Human-Centred AI in Military Cyber Operations

Clara Maathuis
Open University of the Netherlands, Heerlen, The Netherlands

clara.maathuis@ou.nl

Abstract: Military Cyber Operations are an integral part of modern warfare and national security strategies as they crossed the science fiction realm, represent a real operational battlefield, and developed into an option in military toolboxes. Seeing ongoing technological advancements that allow the creation and use of complex mechanisms and technologies, the increasing digitalization of critical infrastructure, the growing abundance of data collected, generated, and exchanged between multiple parties, and the rise of stakeholders engaging in building and/or executing military Cyber Operations together with the increased number of such operations being conducted all over the globe reflects important lessons that need to be learned: the core element of such operations are humans: they build, acquire, execute, and assess them while also being impacted entities by them, e.g., through (psychological or physical) injury or death, or damage or destructions of human infrastructure. Moreover, the key process governing all the life cycle phases of military Cyber Operations is human decision-making or humanly assisted/augmented decision-making relying on advanced intelligent methods built with AI. Nevertheless, building and conducting military Cyber Operations should be done in a legal, responsible, and effective way implying a deep understanding of the context and adversary, proper target and cyber weapon selection, development, and use, and a clear overview of potential effects produced. These represent important aspects that should be properly defined and tackled in this domain. Hence, this research aims to introduce the Human-Centred AI concept and approach in the military cyber domain to illustrate ways to prioritize human involvement and interaction, human understanding, effective decision-making, and ethical considerations when building and conducting military Cyber Operations. To this end, an extensive literature review is conducted in the military, cyber, and AI domains together with instantiation on military Cyber Operations.

Keywords: Human-centred AI, Trustworthy AI, Responsible AI, Cyber operations, Military operations

1. Introduction

motto: "Machines take me by surprise with great frequency." (Alan Turing)

AI emerged as a transformative force in the military domain, redefining the nature of modern warfare. Its unmatched capacity to process massive datasets and combine it with field knowledge and expertise empowers situational awareness, facilitates data, knowledge, or hybrid-driven decision-making, and optimizes resource allocation, not only in traditional military operations but also in the intricate military cyber domain. In the cyberspace battlefield, military Cyber Operations are conducted as independent or part of broader military operations to achieve military goals against adversaries using cyber weapons/capabilities (Maathuis, Pieters & van den Berg, 2018a). Nevertheless, to integrate AI in this domain and benefit from its advantages while dealing with its challenges, the stakeholders involved need to assure that AI not only that it supports or enhances human decision-making (Ahmed, 2022), but that is able to facilitate cyber capabilities that are safe, responsible, and trustworthy, thus human centered. This can be done by assuring a harmonious balance between harnessing AI's potential and upholding international social and ethical norms and values since the design of AI systems and going forward in their life cycle when building and conducting military Cyber Operations while actively involving the stakeholders in this process through a socio-technical approach (Makarius et al., 2020). This perspective is called Human Centered AI (HCAI), and in the AI domain as well as direct critical application domains like defense, this area is surrounded by open questions and challenges (Vorm, 2020). Therein, humans are placed at the core of decision-making processes while AI and human agents form teams for achieving well-defined objectives and adhering to socio-ethical norms and values like international law, transparency, accountability, and safety (Shneiderman, 2020a).

Nonetheless, the focus on employing a HCAI approach within military Cyber Operations highlights a critical knowledge gap that requires immediate consideration from both researchers and practitioner efforts. This calls for creating and building corresponding concepts, methodologies, and techniques, effectively bridging the gap between theory and practical implementation. Addressing this knowledge gap is essential to unlock the full potential of AI techniques while upholding social and ethical norms, principles, and values. In pursuit of this objective, this research seeks to establish a clear understanding of the HCAI concept within this context and introduce a design framework for its responsible development and use. To this end, a transdisciplinary research approach is employed, drawing upon insights from the fields of AI, AI ethics, military studies, cybersecurity, and military Cyber Operations. This approach follows the Design Science Research methodology while incorporating perspectives rooted in the Values Sensitive Design approach. Accordingly, a

comprehensive literature review is conducted, complemented by illustrative scenarios for exemplifying the assessment of proportionality in military Cyber Operations, serving as concrete instantiation to support both the definition and framework proposed.

The outline of this article is organized as follows. Section 2 sets the background of this research and discusses related studies. Section 3 presents the definition proposed for HCAI in military Cyber Operations. Section 4 advances a design framework for HCAI in this domain with exemplification on the proportionality principle. At the end, Section 5 briefly discusses the findings and future perspectives of this research.

2. Research Approach and Background

This research aims to define the concept of HCAI in the military Cyber Operations domain and propose a framework for building and using HCAI systems in a responsible way. Accordingly, the following research questions are formulated:

- How to define the HCAI concept in the context of military Cyber Operations?
- How to design a framework for building and using HCAI systems in military Cyber Operations in a responsible way?

To achieve this aim, a transdisciplinary research approach is considered by including knowledge, methods, and techniques from the AI, AI ethics, military, cyber security, and military Cyber Operations domains using the Design Science Research methodology while respecting the Values Sensitive Design perspective (Peffers et al., 2007; Umbrello & Van de Poel, 2021; Maathuis, 2022a). An extensive literature review is conducted in the ACM Digital Library, IEEE Digital Library, Scopus, and Google Scholar scientific databases using combinations of keywords like military operations, Cyber Operations, human centered, and HCAI.

A socio-technical approach is promoted by (Makarius et al., 2020) where the authors state that bringing AI into organizations, systems, and processes implies understanding the novelty of AI systems and its scope dimensions. The authors argue that AI positions itself apart from its technological predecessors as it implies interaction between the physical, digital, and biological worlds while implying cognitive, relational, and structural issues like strategic decision making and knowledge sharing, trust and identification, coordination, training and development, and socialization. Sperrle et al. (2021) provide a comprehensive overview of evaluations for human-centred machine learning focusing on human-related factors that influence trust, interpretability, and explainability. Usmani, Happonen & Watada (2023) argue that HCAI systems should be developed for working together with humans, involving them in the decision-making processes, and leverage human expertise, intuition, and creativity through ethical considerations and effective human-AI collaboration.

Van den Bosch (2018) identify three characteristics of successful joint activities in human-AI collaboration: be mutually predictable in their actions, be mutually directable, and maintain common ground. Geng & Varshney (2019) model human-machine collaborative decision-making in settings that include random local thresholds, decision fusion in integrated human-machine networks, and binary decision making under cognitive biases. (Shneiderman, 2020a) propose a generic framework for HCAI for building reliable, safe, and trustworthy systems upon the following pillars adopted in the present research: design for high level of human control and high levels of computer automation for increasing human performance, understand the situations where full human control or full computer control and necessary, and avoid the dangers of excessive human control or excessive computer control. The classical OODA (Observe, Orient, Decide, Act) loop is used by integrating the following elements: human values and knowledge, trust and needs, capabilities and intentions, and environment, AI internal model, and the external world. In the same lines, Shneiderman (2020b) provides a series of guidelines for building reliable, safe, and trustworthy HCAI systems, i.e., sure that the systems are reliable based on sound software engineering practices, promote a safety culture through business management strategies, and trustworthy certification assured through independent oversight. Cronholm & Göbel (2020) gather design principles for developing HCAI solutions in decision-making processes: (i) design for amplified decision-making and possible responses from both humans and AI, i.e., a positive response from human and AI, a negative response from both human and AI, a positive response from human and a negative response from AI, and a negative response from human and a positive response from AI, (ii) design for unbiased decision-making through the involvement of several roles and consensus-based decision-making by including multiple perspectives for reducing biases, uncertainty, and vagueness of human judgement, and (iii) design for human and AI learning through procedural support that enables both human and machine learning which will improve the overall performance and capability of the system.

Ozmen Garibay et al. (2023) identify the following six challenges for building HCAI systems: human well-being that asks for considering the unique features of AI that might impact well-being while applying foundational and actionable characteristics of well-being oriented AI; responsible design of AI including values like transparency and fairness; privacy arguing for the right to be alone; limit access to the self/shields self from access, secrecy, and personhood; design and evaluation framework; governance and independent oversight; and human-AI interaction. For autonomous driving, De Caro et al. (2022) proposes an AI-as-a-Service toolkit for human-centred intelligence integrating a data gathering ecosystem from wearables for automatizing stress recognition. The core element is human behaviour, which is modelled in relation to vehicle's state, environment's state, and driver's state.

For building trustworthy robots and autonomous systems, He et al. (2021) investigate challenges for HCAI like safety when dealing with dynamic environments, and security issues from unknown software, hardware, and communication vulnerabilities. The authors call for tackling these issues by adopting a temporal vision that focuses on immediate, near-and-medium-term, and longer-term concerns. In the military domain, (He et al., 2023) identify the following open challenges for implementing HCAI solutions: the use of automated planning algorithms to aid Commanders in complex and time-critical decisions, the need for efforts that enable huma users to provide feedback in ways that are natural, flexible, and that result in proper updates and future learning for the algorithms, and the use of predictive dashboards for unit readiness in activities like acquisition program planning, manning, and recruiting. Ahmed (2022) investigate the role of ML models and the collaboration between humans and ML models during activities taken in the military intelligence operation process, e.g., when establishing intelligence requirements, during estimation and analysis, and preparing intelligence reports. Such activities are integrated when building and deploying, for instance, video surveillance and acoustic detector systems.

In the cyber security domain, Ansari et al. (2018) build a problem-solving ontology for cyber physical production systems identifying as problem characteristics: isolation (unconnected), connected (related), repetitive, unique (singular), transformable into a standard problem, ex ante (based on forecasts), and ex post (based on actual results). The study identifies complementary situations between humans and machines like fulfilment of various tasks, information processing, problem solving, and decision making. Dustdar, Nastic & Scekic (2016) bring a new vision of cyber-human smart city built upon a values architecture structured as follows: core technologies, enablers (provisioning and governance, incentive management, activity coordination and social orchestration, and monitoring with data analytics), value generation, added-value services, and smart city goals (quality of life, sustainability, and development). Maathuis (2023) introduces the concept of Human Centred Explainable AI (HCXAI) in the military cyber domain and proposes a framework for building and integrating HCXAI solutions in military Cyber Operations.

Governance efforts are also directed to HCAI by different stakeholders. The EU Commission published its vision on building trust in human-centric AI stressing the consideration of the values of respect for human dignity, freedom, democracy, equality, the rule of law and respect for human rights. Accordingly, the following seven requirements should be accounted for developing and achieving 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, (EU Commission, 2023) provides a clear overview of efforts for building HCAI solutions by EU countries while stressing the importance of respecting and implementing social and ethical norms and values like transparency, safety, privacy, and inclusion. UNESCO argues that for ensuring ethical AI development and use, laws, cooperation and collaboration, impact assessment, governance mechanisms should be considered (UNESCO, 2021). IEEE, through its Global AI systems (AIS) well-being initiative (IEEE, 2020) aims to propose recommendations and resources for building AI systems now and in the future that align and prioritize the human well-being and ecological sustainability.

This extensive review illustrates that several research, practitioner, and governmental efforts recently started for building and using HCAI systems in a responsible way. Nevertheless, a limited range of studies were conducted in the military and cyber domains, while in the domain of military Cyber Operations this represents a knowledge gap that this research aims to tackle.

3. Definition

HCAI represents a paradigm shift in the AI realm and recently became an important AI research area as it aims at building system that serve a greater purpose than just innovation – they are directed to supporting or

enhancing the overall human experience. This implies that when AI systems encounter humans, they need to understand how humans behave and what they want while humans need to make sure that AI systems embed methods that help humans understanding them (Riedl, 2019). Humans and AI become teammates where AI provides input to humans to make decisions and act (bring AI in the loop of human intelligence) while human intelligence is fundamental to building AI (human intelligence in the AI loop): this relates to combining human intelligence with AI into hybrid intelligence for reaching superior results then could have accomplished separately (Dellermann et al., 2019). Moreover, HCAI implies a socio-technical perspective and defines systems that continuously improve (Cronholm & Göbel, 2019) focusing not only on human input, but also on interactions, collaboration, and the involvement of diverse stakeholders while enabling feedback and model development (Sperrle et al., 2021). Nowak, Lukowicz & Horodecki (2019) relate HCAI to reasoning implying that HCAI needs to understand human reasoning through morals, motivations, and emotions.

The stakeholders involved consider a user-centered participatory design methods for engagement with diverse stakeholders while putting humans at the center of design, thinking, emphasizing user experience design by measuring aspects like satisfaction, interests, needs, and ensuring meaningful control (Shneiderman, 2020b; Mhlanga, 2022). This would facilitate benefits like making well-informed decisions, scalability, and dependability, building highly effective software and products, increasing inclusivity, and customized user experiences. Schmager & Vassilakopoulou (2023 propose the following HCAI guidelines: build reliable and transparent systems based on software engineering practices, pursue safety through effective business management strategies, and increase trust through certification and independent oversight to assure transparency and accountability enhance innovation, public confidence, and societal values.

Table 1: Key elements in HCAI definitions

Key element in existing HCAI definitions	Source
Contact	Riedl (2019)
Connection	Dellermann et al. (2019)
Engagement	Shneiderman (2020b)
Reliability, transparency, safety, and trust	Schmager & Vassilakopoulou (2023)
Decision, control, inclusion	Mhlanga (2022)
Socio-technical perspective	Cronholm & Göbel (2022)
Feedback	Sperrle et al., (2021)
Reasoning	Nowak, Lukowicz & Horodecki (2019)

These definitions tackle important features or key elements that characterize HCAI and are depicted in Table 1. As they provide a general perspective of this concept, for making sure that the development of HCAI systems is done in a responsible way while considering their possible impact and for preventing confusion between the stakeholders involved, this concept is tailored and defined in the military domain as follows:

HCAI in MCO = a sub-field of AI that deals with centring human needs, reasoning, input, control, and feedback in the design, development, deployment, and use of AI systems built and/or used for conducting military Cyber Operations and building and/or using cyber weapons/capabilities.

Its elements are defined as follows:

- *Centring human needs, reasoning, input, control, and feedback* capture the major socio-technological features that characterize HCAI.
- *Design, development, deployment, and use* of AI systems refer to the life cycle phases of AI systems considering a software engineering approach.
- *Military Cyber Operations* are independent military operations or part of other broader military operations that make use of cyber weapons/capabilities for reaching military goals (Maathuis, Pieters & Van den Berg, 2018b).

4. Framework

The proposed definition brings together the socio-technical elements of this concept and paradigm. This holistic approach positions humans in a dimension where AI is a trusted ally helping human stakeholders to navigate and tackle the complexities of this domain while achieving their goals and preserving the socio-ethical norms and values. For enhancing awareness, trust, and resilience of stakeholders and AI systems involved when conducting military Cyber Operations, the HCAI triad and framework are further introduced and

presented in Figure 1, containing three layers (technical, socio-technical, and social) guided by three core actions of building intelligent systems (understanding, building, and deciding) while embedding human-interaction ke- elements (involvement, interaction, and control).

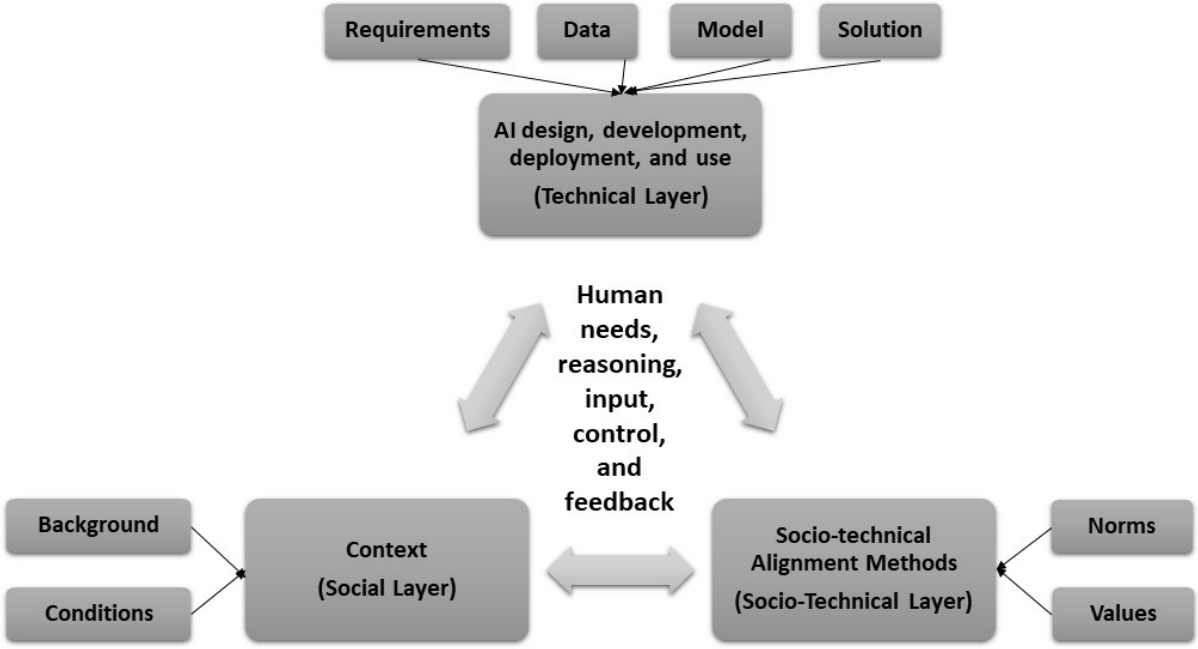

Figure1: The HCAI triad and framework

Human needs, reasoning, input, control, and feedback core which represent the core of HCAI being expressed by (Ansari et al., 2018; Shneiderman, 2020a; Maathuis, 2023; Manzoor et al., 2023):

- *needs* that include human aims, vision, experiences, and expectations of stakeholders involved. These should be directly implemented in context conditions and technical requirements, and be aligned with socio-ethical norms and values.
- *reasoning* that enables AI systems to navigate, capture, and model human needs, moral dilemmas, and make choices that prioritize human well-being and achievement of the objectives defined.
- *input* which can be provided in all phases of AI systems development, thus from defining the objective and design up to the deployment of AI systems. Given current and future developments in the field of generative AI, the input is multimodal, i.e., has different types of data (e.g., text, image, audio) that the model can understand, process, or further generate.
- *control* which embeds the level of control that humans have when building and using AI systems thinking in a scheme of problem and solution profile, problem-solver profile, and problem-solving profile that includes different levels of autonomy, from no autonomy to partial or conditional, to full autonomy while including different forms and characteristics of assistance, e.g., warnings, indicators, and notifications. In an ideal situation, building a reliable, safe, and trustworthy AI system implies having a high level of human control while having a high level of automation.
- *feedback* implies central role as it sustains continuous improvement, the effectiveness of AI systems, and assures the alignment between AI and human needs plus socio-ethical norms and values.

A direct example that can be considered is the proportionality assessment in military Cyber Operations (Maathuis & Chockalingam, 2023). The stated military objectives to be achieved *(i.e., needs)* are merged with assessments for components like collateral damage to assure that the unintended effects of using a specific cyber weapon (i.e., malware) would not be excessive in relation with the anticipated military advantage *(i.e., reasoning).* Accordingly, different types and sources of data are used (i.e., input) for aspects like target geolocation and cyber weapon characteristics. Concrete levels of autonomy and interaction of the AI-based cyber weapon are established since its design phase and corresponding methods and techniques are implemented to build a reliable, safe, and trustworthy AI solution while assuring real interaction between it and the stakeholders involved (i.e., control and feedback) for aspects like vulnerability identification, weapon extraction, and impact assessment.

Context which represents the social layer where the environment, and circumstances of building the AI system are considered (Makarius et al., 2020; He et al., 2021; Gao et al., 2021):

- *background* includes information about the military Cyber Operation, current strategic, economic, civilian, cultural, and media situations, previous lessons learned from former assessments, strategic, operational, and tactical requirements, and information about the stakeholders involved or possibly impacted through its execution.
- *condition*s refer to specific aspects or requirements that should be accounted when building the AI systems and conducting the military Cyber Operation, e.g., ROE and ISR reports, to assure that informed decisions are made in an adaptive manner taking into consideration the dynamism and uncertainty surrounding this battlefield.

For the same example, previous insights about civilian objects in the area and current information about a zero-day vulnerability exist *(i.e., background)* and can be further used while respecting the precise military-legal and political conditions stated in the corresponding ROE defined for this operation *(i.e., conditions).*

Socio-technical alignment methods which represents the socio-technical layer where social-ethical norms and values are included to embed the implementation of ethical standards, legal compliance, and codes of conduct of AI systems used (Ansari et al., 2018; Shneiderman, 2020a; Shneiderman, 2020b; Dustdar, Nastic & Scekic, 2016; Chockalingam & Maathuis, 2022):

- *norms* refer to the adherence to applicable laws and regulations that govern the development and use of AI systems and data privacy plus military cyber professional codes of conduct established by the stakeholders involved.
- *values* point to respecting ethical and moral aspects like transparency, accountability, fairness, beneficence and non-maleficence, privacy, and fairness.

In the example above, IHL compliance is mandatory by respecting the principles of distinction, precaution, and proportionality while respecting Human Rights by protecting civilians and the right of life *(i.e., norms).* Accordingly, dedicated efforts should be developed for engaging the desired military target with the cyber weapon that would not produce disproportionality and would make sure that the right of life is guaranteed while the avoidance or minimization of collateral damage is assured before its use *(i.e., values).*

AI design, development, and use which represents the technical layer that encompasses the AI life cycle phases used in a military Cyber Operation including the following core elements (Maathuis, 2022b):

- *requirements* are technical functional and non-functional guidelines for stakeholders involved and assure the successful development and deployment of AI systems.
- *data* points to multi-source and multi-nature data, real from the field or previous incidents, simulated from current mirroring settings or previous operations, or synthetically generated based on well-established profiles.
- *model* represents the AI model that can imply either a knowledge-based, data-based, or hybrid AI-based approach by relying just on knowledge or expertise, data, or a combination thereof.
- *solution* refers to the specific AI system that will be integrated and used in a military Cyber Operation.

For the same example, the technical requirements are established in terms of infrastructure at all levels, i.e., software, hardware, communication lines, and data (i.e., requirements and data), the paradigms, methods, and techniques are selected for designing the AI system (i.e., model) which is further built and prepared for use for assessing the proportionality in military Cyber Operations while assuring the avoidance or reduction of collateral damage (i.e., solution).

5. Conclusions

By prioritizing the development and integration of AI systems in the military domain that are designed with a deep understanding and centralization of human aims, needs, capabilities, and ethical considerations, the stakeholders involved when building and using them can ensure that these systems become force multipliers rather than risks (Vorm, 2020). Taking into consideration that the elements of the cyberspace battlefield are in a continuous movement and the adversary's capabilities constantly evolve, plus the agility, precision, and effectiveness offered by AI systems when building and conducting military Cyber Operations are invaluable (Usmani, Happonen & Watada, 2023; Maathuis, Pieters & van den Berg, 2018c). For instance, enhanced cyber threat detection, improved overall resilience to unknown cyber and even hybrid threats, and optimization of

cyber weapons effectiveness while minimizing its unintended effects while fostering values like transparency, trust, and safety all along with centering on the stakeholders involved and/or impacted by their actions, facilitates, or strengthens the successful collaboration between humans and AI systems on this battlefield as fundament of assuring a reliable, safe, responsible, and trustworthy human-machine partnership. Nevertheless, the emphasis on a HCAI approach in military Cyber Operations underscores a crucial knowledge gap that demands urgent attention from researchers and practitioners through a bridge between theory and practice by developing corresponding artefacts-concepts, methods, and techniques.

Tackling this knowledge gap is imperative to harness the full potential of AI while safeguarding the principles of humanity and socio-ethical norms, principles, and values in military Cyber Operations. Accordingly, this research aims to define the concept of HCAI in this context and propose a corresponding responsible framework for its development and use. To achieve these goals, a transdisciplinary research approach combines knowledge and methods from the AI, AI ethics, military, cyber security, and military Cyber Operations domains, using the Design Science Research methodology while incorporating the Values Sensitive Design perspective. Extensive literature review is conducted in combination with illustrative scenarios on the proportionality assessment in military Cyber Operations provided as exemplification for the definition and framework proposed. This research continues by integrating other aspects like cognitive and psychological ones in the decision-making processes, and by analyzing mechanisms for facilitating cross-domain collaboration and human-AI teaming for building responsible, safe, trustworthy military Cyber Operations.

References

Ahmed, N. U. (2022). Integrating machine learning in military intelligence process: study of futuristic approaches towards human-machine collaboration. *NDC E-JOURNAL*, *2*(1), 59-89.

Ansari, F., Khobreh, M., Seidenberg, U., & Sihn, W. (2018). A problem-solving ontology for human-centered cyber physical production systems. *CIRP Journal of Manufacturing Science and Technology*, *22*, 91-106.

Chockalingam, S., & Maathuis, C. (2022). An ontology for effective security incident management. In *International Conference on Cyber Warfare and Security* (Vol. 17, No. 1, pp. 26-35). Academic Conferences International Limited.

Cronholm, S., & Göbel, H. (2022). Design principles for human-centred AI.

De Caro, V., Bano, S., Machumilane, A., Gotta, A., Cassarà, P., Carta, A., ... & Bacciu, D. (2022). AI-as-a-Service Toolkit for Human-Centered Intelligence in Autonomous Driving. In *2022 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)* (pp. 91-93). IEEE.

Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J. M. (2019). Hybrid intelligence. *Business & Information Systems Engineering*, *61*, 637-643.

Dustdar, S., Nastic, S., & Scekic, O. (2016). A novel vision of cyber-human smart city. In *2016 fourth ieee workshop on hot topics in web systems and technologies (hotweb)* (pp. 42-47). IEEE.

EU Commission (2019). Building Trust in Human-Centric Artificial Intelligence.

E.U. Commission (2023). EU AI Act: The European approach to AI.

Geng, B., & Varshney, P. K. (2019). On decision making in human-machine networks. In *2019 IEEE 16th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)* (pp. 37-45). IEEE.

He, H., Gray, J., Cangelosi, A., Meng, Q., McGinnity, T. M., & Mehnen, J. (2021). The challenges and opportunities of human-centered AI for trustworthy robots and autonomous systems. *IEEE Transactions on Cognitive and Developmental Systems*, *14*(4), 1398-1412.

IEEE (2020). Global AI systems (AIS) well-being initiative.

Maathuis, C., Pieters, W., & van den Berg, J. (2018a). Developing a cyber operations computational ontology. *Journal of Information Warfare*, *17*(3), 32-49.

Maathuis, C., Pieters, W., & Van den Berg, J. (2018b). Assessment methodology for collateral damage and military (Dis) Advantage in cyber operations. In *MILCOM 2018-2018 IEEE Military Communications Conference (MILCOM)* (pp. 1-6). IEEE.

Maathuis, C., Pieters, W., & van den Berg, J. (2018c). A knowledge-based model for assessing the effects of cyber warfare. In *Proceedings of the 12th NATO Conference on Operations Research and Analysis*.

Maathuis, C., & Chockalingam, S. (2023). Modelling the Influential Factors Embedded in the Proportionality Assessment in Military Operations. In *International Conference on Cyber Warfare and Security* (Vol. 18, No. 1, pp. 218-226).

Maathuis, C. (2022a). On the Road to Designing Responsible AI Systems in Military Cyber Operations. In *European Conference on Cyber Warfare and Security* (Vol. 21, No. 1, pp. 170-177).

Maathuis, C. (2022b). An Outlook of Digital Twins in Offensive Military Cyber Operations. In *European Conference on the Impact of Artificial Intelligence and Robotics* (Vol. 4, No. 1, pp. 45-53).

Maathuis, C. (2023). Human Centered Explainable AI Framework for Military Cyber Operations. In *MILCOM 2023 IEEE Military Communications Conference (MILCOM)*. IEEE.

Makarius, E. E., Mukherjee, D., Fox, J. D., & Fox, A. K. (2020). Rising with the machines: A sociotechnical framework for bringing artificial intelligence into the organization. *Journal of Business Research*, *120*, 262-273.

Manzoor, M. A., Albarri, S., Xian, Z., Meng, Z., Nakov, P., & Liang, S. (2023). Multimodality Representation Learning: A Survey on Evolution, Pretraining and Its Applications. *arXiv preprint arXiv:2302.00389.*

Mhlanga, D. (2022). Human-centered artificial intelligence: the superlative approach to achieve sustainable development goals in the fourth industrial revolution. *Sustainability*, *14*(13), 7804.

Nowak, A., Lukowicz, P., & Horodecki, P. (2018). Assessing artificial intelligence for humanity: Will ai be the our biggest ever advance? or the biggest threat [opinion]. *IEEE Technology and Society Magazine*, *37*(4), 26-34.

Ozmen Garibay, O., Winslow, B., Andolina, S., Antona, M., Bodenschatz, A., Coursaris, C., ... & Xu, W. (2023). Six human-centered artificial intelligence grand challenges. *International Journal of Human–Computer Interaction*, *39*(3), 391-437.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems*, *24*(3), 45-77.

Riedl, M. O. (2019). Human-centered artificial intelligence and machine learning. *Human behavior and emerging technologies*, *1*(1), 33-36.

Schmager, S., P., I., & Vassilakopoulou, P. (2023). Defining Human-Centered AI: a Comprehensive Review of HCAI Literature.

Shneiderman, B. (2020a). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction*, *36*(6), 495-504.

Shneiderman, B. (2020b). Bridging the gap between ethics and practice: guidelines for reliable, safe, and trustworthy human-centered AI systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *10*(4), 1-31.

Sperrle, F., El-Assady, M., Guo, G., Borgo, R., Chau, D. H., Endert, A., & Keim, D. (2021). A Survey of Human-Centered Evaluations in Human-Centered Machine Learning. In *Computer Graphics Forum* (Vol. 40, No. 3, pp. 543-568).

Umbrello, S., & Van de Poel, I. (2021). Mapping value sensitive design onto AI for social good principles. *AI and Ethics*, *1*(3), 283-296.

Usmani, U. A., Happonen, A., & Watada, J. (2023). Human-Centered Artificial Intelligence: Designing for User Empowerment and Ethical Considerations. In *2023 5th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)* (pp. 01-05). IEEE.

UNESCO (2021). The Unesco recommendations on the ethics of AI: shaping the future of AI: shaping the future of our societies.

Van Den Bosch, K., & Bronkhorst, A. (2018, May). Human-AI cooperation to benefit military decision making. NATO.

Vorm, E. S. (2020). Computer-centered humans: why human-AI interaction research will be critical to successful AI integration in the DoD. *IEEE Intelligent Systems*, *35*(4), 112-116.