socio-ethical values and principles.

Taking a responsible stance implies incorporating ethics in AI systems i) in design through regulatory and engineering processes that support the design and evaluation, ii) by design through established behaviour of AI systems, and iii) for designers through codes of conduct, regulatory requirements, standards, and certification processes (Dignum, 2019). The author defines RAI as 'the development of intelligent systems according to fundamental principles and values.' Similarly, Agarwal & Mishra (2021) consider that to assure the applicability, repeatability, and success of RAI systems, corresponding aspects should be integrated during their whole life cycle. Therefore, we formulate the following definition for RAI in military Cyber Operations respecting existing studies (Dignum, 2019; Agarwal & Mishra, 2021; Cheng, Varshney & Liu, 2021; Maathuis, 2022):

RAI in the military cyber domain = a sub-field of AI that deals with the integration of socio-ethical and legal principles, norms, and values when designing, developing, deploying, and using AI methods, techniques, and technologies embedded in different military cyber systems and processes.

This means that a series of agents communicate and collaborate for building military cyber tools for developing and conducting military Cyber Operations, process depicted in an analytical model in Figure 1 with its components addressed below (DARPA, 2016; Dignum, 2019; Maathuis, 2022):

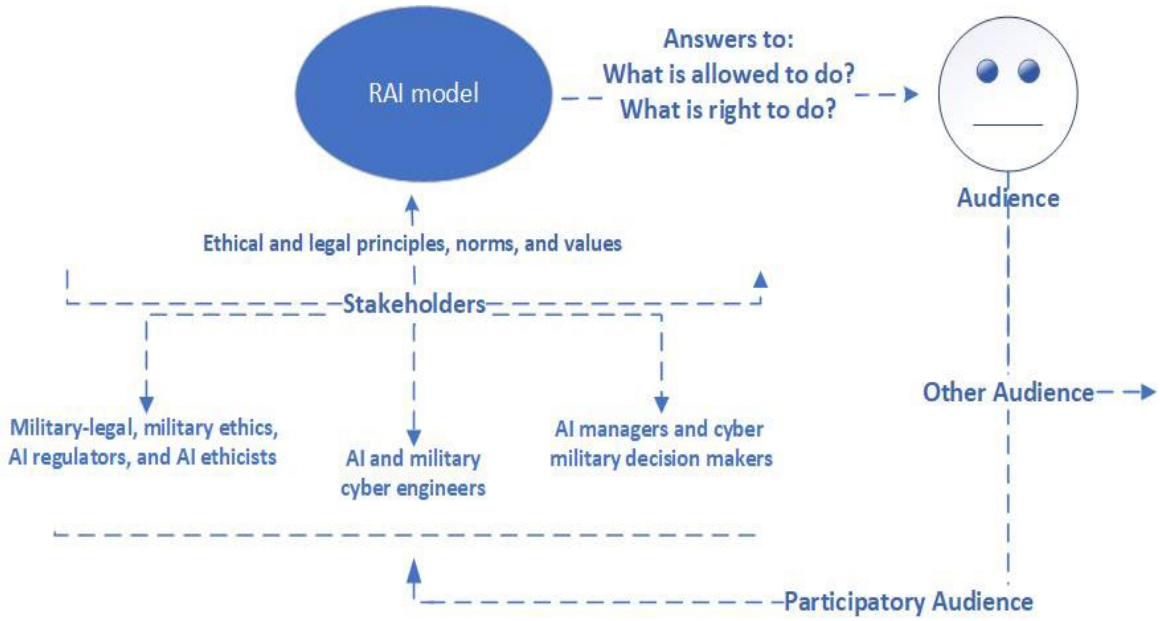

Figure 1: RAI in military cyber operations

Agents: entities participating in the design, development, deployment, and use of RAI solutions in military Cyber Operations. They are further classified considering their position:

4. *Stakeholders*: agents involved either in the process of i) design, deployment, use, standardization, and certification of the model i.e., military-legal, military-ethics, AI regulators, and AI ethicists, ii) theorizing, designing, developing, evaluating, upgrading, deploying the model, i.e. AI and military cyber engineers, or iii) design, development, deployment, and use while making sure that the model is compliant with external requirements i.e. AI managers and military cyber decision makers.
5. *Audience*: agents involved in stakeholders' processes i.e., are participatory audience or end-users, so the other audience.

The RAI model: developed AI model by corresponding agents whose life-cycle process and aim answers the following questions: What is allowed to do? and What is right to do? Important to mention is that agents have the responsibility, opportunity, and power to positively influence model's behaviour (Galliott, MacIntosh & Ohlin, 2020).

4. Challenges and recommendations

AI became an important strategic topic, and many countries are investing significant budgets in different R&D programs concerning upcoming and future trends and systems (EDA, 2021). Special attention is showed to building trustworthy, accountable, and responsible AI systems while reflecting on existing and foreseeable challenges (possibly) occurring in the lifecycle of RAI systems. This is vital when applied to military Cyber Operations since would mean e.g., mismatching a target implies spreading an intelligent cyber weapon in fractions of seconds at global scale and producing massive collateral damage on civilian infrastructure and vital processes, or directly affecting operational military processes of friendly and neutral countries. Thusly, it is necessary to understand and assess what are the challenges encountered when building RAI systems in military Cyber Operations, and from there analyse recommendations to tackle them. Hence, we consider the following categories of challenges and corresponding recommendations:

Education and Expertise: the lack of expertise, and implicitly education for properly integrating the aspects required at technical, social, and ethical levels into the design, implementation, and use of RAI systems in military Cyber Operations, could have (in)direct consequences on the built-system and other (un)related systems. It all begins with education, and more exactly, relevant and effective education. Then, the agents involved (e.g., military decision-makers) would benefit from individual and/or collective tailored training and education during mandatory training, as modular curriculum when joining/partnering with military forces, or as exchange curriculum between defence and commercial partners using gaming and simulation tools e.g., VR, AR, digital twins, or agent-based for target selection and engagement, that would allow understanding, capturing, and learning human's behaviour and values while building military Cyber Operations scenarios for effectively orienting, understanding, and acting in settings mirroring real-live contexts and environments (Dubber, Pasquale & Das, 2020; Meier et al, 2021; Reding & Eaton, 2020; Slayer, 2020).

Data: while symbolic AI systems rely on knowledge, non-symbolic AI system rely on data. Accordingly, knowledge and data are structured, represented, and further worked with as basis for understanding and tackling existing/future problems and unpredictable events that could occur considering e.g., network dynamism of cyberspace or unpredictable behaviour of AI systems as data might be errored, biased, or manipulated. Such facts could alter AI system's behaviour through unknown backdoors that allow e.g., disrupting own communication systems or improperly localizing a target (Dignum, 2019; Krasev, 2020). Then, aspects like data quality should be assured for solutions with sufficient data, and correctly balancing data e.g., oversampling with qualitative and representative technical and human-value data for solutions with scarce data like targeting decisions in military Cyber Operations.

Security: the over-reliance on AI systems conducts to an adversarial AI arms race and introduces new types of vulnerabilities (Reding & Eaton, 2020). AI-cyber vulnerabilities reflect combined and even extended cyber and AI risks to systems implemented e.g., data poisoning using open-source data possibly intentionally corrupted used for detecting advanced forms of cyber threats on military cyber systems or intelligent malware that changes its behaviour to be perceived as a legitimate behaviour and strike back into the network from where an intelligent cyber weapon was launched (Martellini & Trapp, 2020). As Norbert Wiener said: 'We had better be quite sure that the purpose put into the machine is the purpose which we really desire'. Then, intelligent systems able to both strike and defend themselves through online or hybrid learning and adaptive behaviour going through intense verification and testing processes at software, hardware, communication, and human levels represent a solution.

Cyberspace particularities: as these operations are conducted inside/through a multi-cross domain i.e., cyberspace which is dynamic, volatile, and still anonymity-friendly, then multi-domain and multi-source behavioural and value data are necessary to creating the proper picture to the agents involved in their execution along with a solid understanding on the processes involved and the effects assessed to the policy makers involved (Branthly, 2016; Slayer, 2020; Maathuis, 2021)**.**

Trust: are issues between humans while building AI systems due to unclear, unfair, or unexpected ways of tackling the aspects and values that should be integrated, and trust issues between the humans and the AI systems implying the reliability and power of predictability of AI systems. Hence, too much trust might expose to strong unexpected behaviour and too less trust might imply using too exigent control mechanisms which would still be exposed to unexpected behaviour of AI systems (Martellini & Trapp, 2020; Bartneck et al, 2021). Then, communication and collaboration between the agents involved when building AI systems for conducting military Cyber Operations, are needed while actively integrating in a fair process all their relevant elements e.g., researchers, developers, manufacturers, technologies (Cox, 2019; Dignum, 2019; Maathuis, 2022).

The Tragedy of Metrics: statement that aims to capture and extend the classical meaning of the word 'metrics' by adding socio-ethical norms and values that AI systems should respect (Dignum, 2019). The metrics should consider not only technical and military-(ethical and legal) dimensions when building solutions for conducting military Cyber Operations, but also other social and ethical dimensions and aspects e.g., the context, aim, environment, human behaviour, rules and regulation.

Governance and Regulation: currently no specific/dedicated regulation exist for building and conducting AI-based military Cyber Operations, and this is necessary as AI is a dual-use technology that requires and impacts not only defence and industry stakeholders, but society as a whole. However, considering the tendency in the

AI domain and the upcoming awareness in the military domain towards building, using, and assessing intelligent systems on e.g., cognition, interaction, well-being, dedicated incentives in programs for analysing the suitability of current legal frameworks to intelligent systems, and their interpretation and possible adaptation to them should be considered through constructive and positive lenses while seeing intelligent systems as artefacts (Dignum, 2019). This implies collaboration between agents involved when building and using AI systems using a human-centred approach, and the further consideration of third-party RAI certification, RAI auditing, and risk management processes for implementation, testing, and approving AI systems while adopting specific principles, norms, and values in each phase of the life cycle of AI systems, plus sharing problematic incidents involving RAI systems. This further calls for diplomatic solutions for establishing international dialogue and joint of forces for developing common/compatible legal frameworks for RAI systems with defence and industry partners respecting frameworks like IHL, Human Rights Law, and societal norms and values (Hallaq et al, 2017; Petrozzino & Shapiro, 2020; Shneiderman, 2020; EU Commission, 2021a ; Schmidt et al, 2021).

Design: since existing AI systems integrated in military Cyber Operations do not consider yet a responsible approach, for the upcoming ones, to assure the effectiveness of their implementation, responsible considerations should be integrated using methods like Value Sensitive Design, Data/Design Science Research while developing and adopting a code of conduct for AI systems respecting human values and ethical considerations captured both qualitatively and quantitatively in the design, development, deployment, and use of AI systems (Dignum, 2019; Agarwal & Mishra, 2021; Zhu et al, 2022) while being protective to environment (Galliott, MacIntosh & Ohlin, 2020). This allows translating agents' values into AI development and establishing concrete features like integrating conditions or duties for limiting civilian harm, required actions like target engagement only if the conditions are satisfied, and preferences like system training for a good purpose with realistic cases. Moreover, this allows going back to a particular step if a test case (e.g., bias) fails and update the system (Anderson & Anderson, 2014; Burkhardt, Hohm & Wigley, 2019; Agarwal & Mishra, 2021).

Developments: the fact that the AI research community is somehow divided between current technologies focusing on the now and near-term AI, and future implications and technologies focusing on long-term AI based on AGI and superintelligence i.e., radical transformation of AI, creates a gap between these communities which calls for joint effort for tackling existing and emerging security problems having an eye on near and long-term future (Prunkl & Whittlestone, 2020). Hence, what would that imply and mean for targeting decisions and effects assessment in miliary Cyber Operations?

5. Conclusions

Approaching AI systems in military Cyber Operations through techno-ethical lenses allows the stakeholders involved to understand the difference between what they have right to do and what is right to do (Pottery Stewart). In this digital decade (EU Commission, 2019) and further from here since these operations are carried out at fast speeds, in silence, and embed solutions with different autonomy degrees while assessing potential risks and taking corresponding precautions (Morgan et al, 2018), it is important to accelerate education, investments, democratization, and adoption of AI systems from their design to incorporate relevant norms and values while having realistic military objectives that imply avoiding/limiting harm and embracing good purposes (Maathuis, 2022).

Hence, we present a definition and analytical model for RAI applied in military Cyber Operations, and from there analyse the challenges encountered by the agents involved and further draw recommendations that would facilitate the adoption, support, and strengthening of RAI systems in military Cyber Operations focusing on their development and execution. However, as this research focuses on the theoretical foundation of and corresponding instantiations of this topic, it further argues for involvement of academic and industry communities for properly implementing AI-based military Cyber Operations in respect to legal and ethical dimensions, and continues by assessing them for their integration in targeting decisions and controlling, limiting, and avoiding unintended effects of military Cyber Operations on military and civilian stakeholders for assuring the design, implementation, and use of trustable, accountable, and responsible intelligent systems having in mind that 'humans cannot be everywhere at once, but software can" (Schmidt et al, 2021).

References

Agarwal, S. and Mishra, S. (2021). Data and Model Privacy. In *Responsible AI*, pp. 153-170.

Anderson, M. and Anderson, S. L. (2018). GenEth: A general ethical dilemma analyzer. *Paladyn, Journal of Behavioral Robotics*, Vol. 9, No. 1, pp. 337-357.

Baker-Brunnbauer, J. (2021). Management perspective of ethics in artificial intelligence. *AI and Ethics*, Vol. *1, No.* 2, pp. 173-181.

Bartneck, C., Lütge, C., Wagner, A. and Welsh, S. (2021). Risks in the Business of AI. In *An Introduction to Ethics in Robotics and AI*, pp. 45-53.

Brantly, A. F. (2016). *The decision to attack military and intelligence cyber decision-making*. University of Georgia Press.

Brundage, M. et al. (2020). Toward trustworthy AI development: mechanisms for supporting verifiable claims. *arXiv preprint arXiv:2004.07213*.

Burkhardt, R., Hohn, N. and Wigley, C. (2019). Leading your organization to responsible AI. *McKinsey Analytics*.

Canca, C. (2020). Operationalizing AI ethics principles. *Communications of the ACM*, Vol. 63, No. 12, pp. 18-21.

Cheng, L., Varshney, K. R. and Liu, H. (2021). Socially responsible ai algorithms: Issues, purposes, and challenges. *arXiv preprint arXiv:2101.02032.*

Cox, J., Bennett, D., Lathrop, S., Walls, C., LaClair, J., Tracy, C. and Esquibel, J. (2019). The Friction Points, Operational Goals, and Research Opportunities of Electronic Warfare and Cyber Convergence. *The Cyber Defense Review*, Vol. *4*, No. 2, pp. 81-102.

DARPA. (2016). "Explainable Artificial Intelligence", [online], https://www.darpa.mil/program/explainable-artificial-intelligence.

Dignum, V. (2019). *Responsible artificial intelligence: how to develop and use AI in a responsible way*. Springer Nature.

Dobos, N. (2020). *Ethics, Security, and the War Machine: The True Cost of the Military*. Oxford University Press.

Dubber, M. D., Pasquale, F. and Das, S. (Eds.). (2020). *The Oxford handbook of ethics of AI*. Oxford Handbooks.

European Defence Agency.

EU Commision, E. (2019). Communication from the Commission of the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions Empty.

EU Commision, E. (2021a). Europe Fit for the Digital Age: Commission Proposes New Rules and Actions for Excellence and Trust in Artificial Intelligence.

EU Commision, E. (2021b). A European Approach to Artificial Intelligence.

Finney, N. K. and Mayfield, T. O. (Eds.). (2018). *Redefining the modern military: The intersection of profession and ethics*. Naval Institute Press.

Galliott, J., MacIntosh, D. and Ohlin, J. D. (Eds.). (2020). *Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare*. Oxford University Press.

Hallaq, B., Somer, T., Osula, A. M., Ngo, K. and Mitchener-Nissen, T. (2017). Artificial intelligence within the military domain and cyber warfare. In *Proceedings of the 16th European Conference on Cyber Warfare and Security*, pp. 153-157.

Hartmann, K. and Giles, K. (2020). The Next Generation of Cyber-Enabled Information Warfare. In *2020 12th International Conference on Cyber Conflict,* pp. 233-250. IEEE.

IEEE (2021). IEEE Ethics in Action in Autonomous and Intelligent Systems.

Karasev, P. A. (2020). Cyber Factors of Strategic Stability. *Russia in Global Affairs*, Vol. 18, No. 3, pp. 24-52.

Kaurin, P. M. (2014). *The warrior, military ethics and contemporary warfare: Achilles goes asymmetrical*. Ashgate Publishing, Ltd..

Maathuis, C., Pieters, W. and Van Den Berg, J. (2018). A computational ontology for cyber operations. In *Proceedings of the 17th European Conference on Cyber Warfare and Security*, pp. 278-288.

Maathuis, C., Pieters, W. and Van Den Berg, J. (2021). Decision support model for effects estimation and proportionality assessment for targeting in cyber operations. *Defence Technology*, Vol. *17*, No. 2, pp. 352-374.

Maathuis, C. 2022. On Explainable AI Solutions for Targeting in Cyber Military Operations. *In Proceedings of the 17th International Conference on Cyber Warfare and Security*.

Martellini, M. and Trapp, R. (Eds.). (2020). *21st Century Prometheus: Managing CBRN Safety and Security Affected by Cutting-Edge Technologies*. Springer Nature.

Meier, R., Lavrenovs, A., Heinäaro, K., Gambazzi, L. and Lenders, V. (2021). Towards an AI-powered Player in Cyber Defence Exercises. In *2021 13th International Conference on Cyber Conflict,* pp. 309-326. IEEE.

Meritalk (2021). "Austin: DoD to Invest $1.5 Billion in DARPA AI Projects Over Five Years", [online], https://www.meritalk.com/articles/austin-dod-to-invest-1-5-billion-in-darpa-ai-projects-over-five-years/

Morgan, F. E., Boudreaux, B., Lohn, A. J., Ashby, M., Curriden, C., Klima, K. and Grossman, D. (2018). Military Applications of Artificial Intelligence. *Ethical Concerns in an Uncertain World. RAND Corporation.*

NATO (2020a). "Cooperation on Artificial Intelligence will boost security and prosperity on both sides of the Atlantic", [online], https://www.nato.int/cps/en/natohq/news_179231.htm

NATO (2020b). "Artificial Intelligence at NATO: dynamic adoption, responsible use", [online], https://www.nato.int/docu/review/articles/2020/11/24/artificial-intelligence-at-nato-dynamic-adoption-responsible-use/index.html

NATO (2020c). NATO Advisory Group on Emerging Disruptive Technologies. NATO Annual Report 2020.

NATO (2021). "Emerging and Disruptive Technologies", [online], https://www.nato.int/cps/en/natohq/topics_184303.htm

Peters, D., Vold, K., Robinson, D. and Calvo, R. A. (2020). Responsible AI—two frameworks for ethical design practice. *IEEE Transactions on Technology and Society*, Vol. *1*, No. 1, pp. 34-47.

Prunkl, C. and Whittlestone, J. (2020, February). Beyond near-and long-term: Towards a clearer account of research priorities in AI ethics and society. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, pp. 138-143.

Petrozzino, C. and Shapiro, S. (2020). *Actionable Ethics for Fairness in AI*. MITRE CORP MCLEAN VA.

Reding, D. F. and Eaton, J. (2020). *Science and Technology Trends 2020 2040: Exploring the S and T Edge*. NATO S and T Organization.

Russell, S. and Norvig, P. (2021). Artificial Intelligence: A Modern Approach, Global Edition 4th.

Schmidt, E. et al (2021). *National Security Commission on Artificial Intelligence (AI)*. National Security Commission on Artificial Intelligence.

Shneiderman, B. (2020). Bridging the gap between ethics and practice: Guidelines for reliable, safe, and trustworthy Human-Centered AI systems. *ACM Transactions on Interactive Intelligent Systems*, Vol. 10, No. 4, pp. 1-31.

Slayer, K. M. (2020). *Artificial Intelligence and National Security*. Congressional Research SVC Washington United States.

Smeets, M. (2018). Integrating offensive cyber capabilities: meaning, dilemmas, and assessment. *Defence Studies*, Vol. 18, No. 4, pp. 395-410.

U.S. Department of Defense (2020). "DoD Adopts Ethical Principles for Artificial Intelligence", [online], https://www.defense.gov/News/Releases/Release/Article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/

Zhu, L., Xu, X., Lu, Q., Governatori, G. and Whittle, J. (2022). AI and Ethics—Operationalizing Responsible AI. In *Humanity Driven AI*, pp. 15-33.