Artificial Intelligence, Simulation and Society

Kairi Talves
Dierk Spreen *Editors*

Artificial Intelligence in Military Technology

Sociological, cultural and ethical perspectives

OPEN ACCESS

Springer

Artificial Intelligence, Simulation and Society

Volume 192

Series Editor

Petra Ahrweiler, TISSS Lab, Institute for Sociology, Johannes Gutenberg University of Mainz, Mainz, Germany

This book series brings into its fold key and emerging topics on the interactions between growing artificial intelligence technologies and their social impacts. It addresses various aspects of the relationship between AI, simulation, and society and provides insights into their intersections and stimulates discussions on the opportunities and challenges they present. The series is multi- and transdisciplinary in scope, and dynamic. It invites academic contributed volumes and monographs, but also more popular work suitable for lay readership, and innovatively includes some science fiction to initiate readers into the scope and aims of this novel series.

The specific themes and topics covered under the series are:

- **The ethical and societal implications of AI**: The series delves into the ethical considerations and societal impacts of AI technologies. It explores topics such as privacy, bias, job displacement, and the role of AI in shaping social structures from a social science point of view (sociological, political, economic, cultural, legal).
- **Simulation and modeling of social systems**: The series explores how simulation techniques are used to model and understand complex social systems and create artificial societies in silico. It covers topics such as social network analysis, agent-based modelling (ABM), and the simulation of collective behaviour.
- **AI and social simulation**: The series explores how AI technologies are used in social simulation, for example, modelling intelligent agents in agent architectures of ABM, or calibrating and validating models using intelligent data mining and analysis techniques.
- **AI and simulation in social philosophy**: It looks at how AI and simulation are depicted in social philosophy, for example, the role of AI and simulation in socio-technical evolution, the position of AI and simulation in Western rationalism, philosophical counter-designs of current developments, ontological and epistemological limitations and barriers of AI and simulation.
- **AI, simulation and society in fiction**: The series also innovatively examines the portrayal of AI and simulation in and as fiction, demonstrating how these themes reflect societal fears, aspirations, and ethical dilemmas. The series contains both original fiction and second-order analyses.
- **AI and simulation in entertainment**: It covers simulation techniques, combined with AI, that are used to create virtual worlds and characters that mimic human behaviour. Such simulations are used, for example, in video games, virtual reality experiences, and entertainment applications.
- **AI and simulation in various disciplines**: The series discusses the applications of AI and simulations that are/will be transforming various disciplines and domains such as healthcare (e.g., in medical diagnosis, drug discovery, and patient care), work (e.g., automation, Industry 4.0, workforce dynamics), or education (e.g., virtual reality, personalised learning systems, intelligent tutoring systems). It discusses the potential benefits and challenges of integrating these technologies into the conventional space.
- **AI, simulation, and policy**: The series analyses how AI and simulation techniques can inform the policy cycles. It discusses the use of predictive modelling, analysis of what-if scenarios, and decision support systems in shaping policies in various policy domains such as public policy, technology policy or environmental policy.

Kairi Talves • Dierk Spreen
Editors

Artificial Intelligence in Military Technology

Sociological, cultural and ethical perspectives

Springer

Editors
Kairi Talves
Estonian Military Academy
Tartu, Estonia

Dierk Spreen
Institute for Organizational Communication
University of the Bundeswehr Munich
(UniBwM)
Neubiberg, Germany

Department of Business and Economics
Berlin School of Economics and Law
Berlin (HWR)
Berlin, Germany

This book was funded by the Estonian Military Academy (EMA) and published with the kind support of the Bundeswehr Centre for Military History and Social Sciences (ZMSBw).

ISSN 3004-9822 ISSN 3004-9830 (electronic)
Artificial Intelligence, Simulation and Society
ISBN 978-3-031-95577-8 ISBN 978-3-031-95578-5 (eBook)
https://doi.org/10.1007/978-3-031-95578-5

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Contents

Challenges for Communication

About the Authors

Franz Beitzinger is Senior Research Consultant at the Institute for Organizational Communication at the University of the Bundeswehr Munich. His research focuses on communication effects in public space, the framework conditions for Strategic Communication, and the impact of digitalization and new technologies on the domain of organizational communications.

- Zowislo-Grünewald, N., & Beitzinger, F. (2021). *Lehrbuch Strategisches Kommunikationsmanagement* (2nd ed.). LIT Verlag.
- Zowislo-Grünewald, N., & Beitzinger, F. (2021). Attirer des recrues. Informations contextuelles sur la communication de recrutement en ligne par la Bundeswehr. *Allemagne d'aujourd'hui*, (235), 128–140. https://shs.cairn.info/revue-allemagne-d-aujourd-hui-2021-1-page-128
- Beitzinger, F., & Zowislo-Grünewald, N. (2020). Zur Doppelrolle Strategischer Kommunikation von Streitkräften. Strategische Narrative als Instrument der Rollenkoordination. In M. Holenweger (Ed.), *Anwendungsgebiete und Grundlagen Strategischer Kommunikation* (pp. 347–364). Nomos. https://doi.org/10.5771/9783748904717-373
- Beitzinger, F. (2004). *Politische Ökonomie des Politikbetriebs. Die konzeptionellen Unterschiede verschiedener ökonomischer Theorietraditionen in Analyse und Bewertung politischer Ordnungen*. Lucius & Lucius.

Gabriela Beltrão is a Junior Research Fellow and PhD candidate at the School of Digital Technologies, Tallinn University, Estonia. She holds master's degrees in Human–Computer Interaction and Arts, Culture, and Society. Her research explores the interplay between cultural backgrounds and trust, focusing on interdisciplinary approaches to understanding how cultural contexts shape individuals' trust in technology. https://orcid.org/0000-0003-2852-2348

- Beltrão, G., Sousa, S., & Lamas, D. (2025). Assessing the Measurement Invariance of the Human–Computer Trust Scale. *Electronics, 14*(9), Artikel 1806. https://doi.org/10.3390/electronics14091806

- Beltrão, G., Sousa, S., & Lamas, D. (2024). Unmasking trust: Examining users' perspectives of facial recognition systems in Mozambique. In *AfriCHI'23: Proceedings of the 4th African Human Computer Interaction Conference* (pp. 38–43). https://doi.org/10.1145/3628096.3628746
- Bach, T. A., Khan, A., Hallock, H., Beltrão, G., & Sousa, S. (2022). A systematic literature review of user trust in AI-enabled systems: An HCI perspective. *International Journal of Human–Computer Interaction*, *40*(5), 1251–1266. https://doi.org/10.1080/10447318.2022.2138826
- Beltrão, G., & Sousa, S. (2021). Factors influencing trust in WhatsApp: A cross-cultural study. In C. Stephanidis, M. M. Soares, E. Rosenzweig, A. Marcus, S. Yamamoto, H. Mori, P.-L. P. Rau, G. Meiselwitz, X. Fang, & A. Moallem (Eds.), *HCI International 2021–Late Breaking Papers: Design and User Experience. HCII 2021. Lecture Notes in Computer Science* (vol. 13094, pp. 495–508). Springer. https://doi.org/10.1007/978-3-030-90238-4_35

Stuart Casey-Maslen is Extraordinary Professor at the University of Pretoria and teaches international human rights law, counterterrorism law, international criminal law, disarmament law, and the law of armed conflict. He has a doctorate in international humanitarian law and master's degrees in international human rights law and forensic ballistics. https://orcid.org/0000-0001-5181-4002

- Casey-Maslen, S. (2020). *Jus ad Bellum: The law on inter-state use of force*. Hart Publishing. https://doi.org/10.5040/9781509930722
- Casey-Maslen, S. (2024). *Hybrid warfare under international law*. Hart Publishing. https://doi.org/10.5040/9781509979608

Debora C. Firmino de Souza is a PhD candidate at Tallinn University, Estonia, specializing in trust dynamics within collaborative human-robot systems. Her research focuses on enhancing robots' communication mechanisms for successful integration in real-world scenarios. With a background in journalism and human–computer interaction, she investigates how specific robot attributes foster appropriate user trust in cooperative contexts. Debora's work aims to shape the design of intuitive and responsive systems, emphasizing the importance of user-centered, responsible, and ethical human–robot interaction. https://orcid.org/0000-0002-1975-0616

- Correia, N., Souza, D., Nêves, I., & Lobato, J. (2024). Bio-Electron - A multisensory approach to augmenting dance, combining: Biosignals, drawing, sound and electrical feedback. In *ISEA2023 Symbiosis. Proceedings. 28th International Symposium on Electronic Art, Paris* (pp. 336–345). https://doi.org/10.69564/ISEA2023-49-full-Correia-et-al-Bio-Elektron
- Campos, I., Brito, M., De Souza, D., Santino, A., Luz, G., & Pera, D. (2022). Structuring the problem of an inclusive and sustainable energy transition – A pilot study. *Journal of Cleaner Production*, 365, Article 132763. https://doi.org/10.1016/j.jclepro.2022.132763

- C. Firmino De Souza, D., Tikka, P., & Ajenaghughrure, I. B. (2023). Seeking emotion labels for bodily reactions: An experimental study in simulated interviews. In C. Biele, J. Kacprzyk, W. Kopeć, J. W. Owsiński, A. Romanowski, & M. Sikorski (Eds.), *Digital interaction and machine intelligence. MIDI 2022. Lecture Notes in Networks and Systems* (Vol. 710, pp. 127–138). Springer. https://doi.org/10.1007/978-3-031-37649-8_13

Frank Ole Flemisch is a Principal Scientist–Human Systems–Evangelist at the Fraunhofer Institute für Communication, Information Processing und Ergonomics in Wachtberg/Bonn, Professor for Human Systems Integration RWTH Aachen University, and Vice-Chair of NATO STO Human Factors & Medicine Panel. Starting as an aerospace engineer with a specialization in system dynamics, and some years as Tactical Control Officer and Deployment Officer in the German Luftwaffe, he is since 1994 continuously investigating AI-based assistance and automation systems and their interaction and cooperation with humans, organizations, and environments, in application domains like ground and air vehicles, Industry 4.0 and defense systems. Frank Flemisch is a principal researcher in the DFG-cluster of excellence "Internet of Production" and the speaker of the DFG-research group "MiRoVA" on migration of road vehicle automation.

- Flemisch, F., Baltzer, M., Abbink, D., Siebert, L. C., Diggelen, J. v, Herzberger, N. D., Draper, M., Boardman, M., Pacaux-Lemoine, M., & Wasser, J. (2024). Holistic bow-tie model of meaningful human control over effective systems: Towards a dynamic balance between humans and AI-based systems within our global society and environment. In G. Mecacci, D. Amoroso, L. C. Siebert, D. Abbink, M. J. van den Hoven & F. Santoni de Sio (Eds.), *Research Handbook on Meaningful Human Control of Artificial Intelligence Systems* (pp. 309–346). Edward Elgar Publishing. https://www.e-elgar.com/shop/gbp/research-handbook-on-meaningful-human-control-of-artificial-intelligence-systems-9781802204124.html
- Flemisch, F., Preutenborbeck, M., Baltzer, M., Wasser, J., Kehl, C., Grünwald, R., Pastuszka, H.-M., & Dahlmann, A. (2022). Human systems exploration for ideation and innovation in potentially disruptive defense and security systems. In G. Adlakha-Hutcheon, A. Masys (Eds.), *Disruption, ideation and innovation for defence and security. Advanced sciences and technologies for security applications* (pp. 79–117). Springer. https://doi.org/10.1007/978-3-031-06636-8_5
- Flemisch, F., Usai, M., Herzberger, N. D., Baltzer, M. C. A., Hernández, D. L., & Pacaux-Lemoine, M.-P. (2022). Human-machine patterns for system design, cooperation, and interaction in socio-cyber-physical systems: Introduction and general overview. In *IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 1278–1283). https://doi.org/10.1109/SMC53654.2022.9945181
- Flemisch, F., Heesen, M., Hesse, T., Kelsch, J., Schieben, A., & Beller, J. (2012). Towards a dynamic balance between humans and automation: Authority, ability, responsibility and control in shared and cooperative control situations. *Cognition, Technology & Work*, *14*, 3–18. https://doi.org/10.1007/s10111-011-0191-6

Bernd Flessner works as a futurologist at the Competence Center for Interdisciplinary Science Reflection (ZiWiS) at Friedrich-Alexander University Erlangen-Nuremberg (FAU). He completed his doctorate in 1991 with a thesis on technical prognoses in the works of Arno Schmidt and Stanisław Lem. As a scientific advisor to the Deutsches Museum in Munich, he helped design the Future Museum in Nuremberg.

- Flessner, B. (2022). Im emotionalisierten Raum. Human Factors in Hardware- und Softwaredesign von Robotern und Künstlicher Intelligenz. In K. Schäfer, K. Steinmüller, & A. Zweck (Eds.), *Gefühlte Zukünfte. Emotionen als methodische Herausforderung für die Zukunftsforschung* (pp. 199–218). Springer VS. https://doi.org/10.1007/978-3-658-35890-7_9
- Flessner, B. (2022). Destination Moon. Die Privatisierung der Raumfahrt in Science-Fiction und Realität. In D. Spreen, & B. Flessner (Eds.), *Die Raumfahrt der Gesellschaft. Wirtschaft und Kultur im New Space Age* (pp. 125–177). transcript. https://doi.org/10.14361/9783839457627-003
- Flessner, B. (2021). Akzelerator, Nanobots und Medotank. Medizinische Visionen der Science Fiction. In A. Frewer, K. Franzò, & E. Langmann (Eds.), *Die Zukunft von Medizin und Gesundheitswesen. Prognosen – Visionen – Utopien* (series Yearbook Ethics in the Clinic, pp. 21–38). Königshausen & Neumann.
- Flessner, B. (2020). Implizierte Prognosen. Anmerkungen zum Verhältnis von Möglichkeits- und Wahrscheinlichkeitsraum in Science Fiction und Wissenschaft. In M. Jungert, A. Frewer, & E. Mayr (Eds.), *Wissenschaftsreflexion. Interdisziplinäre Perspektiven zwischen Philosophie und Praxis* (pp. 231–250). Brill | Mentis. https://doi.org/10.30965/9783957437372_010

Ferdinand Gehringer is a security Policy Advisor at the Konrad-Adenauer-Stiftung e.V., a lawyer, and a certified mediator. He advises politicians from the German Bundestag and the European Parliament, as well as international organizations and governments, primarily on cybersecurity, hybrid threats, the protection of critical infrastructures, and new and disruptive technologies. His expertise lies at the intersections of geopolitics, technology, digitalization, and regulation.

- Gehringer, F. (2024). Einblicke in globale Desinformations- und Propagandastrategien. *Neue Justiz*, *78*(Beilage 1), B18–B23. https://www.nomos.de/zeitschriften/nj/#nj-beilage-01-2024
- Gehringer, F., & Kramer, J. (2024). Cyber Actors: Iran – Wie Angriffe auf den Staat stark machen. *Analysen und Argumente*, (531). Konrad-Adenauer-Stiftung. https://www.kas.de/de/analysen-und-argumente
- Gehringer, F. (2024). EllaLink – how a submarine cable does more than just connect. In M. Hedrich (Ed.), *As relações Brasil-Europa diante do mundo em transformação. Brazil-Europe relations facing the changing world* (Série Relações Brasil-Europa no. 13, pp. 101–110). Konrad-Adenauer-Stiftung Brasil. https://www.kas.de/pt/web/brasilien/série-brasil-europa
- Gehringer, F., & Strobl, M. (2024). *Influencias híbridas en México y Alemania – La combinación de tácticas híbridas como herramienta explosiva: Así deberían*

protegerse los Estados frente a las amenazas modernas (Policy Brief). Institute for Strategy and Defense Research. https://isdr.mx/policy-briefs

Florian Keisinger is a historian and worked for Airbus from 2013 to 2025. From 2018 to 2023, he was head of the campaign for the tri-national Future Combat Air System (FCAS). He is now Managing Director of the Central Association of German Seaport Operators (ZDS) in Hamburg.

Benjamin J. Knox is a research leader in the Norwegian Armed Forces Cyber Defence, and an affiliated researcher at the Norwegian Defence Research Institute (FFI). He also holds an associate professor position at the Center for Cyber and Information Security (CCIS) within the Norwegian University of Science and Technology (NTNU). His research interests lie in the fields of human factors in cyberspace operations, cognitive warfare, cognitive security, and applied cognitive performance. https://orcid.org/0000-0002-4540-9534

- Knox, B. J., Lugo, R. G., & Sütterlin, S. (2023). Cognitive agility for improved understanding and self-governance: A human-centric AI enabler. In S. K. Katsikas, E. G. Carayannis, & E. Grigoroudis (Eds.), *Handbook of research on artificial intelligence, innovation and entrepreneurship* (pp. 152–172). Edward Elgar. https://doi.org/10.4337/9781839106750.00019
- Ask, F. T., & Knox, B. J. (2023). Cognitive warfare and the human domain: Appreciating the perspective that the trajectories of neuroscience and human evolution place cognitive warfare at odds with ideas of a human domain. In Y. R. Masakowski, & J. M. Blatny (Eds.), *Mitigating and responding to cognitive warfare*. STO Technical Report (STO-TR-HFM-ET-356, pp. 13-1–13-5). NATO Science & Technology Organization. https://www.sto.nato.int/publications/STO%20Technical%20Reports/STO-TR-HFM-ET-356/$$TR-HFM-ET-356-ALL.pdf
- Canham, M., Sütterlin, S., Ask, T. F., Knox, B. J., Glenister, L., & Lugo, R. G. (2023). Ambiguous self-induced disinformation (ASID) attacks: Weaponizing a cognitive deficiency. *Journal of Information Warfare*, *22*(3), 43–58.
- Knox, B. J., Jøsok, Ø., Helkala, K. M., Khooshabeh, P., Ødegaard, T., Lugo, R. G., & Sütterlin, S. (2018). Socio-technical communication: The Hybrid Space and the OLB-Model for science-based cyber education. *Military Psychology*, *30*(4), 350–359. https://doi.org/10.1080/08995605.2018.1478546

Wolfgang Koch, Fellow IEEE, is committed to the German Bundeswehr and the German defense and security industry. His work comprises all aspects of Intelligence, Surveillance, and Reconnaissance (ISR), Electronic/Navigation Warfare, resources management, and Manned-Unmanned Teaming (MuM-T). Being the Chief Scientist and Head of Sensor Data and Information Fusion of the Fraunhofer Institute FKIE, he coordinates on a broader scale R&D activities related to digitalization in the defense and security domain. Wolfgang Koch earned his PhD in theoretical physics

at the RWTH Aachen and his habilitation degree in applied computer science at the University of Bonn, where he teaches as a professor. On the international level, he is active in the International Information Fusion Society ISIF, the IEEE Aerospace and Electronics Systems Society AESS, and the NATO Science and Technology Organization STO. https://orcid.org/0000-0002-5734-3325

- Koch, W. (2014). *Tracking and sensor data fusion. Methodological framework and selected applications* (Springer Mathematical Engineering Series). Springer. https://doi.org/10.1007/978-3-642-39271-9
- Koch, W. (2023). Zur Causa finalis künstlich intelligenter Waffen [On the final cause of artificially intelligent weapons]. In G. M. Hoff, & M. Barth (Eds.), *Digitale Welt – Künstliche Intelligenz – Ethische Herausforderungen* [Digital World–Artificial Intelligence–Ethical Challenges] (pp. 225–254). Karl Alber
- Koch, W., & Keisinger, F. (2024). How can responsible AI be implemented? In J. M. Schraagen (Ed.), *Responsible use of AI in military systems* (pp. 37–58). Chapman and Hall/CRC. https://doi.org/10.1201/9781003410379
- Koch, W., Spreen, D., Talves, K., Wagner, W., Lillemäe, E., Klaus, M., Viidalepp, A., Cooper, C., & Pekarev, J. (2024). On the ethics of employing artificial intelligent automation in military operational contexts. *IEEE Transactions on Technology and Society*, *5*(2), 231−241. https://doi.org/10.1109/TTS.2024.3405309

Igor Kopõtin, PhD, Leading Researcher at the Estonian Military Academy. https://orcid.org/0000-0002-1975-7471

- Kopõtin, I. (2024). Eine auf Deutschland orientierte bewaffnete Neutralität. Die militärpolitische Zusammenarbeit Estlands mit dem Deutschen Reich in den 1930er Jahren. *Militärgeschichtliche Zeitschrift*, *83*(2), 355–387. https://doi.org/10.1515/mgzs-2024-0059
- Kopõtin, I., & Sazonov, V. (2023). The Russian military's use of history to create a post-Soviet identity: The development of conceptual understandings from the 1990s to the mid-2000s. *The Journal of Slavic Military Studies*, *36*(4), 410–434. https://doi.org/10.1080/13518046.2023.2299079
- Kopõtin, I. (2022). Inventing military history teaching in Estonian military education 1919–1940: Approaches, tools, and methods. *Sõjateadlane (Estonian Journal of Military Studies)*, (19), 197−249. https://doi.org/10.15157/st.vi19.24142

Eleri Lillemäe is a researcher at the Department of Applied Research at the Estonian Military Academy. She conducts research related to human resources in the armed forces. She has studied conscription and its relations to societies, relations between reservists and their employers, and public perceptions towards the uses of AI in the military. https://orcid.org/0000-0001-8959-9993

- Lillemäe, E., Talves, K., & Wagner, W. (2023). Public perception of military AI in the context of techno-optimistic society. *AI & Society*, 1–15. https://doi.org/10.1007/s00146-023-01785-z
- Lillemäe, E., Kasearu, K., & Ben-Ari, E. (2024). Making military conscription count? Converting competencies between the civilian and military spheres in a neoliberal Estonia. *Current Sociology*, *72*(5), 909–927. https://doi.org/10.1177/00113921231159433
- Kasearu, K., Lillemäe, E., & Ben-Ari, E. (2023). The military covenant, contractual relations, and social cohesion in democracies: Estonia as an exploratory case study. *Armed Forces & Society*, *49*(3), 729–751. https://doi.org/10.1177/0095327X221100769

Arto Oll, PhD, Research Fellow at the Estonian Maritime Museum.

- Oll, A. (2022). Estonian and Latvian naval collaboration during the Interwar Period of 1920-1940. *Latvijas Vēstures Institūta Žurnāls*, (116), 79−98. https://doi.org/10.22364/lviz.116.05

Marie-Pierre Pacaux-Lemoine is Researcher at the Université Polytechnique Hauts-de-France (France) in the Automatic Control department of LAMIH. She leads research projects on Human-Machine Cooperation, adaptability of shared functions between Humans and Machines, as well as on design and evaluation methodologies. Author of several articles in international journals, books, and conferences, she addresses various domains of application, in transportation (automotive, train, air traffic control), manufacturing, handicap, and military domains (robotics, fighter aircrafts). https://orcid.org/0000-0002-7486-5723

- Pacaux-Lemoine, M.-P., Flemisch, F., & Chaabane, S. (2024). Calibration and resilience of human-AI systems cooperation in industry. In T. Borangiu, D. Trentesaux, P. Leitão, L. Berrah, & J. F. Jimenez (Eds.), *Service oriented, holonic and multi-agent manufacturing systems for industry of the future. SOHOMA 2023. Studies in Computational Intelligence, vol. 1136* (pp 284–294). Springer. https://doi.org/10.1007/978-3-031-53445-4_24
- Pacaux-Lemoine, M.-P., Habib, L., & Carlson, T. (2023). Levels of cooperation in human-machine systems: A human-BCI-robot example. In G. Fortino, D. Kaber, A. Nürnberger, & D. Mendoca (Eds.), *Handbook of human-machine systems*. Wiley, IEEE Press.
- Pacaux-Lemoine, M.-P., Berdal, Q., Guérin, C., Rauffet, P., Chauvin, C., & Trentesaux, D. (2021). Designing human-systems cooperation in Industry 4.0 with cognitive work analysis: A first evaluation. *Cognition, Technology & Work*, *24*, 93–111. https://doi.org/10.1007/s10111-021-00667-y
- Pacaux-Lemoine, M.-P., & Flemisch, F. (2019). Layers of shared and cooperative control, assistance, and automation. Cognition, *Technology & Work*, *21*, 579–591. https://doi.org/10.1007/s10111-018-0537-4

Iuliia Paramonova holds a Master's in Computer Science and Engineering, specializing in Human–Computer Interaction. She is a Junior Research Fellow and PhD candidate at the School of Digital Technologies, Tallinn University. Iuliia's expertise focuses on applying innovative methodologies to design trustworthy, human-centered systems. As a UX Research and Design consultant, she bridges academia and industry, applying insights to real-world applications and enriching academic pursuits with valuable industry experience. As a lecturer, Iuliia shares her knowledge of research and design methods, contributing to the next generation of UX professionals. https://orcid.org/0000-0001-7762-0548

- Paramonova, I., Sousa, S., & Lamas, D. (2023). Heuristics to design trustworthy technologies: Study design and current progress. In J. Abdelnour Nocera, M. Kristín Lárusdóttir, H. Petrie, A. Piccinno, & M. Winckler (Eds.), *Human-computer interaction – INTERACT 2023. Lecture Notes in Computer Science* (vol. 14145, pp. 491–495). Springer. https://doi.org/10.1007/978-3-031-42293-5_60
- Paramonova, I., Sousa, S., & Lamas, D. (2023). Exploring factors affecting user perception of trustworthiness in advanced technology: Preliminary results. In P. Zaphiris, & A. Ioannou (Eds.), *Learning and Collaboration Technologies. HCII 2023. Lecture Notes in Computer Science* (vol. 14040, pp. 366–383). Springer. https://doi.org/10.1007/978-3-031-34411-4_25
- Paramonova, I., Lamas, D., & Sousa, S. (2024). Socio-Technical Trustworthiness (SoTechTrust): A framework to ensure the trustworthiness of socio-technical systems. In J. Ribeiro, C. Brandão, M. Ntsobi, J. Kasperiuniene, & A. P. Costa (Eds.), *Computer supported qualitative research. WCQR 2024. Lecture Notes in Networks and Systems* (vol. 1061, pp. 375–401). Springer. https://doi.org/10.1007/978-3-031-65735-1_21

Kaarel Piirimäe, PhD, Associate Professor of Contemporary History at the University of Tartu. https://orcid.org/0000-0001-6523-1967

- Piirimae, K. (2023). Strategic uses of nationalism and ethnic conflict: Interest and identity in Russia and the post-Soviet space. *Journal of Baltic Studies*, *54*(4), 863−865. https://doi.org/10.1080/01629778.2023.2265220
- Kopõtin, I., & Piirimäe K. (2023). The modern history of the Estonian War of independence or how military history is written in Estonia. *Forschungen zur baltischen Geschichte*, *17*, 188−199. https://doi.org/10.30965/9783657790364_010
- Piirimäe, K. (2023). Contributions of the Baltic Independence Campaigns to Soviet Collapse. In C. Clarke (Ed.), *Understanding the Baltic States: Estonia, Latvia and Lithuania since 1991* (pp. 91−107). Hurst.
- Piirimäe, K. (2014). *Roosevelt, Churchill, and the Baltic Question. Allied relations during the Second World War*. Palgrave Macmillan. https://doi.org/10.1057/9781137442345

Peter Andreas Popp, Dr. phil., retired lieutenant colonel: From June 2020 to October 2023, deputy head of the "Conceptual Development of Innere Führung" department at the Center for Innere Führung (ZInFü) in Koblenz. Freelance lecturer for the Bundeswehr since November 2023.

Alexander Schuster was a policy advisor at the think tank of the Konrad-Adenauer-Stiftung e.V. at the time of writing, where he was responsible for the European Security and Defense portfolio. In this role, he dealt with strategic issues of European defense policy, European arms cooperation, and nuclear deterrence. Since June 2024, he has been working for a leading company in the German defense technology industry.

- Schuster, A. (2023). Einsatzfähige und nachhaltig ausgerüstete Bundeswehr. Wie Deutschland zum Rückgrat der Abschreckung und kollektiven Verteidigung Europas wird. *Monitor Sicherheit*. Konrad-Adenauer-Stiftung. https://www.kas.de/documents/252038/22161843/Einsatzfa%CC%88hige+und+nachhaltig+ausgeru%CC%88stete+Bundesweh.pdf
- Schuster, A. & Bellmann, C. (2024). Tun wir genug? Der deutsche und europäische Beitrag zur NATO. *Auslandsinformationen*. Konrad-Adenauer-Stiftung. https://www.kas.de/documents/259121/30240531/DE_kas_ai_01-2024_bellmann_schuster_web.pdf
- Schuster, A. (2024). Waffenbrüder? Wege zum europäischen Schulterschluss in der Rüstung. *Die Politische Meinung*, (584), 88–92.

Sonia Sousa is an Associate Professor of Interaction Design at the School of Digital Technologies, Tallinn University, Estonia. She holds two PhDs (2006 & 2023) and two postdoc projects (2010 & 2013) in Software Design, Distance Education, and Human–Computer Interaction. Her research explores the interplay between trust, performance, and technology adoption. She leads the Trustworthy HCI lab and has been an active researcher and teacher since 2003. Her portfolio includes participation and coordination of more than 20 funded projects (H2020, CHIST-ERA, and AFOSR). Sonia was nominated by the Estonian Research Foundation (ETAG) as Honorary female academics (AcademiaNet) in 2022. She is also a member of expert panels in Finland (AKA), the European Commission, the USA (AFOSR), and Swiss (SNSF). https://orcid.org/0000-0002-5865-1389

- Gulati, S., McDonagh, J., Sousa, S., & Lamas, D. (2024). Trust models and theories in human–computer interaction: A systematic literature review. *Computers in Human Behavior Reports*, *16*, Article 100495. https://doi.org/10.1016/j.chbr.2024.100495
- Sousa, S., Lamas, D., Cravino, J., & Martins, P. (2024). Human-centered trustworthy framework: A human–computer interaction perspective. *Computer*, *57*(3), 46–58. https://doi.org/10.1109/MC.2023.3287563
- Pilacinski, A., Pinto, A., Oliveira, S., Araújo, E., Carvalho, C., Silva, P. A., Matias, R., Menezes, P., & Sousa, S. (2023). The robot eyes don't have it. The

presence of eyes on collaborative robots yields marginally higher user trust but lower performance. *Heliyon*, *9*(8). Article e18164. https://doi.org/10.1016/j.heliyon.2023.e18164
- Gulati, S., Sousa, S., & Lamas, D. (2019). Design, development and evaluation of a human-computer trust scale. *Behaviour & Information Technology*, *38*(10), 1004–1015. https://doi.org/10.1080/0144929X.2019.1656779

Dierk Spreen is a Visiting Professor at the Berlin School of Economics and Law (HWR) and Senior Project Manager at the Institute for Organizational Communication at the University of the Bundeswehr in Munich (UniBwM). From 2022 to 2024, he was Visiting Researcher at the Estonian Military Academy (EMA). He holds a PhD in sociology and habilitated with a thesis on war and society. He teaches social science, political economy, and communication studies and does research on new technologies, military, security, and Strategic Communication. https://orcid.org/0000-0002-9842-8632

- Spreen, D. (2023). Lethal Autonomous Weapon Systems (LAWS). On the ethics of automation in the military from the perspective of social systems theory. *Sõjateadlane (Estonian Journal of Military Studies)*, (21), 10–40. https://doi.org/10.15157/st.vi21.24177
- Spreen, D., & Flessner, B. (Eds.). (2022). *Die Raumfahrt der Gesellschaft. Wirtschaft und Kultur im New Space Age*. transcript. https://doi.org/10.14361/9783839457627
- Spreen, D. (2015). *Upgradekultur. Der Körper in der Enhancement-Gesellschaft*. transcript. https://doi.org/10.1515/9783839430088
- Spreen, D. (2008). *Krieg und Gesellschaft. Zur Konstitutionsfunktion des Krieges für moderne Gesellschaften* (Sociological writings, vol. 81). Duncker & Humblot. https://doi.org/10.3790/978-3-428-52561-4

Kairi Talves received her doctorate in sociology at the University of Tartu in 2018. She is the Scientific Adviser of the Estonian Ministry of Defence and a visiting scholar at the Estonian Military Academy, where she currently leads the research group of cognitive warfare. Her research interests cover the different aspects of technology in society: societal development and change (e.g., technological developments and their social impact), public acceptance of technology, technology in the military: risks, trust, acceptance, ethical aspects, challenges of human–machine interaction, cognitive warfare. https://orcid.org/0009-0006-6815-7137

- Koch, W., Spreen, D., Talves, K., Wagner, W., Lillemäe, E., Klaus, M., Viidalepp, A., Cooper, C., & Pekarev, J. (2024). On the ethics of employing artificial intelligent automation in military operational contexts. *IEEE Transactions on Technology and Society*, *5*(2), 231−241. https://doi.org/10.1109/TTS.2024.3405309
- Wagner, W., Viidalepp, A., Idoiaga-Mondragon, N., Talves, K., Lillemäe, E., Pekarev, J., & Otsus, M. (2023). Lay representations of artificial intelligence and

autonomous military machines. *Public Understanding of Science*, *32*(7), 926–943. https://doi.org/10.1177/09636625231167071
- Lillemäe, E., Talves, K., & Wagner, W. (2023). Public perception of military AI in the context of techno-optimistic society. *AI & Society*, 1–15. https://doi.org/10.1007/s00146-023-01785-z

Priit Värno has a Master's degree in Educational Innovation from the University of Tartu and a Master's degree in Military Leadership from the Estonian Military Academy. Currently, he is a junior researcher at the Military Academy, participating in the Technosociology Research Group and the Cognitive Warfare Research Group. His research interests are related to aspects of technology acceptance in military organizations: risks, trust, ethical aspects, challenges of human–machine interaction, cognitive warfare, and innovation in military organizations.

- Värno, P. (2024). *Vastuvõtlikkus autonoomsetele tehnoloogiatele väikeriigi kaitseväe näitel* (Acceptance of autonomous technologies by example of small country's defense force), Master's Thesis, Estonian Military Academy

Jörg Vollmer, General (ret.), has been deployed at all levels in leadership roles since joining the Army in 1978: from platoon to two company commander positions, battalion, brigade, and division commander to commander of the field army. Deployments in the former Yugoslavia in 1996/97 and twice for one year in Afghanistan as commander of the Regional Command North in Mazar-e Sharif in 2009 and 2013/14 round out his professional profile. From 2015-2020 he was Chief of the Army. The focus of his work was on the digitization of the army and the restoration of the operational readiness of the land forces. In his last assignment, as Commander-in-Chief of NATO's operational headquarters in Central Europe, the Allied Joint Force Command Brunssum, he was primarily responsible for restoring defense capabilities on NATO's northeast flank. Since his retirement in July 2022, he serves as Chief Advisor Military Affairs at the Fraunhofer Institute for Communication, Information Processing and Ergonomics (FKIE).

Natascha Zowislo-Grünewald is a professor at the Institute for Organizational Communication at the University of the Bundeswehr in Munich. Her research focuses on the effects of communication in a security policy context, the derivation of Strategic Communication for security policy organizations and the analysis of cognitive warfare.

- Zowislo-Grünewald, N. (2023). Verteidigungs- und Sicherheitspolitik im Spannungsfeld zwischen Elitendiskurs und öffentlicher Meinung. In N. Lammert, & W. Koch (Eds.), *Bundeswehr der Zukunft – Verantwortung und Künstliche Intelligenz*. Konrad-Adenauer-Stiftung. https://www.kas.de/de/einzeltitel/-/content/bundeswehr-der-zukunft-5
- Zowislo-Grünewald, N., & Wörmer, N. (Eds.). (2021). *Kommunikation, Resilienz und Sicherheit*. Konrad-Adenauer-Stiftung. https://www.kas.de/de/veranstaltungsberichte/detail/-/content/kommunikation-resilienz-und-sicherheit-1

- Zowislo-Grünewald, N., & Beitzinger, F. (2021). *Lehrbuch Strategisches Kommunikationsmanagement* (2nd ed.). LIT Verlag.
- Zowislo-Grünewald, N., & Hajduk, J. (2020). "Fake News": neue Bedrohung oder alter Hut? Grundlagen für ein Strategisches Diskursmanagement. In R. Hohlfeld, M. Harnischmacher, E. Heinke, L. Sophia Lehner, & M. Sengl (Eds.), *Fake News und Desinformation Herausforderungen für die vernetzte Gesellschaft und die empirische Forschung* (S. 297–310). Nomos. https://doi.org/10.5771/9783748901334-295

Introduction: Artificial Intelligence Is Going to War

Kairi Talves and Dierk Spreen

The robots are here. Artificial Intelligence (AI) or artificially intelligent automation is flanked by discussions in society that are not merely technical in nature. Rather, AI it is highly embedded in social, ethical, legal, and political discourses. Questions arise about the chances and risks. Artificial intelligence raises concerns and fears in public and has a more and more apparent role in conflicts—and thus also in military contexts.

Like any new technology whose development paths and implications are not yet clear, the unknown regarding AI gives rise to perceptions and fears that manifest themselves in extensive debates. It is clear that for making informed decisions we need broader knowledge and discussions. American computer scientist and futurist Jerry Kaplan warns against simplistic, one-sided, and uninformed perceptions of artificial intelligence. He states:

> Benchmarking machine intelligence against human intelligence is the fool's errand. There is the temptation to think about increasingly capable computer programs as embryonic sentient beings, potentially presenting some sort of existential challenge to humans. […] This anthropomorphic framing reinforces the common trope that intelligent machines may suddenly 'wake up' and become conscious, potentially spawning their own intentions,

K. Talves (✉)
Estonian Military Academy, Tartu, Estonia
e-mail: kairi.talves@mil.ee

D. Spreen
Institute for Organizational Communication, University of the Bundeswehr Munich (UniBwM), Neubiberg, Germany

Department of Business and Economics, Berlin School of Economics and Law Berlin (HWR), Berlin, Germany
e-mail: dierk.spreen@unibw.de; https://strategic-communication-management.de

K. Talves, D. Spreen (eds.), *Artificial Intelligence in Military Technology*,
Artificial Intelligence, Simulation and Society 192,
https://doi.org/10.1007/978-3-031-95578-5_1

> goals, judgements, and desires. OMG,[1] when are 'they' coming to take over my job, my home, my life? And what are we going to do if they decide that they do not need us any more? Well, news flash: They are not coming for us because there is no 'they.' Despite appearances, there is no one home. GAIs[2] do not 'think' in the human sense, and they do not have 'minds.' (Kaplan, 2024, p. 6)

It seems that the paradox of the unknown is apparently written into the cultural code of humankind, where, on the one hand, the new and the unknown generate a great excitement and a desire for progress (to see what is around the next corner), while, on the other hand, the new and the unknown is also a source of danger and causes the greatest of existential fears. With this in mind it is important to be aware of the cultural roots that influence attitudes and acceptance of artificially intelligent automation. It is not possible to address these issues in a single discipline. The complexity of the technological and human systems and their interrelationships requires multidisciplinary exchange, taking into account the impacts at different levels, from the individual to the large social systems. Only in this way can knowledge lead to wise and informed decisions.

Artificially intelligent automation in the military is in the spotlight. The AI-based technology is changing the character of war, starting with technology providing situational awareness and aiding in decision-making 'on machine's speed' and ending with the possible lethal weapon systems on the battlefield. The introduction of artificial intelligence and autonomy into warfare will have profound and unforeseen consequences for national security and human society. There seems to be so much at stake—the dangers and risks seem to be infinitely high. Particularly with regard to military applications, many of the problems that artificially intelligent automation raises show up as if in a burning glass. This becomes apparently visible in the debate about Lethal Autonomous Weapon Systems (LAWS)—from the ethical and legal perspectives related to the deployment of such systems.

On the one hand, it is clear that automation happens in every sphere of life and is particularly important in the military. It is naïve to expect that in a domain that is loaded with expectations of achieving superiority and effectiveness over the adversary it does not happen. On the other hand, it puts the pressure on even more careful consideration of possible risks and unwanted outcomes. It should be clear that inappropriate use of technology or use of technically uncontrollable technology in military contexts are immoral and illegal per se. The ethical principles and international laws apply to the emerging technologies in the same way as to conventional technology. The ethical framework should include the principle of responsibly applied technology and a 'digital ethic' that applies to military AI and guides developers and manufacturers, procurement, and military end-users. Only by that it is possible to overcome fears of the 'gray areas' of the use of the technology and of the 'non-precedential scenarios' with military AI that are hard to predict or give them legal force before these have happened or even the technology has developed so far.

[1] Oh My God.

[2] General Artificial Intelligence.

The background to the volume presented here is also the debate about the pros and cons or even a ban on lethal autonomous weapon systems (Reichberg & Syse, 2021). This raises the problem of what is actually meant by 'autonomy' (Koch, 2019, p. 27). In a resolution passed by the European Parliament in September 2018, this was understood to mean weapons "lacking human control in critical functions such as target selection and engagement" (RC-B8-0308/2018, 2018, L.4.) "[M]eaningful human control over the critical functions of weapon systems, including during deployment," should be ensured (RC-B8-0308/2018, 2018, L.2.). However, the individual positions of the EU member states are very different. This is even more true at a global level (Dahlmann & Dickow, 2019, pp. 17–23).

In the ethical debate, a distinction is generally made between deontological and consequentialist argumentation. According to deontological argumentation, "the moral value of acts, principles, laws, or character traits is determined not only by their effects but also by their internal value, which may be influenced, for example, by intentions, motivations, or the type of the assessed act or principle" (Bodziany, 2021, p. 56). In contrast, a consequentialist argument states "that the non-instrumental, internal moral value of acts, rules, laws, character traits, or institutions have only their effects" (Bodziany, 2021, p. 56). The best-known deontological argument against the use of LAWS is the one that sees the use of such systems as a violation of human dignity (Koch, 2019, pp. 30–31). According to this argument, technical automatisms are in principle not capable of making decisions of conscience. The target of such a weapon system is "merely the object of a mathematically calculated decision to kill" (Geiß, 2019, p. 54).[3] "For the autonomous system, it is completely irrelevant whether it is dealing with a sack of potatoes, an animal or a human being" (Koch, 2019, p. 31). A consequentialist argument against LAWS applies, for example, when considerable risks associated with the use of such technologies are pointed out (Koch, 2019, pp. 32–33). One such risk discussed is a destabilizing arms race (Altmann, 2019).

The ban argument is also countered. For example, Dieter Birnbacher (2016, p. 120) argues:

> Of course, machines cannot comprehend the value of human life. But why should this make a difference to their victims if, alternatively, they are threatened with being wounded or killed by manned weapons such as bombers? For the victims whose dignity is at stake, it is a matter of indifference whether the threat they are exposed to comes from manned or unmanned weapons, provided all other parameters of the situation are equal.

From a sociological perspective, however, it is important to contextualize the use of technical automation—however 'intelligent' it is addressed in discourses. Of particular importance here is the context of the organization in which technical automation is used (Luhmann, 1966). Of course, this also applies to the military. In armed forces all 'autonomous' systems should be under command, just like any soldier. He

[3] We have tacitly translated non-English quotations into English.

or she acts 'autonomously' in a chain of command. The same applies to 'autonomous' artificial systems, with the difference that they do not understand what they are doing. They operate in the medium of causality and not in the medium of meaning, as one might say in the sociology of technology (Esposito, 2022; Halfmann, 1996). But nobody needs machines that operate in a permanently deviant manner—no company, no authority, no army, no private user. As is the case for human social actors or organizations, machines should also operate within a normatively defined framework of expectations.

In this sense, these systems are not actually 'autonomous' in a strict sense of the word, but *semi-autonomous*. From such a sociological perspective, it follows that humans (in social roles) are and should be held responsible for the actions of AI weapon systems (Koch et al., 2024; Spreen, 2023). This approach has ethical consequences, too. Under what conditions can soldiers in good conscience assume responsibility for the use of artificially intelligent weapons systems that can have a lethal effect? The idea behind this is that such accountability ensures that the use of artificially intelligent weapons systems is bound to the ethical framework of society, thereby limiting the potential for violence offered by this weapons technology. It's not about a ban, but about containment in the sense of responsible handling.

Against such a sociological background, the use of AI technology in the military is not simply unproblematic; in fact, the opposite is the case. However, this does not justify a general ban—it is much more important to look at the contexts of use, and these are very manifold. They include, for example, law and ethics, military-organizational and social communication, media skills, cultural concepts of robotics, human–machine interaction (HMI), development contexts and, of course, the complexity of decision-making processes on the battlefield. This suggests an interdisciplinary approach, which we would like to encourage with this book.

There is no doubt that artificially intelligent automation, like any new technology, carries worries and risks. One can be deterred by this, but then one should also stop talking about innovation and progress. Ultimately, as Luhmann (1993, p. 89) has pointed out, new technologies can only be made safe by trying them out, implementing them socially, and reflecting on them again and again. One must also see that new technologies allow society to increase its complexity. For example, if there is a bridge, new possibilities for traffic, trade, and communication arise.

This volume brings together multidisciplinary, differentiated, informed, and open approaches to military AI applications from the fields of science, politics, and the military itself. Based on the questions raised with the development of artificially intelligent automation, this volume sets itself a task to critically examine the challenges in the field of current and future military technology. The first part addresses the cultural and societal discourses related to technological developments, including artificial intelligence. It examines the cultural origins of popular, scientific, and political discourses that influence the development of, and more importantly, the attitudes toward and acceptance of, artificially intelligent automation. *Dierk Spreen* and *Kairi Talves* analyze how fear of robots as the discursive motif of technological automation has been reflected in several cultural narratives throughout modern history. The idea about artificially intelligent automation has created great hopes but

also carried a warning sign since its very early developments, which has loaded the public discussions with simultaneous fascination and fear. In the next chapter, *Dierk Spreen* continues on a similar topic by discussing the ethical and societal aspects of human enhancement. He analyzes how the concept of the *cyborg*—a symbiotic relationship between the human body and modern technology—aligns with the development of technology and posthumanist discourse on the human body. Human enhancement has aroused the interest of the military and requires ethical reflection in this societal area. Cultural artifacts could strongly influence the communication of the unknown—the future of humans and technology. It is reviewed in the chapter of *Bernd Flessner*, who discusses the role of science fiction in this communicative arena. The futures sketched out by science fiction authors provide a glimpse of what is to come and are therefore treated as an important source of the cultural outcast about the diversity of transformations. *Ferdinand Gehringer* and *Alexander Schuster* concentrate on the political arena by analyzing the international debates about the potential regulatory limits of the military use of AI. The regulation of AI in military applications is a crucial aspect of both current and future security order, however requiring a dynamic initiative and a dynamic regulatory framework among states. *Kairi Talves*, *Priit Värno*, and *Eleri Lillemäe* introduce the empirical study about the attitudes and acceptance of autonomous technologies in a small state armed forces. They examine how, in Estonia, in a specific context with the influence of global technological and security challenges, attitudes of the military affect the streamline of the technological automation in the armed forces.

The second part of the book turns attention to the challenges and changes posed by the use of AI in military conflicts. It sets up the question of what the critical aspects of change are in the military organization, but also looks at ethical and legal aspects. It examines the role of the military organizations, their basic concepts, and their doctrines, as well as the human factor aspect in the development and implementation of the autonomous technology. *Igor Kopõtin*, *Kaarel Piirimäe*, and *Arto Oll* analyze the role of technology in military conflicts from a historical perspective. They argue that any military technological invention can only be effective if it is successfully integrated into the military organization and its use is linked with the objectives of military operation. The war is still a war between people, and the human factor will always be decisive because the weapon, whatever it may be, will always remain a tool in human hands. In a similar vein, the chapter from *Wolfgang Koch*, *Jörg Vollmer*, and *Florian Keisinger* argue that developing and deploying AI-based defense technologies is more than just the aspect of a technical innovation. It influences the entire way armed forces think and act. The responsible use of AI in the military is the necessity, which requires "digital ethics" from the whole spectrum of the development and deployment of AI—that includes the research, development, and defense planning communities. *Peter Andreas Popp* analyzes the example of the German Bundeswehr leadership concept *Innere Führung* to ask what kind of challenges are posed between the principles forming a "value compass" for the military ethos and military technical developments like artificially intelligent automation. *Stuart Casey-Maslen* examines the legal challenges related to the development of autonomous weapon systems. The compliance of

autonomous weapon systems with the rules of international laws in a situation of armed conflict is the subject of a fast-growing debate.

The third part examines the communication challenges posed by the implementation of AI in the military and its associated opportunities and risks. Communication can build bridges only if the context and impact are properly understood. It is especially the case in technology, where communicative elements depend on and cling between the risk, trust, responsibility, and accountability of the technical dimensions (Koch et al., 2024). *Sonia Sousa*, *Gabriela Beltrão*, *Iuliia Paramonova*, and *Debora C. Firmino de Souza* examine the concept of trustworthy system design as the possibility to integrate and mitigate possible communication challenges related to AI systems and enhance users' trust in the human–machine interrelation. *Natascha Zowislo-Grünewald* and *Franz Beitzinger* tackle another aspect of communicative challenges—about the military technological innovation. Addressing this communication challenge goes beyond mere dissemination of information; it necessitates an understanding of the social and semiotic context influenced by the innovation in military organization. In the next chapter, *Natascha Zowislo-Grünewald* discusses the importance of terminology and unified definitions as the imperative of communication and for achieving interoperability in terms of the new technologies like artificial intelligence. In the final chapter, *Frank Ole Flemisch*, *Dierk Spreen*, *Marie-Pierre Pacaux-Lemoine*, *Benjamin J. Knox*, *Kairi Talves*, and *John Christopher Brill* introduce the "holistic bowtie model," a multi-layered conceptual approach explaining the holistic thinking about AI and automation in defense from a system theory perspective.

The essays collected here are intended to help understand artificial intelligence in social contexts. Military organizations are of particular interest. Naturally, questions of "responsible information processing" arise with particular seriousness in "combat troops" (Luhmann, 1995, p. 176). On the other hand, the military is also one of the functional systems of society (Dammann, 2022, p. 121). This means that the example of military AI systems provides a particularly good focus on issues that can arise from the application of artificially intelligent automation. We are interested in ensuring that AI is used responsibly in the military as well as in other functional contexts of society, i.e., that it remains subject to the accountability and control of human actors in their organizational roles.

This volume was made possible by the Estonian Military Academy (EMA), which funded the project. We gratefully acknowledge Colonel PhD *Raul Järviste*, the former Deputy Commandant of the Estonian Military Academy and current Estonian military attaché in Berlin, for his support in publishing this book. Further support was provided by the Bundeswehr Centre of Military History and Social Sciences (ZMSBw) and Prof. Dr. *Natascha Zowislo-Grünewald* and her team at the Institute for Organizational Communication, University of the Bundeswehr (UniBwM), Munich. We would like to use this opportunity to thank all those who have supported this volume or contributed to it with their articles. Unless otherwise indicated, gender-specific terms used in this book refer to all genders. The authors themselves decide how they wish to handle this in detail.

References

Altmann, J. (2019). Autonome Waffensysteme—der nächste Schritt im qualitativen Rüstungswettlauf? In I.-J. Werkner, & M. Hofheinz (Eds.), *Unbemannte Waffen und ihre ethische Legitimierung* (pp. 111–136). Springer VS. https://doi.org/10.1007/978-3-658-26947-0_6

Birnbacher, D. (2016). Are Autonomous Weapons Systems a Threat to Human Dignity? In N. Bhuta, S. Beck, R. Geiβ, H. Liu, & C. Kreβ (Eds.), *Autonomous Weapons Systems: Law, Ethics, Policy* (pp. 105–121). Cambridge University Press. https://doi.org/10.1017/CBO9781316597873.005

Bodziany, M. (2021). Ethical Conditions of the Use of Artificial Intelligence in The Modern Battlefield—Towards the "Modern Culture of Killing.". In A. Visvizi, & M. Bodziany (Eds.), *Artificial Intelligence and its Contexts, Advanced Sciences and Technologies for Security Applications* (pp. 45–61). Springer. https://doi.org/10.1007/978-3-030-88972-2_4

Dahlmann, A., & Dickow, M. (2019). *Präventive Regulierung autonomer Waffensysteme. Handlungsbedarf für Deutschland auf verschiedenen Ebenen (SWP study 1)*. Stiftung Wissenschaft und Politik. https://doi.org/10.18449/2019S01

Dammann, K. (2022). Vernichtungskrieg, Kampf und Volksgemeinschaft: Empirische Studien mit relationaler *Systemtheorie zur urfaschistischen Begriffswelt. Soziale Systeme*, 27(1–2), 110–143. https://doi.org/10.1515/sosys-2022-0006

Esposito, E. (2022). *Artificial Communication: How Algorithms Produce Social Intelligence*. MIT Press. https://doi.org/10.7551/mitpress/14189.001.0001

Geiß, R. (2019). Autonome Waffensysteme—ethische und völkerrechtliche Problemstellungen. In I.-J. Werkner, & M. Hofheinz (Eds.), *Unbemannte Waffen und ihre ethische Legitimierung* (pp. 41–61). Springer VS. https://doi.org/10.1007/978-3-658-26947-0_3

Halfmann, J. (1996). *Die gesellschaftliche "Natur" der Technik. Eine Einführung in die soziologische Theorie der Technik*. Leske + Budrich.

Kaplan, J. (2024). *Generative Artificial Intelligence. What Everyone Needs to Know?* Oxford University Press.

Koch, B. (2019). Die ethische Debatte um den Einsatz von ferngesteuerten und autonomen Waffensystemen. In I.-J. Werkner, & M. Hofheinz (Eds.), *Unbemannte Waffen und ihre ethische Legitimierung* (pp. 13–40). Springer VS. https://doi.org/10.1007/978-3-658-26947-0_2

Koch, W., Spreen, D., Talves, K., Wagner, W., Lillemäe, E., Klaus, M., Viidalepp, A., Cooper, C. G., & Pekarev, J. (2024). On the ethics of employing artificial intelligent automation inmilitary operational contexts. *IEEE Transactions on Technology and Society, 5*(2), 231–241. https://doi.org/10.1109/TTS.2024.3405309

Luhmann, N. (1966). *Recht und Automation in der öffentlichen Verwaltung. Eine verwaltungswissenschaftliche Untersuchung*. Duncker & Humblot.

Luhmann, N. (1993). *Risk. A Sociological Theory* (R. Barrett, Trans.). de Gruyter.

Luhmann, N. (1995). *Funktionen und Folgen formaler Organisation* (4th ed.). Duncker & Humblot (Original work published 1964).

RC-B8-0308/2018. (2018, September 12). https://www.europarl.europa.eu/doceo/document/TA-8-2018-0341_EN.html

Reichberg, G. M., & Syse, H. (2021). Applying AI on the Battlefield: The Ethical Debates. In J. v. Braun, M. S. Archer, G. M. Reichberg, & M. Sánchez Sorondo (Eds.), *Robotics, AI, and Humanity* (pp. 147–159). Springer. https://doi.org/10.1007/978-3-030-54173-6_12

Spreen, D. (2023). Lethal Autonomous Weapon Systems (LAWS). On the Ethics of Automation in the Military from the Perspective of Social Systems Theory. *Sõjateadlane (Estonian Journal of Military Studies)*, (21), 10–40. https://doi.org/10.15157/st.vi21.24177

Understanding AI

Fear of the Robots: Cultural Perspectives on Technological Autonomy

Dierk Spreen and Kairi Talves

Abstract Since the turn of the nineteenth century, technological autonomy has been critically evaluated and warned against. This article provides an exemplary analysis of this discourse of fear. Several alarming topoi can be identified, among which military robots play a role. Our investigation begins with an interpretation of Goethe's *The Sorcerer's Apprentice*. It then examines the uncanny nature of the automaton in E. T. A. Hoffmann's *The Sandman*, robots' war of extermination against humans in Karel Čapek's *R.U.R.* and the *Terminator* films, and Isaac Asimov's introduction of robot ethics and subsequent fears. Finally, a comparative look at the perception of robots in Eastern and Western cultures is provided. We identify loss of control, replication, superiority, similarity, and destruction/extermination as alarm topoi. These topoi can be combined in various ways to communicate scenarios that trigger fear of robots or technological autonomy. However, this alarmist discourse is not inevitable. There are alternatives. One alternative is the idea of ethical self-control of robots, and another is to look beyond the horizon of Western culture.

D. Spreen (✉)
Institute for Organizational Communication, University of the Bundeswehr Munich (UniBwM), Neubiberg, Germany

Department of Business and Economics, Berlin School of Economics and Law Berlin (HWR), Berlin, Germany
e-mail: dierk.spreen@unibw.de; https://strategic-communication-management.de

K. Talves
Estonian Military Academy, Tartu, Estonia
e-mail: kairi.talves@mil.ee

K. Talves, D. Spreen (eds.), *Artificial Intelligence in Military Technology*,
Artificial Intelligence, Simulation and Society 192,
https://doi.org/10.1007/978-3-031-95578-5_2

1 Introduction

It cannot be said that fear or hostility toward technology is generally widespread among the population. Nevertheless, attitudes differ across different types of technology. According to the 2021 Special Eurobarometer report, in the European Union "almost nine out of ten respondents (86 %) say that the overall impact [of science and technology on society] is positive" (European Commission, 2021, p. 90). Artificial intelligence and nuclear technology achieved the lowest positive approval ratings (European Commission, 2021, p. 95). Widespread acceptance of technology therefore does not preclude the existence of discourses that address specific technological fears. Especially after Niklas Luhmann has worked out that fear combined with morality can be highly effective communicatively and psychologically, it is obvious to ask whether there are discourses or cultural semantics that communicate fear in relation to certain technologies.

In the context of Western modern culture, this certainly applies to robots. Concerns have emerged within the context of human-like automatons, androids, or robots, but also with regard to technological autonomy and artificial intelligence in general. This is surprising insofar as technological progress plays a prominent role in the grand narrative of modernity. But since the turn of the nineteenth century, technological autonomy has been critically assessed and warned against. In the following, we want to break down this discourse of fear with examples. In doing so, we will encounter various topoi of alert. Military robots play a role in this fear discourse, as the possible lethal capability combined with the autonomy is obviously a cause for caution.

On the other hand, there is a discourse that conceives of robots as helpful machine beings. They are ethically programmed and—because they process information independently and act autonomously—are helpful cooperation partners in social contexts. Both discourses are in play today. It is important for us to emphasize that the fear of robots is a discursive motif that by no means arises from their very idea. At other times and in other cultures, technical autonomy is perceived more benevolently and with less fear.

We want to use the term 'technological autonomy' in a broad sense to describe all concepts in which technical devices or artificially intelligent systems act autonomously. Robots are merely a kind of 'discourse classic' for technological autonomy. 'Autonomy' is therefore by no means limited to human-like robotic beings. It also includes, for example, networked superintelligences such as "Skynet" from the *Terminator* films.

We begin our investigation of discourses that evaluate technical autonomy in a meaningful way with an interpretation of Goethe's *Sorcerer's Apprentice*. The broom, animated by a spell, behaves like an automaton that follows an algorithm without a stop condition. Furthermore, we will examine the uncanny nature of the automaton in E.T.A Hoffmann's *The Sandman*, the robots' war of extermination against humans (Karel Čapek's *W.U.R.*, and *Terminator* films), the introduction of robot ethics by Isaac Asimov and subsequent fears. Finally, we take a brief

comparative passage on the perception of robots in Eastern and Western cultures. We do not aim to give a comprehensive account of the discursive construction of technical autonomy—that would go beyond the scope of this paper. Rather, we pick out a few exemplary cases that we consider paradigmatic. However, we want to raise awareness of the fact that seemingly self-evident warnings about technical autonomy arise from this very discourse. Although considerable progress has been made in technological development over the past two centuries, it is striking how similar the fears that have emerged over time are—as seen from *Sorcerer's Apprentice* to the current fear of the super-intelligent singularity.

The question of which social groups promote such horror discourses must also remain open. It is obvious to refer here to C. P. Snow, who identifies literary intellectuals "as natural luddites" and as the promoters of discourses critical of technology (Snow, 1961, pp. 23–29). However, one must not forget that the fear of technical autonomy is also rampant in science fiction literature. But science fiction certainly does not belong to the canon of those "literary intellectuals" that Snow was referring to (Snow, 1961, p. 4). Science fiction is also full of robots that want to eradicate or at least suppress humans.

2 An Early Warning of Technological Autonomy

In the ballad *The Sorcerer's Apprentice*, written by Johann Wolfgang von Goethe in the summer of 1797 and published in 1798, an automaton slips out of human control. The magically activated broom acts, as we would say today, 'autonomously.' Self-controlled and without a 'human in the loop,' it carries out the task assigned to it of fetching water. It moves unassisted to the river, scoops up water, returns to the house, and pours it out again in the bath.

Come, old broomstick, you are needed,
Take these rags and wrap them round you!
Long my orders you have heeded,
By my wishes now I've bound you.
Have two legs and stand,
And a head for you.
Run, and in your hand
Hold a bucket too.

Flow, flow onward
Stretches many,
Spare not any
Water rushing,
Ever streaming fully downward
Toward the pool in current gushing.[1]

[1] Translation by Edwin H. Zeydel (1955, pp. 102–109).

The problem: the broom carries out the task and can no longer be stopped. The action instruction for the broom is *elementary* (organization of a sequence of elementary operations), *determinate* (strict determination of a sequence of basic operations), *general* at least insofar as *The Sorcerer's Apprentice* could also have asked the broom to collect soil in a certain place, and *finite* (expressed in a finite text). Therefore, it is an algorithm. But similar to an algorithm for calculating a square root, it has no stopping condition (Krämer, 1988, pp. 159–160). Like a Bostromian superintelligence that sees its purpose in the production of paper clips and becomes a danger to humanity in the process, Goethe's autonomous broom AI scoops up water without end and floods the house. As Nick Bostrom does, Goethe creates a dystopian horror scenario of machine autonomy that proceeds without stopping conditions:

> An AI, designed to manage production in a factory, is given the final goal of maximizing the manufacture of paperclips, and proceeds by converting first the Earth and then increasingly large chunks of the observable universe into paperclips. (Bostrom, 2014, p. 123)

> Brood of hell, you're not a mortal!
> Shall the entire house go under?
> Over threshold over portal
> Streams of water rush and thunder.
> Broom accurst and mean,
> Who will have his will,
> Stick that you have been,
> Once again stand still!

Did modernity begin with the warning of artificial intelligence and technical autonomy? The fear of the self-acting 'servant' getting out of control has been articulated in mythology,[2] literature, and art. It is still paradigmatic today. The *Süddeutsche Zeitung*, for example, writes with reference to Bostrom "that an artificial superintelligence" could emerge over the course of the next century "that is superior to humans. It could take command of war robots and lock humans in a zoo in species-appropriate proximity to chimpanzees" (Weber, 2015).[3] Here, the loss of control is associated with the topoi of superiority and destruction. Goethe already hints at this, as the water-collecting brooms prove to be impressively efficient, but finally destructive.

> Off they run, till wet and wetter
> Hall and steps immersed are lying.
> What a flood that naught can fetter!

In Goethe's poem, the apprentice tries to stop the autonomous broom by splitting it with a hatchet. However, this only leads to the animated brooms doubling in

[2] In Estonian mythology, the magical creature named "Kratt" ("the treasure-bearer") is present. The Kratt is created from a hay or old household utilities and it will be given three drops of (Devil) blood to bring life into Kratt. It must forever obey its master's commands—mainly to collect various goods and treasures for the master. But if the master fails to give it the command or tasks to fulfill, the Kratt turns against the master and kills him.

[3] We have tacitly translated non-English quotations into English.

number. Even if the broom does not yet cause the multiplication itself, the possibility of uncontrollable replication already appears. Ray Kurzweil, for example, discusses the gray goo scenario: self-replicating nanobots transform carbon, an important building block of life. An out-of-control nanobot plague could turn the biomass on Earth into "gray goo" within a short period of time (Kurzweil, 2005, pp. 399–400).

The animated broom in Goethe's ballad is given a head and legs, i.e., it is given a human-like shape (topos of similarity). An artificial being in a human-like form is what we call an android. The idea of androids and artificial autonomy was by no means new in Goethe's time (Cave & Dihal, 2018). One should think of the story of the sculptor Pygmalion described by Ovid in the Metamorphoses, who falls in love with the very realistic ivory statue of a woman created by himself, which is finally animated by the goddess of love (Ovid, 1922, book 10, line 243–297). In the Iliad, Homer reports that Hephaestus, the god of forging and fire, employed maidens made of gold as assistants (Homer, 1925, book 18, line 417–420). According to Hesiod, Hephaestus is also said to have created Pandora from earth (Kerényi, 1998, p. 172). From ancient China in the third century BC, the Liu-Tzu-yüan tells of the craftsman Yen Shih, who presents the king with an artificial figure that can "sing and do all kinds of things" (Heckmann, 1982, p. 23). In the age of rationalism, which was dominated by the metaphor of the machine "in the form of clockwork" (Stollberg-Rilinger, 1986, p. 34),[4] automata and artificial systems of all kinds found a fascinated audience. Well-known are, for example, the Flute Player (1738) or the Automata Duck (1739) by Jacques de Vaucanson or Wolfgang von Kempelen's chess automaton, which made its first public appearance in Vienna in 1769 (Heckmann, 1982, pp. 219, 258–260). In his writings, Descartes also repeatedly refers to "mechanisms of real existing automata" of his time, "be they fountains, organs, clocks, mills, etc." (Sutter, 1988, p. 62). Sophisticated water games or automata were particularly popular at court, as they were regarded as "insignia of power" (Heckmann, 1982, p. 208).

The Enlightenment, mechanization of calculation, and innovations in machine tools brought a new twist to the idea of the automaton. Descartes discusses the automaton as a "model" of living bodies (Sutter, 1988, p. 54), even though meaning and sense still appear to him to be mechanically inimitable (Sutter, 1988, pp. 66–67; Richtmeyer, 2020, p. 103). In the *Discours de la Méthode*, he sees the difference between humans and androids in the human capacity for language and the situational adaptability of reason (Descartes, 2006, pp. 45–49)—an explanation that is no longer convincing in the age of generative language models and machine learning. Pascal and Leibniz made attempts to "develop mechanical computing devices" (Zimmerli & Wolf, 1994, pp. 10–11), because computational processes also appeared to be machinable. Thinking and calculating move closer together. Already in Descartes, "the idea of organizing linguistic operations in terms of calculation

[4] The machine represents "the fundamental metaphorical resource of rationalist thought" (Stollberg-Rilinger, 1986, p. 31).

within the framework of a *langue universelle*" (Mittelstraß, 1978, p. 180) emerges in passing. Hobbes identifies reasoning and calculation (Hobbes, 1656, pp. 2–3). Calculability becomes the "criterion of reason" (Zimmerli & Wolf, 1994, p. 10). With calculating machines, artificial thinking machines also become conceivable as something feasible in principle. If these are built into a machine body, the result is an artificially intelligent android.

Technical innovations in the field of machine tools, such as the fly-shuttle for looms (1733) or the *Spinning Jenny* (1765), showed in the eighteenth century that mechanical devices could also be productive industrial tools. New, significantly more productive machine tools put the entire production chain under innovation pressure (Landes, 1969, pp. 84–85). The mechanical clockwork, which served as a model for the machine metaphor for a long time, is also intertwined with the development of wage labor, increased productivity, and social and body discipline (Thompson, 1967; Treiber & Steinert, 1980, pp. 23–52).

The first fully automatic loom was designed by the aforementioned Vaucanson during his time as inspector of the French silk mills in France (1745–1748). But it was not more than a prototype. Vaucanson's loom was controlled by a punched card system, which consisted of a perforated cardboard that was wound around a roller that was also perforated (Schneider, 2003, p. 194). The flute player was also controlled by a program that was stored on a roller. In this case, it was a pin roller (Schneider, 2003, p. 193). In her technical-historical reconstruction and evaluation of Vaucanson's work, Birgit Schneider summarizes:

> If, on the one hand, the salons dreamed of hypothetical androids that imitated human speech, feeling and thinking and conquered courtly society in the human form of scribes, piano players, shepherds and chess players, on the other hand, the knowledge and physiology of the worker found its way into the construction of power looms without it being possible to identify their human model. Workers saw their skills translated into gears, rollers and levers, which were thus [...] part of the machine as a whole and no longer part of the thinking, feeling and desiring worker. (Schneider, 2003, pp. 194–195)

Goethe's tool ("broom"), equipped with intelligent self-control ("head") and moving autonomously ("legs"), draws on such ideas, summing them up to a certain extent. At the turn of the First Industrial Revolution, his warning about the autonomous automaton servant does not come out of nowhere.

3 The Uncanny Doppelganger Motif

It is E.T.A. Hoffmann who, in the story *The Sandman*, plays out the possibility of the indistinguishability of humans and androids and thus the topos of similarity. This story was published in 1815 and is considered a work of Romanticism. It marks the human-like automaton as an uncanny doppelganger and thus transforms the automaton into a being that evokes fear.

Hoffmann's main character is the student Nathanael, who goes insane at the end of the story and commits suicide after unsuccessfully trying to push his girlfriend

Clara off the top of the Council Tower. Before the reader's eyes, Nathanael gradually changes and comes ever closer to madness as he undergoes a series of experiences that seem uncanny not only to him, but also to the reader. It begins with an encounter with the optician ("weatherglass dealer") and automaton maker Coppola, whom Nathanael believes to be a sinister person from his childhood, namely the advocate Coppelius. As it turns out, this perception is correct, but for a long time Nathanael is not sure about this.

For us, of course, the uncanny effect of the automaton doll Olimpia is at the center of interest. This is a fully mobile and rudimentarily speaking automaton beauty constructed by Nathanael's physics professor Spallanzani and Coppola, but who enters the plot as Spallanzani's daughter. Nathanael falls in love with her because he has the feeling that he can talk to her and be understood by her, although she can essentially only say "Ah, ah!" (Hoffmann, 1844, p. 158). In doing so, however, she allows Nathanael all kinds of projections and transferences that seem to confirm his romantic feelings.

Hoffmann's reference to the "Ah, ah" is, of course, an ironic comment on the Romantic Era, which he thus summarizes in its (shortest possible) term.[5]

The emerging connection between Nathanael and Spallanzani's 'daughter' is abruptly ended when Nathanael witnesses a struggle between the android's two 'fathers' in which she is fragmented. Coppola escapes with the body, while "a pair of eyes, which lay upon the ground, were staring at him" (Hoffmann, 1844, p. 162). Nathanael is shocked to realize the nature of Olimpia, for "he had seen but too plainly that Olimpia's waxen, deadly pale countenance had no eyes, but blank holes instead—she was, indeed, a lifeless doll" (Hoffmann, 1844, p. 162). Now that he understands that Olimpia is an automaton, Nathanael goes mad for the first time (Hoffmann, 1844, p. 162).

A century later, in 1919, Sigmund Freud, in his treatise on the uncanny, considered Hoffmann's story in detail. Freud asks how the impression of the uncanny arises in relation to an everyday experience or in relation to a literary narrative. He concludes that the uncanny can be traced back to the return of animistic ideas or repressed complexes. The "uncanny is in reality nothing new or alien, but something which is familiar and old-established in the mind and which has become alienated from it only through the process of repression" (Freud, 1955, p. 241). Since the return of the repressed is in a sense a doubling, in Hoffmann's story, he focuses particularly on the "phenomenon of the 'double'" as a means of creating uncanny impressions (Freud, 1955, p. 234).

According to Freud, animistic ideas and the omnipotence of thought are typical of primary narcissism. At this stage the child "do not distinguish at all sharply between living and inanimate objects" (Freud, 1955, p. 233). Children "are especially fond of treating their dolls like live people" (Freud, 1955, p. 233). The animistic stage is overcome in the course of psychic development, whereby it always

[5] In German "Ach–Ach–Ach!" The only other phrase Olimpia manages to utter is: "good night, dearest!" (Hoffmann, 1844, p. 161).

leaves behind "certain residues and traces of it which are still capable of manifesting themselves" (Freud, 1955, p. 240). Freud interprets the fear of injury to the eyes as "a substitute for the dread of being castrated" (Freud, 1955, p. 231). The uncanny effect emanating from the sandman Coppelius/Coppola and the placeholders of the fear of the eyes (such as Olimpia's staring eyes or the view into Clara's eyes conveyed by Coppola's telescope) thus refers to the Oedipus or castration complex. In this context, the uncanniness of the automaton is of particular interest, as a cultural shift can be seen here. In the previous century, automatons did not have a sinister effect on their audience.

Freud explains Olimpia's uncanny effect on Nathanael and the contemporary reading public from her character as a doppelganger. The doubling or mirroring "was originally an insurance against the destruction of the ego" (Freud, 1955, p. 235). Freud speculates that the idea of an immortal soul may have been "the first 'double' of the body" (Freud, 1955, p. 235). But these "ideas [...] have sprung from the soil of unbounded self-love, from the primary narcissism which dominates the mind of the child and of primitive man" (Freud, 1955, p. 235). In more enlightened times or at later stages of psychic development, however, the doppelganger "becomes the uncanny harbinger of death" (Freud, 1955, p. 235).

The human-like automaton or android is a doppelganger and therefore uncanny as long as its true character remains unclear and doubtful. If, on the other hand, it is introduced from the outset as a feature of the fictional reality of a story or novel, then, according to Freud, the uncanny effect no longer applies, because "that feeling cannot arise unless there is a conflict of judgement as to whether things which have been 'surmounted' and are regarded as incredible may not, after all, be possible" (Freud, 1955, p. 250).

The fact that the doppelganger motif also harbors a threat of annihilation becomes even clearer when Jacques Lacan's reflections on the mirror stage from 1949 are taken into account. Lacan notes that when the human child sees and recognizes itself in the mirror, it falls into a state of jubilatory activity. He explains this by the fact that the child recognizes itself in the mirror image in an ideal gestalt and can perceive itself as identical. It is a "total form of his body, by which the subject anticipates the maturation of his power in a mirage" (Lacan, 2006, p. 76). At the same time, the mirror image is always also an "armor" (Lacan, 2006, p. 78) against the disintegration into the partial objects that characterize the early childhood experience of the body and the environment. The mirror phase of subject formation constitutes an "alienating identity," because the "lure" is spatially located elsewhere (Lacan, 2006, p. 78). Lacan explicitly states that the mirror image is a manifestation of the "appearance of *doubles*" (Lacan, 2006, p. 77, emphasis in original). Phantasms of fragmentation are thus always kept alive in the mirror image or in the motif of the doppelganger (Lacan, 2006, p. 78). Hoffmann's story makes use of such fantasies in abundance—not least Olimpia is effectively fragmented during the struggle between Coppola and Spallanzani, which triggers Nathanael's first fit of madness. It should be noted that when the child Nathanael meets Coppelius, he is threatened with more than just castration. After the father is able to avert this threat, Coppelius treats Nathanael like an automaton, whereby a fragmentation into partial objects is carried

out (Hoffmann, 1844, p. 144). The automaton is therefore a doppelganger of Nathaniel.

Although Freud notes an "identity of Olimpia and Nathanael" (Freud, 1955, p. 232, footnote 1), he overlooks the fact that Olimpia is also Nathanael's sister. Both have two automaton-makers as fathers, one of whom is Coppola-Coppelius. It is hard to overlook the fact that Spallanzani, as a teacher and future father-in-law, is also a surrogate father for Nathanael. However, since the rationalist and prosaic Clara sometimes seems like an automaton to him, the same applies to her as to Olimpia—especially when one considers that Olimpia takes Clara's place, i.e., is her doppelganger. So, there is incest in Nathanael's love story. Nathanael is not allowed to marry either Clara or Olimpia. He commits suicide when he realizes the identity of Clara and Olimpia and thus their true nature as (his) sisters. After Nathanael's suicide, Clara finally marries someone else and from then on lives in the socially desirable female role that Olimpia embodied so perfectly. Here it becomes clear that Clara is Olimpia's doppelganger and sister, just as Olimpia is the doppelganger of the pre-emancipatory bourgeois female role.

Why does the human-like automaton suddenly become uncanny? Psychoanalysis explains this with reference to the automaton's doppelganger motif, but this motif did not yet have an uncanny effect in the previous century. Nor can the uncanniness of technical autonomy be explained by the impact of the Enlightenment, which was already at work in the seventeenth and eighteenth centuries. After all, the fascination with mechanistic models of the world and the body is an expression of the Enlightenment. However, an indication of this transformation of the doppelganger motif into something threatening can be extracted from the reflections of Freud and Lacan. There is a threat in them insofar as they can replace the human being—and not only the human being as a factory worker. In the circles of the (bourgeois) owners of the means of production, this was probably of less interest overall. Does E.T.A. Hoffmann give us a helpful hint with Olimpia's romantic ability to speak? Does the looming possibility that machines will become capable of speech, as well as successfully simulate body movements, intellect, and emotions, make them seem uncanny? In any case, it is striking that the scenario of Turing's imitation test appears in both Descartes' and Hoffmann's works. According to Turing, a computing machine 'thinks' when an observer can no longer distinguish its output from that of a human being (Turing, 1950). With Descartes (2006, p. 46) and Hoffmann, the reference to an elaborate ability to speak still serves as proof of the inferiority of automata. But it should be clear that this also raises the possibility of this bastion of the living human falling. Does the possibility of a speaking and thinking automaton make technical autonomy in the bourgeoisie seem uncanny?

Another reason may have been the shift from mechanism as the fundamental ideological paradigm to organism and vitalism, which was popularized by the Romantic movement. This shift has to do with the discovery of "society" as a system of interacting productive forces (Spreen, 1998). Since Romanticism, the mechanical android can be used in texts and images as a "symbol of a cold, rational machine world" (Schlich, 1998, p. 548).

The question remains as to why Hoffmann gave the automaton a female role? Is it because females appear to be the perfect servant, complementary and uncontradictory to the male identity? "Ah, ah" and "good night, dearest" (Hoffmann, 1844, p. 161)? Nevertheless, female automatons remain uncanny, because they keep alive the threat that the dialectic of lord and bondsman will 'turn around' in historical-practical terms—in Freudian terms, this corresponds to the threat of castration.

4 Robot War Against Humans

About a century after *The Sandman*, the Czech writer Karel Čapek published the play *R.U.R. Rossum's Universal Robots*. The play was published just 2 years after the end of the First World War, which entered the collective memory as the "war of machines" (Weber, 1978, p. 981). It is considered a masterpiece of science fiction (Suvin, 1979, pp. 270–283) and was translated into 30 languages. Shortly after its publication, the play was performed worldwide to considerable acclaim, including in Prague, Warsaw, Riga, Moscow, Berlin, Paris, London, and New York. It was the first Czech play on Broadway. "R.U.R. opened at the Garrick Theatre on West 35th Street on 9 October 1922 and met with tremendous success with an unusually long run of 184 performances" (Reilly, 2011, p. 151). The BBC adapted the play for its 1938 television program, which began broadcasting in 1936 (Bould, 2008, p. 210).

This theater play introduced the term 'robot' into the discussion and popularized it. The word is derived from the Czech 'robota', which means 'drudgery' or 'servitude' and refers to feudal labor and social structures (Reilly, 2011, p. 148; Suvin, 1979, p. 270).

The events depicted in the play take place on an island where the Rossum factories are located. Here, human-like robots are produced for all kinds of purposes, but "they've got no will of their own. No passions. No hopes. No soul" (Čapek, 2015, p. 22). "They learn how to speak, write and do arithmetic, as they've got amazing memories. If you read a twenty-volume encyclopedia to them they could repeat it back to you word for word, but they never think of anything new for themselves. They'd make very good university lecturers" (Čapek, 2015, p. 17).

The production process is based on a division of labor and is itself carried out by robots. There are "mixers" in which a kind of "dough" is produced, a "bone factory," a "spinning-mill" for nerve fibers and veins and an "intestine mill, where kilometers of tubing run through at a time." In an "assembly room," everything is put together in a Tayloristic process "like making a car" (Čapek, 2015, p. 16). So Čapek's robots are "essentially organic human beings who were reduced through technological means to the status of machines" (Shelley, 2024, p. 81). Only at the reception did they become mechanical beings. "This alteration of Čapek's *R.U.R.* embodied concerns about technological determinism and introduced the idea of the robot uprising as a result of out-of-control technology that continues to be a theme in science fiction and policy-making today" (Shelley, 2024, p. 82).

The purpose of robots is to replace people because human labor is too expensive (Čapek, 2015, p. 21). In society, they are used as will-less but desirable labor slaves for every conceivable purpose—including as soldiers. As the central director explains: "it's more like working in the way a new piece of furniture works" (Čapek, 2015, p. 17). The company has a monopoly on robotics and does a lot of business with them.

But the lack of labor and effort makes people infertile, according to the story, an effect of general decadence. Helene, who originally came to the factory as an activist of the "League of Humanity" to campaign for the rights of robots, learns from the pacifist and engineer Alquist that human labor has become superfluous, because "there's no need for anyone to work, no need for pain. No-one needs to do anything, anything at all except enjoy himself. This paradise, it's just a curse! [...] And us human beings, the pinnacle of creation, we don't have to take care of work, we don't have to take care of children, we don't have to take care of the poor! Bring in all the fun, quick! [...] There's no need for men any more, Helena, women aren't going to give them any children" (Čapek, 2015, p. 38). Mankind "is actually no longer needed" (Čapek, 2015, p. 42).

So, humanity will go extinct. At the same time, the robots are revolting. This is caused by secret experiments by the head of the experimental department (Čapek, 2015, pp. 58–59). A violent showdown follows, which ends with the destruction of the humans—only Alquist remains alive because the robots recognize a kinship between him and themselves (Čapek, 2015, p. 72). Alquist is also given the task of reconstructing the production knowledge that has been lost in the course of the conflict. Without this knowledge, the robots cannot reproduce and would die out too within 20 years. However, it is not possible to reconstruct the lost "secret of life" (Čapek, 2015, p. 78). At the end of the play, the robots discover the secret of life by themselves, namely—romantic twist—in the form of love, so that there is a new Adam and Eve who initiate "a new cycle of creation or civilization" (Suvin, 1979, p. 272). Here we find the topos of self-replication combined with the topoi of superiority, destruction/extermination, and similarity.

The topos of destruction and extermination that Čapek unfolds is very familiar to us today. We also know it as the *terminator scenario,* because it is copied in the very popular *Terminator* films with Arnold Schwarzenegger in the role of a T-800 (Cameron, 1984, 1991). Alongside other well-known film productions, the first two *Terminator* films from 1984 (*T1*) and 1991 (*T2*) reflect the cultural stereotype of the tool that emancipates itself and becomes the master in a particularly condensed way (Rushing & Frentz, 1995). This stereotype is particularly impressive when the tool is a weapon. "The weapon eventually breaks free from [...] control, becomes technologically perfected, and, in a final profane reversal, turns against the very hand that used to wield it" (Rushing & Frentz, 1995, p. 5). *T1* and *T2* present us with a dystopia in which the artificial superintelligence "Skynet" attempts to wipe out humanity with a coordinated nuclear strike. Machines and humans then fight their battle for survival on a post-nuclear battlefield.

The terminator scenario involves culturally reproduced phantasms of fear that deal with the destruction of humanity by artificially created actors, whereby an appeal is made to human responsibility and human hubris is criticized (Rushing & Frentz, 1995, pp. 184, 187, 188–192). Ultimately, the scenario formulates a warning against a kind of *mad science*, as clichéd by Alquist in *R.U.R.*:

> It's science I blame! Technology I blame! [...] Myself! All of us! It's us, we're the ones to blame! We thought we were doing something great, giving some benefit, making progress. I don't know what magnificent ideas it was for that we've destroyed mankind! And now all our greatness is bursting like a bubble! Not even Genghis Khan built up a heap of human bones like we've done. (Čapek, 2015, p. 58)

Numerous stories more or less reproduce the cliché image and horror scenario of artificial creation turning against humanity.[6] It is already at the heart of Goethe's *Sorcerer's Apprentice*. As a result, modern Western technology culture, which relies on the grand narrative of technological progress, has to cope with an image of computational automation that characterizes android robots as sinister, fears loss of control, and conjures up terminator scenarios.

But is it this culturally mediated and uncanny technology cliché that triggers our fears or are they actually rooted in the technology? It is obvious that the use of AI cannot take place without ethical reflection and legal regulation, as this applies in general (and therefore also in military contexts). In the ethics of AI, everything revolves around accountability, as Niklas Luhmann pointed out in his basic work on automation (Luhmann, 1966; Esposito 2021; Spreen, 2023). "Automation is no more an excuse [...] than human failure" (Luhmann, 1966, p. 81). Accountability can only be held by people. Even if machines act incorrectly, someone has to 'take the rap' within the framework of the organizational role structure. For Luhmann, however, concepts like 'responsibility' and 'accountability' do not refer to personal virtue, but rather to communication problems that depend "decisively on the structure of the social system in which communication takes place" (Luhmann, 1995, p. 175). A person (in an organization) can be accountable for automatic decisions only if he or she can trust the technology and if it operates in an ethically neutral manner, i.e., does what it is supposed to do (Luhmann, 1966, pp. 19–20). If the conditions for an accountable use of technology are met, then there is no reason to develop fundamental fears about technology. However, these conditions are societal because they are political, legal, and organizational.

[6] Just a couple of examples: Mary W. Shelley (1818), *Frankenstein or The Modern Prometheus*; John Wyndham (1951), *The Day of the Triffids*; Philip K. Dick (1953), *Second Variety*; Edmund Cooper (1958), *The Uncertain Midnight*; Charles Eric Maine (1966), *B.E.A.S.T.*; Stanisław Lem (1983), *Weapon Systems of the twenty-first Century or the Upside Down Evolution*.

5 Robot Ethics and the Fear of Losing Responsibility

In 1942, in the middle of the high-tech Second World War and 14 years before the Dartmouth Conference elevated "artificial intelligence" to a research program (McCarthy et al., 2006), Isaac Asimov published the robot story *Runabout*,[7] in which he set out his ethical guidelines for the behavior of robots in social contexts for the first time. These are Asimov's three robot laws, which are still being discussed today. The robot laws read (in the 1942 formulation):

> One, a robot may not injure a human being under any conditions—and, as a corollary, must not permit a human being to be injured because of inaction on his part. [...]
> Two, [...] a robot must follow all orders given by qualified human beings as long as they do not conflict with Rule 1. [...]
> Three: a robot must protect his own existence, as long as that does not conflict with Rules 1 and 2. (Asimov, 1942, p. 100)

The story is about a robot that works on the sunny side of Mercury, with the third law emphasized more strongly for reasons of economy. The robot "Speedy" is supposed to collect liquid selenium on the planet close to the sun. Asimov himself describes the plot in a later meta-reflection as follows:

> Unfortunately, the robot was given his order casually so that the Second Law circuit set up was weaker than usual. Still more unfortunately, the selenium pool to which the robot was sent was near a site of volcanic activity, as a result of which there were sizable concentrations of carbon monoxide in the area. At the temperature of Mercury's sunside, I surmised that carbon monoxide would react fairly quickly with iron to form volatile iron carbonyls so that the robot's more delicate joints might be badly damaged. The further the robot penetrates into this area, the greater the danger to his existence and the more intensive is the Third Law effect driving him away. The Second Law, however, ordinarily the superior, drives him onward. At a certain point, the unusually weak Second Law potential and the unusually strong Third Law potential reach a balance and the robot can neither advance nor retreat. He can only circle the selenium pool on the equipotential locus that makes a rough circle about the site. (Asimov, 1954, p. 23).

The endless circular walk, which Speedy is forced to do by the circumstances and his programmed ethics of action, is ended by the approaching engineers, as one of them deliberately puts his life in danger and thus activates rule 1. Speedy saves him, thereby breaking the algorithmic loop that led to the hike (= activation of a stopping condition).

Here, the robot saves the human. Asimov (1954) refers to this story to show that robot ethics offer many opportunities to create exciting plots. In contrast, he considers the terminator scenario to be an "old, tired plot" (Asimov, 1954, p. 22). He describes the turn toward the uncanny that so many robot and AI stories take as "Frankenstein complex" (Asimov, 1954, p. 23). Asimov opposed the idea of the robot as "a sinister form" (Asimov, 1954, p. 22).

> Only one robot-plot seemed available to the average author: the mechanical man that proved a menace, the creature that turned against its creator, the robot that became a threat to humanity. And almost all stories of this sort were heavily surcharged, either explicitly or

[7] Also included in *I, Robot* (1950).

> implicitly, with the weary moral that "there are some things mankind must never seek to learn." (Asimov, 1954, p. 22)

Asimov is not the only author to have made this paradigm shift.[8] But he is certainly one of the most important ones to initiate it. This opens up a new realm of possibilities for thinking about artificially intelligent automation and technical autonomy. This space of possibility is not dominated by scary Terminator scenarios. Artificially intelligent systems appear here as "moral agents" (Floridi & Sanders, 2004) that do not harm humanity, but benefit it. Asimov suggests equipping these systems with "laws" that determine their behavior.

This question is highly topical (Loh, 2019)—but was raised by Asimov long before the Dartmouth conference. At present, Ron Arkin, for example, argues that military AI systems and robots are the "better soldiers" because they have a programmed "ethical governor" that can reliably prevent war crimes (Arkin & Marsiske, 2012). "I believe that if these systems were properly inculcated with a moral ability to adhere to the laws of war and rules of engagement they could outperform human soldiers with respect to humaneness. The end product then could be [...] a saving of non-combatant lives when compared to human warfighters' behavior" (Arkin, 2011). Arkin also envisions AI advisory systems—"an ethical advisor suitable for enhancing human performance" (Arkin et al., 2012, p. 587)—that could provide a second opinion in complex decision-making situations. This also appears to make sense for civilian contexts. Such robotic advice needs by no means be limited to the ethical dimension, but can support the evaluation of facts. Arkin thus takes up an idea of the science fiction author and futurologist Arthur C. Clarke, who predicted in the mid-1980s that in the future "the computer will be used as a consultant of sorts" (Clarke, 1986, p. 39), whereby Clarke had medicine in mind.

The editor of the magazine[9] in which Asimov's robot laws were first published saw a need for automatic control of weapon systems against the backdrop of "scientific, mechanized war." "The dive bomber and torpedo plane approach so rapidly, and change their angle with respect to the gunner so rapidly, that no manually controlled gun con follow their line of light" (The Editor, 1942).

Asimov draws attention to ethical robotics. Of course, the question of robots that operate under their own control according to ethical guidelines affects all areas of society. This also includes the military, which is one functional system among others. But it is precisely in relation to "combat troops" that questions of "responsible information processing" arise with particular seriousness, because they are significantly less "error-sensitive" than, for example, "casual social clubs" (Luhmann, 1995, p. 176). This area can therefore also be seen as a burning glass for the ethical questions and problems associated with automation and, in particular, with artificially intelligent automation. Through Luhmann's lens, however, it must also be clear that even behind highly autonomous decisions made by AI systems there are accountabilities that can be attributed to people in social roles. Even without a

[8] For example, Ray Bradbury (1969), *I Sing the Body Electric!*

[9] John W. Campbell, Jr. (1910–1971).

permanent link, these systems are *commanded* and in this sense are *only semi-autonomous*. This also means that space must be given to ongoing organizational ethical learning processes (Schuchter et al., 2021) and that the means of diffusion of responsibility in organizations that are often used in social reality (cf. Luhmann, 1995, pp. 180–190) must be limited in relation to the decisions of AI systems—especially when it comes to technical systems that are concerned with existential decisions.

The implementation of ethics in machines, which Asimov envisages, is also seen critically and gives rise to fears. The philosopher Sybille Krämer, for example, who wrote a fundamental work on the history of ideas of symbolic machines (Krämer, 1988), apprehends a "relief from responsibility" through the increasing use of computers in society (Krämer, 1992, p. 335).

> The human being is regarded as a being capable of and obliged to take responsibility. We therefore characterize human activity as an action and the actor of this action as a person. All mechanization now aims to transform an action into an operation. The ideal of this transformation is to replace the actor with a mechanism. A mechanism, however, is an actor for which the quality of personality, i.e., the ability and duty to take responsibility, can be disregarded. (Krämer, 1992, p. 335)

According to Krämer, the use of technology creates "a moral distance to the results of our actions. The bomber pilot, for whom the operational field of the computer screen makes the distinction between merely simulated and actual bombing obsolete, is relieved of much of the remorse of killing and destroying with his own hands" (Krämer, 1992, p. 336). Krämer notes a general trend toward the "dissociation of technology and ethics," which arises from the separation of the knowledge of "how something is done" (problem-solving competence) from the knowledge of "what we are doing and how this action can be justified" (justification competence) (Krämer, 1992, p. 337).

What Krämer has in mind in this article, which is now over three decades old, is something like a critique of instrumental thinking, which draws its justification from functionality and threatens to suppress the ethical dimension. Through programming and informatics, "mental activities" are restructured in such a way that "they can be accomplished without the power of judgment," that they can be "detached" from "the personality of the actor." They should be "stripped of their ethical dimension" (Krämer, 1992, p. 338).

Following this line of argument, artificially intelligent ethical advisory systems could also be seen as a relief from the competence of justification. This raises the fear that such systems would lead to a delegation of accountability to machines. They would not contribute to ethical enhancement, but rather have the opposite effect.

We believe that such arguments should be taken seriously, but therefore consider it all the more important to embed the use of artificially intelligent systems sociologically. In this case, this would mean that they should be contextualized in terms of organizational sociology. Luhmann provided an example of such a contextualization in his work on automation in public administration (Luhmann, 1966; Esposito, 2021; Spreen, 2023). Only then can implicit social romanticism be avoided,

according to which responsibility can only be perceived in direct physical interaction. Already the argument that the dissociation of weapon triggering and weapon effect—see the example of the bomber pilot—leads to an ethical and moral disengagement and irresponsibility must be empirically questioned. For example, drone operators run the risk of psychological damage resulting from a paradoxical combination of contradictory social forms. Due to the high-resolution cameras on their drones, drone operators are 'close to the scene.' However, like other employees, but not like soldiers on deployment, they also return to their families every day, look after their children's homework, etc.[10] This daily alternation between a performance role burdened with existential decisions and the family environment represents a psychological stress factor (Singer, 2010a, pp. 344–347; 2010b, pp. 62–63). From such a counterexample, we can learn that it is important to consider the social contextualization of technological systems.

No ethical expert system relieves its users of responsibility and accountability for their decisions. However, potential users should also be made aware of this. The 'ethical use of ethical counsellors' can be supported by teaching appropriate media competencies. Krämer's argument should therefore be understood as a demand for appropriate contextual conditions for the use of artificially intelligent and ethical automation—contextual conditions that prevent moral avoidance practices.

6 Eastern-Western Cultural Differences in Paradigms of Robots

Previously, we looked at Western cultural discourses that have underpinned perceptions of artificially intelligent systems. However, we cannot ignore the culture of the Orient, which has come a long way and had significant successes in terms of technological development. We will therefore take a closer look at the perception of robots in Eastern culture.

In 1970, when Masahiro Mori, the robotics professor in Tokyo Institute of Technology published an essay *Uncanny Valley* (Mori, 2012), where he envisioned people's reactions to robots that looked and acted like humans, it received almost no attention. However, the interest toward his ideas has quickly increased, as technology evolves and human-like robots are developed by many companies and researchers. And more importantly, his ideas formulate a ground for understanding the fear of robots and pave the way for understanding the differences between the Eastern and Western attitudes to robots.

Mori hypothesizes that a person's response to the robot shifts from empathy to aversion or even disgust depending on the human-likeness of the robot. We tend to

[10] Singer's observation refers to US drone operators carrying out missions from American soil on the other side of the world, e.g., in the Middle East, Afghanistan, Iraq or the Balkans. In the war in Ukraine, however, Ukrainian drone pilots are as close to the front line as regular soldiers.

feel sympathetic about the robots that are similar to us, but only up to certain point. Upon reaching to the uncanny valley, our affinity descends into feeling of strangeness, a sense of unease, and tendency to be scared or freaked out. The characteristics like motion (possible sense of threat), features of the face and body, and also the voice could raise eeriness, as also subsequently confirmed by later research (Gray & Wegner, 2012). The eeriness is closely related to the feeling that a robot could sense and feel—called "attribution of mind" (Gray & Wegner, 2012, p. 125). However, application of human-likeness and normalization of artificial human have different roots in Western and Eastern cultures. Here, it is interesting to note that the attitudes are much more universal when it comes to animal-like robots. For example, the experiments with "Paro," a cuddly baby seal-like therapeutic robot, have shown its rather similar acceptance in different cultures. Probably due to its pet-likeness and widespread perception of pets like family members, it is more widely accepted cross-culturally (Shibata et al., 2009).

In regard to human-like robots, we first have to look at how technology is perceived in general in both cultures. Part of explanation here lays in ontological difference between the notion of the 'nature' and the 'human being' (Kaplan, 2004). For example, Shintoism and Buddhism encourage the blurring between the realization of nature and production of man. This goes back to the principles of animism in Shinto religion, i.e., the belief that inanimate objects can have spirit, soul, or consciousness. In contrast, Western culture favors a clear distinction between man-made and natural effects (Nakamura, 1992). The idea is to categorize the world in a systematic and precise way. The understanding of technology is influenced by the same thinking. For Westerners, technology symbolizes the culture of *man-made* and is fundamental for defining what humans are. Technology is an integral part of the development of modern Western civilization and Western modes of behavior and production. Therefore, the possible convergence of humans and machines is an important topic, equally fascinating and frightening. For Eastern culture, technology is not so 'important' in a sense of giving it the meaning of a fundamental question. Technology plays a more external role and does not define human specificity.

The difference between the natural and the artificial is not so crucial in Eastern culture. Building machines is a positive activity in the search of the natural laws that govern the world. For example, in analyzing Japanese culture of technology, it seems that technology represents rather aesthetic and harmonious forms to exemplify what is essential in Japanese culture. Foreign technology could be tamed without necessarily melding with it. It means that technologies that may one day have been 'wild' and 'unknown' could be one day well mastered and harmoniously integrated in society. The popular culture is reflecting this notion. "Tetsuwan Atom," the TV cartoon character, invented in 1951, is a small, infant-like robot equipped with an "atomic heart" that defends humanity against various threats often coming from outer space. From a Western perspective, the social milieu in which Atom operate is surprising, as it is a robot "that lives in a traditional Japanese family and goes to school like the other children." This robot "thus showed that robots and humans can live together in a quite natural way" (Ichbiah, 2005, p. 86). Another feature that may seem odd for a Western audience is the use of the nuclear energy providing a heart

for the robot. It plays the role of a vital force. At the end of the Second World War, one could have expected that nuclear energy would be associated by Japan with death and defeat. But instead of being diabolized, the destructive energy was reintegrated into fiction as a positive life principle. And more than that, Tetsuwan Atom was exported in the West under the name "Astro Boy," suppressing the reference to nuclear energy to be better accepted by a Western audience (Kaplan, 2004, p. 466).

Another aspect in Western culture that is influencing our perception of robots is the superiority of the human being (Kaplan, 2004). Western culture considers humans to be superior to all other creatures, which in turn means that it fears machines that might become better or more competent than humans. Eastern culture, lacking a sense of superiority over machines or nature, does not experience this fear. Therefore, it is unafraid of newer, more capable robots or artificially intelligent doppelgängers. It means that we see ourselves in the mirror of the machines that we build. Because of the Western notion that human can be considered as the most advanced machinery, it is difficult to accept the more intelligent and more capable inventions. New machines can potentially force us to redefine ourselves, challenging our specificity. Many science fiction stories describe how robot armies conquer the world. Are we afraid of that, or is it more the case that we are afraid that robots will force us to change our self-image? We like ourselves the way we are and don't want that to change—maybe that's our real fear.

7 Conclusions

We find two major strands of discourse in relation to communication about technical autonomy—by which we mean fictions of machines that act and communicate independently to a high degree. On the one hand, there is the construction of an uncanny and threatening technical autonomy. The spectrum of cases analyzed here ranges from loss of control and fear of castration to a war between machines and humans. 'Horror scenarios' are constructed. On the other hand, there is the image of a supportive, albeit not risk-free, technical autonomy. Moral programming is discussed here. Asimov primarily discusses control through a set of rules, while Arkin is more concerned with control through boundary conditions.[11] These rules are intended to prevent misuse, but can lead to unexpected and risky scenarios. The robot Speedy's wanderings on Mercury are a good example of this.[12]

Specifically, we were able to identify various discourse topoi that motivate the fear of robots. These are: loss of control (*Sorcerer's Apprentice*, *R.U.R.*, paperclip scenario, gray goo scenario, superintelligence, relief from responsibility),

[11] In Asimov's 1985 novel, *Robotics and Empire*, the robot "R. Daneel Olivaw" expands the laws of robotics and reprograms himself. The machine learns from its observations and reflects on them morally (Asimov, 1986, esp. pp. 353, 426–427, 463).

[12] The short story *Internal Combustion* (1980) illustrates what can happen if "robotic inhibitions" are not looked after carefully. The story, published by de Camp (1956), also jokes about the fear of the robotic will to take over the world.

replication (*Sorcerer's Apprentice*, *R.U.R.*, gray goo, *Terminator*), superiority (*Sorcerer's Apprentice*, *R.U.R.*, superintelligence/singularity), similarity (*Sorcerer's Apprentice*, doppelganger motif, *Sandman*, *R.U.R.*, *Terminator*), and destruction/extermination (*Sorcerer's Apprentice*, *Sandman*, *R.U.R.*, replaceability/being superfluous, paperclip scenario, gray goo scenario, *Terminator*, i.e., war robots). These discourse topoi can be combined with each other in various mixtures to communicate scenarios of fear.

Discourses that rely on fear have many communicative advantages. They make it possible to be on the 'right' side morally. "Whoever suffers anxiety is morally in the right, particularly if it is anxiety on behalf of others and this can be assigned to a recognized non-pathological type" (Luhmann, 1989, p. 130). Both of the communicative conditions mentioned by Luhmann are fulfilled in relation to threatening technical autonomy: (1) One is afraid for humanity or for human dignity. (2) Risks are thematized, even by experts. One can think here of the open letter launched in 2023 by a number of prominent players from AI research and Silicon Valley to call for a pause in AI training (Future of Life Institute, 2023).

Horror scenarios of technical autonomy enable the communication of fear and generate attention. Examples of communication that derives evidence from the fear of technical autonomy are easy to find in military affairs. For example, Armin Laschet, the Conservative candidate for chancellor in the 2021 elections in Germany and Chairman of the Foreign Affairs Committee in the Bundestag since May 2025, responded to the question of red lines in defense policy: "We don't want a robot army. We don't want weapons that kill blindly according to an algorithm. That's a horror idea." By robots, he means "autonomous weapons systems, i.e., weapons that work without human intervention" (Busch, 2024).

Laschet is not alone in his fears. German social democrats also fear a "dehumanization of warfare" as a result of "increasing algorithmization"—at least that's what the position paper of a working team of the SPD parliamentary group in the German Bundestag said. The paper formulates framework conditions for technical "autonomy"—particularly with regard to the question of responsibility. If these framework conditions were fulfilled or regulated with legal certainty, a "categorical" rejection would by no means be justified, which raises the question of whether the position was not motivated by uncanniness and horror stories (SPD Bundestagsfraktion, 2019).[13]

The tech experts' open letter must itself be seen as an expression of overblown communication of fear (Leisegang, 2023). "*Should* we let machines flood our information channels with propaganda and untruth? *Should* we automate away all the jobs, including the fulfilling ones? *Should* we develop nonhuman minds that might eventually outnumber, outsmart, obsolete and replace us? *Should* we risk loss of control of our civilization?" (Future of Life Institute, 2023; emphasis in original).

By the way, do 'we' or could 'we' ever have control over modern civilization? A bizarre idea from a systems theory perspective, because all functional systems

[13] The working team in question has been dissolved.

control themselves only through specifically coded communication. There is no meta-control (Luhmann, 1989, pp. 106–114).

After all, in these fear scenarios, we are dealing with the discursive construction of an *alienated other*. The machine becomes independent and emancipates itself from its creators, the humans, in order to ultimately turn against them. What is being communicatively used here is the fear of the other.

In the age of enlightenment and science, highly developed, opaque, and autonomous technology may resemble magic. Arthur C. Clarke has pointed this out. His Third Law states that "[a]ny sufficiently advanced technology is indistinguishable from magic" (Clarke, 1999, p. 2). Ernst Jünger formulates the same idea in his science fiction novel *The Glass Bees* (Jünger, 1991, pp. 28, 66). Does this mean that the Enlightenment is turning against itself because its own products seem magical? One can think of the dialectic of enlightenment diagnosed by Adorno and Horkheimer. In this case, one might also expect the return of "terror" and "horror"—this time in relation to a machinic magic (Horkheimer & Adorno, 2002, p. 10). Freud argues similarly, seeing in technological autonomy a kind of return of animism that makes this technology seem uncanny.

But fear of robots is not absolute. A look at the seventeenth and eighteenth centuries reveals the contingency of this view, as does a look beyond the cultural horizon of "the West." And even within this Western cultural framework, the discourse on helpful technical autonomy offers a different perspective.

- The roots of the fear of technological autonomy go back to the beginning of the modern age.
- This discourse of fear does not run parallel to technological development.
- The same discourse topoi are used repeatedly. These are loss of control, replication, superiority, similarity, and destruction or extermination.
- In the age of the mechanism and in Far Eastern culture, technological autonomy is viewed less critically.
- In modern Western culture, too, there is a view that conceives of technological autonomy as helpful support. This view calls for an ethical programming of machines (Asimov) and the attribution of accountability to human actors (Luhmann).

References

Arkin, R. C. (2011). Military Robotics and the Robotics Community's Responsibility. *Industrial Robot, 38*(5). https://doi.org/10.1108/ir.2011.04938eaa.001

Arkin, R. C., & Marsiske, H.-A. (2012). Sind Roboter die besseren Soldaten? (Interview). In H.-A. Marsiske (Ed.), *Kriegsmaschinen. Roboter im Militäreinsatz* (pp. 141–144). Heise.

Arkin, R. C., Ulam, P., & Wagner, A. R. (2012). Moral Decision Making in Autonomous Systems: Enforcement, Moral Emotions, Dignity, Trust, and Deception. *Proceedings of the IEEE, 100*(3), 571–589. https://doi.org/10.1109/JPROC.2011.2173265

Asimov, I. (1942). Runabout. *Astounding Science-Fiction, 29*(1), 94–103.

Asimov, I. (1954). Robots I Have Known. *Computers and Automation, 3*(8), 22–26.

Asimov, I. (1986). *Robots and Empire*. Ballentine, Del Rey.
Bostrom, N. (2014). *Superintelligence. Paths, Dangers, Strategies*. Oxford University Press.
Bould, M. (2008). Science Fiction Television in the United Kingdom. In J. P. Telotte (Ed.), *The Essential Science Fiction Television Reader* (pp. 209–230). University Press of Kentucky.
Bradbury, R. (1969). I Sing the Body Electric! In R. Bradbury (Ed.), I Sing the Body Electric! (pp. 150–190). Knopf.
Busch, F. (2024, February 18). *Armin Laschet: "Wir wollen keine Roboter-Armee."* GMX Aktuelle News. https://www.gmx.net/magazine/politik/armin-laschet-roboter-armee-39331582
Cameron, J. (Director). (1984). *The Terminator* [Film]. Hemdale Film Corporation; Orion Pictures.
Cameron, J. (Director). (1991). *Terminator 2: Judgment Day* [Film]. Carolco Pictures; Tri-Star Pictures.
Čapek, K. (2015). R. U. R. (*ROSSUM'S UNIVERSAL ROBOTS*) (D. Wyllie, Trans.). Wildside Press (Original work published 1920).
Cave, S., & Dihal, K. (2018). Ancient Dreams of Intelligent Machines: 3,000 Years of Robots. *Nature, 559*, 473–475. https://doi.org/10.1038/d41586-018-05773-y
Clarke, A. C. (1986). *July 20, 2019. Life in the 21st Century*. MacMillan.
Clarke, A. C. (1999). *Profiles of the Future. An Inquiry into the Limits of the Possible* (millennial ed.). Victor Gollancz (Original work published 1962).
De Camp, L. S. (1956). Internal Combustion. *Infinity Science Fiction, 1*(2), 30–47.
Descartes, R. (2006). *A Discourse on the Method of Correctly Conducting One's Reason and Seeking Truth in the Sciences* (I. Maclean, Trans.). Oxford University Press (Original work published 1637).
Esposito, E. (2021). Was Luhmann von der Digitalisierung und von Algorithmen schon wusste. In T. Beyes, M. Warnke, W. Hagen, & C. Pias (Eds.), *Niklas Luhmann am OVG Lüneburg: Zur Entstehung der Systemtheorie* (pp. 127–132). Duncker & Humblot. https://doi.org/10.3790/978-3-428-55932-9
European Commission, Directorate-General for Communication. (2021). *European Citizens' Knowledge and Attitudes towards Science and Technology–Report* (Special Eurobarometer 516, April–Mai 2021). Publications Office of the European Union. https://data.europa.eu/doi/10.2775/071577
Floridi, L., & Sanders, J. (2004). On the Morality of Artificial Agents. *Minds and Machines, 14*, 349–379. https://doi.org/10.1023/B:MIND.0000035461.63578.9d
Freud, S. (1955). The 'Uncanny.' In *The Standard Edition of the Complete Psychological Works of Sigmund Freud, volume XVII (1917–1919): An Infantile Neurosis and Other Works* (J. Strachey, A. Freud, A. Strachey, & A. Tyson, Trans., pp. 219–252). Hogarth Press (Original work published 1919).
Future of Life Institute. (2023, March 22). *Pause Giant AI Experiments: An Open Letter. We Call on all AI Labs to Immediately Pause for at Least 6 Months the Training of AI Systems more Powerful than GPT-4*. https://futureoflife.org/open-letter/pause-giant-ai-experiments
Gray, K., & Wegner, D. M. (2012). Feeling Robots and Human Zombies: Mind Perception and the Uncanny Valley. *Cognition, 125*(1), 125–130. https://doi.org/10.1016/j.cognition.2012.06.007
Heckmann, H. (1982). *Die andere Schöpfung. Geschichte der frühen Automaten in Wirklichkeit und Dichtung*. Umschau.
Hobbes, T. (1656). *Elements of Philosophy the First Section, Concerning Body* (Unknown, Trans.). R., & W. Leybourn. http://quod.lib.umich.edu/e/eebo/A43987.0001.001
Hoffmann, E. T. A. (1844). The Sandman. In *Tales from the German. Comprising Specimens from the most celebrated Authors* (J. Oxenford, & C. A. Feiling, Trans., pp. 140–165). Chapman And Hall (Original work published 1816).
Homer. (1925). *The Iliad.* Volume II (A. T. Murray, Trans). Harvard University Press. https://www.perseus.tufts.edu/hopper
Horkheimer, M., & Adorno, T. W. (2002). *Dialectic of Enlightenment. Philosophical Fragments* (E. Jephcott, Trans.). Stanford University Press (Original work published 1947).
Ichbiah, D. (2005). *Roboter: Geschichte–Technik–Entwicklung*. Knesebeck.

Jünger, E. (1991). *The Glass Bees* (L. Bogan, & E. Mayer, Trans.). New York: Noonday Press (Original work published 1957).

Kaplan, F. (2004). Who is Afraid of the Humanoid? Investigating Cultural Differences in the Acceptance of Robots. *International Journal of Humanoid Robotics, 1*(3), 1–16. https://doi.org/10.1142/S0219843604000289

Kerényi, K. (1998). *Die Mythologie der Griechen. Bd. 1: Die Götter- und Menschheitsgeschichten* (19th ed.). dtv.

Krämer, S. (1988). *Symbolische Maschinen. Die Idee der Formalisierung in geschichtlichem Abriß*. Wissenschaftliche Buchgesellschaft.

Krämer, S. (1992). Symbolische Maschinen, Computer und der Verlust des Ethischen im geistigen Tun. In W. Coy, F. Nake, J.-M. Pflüger, A. Rolf, J. Seetzen, D. Siefkes, & R. Stransfeld (Eds.), *Sichtweisen der Informatik. Theorie der Informatik* (pp. 335–341). Vieweg. https://doi.org/10.1007/978-3-322-84926-7_23

Kurzweil, R. (2005). *The Singularity is Near: When Humans Transcend Biology*. Viking Penguin.

Lacan, J. (2006). The Mirror Stage as Formative of the Function of the I as Revealed in Psychoanalytic Experience. In J. Lacan (Ed.), *Écrits: The First Complete Edition in English* (B. Fink, H. Fink, & R. Grigg, Trans., pp. 75–81). W.W. Norton & Company.

Landes, D. S. (1969). *The Unbound Prometheus. Technological Change and Industrial Development in Western Europe from 1750 to the Present*. Cambridge University Press.

Leisegang, D. (2023, March 31). *Offener Brief zu KI: Opfer des Hypes*. netzpolitik.org. https://netzpolitik.org/2023/offener-brief-zu-ki-opfer-des-hypes/

Loh, J. (2019). Maschinenethik und Roboterethik. In O. Bendel (Ed.), Handbuch Maschinenethik (pp. 75–115). Springer VS. https://doi.org/10.1007/978-3-658-17484-2_6-1

Luhmann, N. (1966). *Recht und Automation in der öffentlichen Verwaltung. Eine verwaltungswissenschaftliche Untersuchung*. Duncker & Humblot.

Luhmann, N. (1989). *Ecological Communication* (J. Bednarz, Jr., Trans.). Chicago: University of Chicago Press.

Luhmann, N. (1995). *Funktion und Folgen formaler Organisation* (4th ed., epilogue 1994). Duncker & Humblot.

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. *AI Magazine, 27*(4), 12–14. https://doi.org/10.1609/aimag.v27i4.1904

Mittelstraß, J. (1978). Die Idee einer Mathesis universalis bei Descartes. *Perspektiven der Philosophie, 4*, 177–192. https://doi.org/10.5840/pdp1978412

Mori, M. (2012). The Uncanny Valley (K. F. MacDorman, & N. Kageki, Trans.). *IEEE Robotics and Automation Magazine, 19*(2), 98–100. https://doi.org/10.1109/MRA.2012.2192811 (Original work published 1970).

Nakamura, H. (1992). The Idea of Nature in the East in Comparison with the West. *GeoJournal, 26*(2), 113–128. https://doi.org/10.1007/BF00241205

Ovid. (1922). *Metamorphoses* (B. More, Trans.). Cornhill Publishing. https://www.perseus.tufts.edu/hopper

Reilly, K. (2011). *Automata and Mimesis on the Stage of Theatre History*. Palgrave Macmillan. https://doi.org/10.1057/9780230347540

Richtmeyer, U. (2020). René Descartes (1596–1650). In M. Heßler, & K. Liggieri (Eds.), Technikanthropologie. Handbuch für Wissenschaft und Studium (pp. 95–106). Nomos. https://doi.org/10.5771/9783845287959-95

Rushing, J. H., & Frentz, T. S. (1995). *Projecting the Shadow. The Cyborg Hero in American Film*. University of Chicago Press.

Schlich, T. (1998). Vom Golem zum Roboter – Der Traum vom künstlichen Menschen. In R. V. Dülmen (Ed.), *Erfindung des Menschen. Schöpfungsträume und Körperbilder 1500–2000* (pp. 543–557). Böhlau.

Schneider, B. (2003). Clothes for Automata. Patterns and Cards in Punch Card Weaving of the 18th Century with Special Consideration of the Loom of Jacques Vaucanson. *Technikgeschichte, 70*(3), 185–205. https://doi.org/10.5771/0040-117X-2003-3-185

Schuchter, P., Krobath, T., Heller, A., & Schmidt, T. (2021). Organisationsethik. Impulse für die Weiterentwicklung der Ethik im Gesundheitssystem. *Ethik in der Medizin, 33*(2), 243–256. https://doi.org/10.1007/s00481-020-00600-3

Shelley, C. (2024). Rossum's Universal Robots: A Technology Studies Perspective. *IEEE Technology and Society Magazine, 43*(2), 78–87. https://doi.org/10.1109/MTS.2024.3400033

Shibata, T., Wada, K., Ikeda, Y., & Sabanovic, S. (2009). Cross-cultural Studies on Subjective Evaluation of a Seal Robot. *Advanced Robotics, 23*(4), 443–458. https://doi.org/10.1163/156855309X408826

Singer, P. W. (2010a). *Wired for War. The Robotics Revolution and Conflict in the Twenty-first Century*. Penguin books.

Singer, P. W. (2010b). War of the Machines. *Scientific American, 303*(1), 56–63.

Snow, C. P. (1961). *The Two Cultures and the Scientific Revolution. The Rede Lecture 1959*. Cambridge University Press.

SPD Bundestagsfraktion–AG Abrüstung, Rüstungskontrolle und Nichtverbreitung. (2019, January 29). *Für ein Verbot Letaler Autonomer Waffensysteme* (Position paper). Retrieved February 12, 2024, from https://www.spdfraktion.de/system/files/documents/verbot_waffensysteme-positionspapier_spd-20190129.pdf

Spreen, D. (1998). *Tausch, Technik, Krieg. Die Geburt der Gesellschaft im technisch-medialen Apriori*. Argument.

Spreen, D. (2023). Lethal Autonomous Weapon Systems (LAWS). On the Ethics of Automation in the Military from the Perspective of Social Systems Theory. *Sõjateadlane (Estonian Journal of Military Studies)*, (21), 10–40. https://doi.org/10.15157/st.vi21.24177

Stollberg-Rilinger, B. (1986). *Der Staat als Maschine. Zur politischen Metaphorik des absoluten Fürstenstaats*. Duncker & Humblot.

Sutter, A. (1988). *Göttliche Maschinen. Die Automaten für Lebendiges bei Descartes, Leibniz, La Mettrie und Kant*. Athenäum.

Suvin, D. (1979). *Metamorphoses of Science Fiction. On the Poetics and History of a Literary Genre*. Yale University Press.

The Editor [Campbell, J. W.]. (1942). Science-fiction and War. *Astounding Science-Fiction, 29*(1), 6.

Thompson, E. P. (1967). Time, Work-discipline, and Industrial Capitalism. *Past & Present, 38*(1), 56–97. https://doi.org/10.1093/past/38.1.56

Treiber, H., & Steinert, H. (1980). *Die Fabrikation des zuverlässigen Menschen. Über die 'Wahlverwandtschaft' von Kloster- und Fabrikdisziplin*. Heinz Moos.

Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind, LIX*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

Weber, M. (1978). Economy and Society. *An Outline of Interpretive Sociology* (G. Roth, & C. Wittich, Eds.). University of California Press.

Weber, C. (2015, May 31). *Büroklammern bis zum Ende der Menschheit*. Süddeutsche Zeitung. https://www.sueddeutsche.de/wissen/kuenstliche-intelligenz-bueroklammern-bis-zum-ende-der-menschheit-1.2498719

Zeydel, E. H. (Ed.). (1955). *Goethe, the Lyrist. 100 Poems in New Translations facing the Originals, with a Biographical Introduction* (E. H. Zeydel, Trans.). University of North Carolina Press. https://catalog.hathitrust.org/Record/001030260

Zimmerli, W. C., & Wolf, S. (1994). Einleitung. In W. C. Zimmerli, & S. Wolf (Eds.), *Künstliche Intelligenz. Philosophische Probleme* (pp. 5–37). Reclam.

Cyborg Soldiers and Ethical Enhancement

Posthumanism, Human Dignity, and Military Ethics with Helmuth Plessner and Niklas Luhmann

Dierk Spreen

Abstract This chapter argues that the symbiosis of humans and modern technology ('cyborg') does not represent a transitional stage to a machine civilization in a transhumanist sense. In this respect, the following argumentation can be seen as a contribution to a posthuman debate. However, the perspective here differs from an understanding of posthumanism that, in the wake of a 'symmetrical anthropology,' mitigates fundamental differences in competence levels through symmetrization. In detail, the regulator model is presented first. Then the concept of human cyborgs is developed. Based on this, questions for posthumanist social theory are formulated and an idea of human dignity is explored. In the following chapters, the developed perspective is related to the military context. On the one hand, this should make it easier to understand what ethical enhancement can mean. On the other hand, ethical problems of posthuman social theory become evident. In addition to Plessner, the article also draws on Niklas Luhmann, but does not attempt to synthesize the two approaches. The focus is on understanding the phenomenon of the cyborg soldier, using both models as tools.

1 Machines in the Body

The word 'cyborg' is an acronym combining the terms 'cybernetics' and 'organism.' The 'cybernetic organism' or 'cyborg' was introduced at a military conference on space medicine in the context of the Cold War and the upcoming Space Race. The creators of the new wording were Manfred Clynes, chief research scientist at Rockland State psychiatric hospital in New York, and Nathan S. Kline, director of

D. Spreen (✉)
Institute for Organizational Communication, University of the Bundeswehr Munich (UniBwM), Neubiberg, Germany

Department of Business and Economics, Berlin School of Economics and Law (HWR), Berlin, Germany
e-mail: dierk.spreen@unibw.de; https://strategic-communication-management.de

K. Talves, D. Spreen (eds.), *Artificial Intelligence in Military Technology*, Artificial Intelligence, Simulation and Society 192,
https://doi.org/10.1007/978-3-031-95578-5_3

research at the same hospital and a specialist in therapeutic drugs (Kline, 2009). The corresponding publication appeared in 1960 in the journal *Astronautics* (Clynes & Kline, 1960) and had a considerable discursive impact, not least due to its wide reception in science fiction—a reception that initially put a dampener on the recognition of cybernetics as a meta-science in the USA and Great Britain, while it found resonance in the counterculture (Kline, 2009). But it is precisely this counterculture that is now seen as one of the sources of the new capitalist spirit (Boltanski & Chiapello, 2007) and of Californian ideology, which means that cybernetics has made its way after all (Barbrook & Cameron, 1995; Rosenberg, 2008).

Clynes and Kline's idea was to technically optimize and expand human biology in such a way that it would be possible for humans to live "*qua natura*" in space (Clynes & Kline, 1960, p. 27, emphasis in original). The body is artificially adapted to the conditions in space: It must, for example, be able to live without breathing. This 'new' body is an organic-technological overall system. It contains technological components integrated into its organic functional circuits that expand and improve the physical capabilities of the human being. The living body and technology synthesize.

At its core, the concept of a space cyborg applies an idea from the cyberneticist Norbert Wiener, who already pointed out in the 1950s that we "have modified our environment so radically that we must now modify ourselves in order to exist in this new environment" (Wiener, 1954, p. 46). In the case of the problem posed by Clynes and Kline, the "new environment" is the space outside the Earth, which can only be reached with technological assistance (Spreen, 2014a, pp. 99–107; 2022a).

Traces of the reception of the space cyborg can also be found in the philosophical anthropology[1] of Helmuth Plessner—for example, in a lecture given in 1966. In this lecture, Plessner deals with the question of inhumanity. He first discusses how to interpret alienation and "presumptuousness." By this he means the "abandonment of venerable standards" and the "ideals of 'humanity' and gentleness, of deference and noblesse" formed in the classical and romantic periods (Plessner, 1983b, p. 334). He concludes that the "Promethean and democratic world," which is subject to an imperative of productivity and optimization—he speaks of the "compulsion to manufacture"—, must apply different standards than the classical and romantic periods. The "supermen of science fiction have little to do with Nietzsche's superman, who was conceived as an ultimate, an end and a fulfilment" (Plessner, 1983b, p. 334). Every "societal transformation" is linked to a renunciation of traditional standards, but at the same time grants "the chance of rebirth" (Plessner, 1983b, p. 334). This is why society reacts with widespread acceptance of new technology:

[1] Following Joachim Fischer, philosophical anthropology is understood here as a "paradigm" (Fischer, 2022, p. 209), which starts from the "natural conditions of peculiar humanity" and introduces the "distinction between *hominitas* and *humanitas*" (Fischer, 2022, p. 210, emphasis in original). "*Hominitas* refers to the special human position in natural history, while *humanitas* refers to what human organisms make of this special position" (Fischer, 2022, p. 210, emphasis in original). Quotations from German references have been translated into English here.

> If it is said that astronauts' digestive tracts will be replaced by artificial organs in order to improve their ability to adapt to the conditions of long-duration interplanetary journeys, or that their metabolism will be kept low through periods of artificial hibernation, the public will at best react with amazement and admiration for the medical skills involved, but will not be offended by the idea that humans are to become machines. (Plessner, 1983b, p. 329).

Plessner thus perceives the acceptance of technology as a basic tendency in public opinion and links it to the modern type of society. Technology acceptance is quite widespread (Petersen, 2011; Renn, 1986, 2005). According to the 2021 Special Eurobarometer report, in the European Union "almost nine out of ten respondents (86 %) say that the overall impact [of science and technology on society] is positive" (European Commission, 2021, p. 90), with artificial intelligence and nuclear technology achieving the lowest positive approval ratings (European Commission, 2021, p. 95). This means, despite generally positive expectations, perceptions of danger or risk can vary significantly in relation to certain technologies (Beck, 1992; Luhmann, 1993). With regard to artificial intelligence, it is mainly the fear of "loss of control" that worries people (Ortiz, 2022), and the fear that jobs will be lost (European Commission, 2021, p. 140). Regarding cyborgs, there are concerns that "humans and their bodies are mutating into an animated artifact of the technical system" (Schneider, 2005, p. 372). The positive expectations in relation to technologies aimed at increasing brain performance and in relation to biotechnology and genetic engineering are also quite high (71% and 70%, respectively, European Commission, 2021, p. 95).

The cyborg idea, which received its social-theoretical knightly accolade at the latest with the *Manifesto for Cyborgs* written by the science historian and left-wing feminist Donna Haraway (1985), now reflects the increasing nearness of humans and technology in everyday and working life, with recourse to contemporary high-tech. This nearness is not merely accidental, but normal. Already in the age of industrialization, new social institutions took over the instrumental coding of the body. Educational and training institutions formed a societal interface between technology and the body. "Over the whole surface of contact between the body and the object it handles, power is introduced, fastening them to one another. It constitutes a body-weapon, body-tool, body-machine complex" (Foucault, 1979, p. 159). It is a matter of "synthesis" (Foucault, 1979, p. 159). Today, body-technology syntheses are becoming normal phenomena in the world of life and work.

The cyborg discourse in social theory, which follows Foucault and Haraway, focuses primarily on the bodily-material dimension and its subjectivation effects. The recourse to philosophical anthropology can prove to be a valuable addition, because it focuses on the specifically human form of the intentional and meaningful relations to the world, society, and the self. As a result, ethical aspects and human dignity as a guiding concept of intersubjective and societal communication can be discussed anew in the context of body–technology symbioses. With Plessner (and Luhmann), it is less about follow-up effects of technology in 'dispositives' and more about dealing with technology in social contexts, whereby the 'human side' does not appear as a mere appendage to technological and material arrangements.

2 Understanding Cyborgs

How is the increasingly condensed synthesis between the living human body and modern technology, which is addressed by the cyborg idea and the cyborg discourse, to be understood anthropologically? Are we, as cyborgs, in a transitional stage to a posthuman society, as many authors suggest? On the one hand, 'posthuman' can mean that material factors must be considered in social theory (Henkel, 2016). And there is no doubt that in the society of cyborgs, close-bodily and intra-bodily technology will be a subject of discourse.

On the other hand, 'posthumanism' could also mean that the cyborg transformation indicates the transition to a transhuman society. With the transhumanist Hans Moravec, cyborgization could be interpreted as a transitional stage into a machine civilization. Moravec describes this process as one that transforms human beings into artificial intelligences. Initially, the body is increasingly replaced by artificial systems, and eventually the brain as well, so that people then only exist as software. To improve functionality, the third step is to optimize this software with programs that are adapted to disembodied cyberspace, so that the machine man can do without a representation of a body (Moravec, 1993, pp. 81–89; de Mul, 2010, S. 248–251; Spreen, 2024).

Following Helmuth Plessner and philosophical anthropology, this paper will argue that the close-bodily and intra-bodily synthesis of human beings and modern technologies *does not* represent a transitional stage to a machine civilization in a transhumanist sense. Rather, it can be argued with Plessner that human cyborgs *are still human beings*. To grasp the close-bodily and intra-bodily technology, the regulator model is presented. This states that it is not possible to specify a definitive threshold that, if transgressed, would turn a human being into a cyborg. Rather, we have already entered such close syntheses with modern technology in our lifeworlds that Donna Haraway's provocative dictum that "we are all [...] theorized and fabricated hybrids of machine and organism"—or "we are cyborgs" for short—must be understood as a description of reality (Haraway, 1985, p. 66). In relation to this continuum of close syntheses of body and technology, we will speak of 'bioartificial symbioses' in the following. According to the cultural sociologist Wolfgang Eßbach, this should be understood as artifact or technology relationships in which the technology used "cannot be put down again like tools of the trade. The instrumental character becomes weaker here. A multitude of modern artifacts have become more the reason than the means of our lives" (Eßbach, 2011, p. 73).

However, this makes it clear that technology must be taken seriously in terms of social theory. In this respect, this chapter can be seen as a contribution to a posthuman debate—albeit one that locates cyborgization within the framework of what is possible for humans. The term 'posthuman' then means little more than thinking about the relationship between human beings and machines. Philosophical anthropology is interesting here because it fundamentally reflects on the comparisons between human/animal, human/human (ethnology), and human/technology (Fischer, 2019). But it differs from 'symmetrical anthropology' in the tradition of Bruno

Latour (1993) because it encounters fundamental differences in levels of competence or complexity in these comparisons and does not soften them through symmetrization. Regarding the animal/human comparison, this becomes clear through the determination of different positionalities in relation to the environment, which for humans becomes a 'world' that they can objectively grasp, change, and use.

After presenting the regulator model, the concept of human cyborgs is developed following Plessner. Based on this, questions for posthuman social theory are formulated, and the concept of human dignity is developed. In the following chapters, the perspective developed is related to the military context. On the one hand, this should make it easier to understand what ethical enhancement can mean. On the other hand, ethical problems of posthuman social theory become apparent.

In addition to Plessner, Niklas Luhmann is also consulted to understand the phenomenon of cyborgs in society and in the military. There are several reasons for this. Firstly, there are parallels in the concept of human dignity between the two authors. Secondly, they can be read as mutually complementary because each of them focuses on aspects that the other does not primarily address. Plessner focuses on the bodily side of human actors, while Luhmann on the self-reference of society and communication, which is helpful for the analysis of human–machine cooperation. Both sides bring benefits for the analysis of the cyborg phenomenon. The third reason is that a 'posthuman' synthesis of both theories is being discussed; this will be critically questioned in an excursus. In this chapter, however, a social-theoretical synthesis of both approaches is avoided. The focus lies on an appropriate understanding of the phenomenon, using Plessner's philosophical anthropology and Luhmann's sociological system theory as tools.

3 The Regulator Model

The guiding principle of technology development in the cyborg sector is the neurochip or biochip implant (Schneider, 2005, pp. 384–385). This refers to a connection between nerves and electrical conductors that is comparable to a 'soldered' connection. This vision aims to develop a technology that allows lossless and frictionless two-way translations between organic and digital information processing (Bothe & Engel, 1993, p. 175). By means of such a technology, it should be possible, for example, to translate sensor information into sensual information. Neural prostheses no longer have to be seen "as technical appendages" because they become "part of the patient" through "neuronal integration" (Bothe & Engel, 1993, p. 176). According to the model of the neurochip cyborg technology is a kind of technology that adapts to the skin or even 'amalgamates' with the body. The guiding idea of the neurochip also illustrates that cyborg technology is often highly advanced information processing technology. This chapter, therefore, limits the cyborg-term to technical-organic hybrid bodies that operate in a highly artificial modern society. If we follow Armin Nassehi and sociological systems theory, then "the reference problem of digitalization is […] rooted in the social structure of modern society" (Nassehi,

2024, p. 125) because its functionally differentiated communication has a binary structure. The context of Clynes and Kline's creation of the c-term—i.e., space travel—also speaks for a restriction of cyborgization to modernity. Space travel as a technically feasible project became imaginable only since the second half of the nineteenth century.

The inclusion of close-bodily technologies in the concept makes sense if they function in a similar way as intra-bodily cyborg systems. Examples would be cybernetic gloves or non-invasive methods of reading nerve impulses. Prostheses are being developed which provide sensory feedback in real time by means of nerve stimulation and approach a "life-like" quality (Raspopovic et al., 2014). Research is also being carried out on non-invasive exo-gloves (In et al., 2015) or flexible exoskeletons (Shein, 2019). These are technical organ extensions that can be put on or pulled over and then follow movements. Control and regulation algorithms can be used to support movement sequences (Grote, 2012). Not only are such organ extensions helpful for therapeutic purposes, but they can also improve strength and endurance or be used to control avatars. "An exoskeleton is a mechanical device or soft material worn by a patient/operator, whose structure mirrors the skeletal structure of the operator's limbs (joints, muscles, etc.). The structure works in tandem with the person wearing it, and it is utilized to amplify their capabilities, serving as an assistive device, haptic controller, or for rehabilitation purposes" (Shein, 2019, p. 14).

Bioartificial symbioses are no longer mere tools, kitchen utensils, or industrial machines; they are now more closely related to the human body. They cling to the body or become 'intimate' aspects of the body. "[M]achines can be prosthetic devices; intimate components, friendly selves" (Haraway, 1985, p. 97). At the same time, prostheses, implants, exo-gloves, and the like are no longer regarded as mere therapy support. Rather, they are integrated into the imperatives and discourses of an "upgrade culture" (Spreen, 2015a) aimed at optimization, as the example of exoskeletons makes clear. Prostheses increasingly appear as an improvement to an inherently deficient body that can and should be optimized. There is now "a continuum of bodies that are enhancable and worthy of improvement" (Harrasser, 2013, p. 95).

This extension of the cyborg concept makes sense because close-bodily systems also form an integrated overall system together with the living body. In contrast to implants with a neurointerface, however, it is possible to temporarily discard such close-bodily systems. However, as this means that opportunities for optimization and enhancement are lost, such disarmament results in a reduction in access to the world. Compared to the enhanced cyborg, the mere bio-body is transformed into a "deficient being."[2]

Nomophobia would be a common example of psychological effects that could result from the removal of symbiotic technology. No-Mobile-Phone-Phobia

[2] The sociologist Arnold Gehlen, who belongs to the field of philosophical anthropology, describes humans as "deficient beings" in "comparison to animals" (Gehlen, 1988, pp. 13, 26).

"describes the fear of being disconnected and unavailable without one's smartphone. This can happen especially when the battery is empty, there is no network reception, or the smartphone has been forgotten" (Coenen & Görlich, 2022, p. 2). Nomophobia describes experiences that can be associated with the removal or malfunction of a bioartificial symbiosis that is commonplace today—the cell phone. The phenomenon is described in the corresponding psychological tests by four dimensions: "not being able to communicate, losing connectedness, not being able to access information and giving up convenience" (Yildirim & Correia, 2015, p. 130). These four dimensions describe decreases in access to the world.

The decrease in access to the world that human cyborgs can experience becomes even clearer if we take Neil Harbisson's "eyeborg" as an example (Harbisson, 2012). The color-blind musician wears a spectral sensor on a wire curved over his head that converts color information into sound. The sensor is firmly connected to the skull bone at the back of the head and creates a soundscape of the world drawn in gray shades that Harbisson's eyes can perceive. According to Harbisson, the "cybernetic device" is no longer a "device" in the usual sense. "It had become a part of my body, an extension of my senses." The implanted sensor has become a part of Harbisson's body and an extension of his senses to an extent that he can dream "in colors," because software and brain "had united." In Harbisson's case, not only does the implanted technology prompt a new form of self-description—as a "cyborg,"[3]—but also shows the transition between medical repair and optimizing expansion. The eyeborg not only makes it possible to compensate for a limitation. Rather, this extension can also translate ultraviolet and infrared into sounds. Consequently, Harbisson advocates an optimization of the senses that goes beyond the scope of human nature. This means that environmental information that otherwise remains hidden from people in their natural environment can now also be processed. Medical compensation has become an optimizing extension. On the other hand, if the eyeborg malfunctions or is removed, access to the world would be restricted.

To capture the body-related mechanization process reflected in the cyborg discourse, the regulator model has been proposed (Spreen, 2014b, 2015a, pp. 27–38). This model intends to describe processes of the embodiment of technology without focusing on the transgression of the boundary of the skin. Rather, the connection between the technical pervasion of the close-bodily space and of the communication sphere should be kept in mind. Cyborgization can thus be described as shifting a regulator on a continuous scale of mechanization. Accordingly, we can speak of the 'cyborgization of the body' when technology enters an 'intimate' functional relationship with the organism, i.e., when it combines with the body close to or under the skin to form an extended body system. Corporeality becomes organic-technical. Specific bioartificial symbioses may well be limited in time, as the technological range is constantly being expanded and improved according to the logic of innovation (Arthur, 2009), and implants or exo-organs can therefore be replaced by new and better ones (Fig. 1).

[3] Harbisson was able to ensure that his passport photo shows him with the extension.

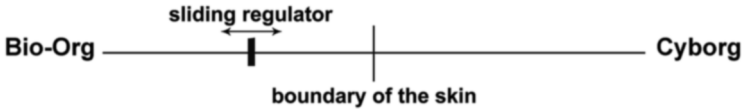

Fig. 1 Sketch of the regulator model (*source*: own visualization)

On the one hand, the regulator model is intended to illustrate that the modern self and its social relations are to be understood in relation to material technologies and technical media. On the other hand, the model makes it possible to mark the new that is taking place by way of the mechanization of human body. There is no doubt that the systematic development of close-bodily and intra-body cybernetic technologies represents a cultural innovation that can be traced back to impressive advances in the life sciences and computer technology. Human use of technology in itself does not justify talking about 'cyborgs.' What would be the significance of an expression that, under a spectacular label, merely restates the old insight that "technical activity" must be counted among the "human constitutional characteristics?" (Gehlen, 1961, p. 95).

The regulator model describes a scale of possible cyborgizations and sees these in turn in connection with the development of technology and media. However, it avoids a definitive criterion that distinguishes between humans and cyborgs. There are two reasons for this: Firstly, cyborgs are not at all a new, non-human species, as long as we are talking about human-related bioartificial symbioses. Secondly, it is difficult, if not impossible, to specify a definitive criterion for becoming a cyborg (Heilinger & Müller, 2007, p. 25).

4 Human Cyborgs

The fact that cyborgization represents something new invites speculation and promises that herald a new age and a new society. Cyborgization is interpreted as a sign of the dawn of a 'posthuman' future, in which a new stage of evolution and an age 'after humankind' is asserted either with euphoric approval or with cultural-critical gestures. Based on the regulator model represented here, such assertions appear unconvincing. Rather, by drawing on Helmuth Plessner's philosophical anthropology which is fully compatible with this model, it is easy to show that cyborgization does not go beyond the possibilities of human civilization and societal association. Not only do human cyborgs *have* technical means, i.e., means that they could stop using. Rather, they *are also* technology. People as cyborgs are characterized by an "also-being-technology" (Spreen, 2015a, p. 34).

Following Plessner, the human beings can be defined topologically, i.e., by describing their position in the world. Plessner outlines a bodily relationship to the world that is always already beyond the body; it has its center outside the body, and

therefore he speaks of the "excentric positionality" of the human being. This version of the human being is not only compatible with the technical development of external spaces and environments, but also with technical access to the inner space of the body (Fischer, 2002, pp. 236–239). As a natural-artificial living being, humans are beings that have always been able to leave the framework of the biological and inhabit a world of culture, history, art, technology, and language. Why should the human potential to open up to the world and make it accessible stop at one's own body? The anthropology of openness to the world can follow human beings' gestalt-switch respectively metamorphosis and thus also the mechanization and transformation of the human body. In his main work, published in 1928, Plessner (2019, p. 272) writes:

> Being human is not tied to any particular gestalt and [...] could just as well take on a variety of gestalts that do not correspond with our own. The human is tied only to the centralized form of organization, which forms the basis of his excentricity.

And what about the problem of the attributability of actions ('agency') discussed in the context of the mechanization of the body? Programmed reactions carried out by a cyborg organ controlled by neurochip interfaces or similar systems raise the question of agency. Are actions in a high-tech "artificial society" (Popitz, 1995), in which vital bodies and technology synthesize, still attributable to human actors? Or are actions increasingly becoming "an externally controlled process" (Zoglauer, 2003), which can or even must be attributed to cyborg extensions in societal communication?

From a Plessnerian perspective, this problem does not appear too serious, as the eccentric positionality of humans does not imply absolute autonomy of will and sovereignty anyway. Unbound decision-making sovereignty is an anthropological illusion. Located between being a body and having a body, people are necessarily confronted with the possibility that their body slips out of their control. People can laugh and cry. According to Plessner, this is an "autonomous process" that is characterized by a "loss of control" (Plessner, 1970, p. 65). He also calls this a "catastrophe" (Plessner, 1970, p. 66), but this is not why a person suddenly becomes non-human. This is because such a loss of control is perceived as an appropriate "expression of [...] and answer to" the situation in question (Plessner, 1970, p. 66).

With Plessner, the derailment of the body can also be understood as a constitutive condition for the creation of new meaning: the "opacity" and "meaninglessness" (Plessner, 1970, p. 68) of bodily expressions demand interpretation, attribution of meaning, discourse, and discipline. Such experiences therefore play an important role in the development of the self. Total sovereignty over the will is not a characteristic of human existence. Experiences of self-surprise are genuinely human and constitutive of the self.

So, when human cyborgs are irritated by the reactions of their 'smart' extensions, is something taking place that lies outside the horizon of human experience? Rather, following Plessner, it can be assumed that experiences with techno-bodily expressions can be mediated by discourses, so that these experiences can be put into words, produce meanings, and thus be culturally and discursively absorbed. For example,

they can become an occasion to optimize the corresponding technologies. Modern high technology does not enter the world "perfectly," as Georg Friedrich Jünger (1953) assumed, but requires constant technological readjustment due to the "growth of causal complexity" (Luhmann, 1993, p. 91).

In addition to vital forms of expression, technical cyborg organs can also irritate consciousness. An AI support of these organs could even communicate language-based information. In the context of "artificial communication" (Esposito, 2022), technically generated but textual or verbal interpretations of the irritation would be provided and thus support the emergence of a discourse on the problem.

As long as this excentric positionality, i.e., the simultaneous openness to the world and being bound to it, is not eliminated by internal mechanization, cyborgs will therefore remain humans—regardless of their organ configuration or gestalt. Unlike cyborg philosopher Chris Hables Gray, for example, claims, cyborgs are not a "successor species" (Gray, 2001, p. 2). However, if humans were to be genetically reprogrammed in such a way that they were controlled by programmed behavioral instincts—for example, by inscribing them with a peace gene that bans aggression from their behavioral repertoire—, then we would indeed be dealing with a posthuman species. It would no longer have an openness to the world or freedom of choice. Ethically, such restrictive transformations could not be justified because they "no longer promote human openness, but undermine it" (Müller, 2010, p. 194).

For Plessner, the human being is an open question (Plessner, 2018, pp. 39–46), who enters a "relation of indeterminacy with itself" (Plessner, 2018, p. 74). As a result, the human being "is empowered *to itself*," and "finds itself to be responsible or free" (Plessner, 2018, p. 60, emphasis in original). If cyborg enhancements such as an *Eyeborg*, an exo-glove, or a cell phone with its apps do not systematically restrict this openness and indeterminacy relationship, they should be ethically examined from a philosophical-anthropological perspective, but not rejected across the board (Heilinger, 2010, esp. pp. 266–284). One example of a problematic use is the bubble phenomenon, a communicative self-restriction in the context of the use of social media. The formation of closed virtual communities, however, does not contradict the excentric positionality—the "hybrid nature of man" (Plessner, 1983a, p. 187)—which is to be understood as an entanglement of community and society, body-being and body-having, vital being and person, being bound to the environment and being open to the world (Plessner, 1983a, p. 182). The formation of relatively closed communication communities is indeed an option in the building of social relationships. However, the use of technology and language already indicates the potential for crossing such communicative boundaries (Plessner, 1983a, p. 187; 1999, p. 80).

The conclusion that can be drawn here is that processes of cyborgization do not abolish 'the human being.' On the contrary, the transformation of humans into cyborgs is not per se incompatible with human openness to the world. The use of symbiotic high-tech does not necessarily lead to a post- or transhuman future, but merely to a society that must deal communicatively with the surprises that these technologies hold in store. And such surprises are sure to occur, because the very idea that high-tech processes could be completely controllable quickly proves naive on closer inspection (cf. Luhmann, 1993, pp. 90–95).

5 Excursus: Questions for Posthuman Social Theory

Plessner's observations on the significance of the human being's vital-bodily mode of existence for the constitution of meaning are important in the context of the posthumanism debate. This discussion emphasizes the importance of "materialities." For example, Anna Henkel has proposed to include "corporealized meaning as a medium of meaning" in communication and systems theory (Henkel, 2016, pp. 11–14; Henkel, 2017, pp. 283–286).

Plessner pointed out that expressive or action behavior can be understood without the need for verbal explication (Plessner & Buytendijk, 1982). Consider, for example, the 'doubtful look' of conference participants or the hesitant attention and subsequent flight of an animal. According to Henkel, when the body 'expresses' itself in the mode of irritation, it approaches a negating or at least not simply agreeing action. Henkel identifies such "doing negativity" as the core of the concept of meaning, which, in her opinion, does not necessarily have to be coded linguistically (Henkel, 2016, pp. 4–5). However, such irritations cannot replace the verbal communication on which society and societal complexity are essentially based. Physical or material irritations can only motivate discourses, insofar as they are observed and addressed (Spreen, 2022b).

Henkel blames Luhmann for being fixated on linguistic coding and "shortcomings of a language-centered view of materiality that is dominant in systems theory" (Henkel, 2016, p. 6). In contrast, she argues for a concept of "meaning beyond human understanding" (Henkel, 2016, p. 1). However, Luhmann does not claim that processes of understanding in consciousness must necessarily be based on language. He only claims that such understanding takes place in the medium of meaning, whereby language plays a prominent role (Luhmann, 2012, pp. 123–124). Plessner takes a similar view (Plessner & Buytendijk, 1982, esp. pp. 92–93). Finally, material objects or bodily information can also be understood or grasped in a meaningful way. Laughter, crying, pain, or other bodily impulses become noticeable in consciousness. Sometimes they impose themselves; sometimes they remain on the edge of the perception threshold. And buildings, works of art, or the text output of algorithms can also be understood, although we are not always sure in what sense. The fact that there are non-verbal forms of communication (e.g., gestures) in addition to language is not a problem for Luhmann, as these also operate in the medium of meaning. Does social theory therefore require the living body "as another medium of meaning" (Henkel, 2016, p. 20) to capture "not only visual" forms of meaning but also those of an "olfactory, tactile, auditory or gustatory nature?" (Henkel, 2017, p. 285) After all, these are all also perceptions. But perceptions are—at least according to Luhmann *and* Plessner (Luhmann, 2005, p. 45; Plessner & Buytendijk, 1982, p. 120)—operations of consciousness.

In other words, if living bodies or other materialities shall be adequately considered in social theory, then it will by no means be necessary to promote them to meaning-processing systems which operate as a kind of 'second consciousness.' Rather, a body "can irritate neither consciousness nor social systems other than as a

signified body (observed by these systems)" (Fuchs, 2005, pp. 52–53). But through such irritation vital body is in play.

> Words mean, while sounds can at best express a particular state and thus signal it. A cry, a whoop, sobs and groans, gurgles and grunts belong to states and situations, and therefore have meaning, but they do not carry it. Their infectious power (think of laughing, crying, yawning, coughing) conveys nothing. (Plessner, 1983a, p. 173)

Bodily expressions *have* meaning, but they *do not carry it*. This means that they are not suitable for building up systemic autopoiesis and societal complexity. Corporealized meaning can therefore not be understood as basic medium for a communication that includes metacommunication (Luhmann, 2012, pp. 123, 124). Rather, the functionality of "purely organic selection" is quite limited in contrast to conscious experience, because organic selectivity can neither preserve the complexity of other possibilities nor the contingency of experience (and action) (Luhmann, 1971, p. 33). Luhmann explicitly points out that negation is a "peculiarly human [...] ability" (Luhmann, 1971, p. 35). From the perspective of philosophical anthropology, this also quickly becomes understandable. According to the co-founder of philosophical anthropology Max Scheler, "the human being is the 'Nay-sayer'" (Scheler, 2009, p. 49). In this, human beings differ from plants or animals. The human distances itself from its environment and objectifies it (Scheler, 2009, pp. 27–29). The use of language as a communication medium is of fundamental importance for this (Plessner, 1983a, pp. 180–189; Luhmann, 2012, p. 123).

What is the purpose of the leveling of differences in competence between consciousness, the organic embodiment, and/or technology? If the vital body, which in a sense represents the centric positional form of animals in humans and which as such must be reflected upon and recognized in social theory, is promoted to an equal medium of meaning on the same level as human consciousness, are other living or non-living 'entities' to be treated on the same or a similar level as human communication partners? Do things, artifacts, artificial intelligences, living beings, humans, or pharmaceuticals in their original packaging, which are "observed as 'actors'" (Henkel, 2017, p. 288), together form the social network?—There are considerable differences in competence between these various 'entities,' which should be considered, but which threaten to disappear in a 'symmetrical' theoretical system. AIs, for example, can collect data using sensors, i.e., 'observe' in a way. They can evaluate prompts and link to them communicatively, but they do not understand what they are doing. An algorithm "does not understand content, meaning, or interpretation. It deals only with data" (Esposito, 2022, p. 6). Pets can be communicatively included in certain social systems (e.g., in a family or an interaction system), because (limited) interspecies communication is certainly possible (Plessner & Buytendijk, 1982). However, due to their centric positionality, animals are constrained in their capacity for linguistic communication, comprehension of the technical environment, and the ability to gain societal complexity through historical change. For example, they would never send representatives to a parliament, demonstrate for their rights on the street, or allow themselves to be mobilized for a war based on intrinsic motivation.

But if a posthuman society is to consist of a network of different 'actants' and if human and non-human agency are to be placed in a symmetrical relationship, what about human dignity?

6 Dignity in the Cyborg Society

Beyond "treatment that is clearly contrary to human rights," there is no consensus on "which ways of dealing with people by public authorities or private powers are unbearable and humiliating" (Frankenberg, 2003, p. 277). It is not so easy to say in detail what constitutes a violation of human dignity and what does not. This has led to trivializations and strange juridical effects (Frankenberg, 2003, pp. 272–274, 280–282). Regarding the problem of the cyborgization of human beings, it may be helpful to ask how dignity can be conceptualized in light of the work of Helmuth Plessner. Given the similarities with Niklas Luhmann's approach, his concept of human dignity should also be included in the considerations.

Plessner did not explicitly develop a concept of dignity, so that it must be (creatively) reconstructed (Schumann, 2019, p. 29, n. 90). In his early critique of social radicalism from 1924, however, it becomes clear (Plessner, 1999, pp. 103–127) that dignity is endangered when the unfathomability of the person is removed, when the protective mask is torn away through "invasions of privacy" (Goffman, 1986, p. 24) or one's own expressive behavior. Invasive communication violates dignity because it aims to drag everything into the light, make the person transparent, and 'dissolve' the open question. However, you can also expose yourself, for example, through an embarrassing mishap, a 'Freudian' slip of the tongue, pure affect, or over-the-top spontaneity. Because such situations are usually accompanied by shame, people try to avoid such behavior. However, this cannot always succeed; the social "risk of ridicule" remains unavoidable, as Plessner realistically states (Plessner, 1999, pp. 117, 122–123). Tolerance for the ambivalences and ambiguities within one's own behavior therefore is required to maintain self-image and ego strength. In such cases, the social environment would do well to overlook the incident so as not to reinforce the dissolution of the protective shield. This is then a question of tact (Luhmann, 1965, p. 67; Plessner, 1999, pp. 161–169). Plessner thus understands "human dignity" as a kind of coherence or harmony "between soul and expression, soul and objectual body" (Plessner, 1999, p. 123). For Plessner, dignity is an "ideal constitution to which one strives" (Plessner, 1999, p. 123)—for Luhmann, very similarly a "concept of desire" (Luhmann, 1965, p. 68)—which is linked to a certain "imperative to acquire form" (Plessner, 1999, p. 119), by which is meant tactful behavior "under careful obedience to distance" (Plessner, 1999, p. 166). This involves "an art of not coming too close and of not being too open" (Plessner, 1999, p. 162), i.e., communicative forms that help to save face (Plessner, 1999, p. 161).

Plessner's remarks refer to the social context. He therefore advocates an "ethos of the respect for human dignity" (Plessner, 1999, p. 158). Respect for human dignity is expressed as the right to be respected as a personality in one's "personal way

of life" (Schumann, 2019, p. 36) on the one hand and the right to be "clothed with form" (Plessner, 1999, p. 119) on the other. Plessner conceives the concept of dignity around the ambivalence and ambiguity of the soul or psychic—the need for "validity" on the one hand and for "modesty" on the other (Plessner, 1999, p. 109). "We want ourselves to be seen and to have been seen as we are; and we want just as much to veil ourselves and remain unknown, for behind every determination of our being lies dormant the unspoken possibility of being different" (Plessner, 1999, p. 109). Human dignity is a kind of "real symbol [...] of the balance of the various demands that the individual person must more or less make of themselves for their self-presentation in society," but which is "also due to them without personal achievements per se" (Fischer, 2022, p. 219).

From this point of view, human dignity is not a matter of nature, but only comes about because the shared world treats people "in such a way that they can *thereby* become and 'remain' the 'full-valued actors' that were anticipated in them" (Schumann, 2019, p. 36, emphasis in original). Conversely, people who are addressed with dignity can more easily learn to address others with respect to their dignity.

Luhmann understands dignity as the possibility of self-presentation, which in modern society is released by communication because it has the function of protecting humans as 'individuals.' Under the conditions of functional differentiation, "every human being" is "expected to be able to relate its actions to several social systems and to unite their unbalanced requirements in a personal synthesis of behavior" (Luhmann, 1965, p. 53). This ability to synthesize behavior makes "the individual functionally important for the structural coordination of the social order" (Luhmann, 1965, p. 49), which is why the individual must be granted "special protection" through "fundamental rights" (Luhmann, 1965, p. 52). "Freedom and dignity [...] denote basic conditions for the success of a person's self-presentation as an individual personality. [...] Self-conscious individuality is only gained by presenting oneself" (Luhmann, 1965, p. 61). Luhmann discusses both the problem of self-endangerment of dignity "through inconsistent and therefore embarrassing information" and the problem of external endangerment of dignity through invasive communication (Luhmann, 1965, p. 67). Self-presentation and dignity do not arise by themselves, are not simply there, but are socially constituted: "Self-representation is the process that allows the human being to become a person in communication with others and thus constitutes it in its humanity. Without success in self-presentation, without dignity, it cannot utilize its personality" (Luhmann, 1965, p. 69). Luhmann pursues a functional definition of dignity. It includes a representational aspect aimed at social recognition and validity as well as an aspect aimed at restraint and respect for the "intimate sphere" (Luhmann, 1965, p. 67).

Both Plessner (philosophical anthropology) and Luhmann (sociological systems theory) emphasize that dignity emerges in social communication and interaction and can therefore also be violated through communication and interaction. If people enter a close relationship with technology in Haraway's sense, this relationship, as can be shown with reference to Plessner, remains within the scope of the horizon of possibilities of excentric positionality, so that people's ability to express themselves

and communicate is not fundamentally limited by their cyborgization. Therefore, based on both Plessner and Luhmann, it can be said that if people have the opportunity and are given the opportunity to express and assert their dignity in communication, the use of technologies, however close they may be to the body, cannot be considered a fundamental violation of human dignity. This would only be the case if such technologies were to remove the inner protective sphere. As with any technology that collects and stores data, this would also be conceivable and feasible with technologies that are close to the body. The use of (cyborg) technology that deliberately spies on people's privacy and intimacy and thus communicates indiscretion violates human dignity. It creates a "displeasure in being exposed" (Goffman, 1986, p. 24). However, this does not concern the technology as such, but the way in which it is used. Moreover, being exposed is a communicative power technique that is already present in verbal interaction—for example, as "morbid curiosity" (Goffman, 1986, p. 24).

If human beings remain 'human' even when they become 'intimate' with technology, an 'intimacy' which is likely to be the case for many people today (cochlear implants, pacemakers, prostheses—and in the not-so-distant future probably brain–computer interfaces such as *Neuralink*, etc.), then human cyborgs also have dignity. Against the social-theoretical background discussed here, two aspects are particularly important: (1) Human cyborgs also have the right to 'clothing with form.' Cyborg technologies should not be misused to make people 'transparent.'[4] (2) Human cyborgs must be able to present their self through communication in their own individual way.

To what extent is the verve of endowing the technical additions, enhancements, and prostheses with "agency" (Escobar, 1994, p. 216) and having the human being in the cyborg perform "a nightmarish danse macabre" (Sobchack, 2004, p. 212) as a puppet of techno-actants compatible with the concept of human dignity? If the technical parts in the human cyborg appear as powerful 'actants,' is it not then denied that cyborgs can represent a 'self' in communication at all? After all, such a self-presentation presupposes a certain personal identity on the part of the self-presenter, because "without some level of unitary identity, such a cyborg will not have the ability to act coherently" (Gray, 2001, p. 27). But what can 'human dignity' then mean in the context of symbiotic technologies to which *agency* is attributed?

Media scientist Vivian Sobchack, who uses an intelligent prosthetic leg, complains unequivocally about the widespread technological animism and fetishism. Sobchack is not an anti-technologist. She is not seeking to purge social theory of machines. But she does insist that the experiences and feelings of prosthesis users and the reactions of their social environments should be placed at the center of cultural and technological analyses. She herself, for example, is not interested in post-human enhancement, but simply in "a leg to stand on" (Sobchack, 2004, p. 225, no fetishism). A good prosthesis is also one that disappears in everyday life, not one

[4] Gray (2001, p. 28) also discusses this aspect in his "Cyborg Bill of Rights" under the points "The Right of Electronic Privacy" and "Freedom of Consciousness."

that governs people and takes over control (no animism). There is a tendency in posthuman theory to declare people who use cyborg technology to be a function of their technology. But if we follow this posthuman tendency to the end, are we not then impugning the ability of human cyborgs, or even humans acting in highly technological environments, to present themselves as a fully-valued self?

Ideally, cyborg technology expands access to the world. Its withdrawal can be perceived as a loss. However, it is not the technology that has such an experience and communicates these experiences. Rather, only the humans inside the cyborg can do this. Even technologies capable of speech *cannot understand*. Unlike consciousnesses and communication, they do not operate in the medium of meaning (Esposito, 2022). Even AI language models are only sophisticated aggregations and interconnections of trivial machines (Baecker, 2023, p. 254). Therefore, talking AIs cannot be recognized as equal discourse partners.

In a nutshell: The perspective taken here on *cyborgs as humans* sees no need for cyborgs to experience themselves as posthuman appendages of their artificial organs, with which they are 'intimately' networked. Rather, they can continue to perceive themselves as actors who present themselves communicatively by way of their actions. From the perspective proposed here, they *should*–ethically speaking–indeed do exactly this. Of great importance here are the skills to deal with cybernetic-organic irritations/self-surprises, technological risks, and the enhancement of access to the world. Whereas, if you extend agency indeterminately to all possible 'posthuman entities' and conceptualize society as a network and chain of these actors or 'actants,' then you run into the problem that differences in competence or positional gradations between plants, animals, and humans as well as between machines, artificially intelligent automation, and human consciousness become blurred. This is because not all of these entities' have the same level of communication skills. In the end, it will only be "augmented humans"[5] who are able to speak for them.

Particularly in the case of organizationally bound actions and decisions, accountability for technical operations is in the hands of individuals in social roles. In contrast, artificial 'actants' cannot be held accountable. "If accountability is to be ensured organizationally [...], this presupposes not only a precisely formulated system of expectations and assessment standards, but also the clear attributability of performance and errors. Accountability is necessarily individually related and thus distributed within the system" (Luhmann, 1995, p. 180). In military organizations, this applies even more so (Luhmann, 1995, p. 176; Spreen, 2023, pp. 23–26; Koch et al., 2024). Neither the roles of communicative representatives nor the roles of accountability should be concealed. From an ethical point of view, concealing these roles would mean that decisions taken by humans would no longer be attributed to them, which could lead to an anomic state of society. Any representative or person in charge can wash their hands in innocence then. "Automation is no more an excuse [...] than human failure" (Luhmann, 1966, p. 81). Accountability and the "liability for errors" cannot be abolished; without them, "everyone could act as he pleases"

[5] "Augmented humans" is an apt term for human cyborgs used by Australian SF author Joel Shepherd (2006).

(Luhmann 1966, p. 113). From a social-theoretical point of view, such a concealment of authorship of decisions and actions would simply be what has been called 'reification' in critical theory since Marx.

7 Cyborg Soldiers

To overcome the proverbial fog of war, work has been done on optimizing networked operations management. The aim is to obtain a shared real-time situational picture of the combat area in all spatial dimensions (ground, water, air, space, cyber) by networking the command, control, communication, intelligence, and impact levels to faster and better coordinate the forces and capabilities involved. Ideally, this situational picture would be accessible down to the level of the soldiers operating in the battle to provide them with information that extends beyond the field of perception of their own sensors and senses. This concept is also known as *Network Centric Warfare* (NCW). On the one hand, NCW allows one's own forces to locate enemy troops, even when they are not directly observable. On the other hand, better coordination and considerable acceleration of one's own operations are to be achieved. The aim is to retain the initiative and overwhelm the enemy's military organization. To a certain extent, one wants to fight the opponent "dizzy" (Lange, 2004, p. 7).

To integrate soldiers into NCW, the soldier's body and the digital, AI-enhanced communications network must be structurally coupled. This can be achieved through combat suit systems. Examples of such combat suit systems include the *Infantryman of the Future–Expanded System* (System IdZ[6]) from the German armed forces, the successor system *Gladius 2.0* from *Rheinmetall*, and the *Tactical Assault Light Operator Suit* (TALOS) currently under development for the US Army. The latter is an intelligent combat suit system including armor and exoskeleton support. The former provides the soldier's body with protective equipment and an electronic back. The senses are enhanced by a control and display unit and an alternative helmet display. Orientation is provided by an integrated GPS. Networking is established via a headset connected to the squad leader's command radio. New features compared to the basic system include the ability to exchange voice and data with the next higher command level and the connection to the Army's command information system (Planungsamt der Bundeswehr, 2013).

Such networked combat suit systems represent a form of human enhancement because they aim to improve performance and perception. For example, a *Gladius 2.0* soldier can look behind objects by accessing a real-time digital representation of the battlefield, which is calculated from a variety of perspectives. These perspectives can come from other soldiers, from drones, or from further reconnaissance sensors and are by no means limited to the visible light spectrum.

A few years ago, in the context of the paradigm shift to networked operations, observers still expected flat hierarchies and a collective self-adjustment of combat

[6] Infanterist der Zukunft—Erweitertes System.

units in line with the situation (Kaufmann, 2006). However, NCW and networked operations management have also increased the scope for higher levels of command and accountability to intervene at the immediate operational level (Singer, 2010a, p. 349; Warburg, 2008, p. 311). The extended control options mean that in the context of politically sensitive operations, even the civilian leadership could intervene in military micromanagement and, under certain circumstances, may be able to abort the operation. Soldiers only act independently in the event of an interrupted connection, but even then, they remain 'integrated.' They remain bound by the mission, the rules of the military organization, overarching rules (Rules of Engagement, Law of Armed Conflict), and ethical programs.

Just as the actions of soldiers cannot be understood outside of this framework—although they act and think independently as a vital organism and consciousness, they are included in the organization and in the society via roles, norms, and convictions—, the use of artificially intelligent automation cannot be detached from this framework, as can be shown with recourse to systemic perspectives (Spreen, 2023). An anomic Soldateska cannot be ethically justified, and this also applies to armed automatons. Of course, scenarios are conceivable in which conflict parties create precisely such soldiers and killer systems—for example, to spread fear and terror—, but this cannot be justified either ethically or legally. From this perspective, "there must be no truly autonomous weapon systems [...], but 'only' weapon systems with autonomous capabilities that must be securely embedded in the respective *system of systems*, organization, and society" (Flemisch & Nitsch, 2023, p. 249, emphasis in original). These weapon systems thus ultimately remain 'semi-autonomous.' Only as semi-autonomous, embedded military systems can they meet the overarching requirements of civil society at all (Spreen, 2015b).

Also, against the background of the possibilities of artificial intelligence, close human–machine cooperation in "mixed teams" (Singer, 2010b, p. 60) remains desirable. Wolf Graf von Baudissin already pointed out (in the context of the German Armed Forces' ethical program of 'Innere Führung') that in modern technical combat it is important that "there are people behind the weapons who know what they are doing" (Baudissin, 2006, p. 118). This is where intelligent combat suits come into play. Not only do they improve the embedding of soldiers in networked operations, but conversely also human–machine cooperation and thus the control of machines in relation to political and military objectives as well as in relation to the ethical, normative, and legal framework.

With reference to the controller model and following Chris Gables Gray (2001, pp. 55–65), who describes himself as a representative of posthumanism, one can certainly speak of "cyborg warriors." Gray, however, conceptualizes the cyborg soldier as "part of a weapon system" (Gray, 2001, p. 63). This is a very shortened perception, because cybernetically augmented human soldiers with combat suit systems are not simply absorbed by the networked systems to which they are connected. They do not become a mere appendage or a mere function. This is precisely the point of reflecting on cyborgization from the perspective of Plessnerian anthropology. Even a cyborg remains human and is positioned excentrically. Human beings stand between being a body and having a body—and more recently *also*

between being a technology and having a technology, which indicates the historical character of their world, self, and social relations. Cyborg soldiers also remain integrated into their social roles and the associated norms and expectations. To paraphrase Baudissin: they must be addressed as (responsible) actors and should perceive themselves as acting and experiencing.

Deeper human–machine cooperation is not without risk. This is hardly surprising, as high technology is constitutively risky. As Luhmann (2010, p. 88) explains, "[i]t transforms dangers to risks simply because it creates possibilities of making decisions that had not existed before." This makes it necessary to research these risks to minimize them as far as possible. For example, decision-making cycles are accelerating in high-tech battles. Sun Tzu already knew that "rapidity is the essence of war" (Tzu & Giles, 1910, p. 122). In the age of the "dromocratic revolution" (Virilio & Lotringer, 2008, p. 59), this is all the truer. Acceleration and thus time pressure are virtually a hallmark of modern warfare. However, the promise of digitally penetrating the fog of war remains only a promise. There is always the question of the quality of digital duplication. Although it allows for a much better picture of the situation, it is not identical to analog reality simply because it is digital (Henkel, 2019, pp. 35–36). Moreover, digital perception support should not simply represent, but also evaluate, i.e., be able to discriminate between friend/foe, combatant/non-combatant, for example. Is this possible with certainty? As a rule, uncertainties should also be communicated, e.g., "80 % probability of enemy vehicle" (Koch et al., 2024, S. 236). Time pressure may make it impossible to fully secure the decision. Now cyborg soldiers are in demand as human beings. This is because they have to decide and act in the context of critical situations—and thus take a risk and assume responsibility (Schluchter, 2014, p. 33). In order to draw conclusions about the technical and ethical competencies relevant for soldiers, it is necessary to understand the human-machine interactions (HMI) that occur in the context of AI use. Ideally, this understanding would take into account all relevant system levels because human accountability should frame the entire HMI process (Flemisch & Nitsch, 2023; Spreen, 2015b; 2023).

Technical enhancement also requires ethical enhancement. Cyborg soldiers should not see themselves as 'cogs in the wheel,' but must be able to address themselves as responsible actors and also be addressed as such in communication. "Self-efficacy" is an important requirement in highly technical environments and complex social relationships (Spreen, 2023, p. 29), which Baudissin already clearly recognized. According to him, "democracy, the modern world of work and military craft" demand "rational action." These demands can only be met by the soldier "who feels challenged by the task as an individual, i.e., 'undivided' with his intellect, feelings, and skills" (Baudissin, 1969, p. 126). Only soldiers who are "addressed with respect for their dignity" in the military system can imagine human dignity and respect it (Baudissin, 1969, p. 126, cf. p. 120).[7]

[7] This requires the abandonment of mechanical drill.

8 Ethical Enhancement

Can soldiers who experience themselves as self-presenting and self-effective in their actions, i.e., who are addressed with respect for their dignity, be seen as a kind of ethical enhancement of the machine? If it is true that such addressing helps people to reflect ethically on their own actions even under difficult conditions, then internal military communication and the moral programming of the military would be key factors for ethical action in conflict situations. This is what Baudissin and Innere Führung (as a variant of the moral programming of military organizations) had in mind. Sociologically speaking: It is less about the ethics of the technology and more about the ethics of the organization. If this organization configures its soldiers in a way that systematically promotes amoral action and "dissociation mentality" (Hüppauf, 1993), then these soldiers will also use the technologies available to them in this way—and vice versa.[8]

An "ethical governor," as recommended by the computer scientist Ronald C. Arkin and others, can only prevent unethical behavior by artificially intelligent machines if these machines are deployed by a military that complies with the ethical and legal framework (and not just on paper, but in practice). If the dignity of (cyborg) soldiers is respected in the military organization (and in the society), then they can behave accordingly. Then they can also act as an ethical enhancement of the machines. In military conflict settings, problematic situations will repeatedly and probably quite often arise that remain unclear and dubious even in digital duplication. It is likely that such situations will overexert machines (Lee, 2018, p. 307). In such cases, the decisions of human soldiers are required. They must be prepared for this, which demands adjustments to the ethical program of the military organization. This is precisely what Baudissin pointed out in his reflections on Innere Führung.

Of course, machines can also serve as ethical aids that can be used by soldiers in critical situations or become active on their own to offer their users ethical support when making decisions. Artificially intelligent automation serves here as "an ethical advisor suitable for enhancing human performance" (Arkin et al., 2012, p. 587). Such support "will be able to provide a second opinion for human users operating in ethically challenging areas" (Arkin et al., 2012, p. 587). Such ethical advisors could also be usefully deployed in civilian circumstances "to enhance human-human relationships" (Arkin et al., 2012, p. 587). Arkin thus takes up an idea of Arthur C. Clarke, who predicted in the mid-1980s that in the future "the computer will be used as a consultant of sorts" (Clarke, 1986, p. 39). However, the use of

[8] The Concept of dissociation mentality, introduced by Bernd Hüppauf, refers to collectively shared views and philosophies of life which untied "the ties to the universal ideas of political and aesthetic traditions" (Hüppauf, 1993, p. 74; cf. Spreen, 2011). This refers to a type of soldier who is conceived as a "modern fighting machine" (Hüppauf, 1998, p. 85). This is how the Nazis saw their soldiers. Warfare was for them a "highly organized, amoral and merciless" (Hüppauf, 1998, p. 70) business that demanded a new man: "amoral, cool, functional, experienced, hardened men who no longer needed ideals to identify with, or emotions as a basis for their fighting spirits" (Hüppauf, 1998, p. 84).

ethical computers presupposes that the organization in question is guided by a moral program that is not just one on paper.

However, ethical AI advice does not change the fact that soldiers must take accountability for their decisions as *their* decisions. They should not be able to excuse themselves by saying they are just a cog in the wheel. This implies that they experience themselves as self-effective and respected in their dignity. In highly automated environments, the opportunities to experience self-efficacy and self-expression in action, which could perhaps be understood as the essence of human dignity, may well be lost (Arkin et al., 2012, pp. 586–587). This must be prevented. 'Ethical enhancement' means that, ultimately, human (cyborg) soldiers are able to determine in all conscience whether a decision is justified or not. Such a decision may, for example, also include the use of Autonomous Weapon Systems (AWS, Spreen, 2025, S. 296). If soldiers are to rely on their technical and ethical support, it must also be clear that the machine's ethical advice in turn is subject to human accountability—for example, through its manufacturer and/or the military procurement. There must therefore be liability for errors and responsibility for outcomes that can be attributed to persons—a accountability that cannot be delegated to the operational complexity of data processing.[9] In the end, everything is about accountability—and this is exercised by social actors in organizational roles. "Accountability must always be tailored to persons, as the decisions themselves are not permanent. Even committees, groups, teams that decide by majority cannot be held accountable" (Luhmann, 2000, pp. 197–198).

9 Conclusions

From the perspective of philosophical anthropology, soldiers are not understood as a function or appendage of their technology. At the same time, from this perspective, body enhancement should neither be rejected in principle nor seen as a harbinger of overcoming 'the human being.' However, the opportunities and risks resulting from the technology must be considered. The purpose of the regulator model is to make it possible to address technological developments in the area close to the body without blurring fundamental differences and relationships between consciousness, vital corporality, technology, and communication or relativizing the special status of humans in relation to the living and technical environment. Reflecting on the "natural artificiality" (Plessner, 2019, pp. 287–298) of humans and their social world does not require 'posthumanist' relativizations, but is also possible within the framework of an anthropology that can reconstruct the special status of humans, as Helmuth Plessner, for example, plausibly succeeded in doing.

[9] In proceedings on the "right to be forgotten" before the European Court of Justice, *Google* argued "that it cannot be held responsible because the processing of data is performed by the search engine" (Esposito, 2022, p. 67). But can "the autonomy of the operation of algorithms relieve the company from the responsibility for data management?" (Esposito, 2022, p. 67). The Court took a different view than *Google* and "considers Google accountable and responsible" for the active role of the algorithm (Esposito, 2022, p. 67).

Cyborg technology often has an interface not only to the body, but also to communication. This duplication of the interface is risky, as digital attacks can now have a direct effect on bodies. A virtual virus can become a real projectile. Consider, for example, the potential danger posed by hackers accessing pacemakers, autonomous vehicles, weapon systems or even old-fashioned pagers. In the age of hybrid and cognitive warfare, it is necessary to think about protective measures, especially as AI can be helpful here (Santoso & Finn, 2023). In the military domain, enemy intrusions into the digital sphere can make you dizzy instead of fighting the enemy into dizziness.

- Cyborgization means that technology is moving closer to the human body. It can be described as a continuum of technology–body symbiosis in the near-bodily or even intra-bodily realm (controller model).
- Cyborg enhancement can be effectively supported by artificially intelligent automation. AI generates new opportunities and can improve performance.
- From the perspective of Helmuth Plessner's philosophical anthropology, there is no reason in principle not to understand human cyborgs as 'humans.' Seen in this light, there is no objective need for 'posthumanist' perspectives.
- In the cyborg society and its upgrade culture, a critical and differentiated examination of 'technology' and 'materialities' is essential. However, one does not have to become a 'post- or transhumanist' to know that.
- Technical extensions of the body do not in principle impede the possibility of presenting oneself in social contexts. Nor do they fundamentally dissolve the unfathomability of the person. Rather, they can become a means of self-expression. Nevertheless, there are risks.
- Cyborg technologies are also very interesting for the military. But it remains important to address cyborg soldiers as human beings with respect for their dignity. They should not be cogs in the machine; they should not see themselves as mere machines.
- Networking creates an interface for communication that opens up new possibilities for digital attacks. The connection between the body and technology tends to dissolve the separation between virtual and real effects.
- Cyborg soldiers can also be seen as an ethical enhancement of machines if they are appropriately trained and communicatively addressed. Conversely, artificial ethical advisors can in turn provide support in critical situations. Ideally, this would lead to a process of mutual ethical enhancement.

References

Arkin, R. C., Ulam, P., & Wagner, A. R. (2012). Moral Decision Making in Autonomous Systems: Enforcement, Moral Emotions, Dignity, Trust, and Deception. *Proceedings of the IEEE, 100*(3), 571–589. https://doi.org/10.1109/JPROC.2011.2173265

Arthur, W. B. (2009). *The Nature of Technology. What It Is and How It Evolves*. Free Press.

Baecker, D. (2023). Technik im Datenraum. In D. Baecker, U. Elsholz, M. Locher, & M. Thomas (Eds.), *Post-digitales Management. Arbeit an den Schnittstellen einer Produktionsorganisation* (pp. 243–256). Springer VS. https://doi.org/10.1007/978-3-658-40707-0_19

Barbrook, R., & Cameron, A. (1995). The Californian Ideology. *Mute, 1*(3) https://www.metamute.org/editorial/articles/californian-ideology

Baudissin, W. G. (2006). In A. Dörfler-Dierken (Ed.), *Als Mensch hinter den Waffen*. Vandenhoeck & Ruprecht.

Baudissin, W. G. (1969). *Soldat für den Frieden. Entwürfe für eine zeitgemäße Bundeswehr*. Piper.

Beck, U. (1992). *Risk Society. Towards a New Modernity* (M. Ritter, Trans.). Sage Publications.

Boltanski, L., & Chiapello, È. (2007). *The New Spirit of Capitalism*. Verso.

Bothe, H.-W., & Engel, M. (1993). *Die Evolution entlässt den Geist des Menschen. Neurobionik—Eine medizinische Disziplin im Werden*. Umschau.

Clarke, A. C. (1986). *July 20, 2019. Life in the 21st Century*. Macmillan.

Clynes, M. E., & Kline, N. S. (1960). Cyborgs and Space. *Astronautics, 5*(9), 26–27, 74–76.

Coenen, M., & Görlich, Y. (2022). Exploring Nomophobia with a German Adaption of the Nomophobia Questionnaire (NMP-Q-D). *PLoS One, 17*(12), Article e0279379. https://doi.org/10.1371/journal.pone.0279379

de Mul, J. (2010). *Cyberspace Odyssey. Towards a Virtual Ontology and Anthropology*. Cambridge Scholas Publishing.

Escobar, A. (1994). Welcome to Cyberia: Notes on the Anthropology of Cyberculture. *Current Anthropology, 35*(3), 211–231. https://doi.org/10.1086/204266

Esposito, E. (2022). *Artificial Communication: How Algorithms Produce Social Intelligence*. MIT Press. https://doi.org/10.7551/mitpress/14189.001.0001

Eßbach, W. (2011). *Die Gesellschaft der Dinge, Menschen, Götter*. VS. https://doi.org/10.1007/978-3-531-92835-7

European Commission, Directorate-General for Communication. (2021). *European Citizens' Knowledge and Attitudes towards Science and Technology—Report* (Special Eurobarometer 516, April–Mai 2021). Publications Office of the European Union. https://data.europa.eu/doi/10.2775/071577

Fischer, J. (2002). Androiden – Menschen – Primaten. Philosophische Anthropologie als Platzhalterin des Humanismus. In R. Faber, & E. Rudolph (Eds.), *Humanismus in Geschichte und Gegenwart* (pp. 229–239). Mohr Siebeck. https://doi.org/10.5771/9783845280967-344

Fischer, J. (2019). Philosophische Anthropologie im digitalen Zeitalter: Tier-/Mensch-, Maschine-/Mensch-, Mensch-/Mensch-Vergleich. In J. F. Burow, L.-J. Daniels, A.-L. Kaiser, C. Klinkhamer, J. Kulbatzki, Y. Schütte, & A. Henkel (Eds.), *Mensch und Welt im Zeichen der Digitalisierung. Perspektiven der Philosophischen Anthropologie Plessners* (pp. 231–259). Nomos. https://doi.org/10.5771/9783845293226-229

Fischer, J. (2022). Menschenrechte. Eine Analyse aus der Perspektive der Philosophischen Anthropologie. *Österreichische Zeitschrift für Soziologie, 46*(Suppl 1), 205–223. https://doi.org/10.1007/s11614-022-00489-w

Flemisch, F., & Nitsch, V. (2023). Kooperative Systeme und hybride Intelligenz. Plädoyer für ganzheitliche Mensch-Maschine-Integration. In N. Lammert, & W. Koch (Eds.), *Bundeswehr der Zukunft. Verantwortung und Künstliche Intelligenz* (pp. 237–250). Konrad-Adenauer-Stiftung. https://www.kas.de/de/einzeltitel/-/content/bundeswehr-der-zukunft-5

Foucault, M. (1979). *Discipline and Punish. The Birth of the Prison*. Vintage Books.

Frankenberg, G. (2003). *Autorität und Integration. Zur Grammatik von Recht und Verfassung*. Suhrkamp.

Fuchs, P. (2005). Der Körper als Form. In M. Schroer (Ed.), *Soziologie des Körpers* (pp. 48–72). Suhrkamp.

Gehlen, A. (1961). *Anthropologische Forschung. Zur Selbstbegegnung und Selbstentdeckung des Menschen*. Rowohlt.

Gehlen, A. (1988). *Man. His Nature and Place in the World* (C. McMillan, & K. Pillemer, Trans.). Columbia University Press.

Goffman, E. (1986). *Stigma. Notes on the Management of Spoiled Identity. Simon & Schuster.*

Gray, C. H. (2001). *Cyborg Citizen. Politics in the Posthuman Age*. Routledge. https://doi.org/10.4324/9780203949351

Grote, C. (2012, November 12). *Exoskelett unterstützt Handbewegungen*. elektroniknet.de. https://www.elektroniknet.de/medizintechnik/medtech-komponenten/exoskelett-unterstuetzt-handbewegungen.93273.html

Haraway, D. (1985). A Manifesto for Cyborgs: Science, Technology, and Socialist Feminism in the 1980s. *Socialist Review, 80*, 65–108.

Harbisson, N. (2012, June). *I listen to color* [Video]. TED Conferences. https://www.ted.com/talks/neil_harbisson_i_listen_to_color

Harrasser, K. (2013). *Körper 2.0. Über die technische Erweiterbarkeit des Menschen*. transcript. https://doi.org/10.14361/transcript.9783839423516

Heilinger, J.-C. (2010). *Anthropologie und Ethik des Enhancements*. de Gruyter. https://doi.org/10.1515/9783110223705

Heilinger, J.-C., & Müller, O. (2007). Der Cyborg und die Frage nach dem Menschen. Kritische Überlegungen zum "homo arte emendatus et correctus". *Jahrbuch für Wissenschaft und Ethik, 12*(1), 21–44. https://doi.org/10.1515/9783110192476.1.21

Henkel, A. (2016). Posthumanism, the Social and the Dynamics of Material Systems. *Theory, Culture & Society, 33*(5), 65–89. https://doi.org/10.1177/0263276415625334

Henkel, A. (2017). Die Materialität der Gesellschaft. Entwicklung einer gesellschaftstheoretischen Perspektive auf Materialität auf Basis der Luhmann'schen Systemtheorie. *Soziale Welt, 68*(2–3), 279–299. https://doi.org/10.5771/0038-6073-2017-2-3-279

Henkel, A. (2019). Digitalisierung der Gesellschaft. Perspektiven der reflexiven Philosophischen Anthropologie auf gesellschaftlichen Wandel durch Digitalisierung. In J. F. Burow, L.-J. Daniels, A.-L. Kaiser, C. Klinkhamer, J. Kulbatzki, Y. Schütte, & A. Henkel (Eds.), *Mensch und Welt im Zeichen der Digitalisierung. Perspektiven der Philosophischen Anthropologie Plessners* (pp. 19–45). Nomos. https://doi.org/10.5771/9783845293226-17

Hüppauf, B. (1993). Schlachtenmythen und die Konstruktion des "Neuen Menschen." In G. Hirschfeld, G. Krumeich, & I. Renz (Eds.), *Keiner fühlt sich hier mehr als Mensch... Erlebnis und Wirkung des Ersten Weltkriegs* (pp. 43–84). Klartext.

Hüppauf, B. (1998). Langemarck, Verdun and the Myth of a New Man in Germany after the First World War. *War & Society, 6*(2), 70–103. https://doi.org/10.1179/106980488790304887

In, H., Kang, B. B., Sin, M., & Cho, K.-J. (2015). Exo-Glove: A Wearable Robot for the Hand with a Soft Tendon Routing System. *IEEE Robotics & Automation Magazine, 22*(1), 97–105. https://doi.org/10.1109/MRA.2014.2362863

Jünger, F. G. (1953). *Die Perfektion der Technik* (4th ed.). Klostermann.

Kaufmann, S. (2006). Land Warrior. The Reconfiguration of the Soldier in the "Age of Information." *Science, Technology & Innovation Studies, 2*(11), 81–102. https://doi.org/10.17877/DE290R-717

Kline, R. (2009). Where are the Cyborgs in Cybernetics? *Social Studies of Science, 39*(3), 331–362. https://doi.org/10.1177/0306312708101046

Koch, W., Spreen, D., Talves, K., Wagner, W., Lillemäe, E., Klaus, M., Viidalepp, A., Cooper, C. G., & Pekarev, J. (2024). On the Ethics of Employing Artificial Intelligent Automation in Military Operational Contexts. *IEEE Transactions on Technology and Society, 5*(2), 231–241. https://doi.org/10.1109/TTS.2024.3405309

Lange, S. (2004). *Netzwerk-basierte Operationsführung (NBO). Streitkräfte-Transformation im Informationszeitalter* (SWP study 22/2004). Stiftung Wissenschaft und Politik. https://nbn--resolving.org/urn:nbn:de:0168-ssoar-243490

Latour, B. (1993). *We Have Never Been Modern*. Harvard University Press.

Lee, P. (2018). Armed Drones: Automation, Autonomy, and Ethical Decision-Making. In R. Kiggins (Ed.), *The Political Economy of Robots. International Political Economy Series* (pp. 291–315). Palgrave Macmillan. https://doi.org/10.1007/978-3-319-51466-6_14

Luhmann, N. (1965). *Grundrechte als Institution. Ein Beitrag zur politischen Soziologie*. Duncker & Humblot.

Luhmann, N. (1966). *Recht und Automation in der öffentlichen Verwaltung. Eine verwaltungswissenschaftliche Untersuchung*. Duncker & Humblot.

Luhmann, N. (1971). Sinn als Grundbegriff der Soziologie. In J. Habermas, & N. Luhmann (Eds.), *Theorie der Gesellschaft oder Sozialtechnologie – Was leistet die Systemforschung?* (pp. 25–100). Suhrkamp.

Luhmann, N. (1993). *Risk. A Sociological Theory* (R. Barrett, Trans.). de Gruyter (Original work published 1991).

Luhmann, N. (1995). *Funktionen und Folgen formaler Organisation* (4th ed., Epilogue 1994). Duncker & Humblot.

Luhmann, N. (2000). *Organisation und Entscheidung*. Westdeutscher Verlag.

Luhmann, N. (2005). Wie ist Bewußtsein an der Kommunikation beteiligt? In N. Luhmann, *Soziologische Aufklärung 6. Die Soziologie und der Mensch* (2nd ed., pp. 38–54). VS.

Luhmann, N. (2010). The Morality of Risk and the Risk of Morality. *International Review of Sociology Series 1*, *1*(3), 87–101. https://www.tandfonline.com/doi/abs/10.1080/03906701.1987.9971345

Luhmann, N. (2012). *Theory of Society.* (Vol. 1, R. Barrett, Trans.). Stanford University Press (Original work published 1997). https://doi.org/10.1515/9780804786478

Moravec, H. (1993). Geist ohne Körper – Visionen von der reinen Intelligenz. In G. Kaiser, D. Matejovski, & J. Fedrowitz (Eds.), *Kultur und Technik im 21. Jahrhundert* (pp. 81–90). Campus.

Müller, O. (2010). *Zwischen Mensch und Maschine: Vom Glück und Unglück des Homo faber*. Suhrkamp.

Nassehi, A. (2024). *Patterns. Theory of the Digital Society* (M. Wittwar, Trans.). Polity Press.

Ortiz, M. (2022). *Kontrollverlust und Technologieakzeptanz in der (digitalen) Transformation. Akzeptanz- und Gestaltungsfaktoren eines heuristischen Modells*. Springer VS. https://doi.org/10.1007/978-3-658-35697-2

Petersen, T. (2011). Kein Fortschrittspessimismus. *Frankfurter Allgemeine Zeitung, 115*, 5.

Planungsamt der Bundeswehr – Dezernat Zukunftsanalyse. (2013). *Future Topic Human Enhancement—Eine neue Herausforderung für Streitkräfte?* Bundesamt für Infrastruktur, Umweltschutz und Dienstleistungen der Bundeswehr.

Plessner, H. (1970). *Laughing and Crying: A Study of the Limits of Human Behavior* (J. S. Churchill, & M. Grene, Trans.). Northwestern University Press (Original work published 1941). https://doi.org/10.21985/N2313X

Plessner, H. (1983a). Die Frage nach der Conditio Humana. In H. Plessner, *Gesammelte Schriften VII* (G. Dux, O. Marquard, & E. Ströker, Eds., pp. 136–217). Suhrkamp (Original work published 1961).

Plessner, H. (1983b). Das Problem der Unmenschlichkeit. In H. Plessner, *Gesammelte Schriften VII* (G. Dux, O. Marquard, & E. Ströker, Eds., pp. 328–337). Suhrkamp (Original work published 1967).

Plessner, H. (1999). *The Limits of Community: A Critique of Social Radicalism* (A. Wallace, Trans.). Humanity Books (Original work published 1924).

Plessner, H. (2018). *Political Anthropology* (N. F. Schott, Trans.). Northwestern University Press (Original work published 1931).

Plessner, H. (2019). *Levels of Organic Life and the Human. An Introduction to Philosophical Anthropology* (M. Hyatt, Trans.). Fordham University Press (Original work published 1928).

Plessner, H., & Buytendijk, F. J. J. (1982). Die Deutung des mimischen Ausdrucks. Ein Beitrag zur Lehre vom Bewußtsein des anderen Ichs. In H. Plessner, *Gesammelte Schriften VII* (G. Dux, O. Marquard, & E. Ströker, Eds., pp. 67–129). Suhrkamp (Original work published 1925).

Popitz, H. (1995). *Der Aufbruch zur Artifiziellen Gesellschaft. Zur Anthropologie der Technik*. J.C.B. Mohr.

Raspopovic, S., Capogrosso, M., Petrini, F. M., Bonizzato, M., Rigosa, J., Di Pino, G., & Micera, S. (2014). Restoring Natural Sensory Feed-back in Real-Time Bidirectional Hand Prostheses. *Science Translational Medicine, 6*(222), 19–22. https://doi.org/10.1126/scitranslmed.3006820

Renn, O. (1986). Akzeptanzforschung: Technik in der gesellschaftlichen Auseinandersetzung. *Chemie in unserer Zeit, 20*(2), 44–52. https://doi.org/10.1002/ciuz.19860200203

Renn, O. (2005). Technikakzeptanz: Lehren und Rückschlüsse der Akzeptanzforschung für die Bewältigung des technischen Wandels. *Technikfolgenabschätzung—Theorie und Praxis, 14*(3), 29–38. https://doi.org/10.14512/tatup.14.3.29

Rosenberg, E. S. (2008). Far Out: The Space Age in American Culture. In S. J. Dick (Ed.), *Remembering the Space Age* (pp. 157–184). NASA. https://nss.org/wp-content/uploads/Remembering-the-Space-Age-NASA-SP4703.pdf

Santoso, F., & Finn, A. (2023). Trusted Operations of a Military Ground Robot in the Face of Man-in-the-Middle Cyberattacks Using Deep Learning Convolutional Neural Networks: Real-Time Experimental Outcomes. *IEEE Transactions on Dependable and Secure Computing*, *21*(4), 2273–2284. https://doi.org/10.1109/TDSC.2023.3302807

Scheler, M. (2009). *The Human Place in the Cosmos* (M. S. Frings, Trans., E. Kelly, Intro.). Northwestern University Press (Original work published 1928).

Schluchter, W. (2014). Globalisierung von Risiken und Verflüchtigung von Verantwortung. Zu einem Strukturproblem der wissenschaftlich-technischen Zivilisation. In H. Hahn, T. Holstein, & S. Leopold, S. (Eds.), *Risiko und Verantwortung in der modernen Gesellschaft* (Writings of the Mathematics and Natural Sciences Class, vol. 26, pp. 31–41). Springer Spektrum. https://doi.org/10.1007/978-3-658-06322-1_6

Schneider, W. (2005). Der Prothesen-Körper als gesellschaftliches Grenzproblem. In M. Schroer (Ed.), *Soziologie des Körpers* (pp. 371–397). Suhrkamp.

Schumann, M. (2019). *Extraterrestrische Ex-zentriker. Zur theoriestrukturellen Einarbeitung des Außerirdischen bei Helmuth Plessner als Grundlage einer Philosophischen Anthropologie des Raumfahrtzeitalters* [Master's thesis, University of Potsdam]. Publish.UP. https://doi.org/10.25932/publishup-43420

Shein, E. (2019). Exoskeletons Today. *Communications of the ACM, 62*(3), 14–16. https://doi.org/10.1145/3303851

Shepherd, J. (2006). *Crossover: A Cassandra Kresnov Novel*. Pyr.

Singer, P. W. (2010a). *Wired for War. The Robotics Revolution and Conflict in the Twenty-first Century*. Penguin books.

Singer, P. W. (2010b). War of the Machines. *Scientific American*, *303*(1), 56–63.

Sobchack, V. (2004). *Carnal Thoughts. Embodiment and Moving Image Culture*. University of California Press. https://doi.org/10.1525/9780520937826

Spreen, D. (2011). Cruelty and Total War. Political-philosophical Preconditions of the Dissociation Mentality. In T. v. Trotha, & J. Rösel (Eds.), *On Cruelty – Sur la cruauté – Über Grausamkeit* (Siegener Beiträge zur Soziologie, vol. 11, pp. 231–252). Rüdiger Köppe.

Spreen, D. (2014a). Die dritte Raumrevolution. Weltraumfahrt und Weltgesellschaft nach Carl Schmitt und Niklas Luhmann. In J. Fischer, & D. Spreen (Eds.), *Soziologie der Weltraumfahrt* (pp. 89–127). transcript. https://doi.org/10.1515/transcript.9783839427750.89

Spreen, D. (2014b). Not Terminated: Cyborgized Men Still Remain Human Beings. In J. de Mul (Ed.), *Plessner's Philosophical Anthropology: Perspectives and Prospects* (pp. 425–442). Amsterdam University Press. https://doi.org/10.26530/OAPEN_626454

Spreen, D. (2015a). *Upgradekultur. Der Körper in der Enhancement-Gesellschaft*. Transcript. https://doi.org/10.1515/9783839430088

Spreen, D. (2015b). Die Kriegsautomaten der Zivilgesellschaft. Semiautonome technische Systeme in bewaffneten Sicherheitsoperationen. In N. Leonhard, & J. Franke (Eds.), *Militär und Gewalt: Sozialwissenschaftliche und ethische Perspektiven* (pp. 163–184). Duncker & Humblot. https://doi.org/10.3790/978-3-428-54581-0

Spreen, D. (2022a). Semantik der Kugel. Kugelraumschiffe und andere sphärische Technologien. In D. Spreen, & B. Flessner (Eds.), *Die Raumfahrt der Gesellschaft. Wirtschaft und Kultur im New Space Age* (pp. 269–300). transcript. https://doi.org/10.1515/9783839457627-007

Spreen, D. (2022b). Was sagen uns Kriegserfahrungen? Über Kriegsdiskurse im Anschluss an Niklas Luhmann. *Ästhetik & Kommunikation, 51*(184/185), 157–163.

Spreen, D. (2023). Lethal Autonomous Weapon Systems (LAWS). On the Ethics of Automation in the Military from the Perspective of Social Systems Theory. *Sõjateadlane (Estonian Journal of Military Studies)*, (21), 10–40. https://doi.org/10.15157/st.vi21.24177

Spreen. D. (2024). Transhumanismus. In M. Dederich, & J. Zirfas (Eds.), *Optimierung. Ein interdisziplinäres Handbuch* (pp. 345–350). J.B Metzler. https://doi.org/10.1007/978-3-662-67307-2_49

Spreen, D. (2025). Führung mit künstlicher Intelligenz und Verantwortung. In U. Hartmann, R. Janke, & C. v. Rosen (Eds.), *Jahrbuch Innere Führung 2024/25. Wissenschaft und Bildung für gute Führung in einer kriegstüchtigen Bundeswehr* (pp. 290–300). Miles.

Tzu, S., & Giles, L. (1910). *Sun Tzu on the Art of War. The Oldest Military Treatise in the World* (L. Giles, Trans.). Luzac (Original work published around 500 BC).

Virilio, P., & Lotringer, S. (2008). *Pure War. Twenty-Five-Years later*. Semiotext(e) (Original work published 1983).

Warburg, J. (2008). *Das Militär und seine Subjekte. Zur Soziologie des Krieges*. transcript. https://doi.org/10.14361/9783839408520

Wiener, N. (1954). *The Human Use of Human Beings. Cybernetics and Society* (2nd ed.). Doubleday.

Yildirim, C. F., & Correia, A. P. (2015). Exploring the Dimensions of Nomophobia: Development and Validation of a Self-reported Questionnaire. *Computers in Human Behavior, 49*, 130–137. https://doi.org/10.1016/j.chb.2015.02.059

Zoglauer, T. (2003). Der Mensch als Cyborg? Philosophische Probleme der Neuroprothetik. *Universitas, 58*(12), 1267–1278.

Autonomous Weapon Systems in Science Fiction

Bernd Flessner

Abstract The genre of science fiction (SF) uses literary means to explore the space of possibilities that lies before us. Implicit predictions can thus be found in the texts, especially about future technologies. This article uses this characteristic to show examples of implied predictions concerning autonomous weapons systems. If one follows the texts, they will also be realized, if they are not already.

An essential goal of the Enlightenment and the sciences committed to it, Horkheimer and Adorno explain, is "making the world calculable," because only this guarantees dominance over nature and mankind. As a methodological consequence of this calculability, "number became enlightenment's canon" (Horkheimer & Adorno, 2002, p. 4). This program naturally implies the premise that the world has the property of being calculable and can be described with the formal science of mathematics constructed by humans.

At least one area of our world is definitely at odds. "The future as such eludes knowledge," writes sociologist Elena Esposito. She refers to a widespread, albeit false, assumption that the present develops from the past in a kind of historical logic that is as stringent as it is plausible (Esposito, 2007, p. 29).[1] But it is not only the future that eludes a purely mathematical approach; the present and the past also refuse to be understood, because they were once also a (present) future.

While we no longer perceive the present as a variant in a now past and therefore closed space of possibilities, we continue to grant the future this characteristic of being such a variant, one of an infinite number. The historian Lucian Hölscher states: "The more distant the historical event becomes, the more our awareness of

[1] All quotations from German sources have been translated by the author. In particular, Jules Verne and Stanisław Lem were translated from German editions.

B. Flessner (✉)
Center for Applied Philosophy of Science and Key Qualifications (ZiWis),
Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Bavaria, Germany
e-mail: bernd.flessner@fau.de

K. Talves, D. Spreen (eds.), *Artificial Intelligence in Military Technology*,
Artificial Intelligence, Simulation and Society 192,
https://doi.org/10.1007/978-3-031-95578-5_4

the openness of the situation at that time fades" (Hölscher, 2016, p. 9). The present was once also the future. However, we are gradually blocking out all possible variants and thus everything that was once possible but fictitious, which could just as easily have become the present. This fading out is so easy for us because the once fictitious can now only be identified as such in the first place through the solidification of a possible variant into the present. Lucian Hölscher once again:

> The past world can be divided into factual and fictional events. To classify the future world in this way, on the other hand, would be pointless, because in it we cannot say with certainty whether an event belongs to the realm of fact or fiction. And this is not an incidental, but an essential characteristic of such events. (Hölscher, 2016, p. 8)

And yet there is a tendency in the various academic disciplines to regard the fictional as suspect, with the exception of those disciplines that explicitly deal with the fictional, such as literary studies. The fictional is often regarded as the counterfactual, as an extraterritorial area of the real. Unfortunately, however, the future implies the fictional and thus uncertainty, imponderability, the uncontrollable, which eludes the grasp of knowledge. For this reason, "tools for coping with the uncertainty of expectation" were sought and, how could it be otherwise in the paradigm of the Enlightenment, mathematics was used (Esposito, 2007, p. 29). Probability theory and various methods of modeling on a statistical basis were and are intended to tame the open future, the fictitious and contingency (Esposito, 2007, p. 29). However, the fictitious cannot be shaken off so easily. It gains access through the back door, so to speak, to the very method that is being used against it. Esposito sums up:

> Probabilities can be calculated, predictions can be made based on them. However, it is perfectly clear that these are pure fictions, because the future presents will not be more or less probable, they will not materialize to 40 or 75 percent, but exactly as they will be. [...] Things develop completely independently of all forecasts. No matter how hard you try to calculate the future and remove it from randomness, you can never be sure that the fictitious predictions of probability theory will actually come true. (Esposito, 2007, pp. 31, 34)

1 Portent Search as a Method

This does not make probability calculation obsolete. It merely forfeits the claim that it can make the future known if only enough data is available. Futurology therefore takes a rather skeptical view of forecasts based on probability calculations, without completely abandoning them. It prefers the scenario method, which deals with the variants within the realm of possibility. Instead of the one, "certain" future, this method focuses on several alternative futures that are situated in the possibility space in such a way that they represent it as broadly as possible (Homburg & Krohmer, 2006, pp. 455–460). The openness of the future is thus methodically taken into account. In addition, "a scenario also describes the developments, dynamics and driving forces from which a certain image of the future results" (Kosow & Gaßner, 2008, p. 10). A scenario is therefore *well-founded*. The main sources of scenarios are trend extrapolations, Delphi studies, statistics, influence and

consistency analyses, and ultimately all thematically relevant scientific publications (Kosow & Gaßner, 2008, p. 38).

The fact that scenarios can be well founded at all points to a fundamental characteristic of the future. For despite "the historical openness of the future," it is not characterized by absolute contingency or permanent discontinuity, it does not occur unconditionally and arbitrarily, but is already laid out in the past and present as a variant in a space of possibilities (Hölscher, 2016, p. 9). But which of the countless possible variants the future present will manifest is actually open.

In addition to the sources and methods already mentioned for the foundation of scenarios, there is another, less well-known one that refers directly to the variants created in the space of possibilities. One of the first to attempt to give a name to this diffuse but existing provenance of possible futures was Ernst Bloch. In his main work, Principle of Hope, written between 1938 and 1947, the philosopher writes: "The Real Possible begins with the seed in which what is coming is inherent" (Bloch, 1986, pp. 237–238). In the context of Bloch's work, the metaphor "seed" is sufficient to take away the arbitrary nature of the future. In 1968, the British philosopher Bernard Mayo coined the term "portent" for signs in the sense of omens for the future and distinguished it from the term "trace," which in turn refers to traces of the past:

> [...] the word 'trace' specifically names backward-pointing evidence. We do have the word 'sign,' but this is also used for backward-pointing evidence, besides being allready overworked in other ways. We need to invent a term for the future analogue for 'trace.' I shall use the word 'portent' for this purpose. (Mayo, 1986, p. 290)

The question naturally arises as to where these portents can be found and how they can be identified. The answer is not easy, because ultimately they can be found everywhere, as the whole world is moving toward the future. However, areas such as science and technology, art, media, advertising, and literature are particularly productive. At the top of the list is, how could it be otherwise, science fiction (SF), which has a genre-specific characteristic that is virtually predestined to serve as a source of content, namely novelty. The literary scholar Darko Suvin defines it as a "necessary condition for SF" (Suvin, 1979, p. 65). This refers to the fictional invention that generally results "a change in the whole universe of the tale, or at least of crucially important aspects thereof" (Suvin, 1979, p. 64). Of course, scientific and technological novelties are particularly evident, but they can also be of a cultural, social, or political nature.

The search for portents has gradually established itself as a method of futurology. "Among the millions of signs of the present, the futurologist must identify those that can and may be used as omens for the future—he searches for the 'portents,'" explains futurologist Karlheinz Steinmüller (2007, p. 164). The "utopian potential" has proven to be a key criterion that contributes to the identification of portents. The more improbable an idea of the future appears to be, the more worthwhile it is to analyze it. James Allen Dator, Director at the Hawaii Research Center for Futures Studies at the University of Hawaii in Manoa and a kind of eminence

grise of futurology, has established several laws for dealing with the future. His second law is:

> Any useful idea of the future should seem ridiculous. Always keep this in mind: If a supposed future statement seems useful to you, it probably refers to the present and is therefore not very useful. If it shocks or outrages you or seems like ridiculous science fiction, it may actually relate to the future and therefore be useful. (Dator, 2012, p. 27)

The ridiculousness mentioned by Dator corresponds in turn with the impossibility prognosis, which plays an inglorious role in the history of ideas and technology. This is because it is a constant companion of the conceivable, the new, the utopian counter-design to the established, which, as Kathrin Passig writes, is brought into the field downright "reflexively" (Passig, 2013, p. 8). The so-called involved experts in particular like to attest that disruptive future concepts have no chance of being realized (Flessner, 2020, p. 233). They often use the "science fiction argument," which is diametrically opposed to Dator's "ridiculousness argument" (Flessner, 2020, p. 238). According to many experts, the fact that a future concept is a novelty from science fiction is a clear indication of the impossibility of its realization. From powered flight and the pocket telephone to space travel, nanotechnology, or artificial intelligence, almost every innovation that is taken for granted today was considered impossible by the experts involved before it was realized (Flessner, 2020). Futurologist and science fiction author Arthur C. Clarke gleefully quotes expert opinions on the impossibility of various technologies that are taken for granted today, such as space travel or nuclear power. In this regard, he speaks of the "hazards of prophecy" (Clarke, 1999, p. 9).

2 Autonomous Weapons in SF—An Exemplary Selection

Historical portents are usually easy to identify, as their nature has been proven by a corresponding development. It has long been known that Jules Verne's novels *From the Earth to the Moon* (1865) and *Journey around the Moon* (1870) represent a veritable treasure trove of portents (Flessner, 2020). However, Jules Verne's work is not just about space travel, but also about autonomous weapons systems. His novel *The Astonishing Adventure of the Barsac Expedition*, published posthumously in 1919 and revised by his son Michel Verne, describes a new type of weapon:

> As if obeying this command of its own accord, a bizarre instrument emerged from the base of the tower and detached itself from it. It was a kind of vertical cylinder whose opening facing the ground widened into a cone shape. At the other end, four propellers, one horizontal, the other vertical, moved with dizzying speed. The strange machine rose into the air and at the same time moved away in the direction of the perimeter wall. When it had reached it, or perhaps had already left it a few meters behind, it fell into a horizontal flight and followed exactly the periphery of the factory. But the first machine had already been followed by a second, then a third, and others followed. [...] 'These are my wasps,' Camaret explained, emphasizing the possessive pronoun a little. (Verne, 1978, p. 385)

Today, the Wasps are easily recognizable as drones that attack their opponents from the air with a "hail of cartridges" that does not fail to have an effect. At the end of their mission, each wasp returns to base to pick up "a new load" and continue the mission (Verne, 1978, p. 386). The novel does not explain exactly how the inventor Camaret's wasps are instructed, but it can be assumed that they act largely autonomously.

This also applies to the drones that the comic book hero Anthony "Buck" Rogers, invented by the American author Philip Francis Nowlan, has to deal with in his first adventure (Flessner, 2019, p. 24). The story *The Airlords of Han* appeared in the August 1929 issue of *Amazing Stories* magazine, edited by Hugo Gernsback. In this case, the drones are called "air balls" because they are spherical missiles:

> Repair men who shot up the shafts a few minutes later to bring new broadcast lamps to replace those which had been shattered, reported what seemed to be a sphere of metal, about three feet in diameter, with a four-inch lens in it, floating slowly down the shaft, as though it were some living creature making a careful examination, pausing now and then as its lens swung about like a great single eye. The moment this 'eye' turned upon them, they said, the ball 'rushed' down on them, crushing several to death in its vicious gyrations, and jamming the mechanism of the elevator, though failing to crash through it. Then, said the wounded survivors, it floated back up the shaft, watchfully 'eyeing' them, and slipped off to the side at the wrecked level. The next night several of these 'air balls' were seen, following explosions in various towers and sections of the city roof and walls. In each case repair gangs were 'rushed' by them, and suffered many casualties. On the third night a few of the air balls were destroyed by the repair men and guards, who now were equipped with disintegrator pistols. (Nowlan, 1929, p. 1126)

Since these first appearances, drones of all kinds have been a staple of science fiction and have been used in various contexts, of which military is just one. Drones also find their way into film and appear in *Earth* vs. *the Flying Saucers* by Sears (1956). They are used by the aliens to spy on the development of new types of weapons. In the laboratory, they act so quickly that they could easily be mistaken for autonomous, which also corresponds to the aliens' technical skills.

The "Nomad" probe, which Captain James T. Kirk has to deal with in the *Star Trek* episode *The Changeling* (Daniels, 1967), can also be interpreted as an intelligent and autonomous drone. After wiping out all life in a star system, the *Enterprise* becomes the new enemy of the probe, which, as it turns out, was built on Earth in 2020. Its mission was to discover new life forms in the Milky Way. However, the encounter with the robot "Tan-Ru," whose task is to store sterile soil samples, has led to a kind of update of Nomad's program, which has been destroying all imperfect life forms since the encounter with Tan-Ru. He is capable of doing this militarily. Even the *Enterprise* has nothing to oppose him with. However, Kirk manages to outwit Nomad by proving to the drone that it is imperfect itself, whereupon it eliminates itself (Sander, 1989, p. 73).

A more recent and well-known example is the fifth episode of the *Star Wars* saga, *The Empire Strikes Back* (Kershner, 1980). While the rebels are hiding on the ice planet Hoth, Darth Vader has drones search the planets in question. One of these drones also lands on Hoth and finds the rebels. But it is also discovered and immediately opens fire. Before Han Solo and Chewbacca can destroy it, it reports its

discovery to Darth Vader. The use of such an armed search drone only makes sense if it can operate autonomously. Any kind of remote control would be a hopeless endeavor given the gigantic distances involved. The main tasks of drones in science fiction are reconnaissance and combat missions. As such, they have long anticipated current military deployment concepts and have proven to be portents.

3 Robots and Androids

The classic nova definitely includes the robot, the modern version of the second creation (Flessner, 2000). Its use also repeatedly demands gradual or comprehensive autonomy. The first armed robot in film history impresses with its ability to be autonomous. Based on the novel *Farewell to the Master*, written by Harry Bates in 1940, Robert Wise made the 1951 film *The Day the Earth Stood Still* (Wise, 1951) about the landing of a humanoid alien in Washington. The visitor named "Klaatu" is accompanied by the giant robot "Gort," who represents an absolute power of order that guarantees universal peace. Klaatu explains: "In matters of aggression, we have given them absolute power over us." In order to stand up to any aggressor, the robots are armed with laser weapons and are naturally capable of acting autonomously. This autonomy is even more pronounced in the novel. There, the robot is called "Gnut" and explains to a reporter that Klaatu is not in charge at all: "You misunderstand, I am the master" (Bates, 2013, p. 82).

Another autonomous robot used as a combat robot can be seen in the 1965 episode *The Cybernauts* of the British series *The Avengers* (Hayers, 1965). This time, the opponent of agents John Steed and Emma Peel is electronics entrepreneur Dr. Clement Armstrong, who uses humanoid killer robots to carry out his economic plans. At first, these robots follow a predetermined program, but then Armstrong uses a new type that has a "brain of its own" and thus makes its own decisions.

The American Philip K. Dick presents combat robots for military use in the story *Second Variety* from 1952. The story takes place after a third world war on a largely destroyed Earth where the two opponents, Russians and Americans, continue to fight each other. Their primary weapons are highly developed combat robots that repair and develop themselves. But it is precisely this characteristic that becomes an existential problem for both sides, as the military lose control over the technological evolution of the robots. The robots are increasingly adapting their appearance to that of humans so as not to be recognized as robots. The tricks with which the robots overcome their human opponents, who are soon no longer able to distinguish between man and machine, become ever more subtle. Eventually, they emancipate themselves completely from their creators and defeat them. The only hope left for the humans is the secret moon base. Tasso, the last survivor, is to fly there. But Tasso, of all people, turns out to be the long-sought "variant two" of the robot evolution. However, the mortally wounded Hendricks is left with one satisfaction on Earth. He discovers that the different robot variants have long since begun to fight each other. The story ends with his final thought: "They were already beginning to

design weapons to use against each other" (Dick, 1953, p. 144). Although the autonomous, self-reproducing, and self-developing robots initially prove to be a military success, their autonomy eventually turns against their creators and then against themselves.

Something similar also happens to the humans in Dick's 1968 novel *Do Androids Dream Of Electric Sheep?*, which was made into a film by Ridley Scott in 1982 under the title *Blade Runner* (Scott, 1982). This time they are not machine robots, but products of synthetic biology, called "androids" in the novel and "replicants" in the film. They are bred for use, including combat, on the colonies in space. The lifespan of these androids is limited. As some of these artificial creatures always manage to get to Earth, special android hunters are deployed there to kill them. The hero of the novel, Rick Deckard, is one such hunter. His current mission presents him with a major problem, because "Nexus-6," the latest model, is virtually indistinguishable from a human, which also applies to its intellectual abilities:

> Nexus-6 did have two trillion constituents plus a choice within a range of ten million possible combinations of cerebral activity. In .45 of a second an android equipped with such a brain structure could assume any one of fourteen basic reaction-postures. Well, no intelligence test would trap such an andy. (Dick, 1982, p. 25)

On the contrary, the Nexus-6 model has long since outstripped humans: "The servant had become smarter than his master in some respects" (Dick, 1982, p. 26). Like many of Dick's works, this novel is a multi-layered puzzle that is capable of irritating the reader. In the course of the plot, the suspicion arises that Rick Deckard is so successful despite the superiority of the Nexus 6 model because he hunts his own kind, as director Ridley Scott explains: "Deckard is a Nexus 7, he probably has an unknown life span and therefore is starting to get awfully human" (Greenwald & Scott, 2007). The replicants have been developed into superhumans, but apart from a few specific abilities, they are indistinguishable from humans.

The design of the androids used by the Trade Federation in *Star Wars: Episode I—The Phantom Menace to occupy the planet Naboo* (Lucas, 1999) is completely different. The invaders send an entire droid army, consisting of mass-produced humanoid combat robots, which march into battle in anachronistic-looking phalanges. However, they are armed with modern ray or particle weapons. In this and the subsequent films in the Star Wars series, different humanoid combat robots appear that repeatedly act autonomously. For the humans and aliens involved in the battle, the use of robots has long been a familiar military strategy.

4 Nanorobots and Swarm Intelligence

In the novel *Niezwyciężony* (*The Invincible*) by Polish author Stanisław Lem, published in 1964, the completely autonomous opponent is not even remotely known. The spaceship *Kondor*, which has landed on the planet Regis III, is missing. A rescue mission is launched with the sister ship *The Invincible*. The astronauts find the

missing ship, but no survivors. What's more, the *Kondor* is badly damaged despite all its technical refinements. Only gradually does the enemy reveal itself on the largely lifeless planet. A swarm intelligence of tiny micro-robots controls the entire surface and attacks every living creature and almost all technical artifacts. These techno-organisms are the result of a techno-evolution. Their initiators—the inhabitants of Regis III—have long since died out and were in all probability wiped out by the swarm creatures. As the attacks by the swarm intelligence become more frequent, the crew deploys the Cyclops, an autonomous combat robot equipped with an anti-matter launcher. Like the spaceship itself, it too is considered invincible. Billions of the swarm particles crash into the Cyclops and are destroyed billions of times over.

> If the attackers had been living beings, the massacre to which they fell would probably have caused the following ranks to turn back or at least forced them to stop before the blazing hell. Here, however, the dead fought against the dead; the atomic fire was not extinguished, but only changed the shape and direction of the main attack. (Lem, 1971a, p. 111)

In the end, the sheer mass of the swarm particles forces victory as the rising temperature also affects the Cyclops' AI. Suddenly, the combat robot destroys the drones sent out and approaches the spaceship, which is preparing an alarm launch. The swarm intelligence has eliminated the enemy's AI, whose armor and weapons could not be defeated. Ultimately, one artificial intelligence fought against another, while the humans had to resign themselves to the role of observer. *The invincible* leaves the planet as the loser. The hubris of man is exposed as such.

Lem presents another nanoweapon in his 1958 novel *Eden*, which is used to attack the crew of an earthly spaceship that has made an emergency landing on an unknown planet. The inhabitants of the planet defend themselves by planting automatically acting nanoparticles around the spaceship, which automatically erect an almost impenetrable isolation barrier. "Inorganic germs," the terrestrial cyberneticist states matter-of-factly and goes on to determine that the autonomous particles can process and convert almost any source material (Lem, 1974, p. 204).

In Lem's 1969 story *Opowiadania* (*Night and Mold*), synthetic microorganisms are also at the heart of the story. The pseudobacterium *Whisteria Cosmolytica*, generated in a military laboratory, is capable of dissolving all matter. "Because the Whisteria [...] destroys matter. Synthesis of antiprotons, encapsulation in a force field, division—this is the life cycle of Whisteria," explained one of the scientists (Lem, 1971b, p. 10). However, the autonomous and ultimate weapon is not yet fully developed when it is released in an accident. The revocation command, which is indispensable for such a weapon, has not yet been applied. *Whisteria* is therefore unstoppable. And despite all the immediate measures taken, one specimen escapes and ends up with a hermit who lives not far from the laboratory. The unsuspecting man discovers the pseudobacterium by chance and observes its conspicuous division, which he finds aesthetically pleasing. Only when his house collapses does he realize the danger.

Another variant of autonomous nanoparticles is described by the American writer Christopher Anvil (actually Harry Christopher Crosby) in the short story

Uncalculated Risk, published in 1962. In this case, in the context of the Cold War, the scientist Dr. Green develops specific nanorobots, called "texturing agents," which can remove all forms of dirt and grime or convert them into useful substances such as good soil. However, he makes a small mistake during field trials, which causes his creation to multiply uncontrollably. "As far as I can see, the process would not be stopped. It would be accelerated," explains Dr. Green to the responsible General Lyell Berenger (Anvil, 2010, p. 117). All that will remain of the world is a formless mass, which Anvil describes as a "brown-gray glop" (Anvil, 2010, p. 122). The parallels to Lem's story cannot be overlooked. Anvil's nanobots are also conceivable as autonomous weapons.

Basically, most of the motifs and images had already been assembled and explored when the American nanotechnologist Eric Drexler presented his scientifically sound visions in his 1986 book *Engines of Creation: The Coming Era of Nanotechnology* in 1986 (Drexler, 1986). In essence, it is about the construction of self-reproducing nanomachines, also known as nanobots or assemblers, which can ultimately produce any object from atoms and molecules.

In Lem's 1985 novel *Pokój na ziemi* (*Peace on Earth*), the entire arms race of the bipolar world is outsourced to the Moon and takes place there in autonomous processes. This means that not only the weapons are autonomous, but also their development and further development. What exactly happens in the zones of the respective nations is and remains secret in order to maintain the balance of power. The lunar arms race is monitored by the UN's *Lunar Agency* and is intended to protect the world from surprise attacks. The necessary inspections are carried out by automatic probes and robot.

However, when they suddenly disappear and the situation reports from the Moon fail to materialize, Lem's tried and tested hero Ijon Tichy is tasked with an inspection trip. At first, Tichy only uses delegates or 'sentlings' (*Sendlinge*), remote-controlled avatars that transmit their sensory impressions to the astronauts in the spaceship. After all the available transmitters have failed, Tichy lands on the Moon in person, despite being banned from entering, in order to establish what he already knew. As Tichy reports, the once still reasonably manageable techno-evolution on the Moon has eluded all control, all probability, and all prognosis:

> On the Moon, the programs of different sides have wedged and intertwined, they have mixed, and who started first is unimportant, at least now. To put it simply, the effect there is a kind of cancer covering the surface. Mutual, disorderly annihilation, various phases of different simulation and arms production, everything has penetrated, overlapped, attacked, countered—call it what you will. (Lem, 1986, p. 219)

The moon landing has consequences, however, because a seemingly harmless dust has attached itself to Ijon Tichy's spacesuit, which only later reveals its true identity: "These are not grains of dust, but silicone polymers. As the experts claim, the beginnings of an ordogenesis, a necro-organization" (Lem, 1986, p. 219). These self-reproducing nanorobots, later called "selenocytes," continue the battle begun on the Moon and destroy all modern technologies—with fatal consequences: "We have fallen back into the first half of the twentieth century. In technical terms and in

general" (Lem, 1986, p. 266). This autonomously acting weapon is also extremely effective, as it targets complex, technical structures from which it removes this very complexity. However, since all modern weapons and all modern infrastructures are computer-based, only the mechanical remains intact. All computer-based warfare is suddenly impossible. From drones and satellite-based artillery to modern combat aircraft and tanks, everything freezes into immobility.

The novel *Retour à "0"* (*Inferno Moon*) by French author Stefan Wul (actually Pierre Pairault), published in 1956, features similar motifs. With the help of specific rays, scientists on Earth construct tiny weapons and radio devices to take action against a total regime on the Moon. What is particularly interesting is the ability of these micromachines to reproduce themselves. To liberate and pacify the Moon, vast quantities of these micromachines are by no means necessary:

> A single speck of dust in every city, a single radio, will multiply infinitely and will be enough to infect the entire Moon. This dust will creep in everywhere like microbes. (Wul, 1956, p. 135)

5 Enhancement—The Modified Human

The American science fiction author Ted Chiang presents a completely different kind of autonomous weapon in his short story *Understand*, published in 1991. After an accident, a new type of therapy is tested on the first-person narrator Greco to regenerate his damaged brain. It soon becomes clear that the therapy is capable of greatly enhancing his cognitive abilities. The CIA immediately gets involved "to carry out further tests, perhaps on other patients if the test results are positive" (Chiang, 2014, p. 19). But thanks to his superior and constantly evolving abilities, Greco eludes the grasp of the military and the secret service. Their interest is all too understandable, as therapeutically induced enhancement can be used as an innovative weapon. Greco effortlessly misleads his opponents, which justifies the CIA's interest in weaponry. They can not even begin to catch his scent. After just a few feints and maneuvers, his invincibility becomes apparent. But then the tide turns, because after one of his clever stock deals, he realizes: "There's someone else out there like me" (Chiang, 2014, p. 39).

> Greco soon realizes that Reynolds, the name of the man also being treated with the new drug, is his opponent. This leads to an unusual duel, fought primarily with cognitive means, because Greco loses: "I capitulate to his great genius." (Chiang, 2014, p. 50)

So, the CIA was right. As an autonomous weapon, a person with such an enhancement, who has a perfect command of the digital world and is far superior both strategically and tactically, is almost impossible to stop, especially as he is outwardly indistinguishable from a normal person. Only his peers can defeat him.

6 Conclusions

This small, exemplary, and incomplete selection of novas shows the state of autonomous weapon systems in science fiction. They can all serve as portents, i.e., as signs of possible futures. If this approach is followed, a wide variety of autonomous weapons will be generated and will naturally be used. The futures sketched out in science fiction will therefore come to pass, not one-to-one, but in comparable quality. Science fiction definitely provides a glimpse of what is to come and is therefore also treated as an important source of content by futurology. There are no doubts about the implementation of the technical options and their use in the science fiction drafts presented here. The authors discuss less ethical and moral issues, but rather the potential danger of a loss of control and the emancipation of weapons systems from their creators and commanders.

- Science fiction provides images of the future that deliver portents for futurology.
- The images of the future generated by science fiction have a high degree of probability.
- The range of possible autonomous weapon systems is very large.

References

Anvil, C. (2010). Uncalculated Risk. In C. Anvil (Ed.), *War Games* (pp. 107–132). Baen.

Bates, H. (2013). *Farewell to the Master*. No Place Given: Spastic Cat Press.

Bloch, E. (1986). *The Principle of Hope*. (N. Plaice, S. Plaise, & P. Knight, Trans.). MIT Press.

Chiang, T. (2014). Verstehen (K. Will, Trans.). In T. Chiang (Ed.), *Das wahre Wesen der Dinge* (pp. 7–50). Golkonda.

Clarke, A. C. (1999). *Profiles of the Future. An Inquiry into the Limits of the Possible* (millennial ed.). Victor Gollancz.

Daniels, M. (Director). (1967, September 29). The Changeling (Season 2, Episode 3) [TV series episode]. In G. L. Coon, J. M. Lucas, & F. Freiberger (Producers), *Star Trek*. Desilu Productions; Paramount Television; Norway Corporation.

Dator, J. (2012). Der Blick in die Zukünfte. In A. G. Deutsche Post (Ed.), *Delivering Tomorrow. Logistik 2050. Eine Szenariostudie* (pp. 20–27). Deutsche Post AG. https://www.post-und-telekommunikation.de/PuT/1Fundus/Dokumente/Studien/Postdienste/Zukunftsstudie_Logistik_2050/Zukunftsstudie_Logistik_2050.pdf

Dick, P. K. (1953). Second Variety. *Space Science Fiction, 1*(6), 102–144.

Dick, P. K. (1982). *Blade Runner (Do Androids Dream of Electric Sheep)*. Ballantine, Del Rey (Original work published 1968).

Drexler, E. (1986). *Engines of Creation. The Coming Era of Nanotechnology*. Anchor Books.

Esposito, E. (2007). *Die Fiktion der wahrscheinlichen Realität*. Suhrkamp.

Flessner, B. (2000). Antizipative Diffusion. Science Fiction als Akzeptanzbeschleuniger und Wegbereiter einer multitechnokulturellen Gesellschaft. In B. Flessner (Ed.), *Nach dem Menschen. Der Mythos einer zweiten Schöpfung und das Entstehen einer posthumanen Kultur* (pp. 245–264). Herder.

Flessner, B. (2019). Fabbing, Drohnen, Wunschmaschinen. Comic als Archiv historischer Prognostik. In C. Heydenreich (Ed.), *Comics & Naturwissenschaften* (pp. 17–34). Berlin: Bachmann.

Flessner, B. (2020). Implizierte Prognosen. Anmerkungen zum Verhältnis von Möglichkeits- und Wahrscheinlichkeitsraum in Science Fiction und Wissenschaft. In M. Jungert, A. Frewer, & E. Mayr (Eds.), *Wissenschaftsreflexion. Interdisziplinäre Perspektiven zwischen Philosophie und Praxis* (pp. 231–250). Brill, mentis. https://doi.org/10.30965/9783957437372_010

Greenwald, T., & Scott, R. (2007, September 26). Ridley Scott Has Finally Created the Blade Runner He Always Imagined. Wired, https://www.wired.com/2007/09/ff-bladerunner

Hayers, S. (Director). (1965, Oktober 12/16). The Cybernauts (Season 4, Episode 3) [TV series episode]. In B. Clemens, & J. Wintle (Producers), *The Avengers*. ABC Television; Associated British Picture Corporation; ABC Weekend TV; ABC Television Ltd.

Hölscher, L. (2016). *Die Entdeckung der Zukunft*. Wallstein.

Homburg, C., & Krohmer, H. (2006). *Grundlagen des Marketingmanagement. Einführung in Strategie, Instrumente, Umsetzung und Unternehmensführung*. Springer Fachmedien. https://doi.org/10.1007/978-3-658-13654-3

Horkheimer, M., & Adorno, T. W. (2002). *Dialectic of Enlightenment. Philosophical Fragments* (E. Jephcott, Trans.). Stanford University Press (Original work published 1947).

Kershner, I. (Director). (1980). *Star Wars: The Empire Strikes Back* [Film]. Lucasfilm Ltd.; 20th Century-Fox.

Kosow, H., & Gaßner, R. (2008). *Methoden der Zukunfts- und Szenarioanalyse. Überblick, Bewertung und Auswahlkriterien (ITZ workshop report no. 103)*. Institut für Zukunftsstudien und Technologiebewertung ITZ. https://www.researchgate.net/publication/262198781_Methoden_der_Zukunfts-und_Szenarioanalyse_Uberblick_Bewertung_und_Auswahlkriterien

Lem, S. (1971a). *Der Unbesiegbare*. Fischer.

Lem, S. (1971b). Nacht und Schimmel. In S. Lem (Ed.), *Nacht und Schimmel* (pp. 7–28). Suhrkamp.

Lem, S. (1974). *Eden*. dtv.

Lem, S. (1986). *Frieden auf Erden*. Insel.

Lucas, G. (Director). (1999). *Star Wars: Episode I—The Phantom Menace* [Film]. Lucasfilm Ltd.; 20th Century Fox.

Mayo, B. (1986). Traces and Portents. *The Philosophical Quarterly, 18*(73), 289–298. https://doi.org/10.2307/2217790

Nowlan, P. F. (1929). *The Airlords of Han. Amazing Stories, 3*(12), 1106–1136.

Passig, K. (2013). *Standardsituationen der Technologiekritik: Merkur-Kolumnen*. edition unseld.

Sander, R. (1989). *Das Star Trek Universum*. Heyne..

Scott, R. (Director). (1982). *Blade Runner* [Film]. The Ladd Company; Shaw Brothers; Warner Bros.

Sears, F. F. (Director). (1956). *Earth vs. the Flying Saucers* [Film]. Clover Productions; Columbia Pictures.

Steinmüller, K. (2007). Zeichenprozesse auf dem Weg in die Zukunft: Ideen zu einer semiotischen Grundlegung der Zukunftsforschung. *Zeitschrift für Semiotik, 29*(2–3), 157–176.

Suvin, D. (1979). *Metamorphoses of Science Fiction. On the Poetics and History of a Literary Genre*. Yale University Press.

Verne, J. (1978). *Das erstaunliche Abenteuer der Expedition Barsac*. Diogenes.

Wise, R. (Director). (1951). *The Day the Earth Stood Still* [Film]. 20th Century Fox.

Wul, S. (1956). *Retour á "O"*. Fleuve noir.

The SMART Initiative: Political Limits of Using Artificial Intelligence on the Battlefield of the Future?

Ferdinand Gehringer and Alexander Schuster

Abstract This article examines the use of artificial intelligence (AI) in the military sector and the associated political challenges. It describes the current applications of AI in data analysis, process optimization, decision-making, reconnaissance, as well as cyber and communication security. Additionally, the future use of autonomous weapon systems (LAWS) and swarm technologies is discussed. A central point is the lack of international regulation and the differing positions of organizations such as the United Nations, NATO, the European Union, and individual countries like the USA, China, and Russia. The text emphasizes the need for a new initiative, the Strategic Military AI Regulation Talks (SMART), to achieve a global, dynamic, and transparent regulation. This initiative aims to establish clear definitions, operational scenarios, and responsibilities, and to involve as many countries as possible. SMART aims to regulate the proliferation and use of LAWS and to uphold ethical principles by ensuring human control over lethal decisions. The initiative is to be supported by practical simulations and international cooperation to effectively address the challenges and risks of AI in the military context.

The battlefield is changing. Technologies are developing at a rapid pace. Elements supported by Artificial Intelligence (AI) are increasing. The use of AI in weapon systems and military strategies has the potential to fundamentally change the way wars are fought.

F. Gehringer (✉)
Department International Politics and Security Affairs, Division Analysis and Consulting, Konrad-Adenauer-Stiftung e.V. (KAS), Berlin, Germany
e-mail: ferdinand.gehringer@kas.de

A. Schuster
Former Department International Politics and Security Affairs, Division Analysis and Consulting, Konrad-Adenauer-Stiftung e.V. (KAS), Berlin, Germany
e-mail: alexander.schuster@heckler-koch-de.com

K. Talves, D. Spreen (eds.), *Artificial Intelligence in Military Technology*,
Artificial Intelligence, Simulation and Society 192,
https://doi.org/10.1007/978-3-031-95578-5_5

However, there is no international consensus on the use and regulation of Artificial Intelligence in weapon systems at the political level. The political debate on the use and potential regulatory limits for Artificial Intelligence is complex and multifaceted. It ranges from the question of accountability for AI-controlled actions to the need for international norms and agreements to limit its use on the battlefield.

This chapter examines the international political debate in more detail. First, the previous use of AI in the military is illuminated before presenting the topics of discussion and the positions of selected organizations (United Nations, NATO, European Union, G7, OECD, and APAC) and countries.

Finally, a proposal is made on how to advance the stalled discussions at the international level and reach a consensus on the scope and regulation of the use of Artificial Intelligence in weapon systems. Additionally, the SMART Initiative will be explained.

1 Use of Artificial Intelligence in the Military

Artificial Intelligence is already being used in a variety of ways in the military context (NATO, 2021). It is used in reconnaissance and analysis, process automation, and decision-making. It is no longer only used for pure data analysis and reconnaissance but is increasingly considered in the planning and execution of operations.

1.1 Data Analysis

Artificial Intelligence supports the analysis of large data sets. The use of drones for reconnaissance, the carrying of cameras by soldiers, and the capture of the battlefield by satellites generate a very large amount of data. Data analysis tools collect and consolidate these large data sets (BWI GmbH, 2022). Artificial Intelligence promotes faster processing and more accurate evaluation of data, which can lead to critical information for decision-making (Bundeswehr, 2019, pp. 24–25). Situations are analyzed more quickly. Threats to one's own forces can be identified earlier and damage cases reduced. Strategic and especially operational planning are supported.

1.2 Process Optimization

Artificial Intelligence is also used to optimize processes. The technology enables the simulation of complex scenarios. It can consider historical data and information and use these for the simulation of future possible scenarios. The simulation helps prepare strategic decisions. The accuracy and effectiveness of operation plans can

be increased. This allows for more targeted use of personnel and/or material resources and relieves data analysts.

In logistics, the use of Artificial Intelligence can optimize the logistics chain (Bundeswehr, 2019, p. 25). It can calculate needs, material shortages, or savings potential as well as predict the need for maintenance work on military equipment based on wear and tear. This not only saves material but also optimizes logistical (planning) processes.

1.3 Decision-making

AI systems are increasingly used in decision-making. They assist military leadership (Bundeswehr, 2019, p. 4). Complex scenarios and situations are simulated and presented more simply through AI, and individual courses of action are suggested. Algorithms can evaluate individual options and make recommendations based on these evaluations.

Weapon systems are also being trained through machine learning (Bundeswehr, 2019, p. 12). By continuously feeding new data, these systems can automatically identify targets, formulate action decisions, and execute them independently. This allows for independent decision-making and automated task execution.

1.4 Reconnaissance

AI-equipped drones can capture situations more quickly and accurately. These drones use advanced algorithms and sensors to navigate autonomously, make decisions, and perform complex tasks. They provide a precise and quick way to conduct reconnaissance missions, identify targets, and execute attacks without direct human intervention. They can be used to reconnoiter enemy territory and capture situations without exposing human lives to danger (Scharre, 2018, pp. 152–174).

1.5 Cyber and Communication Security

The use of Artificial Intelligence contributes to increased cyber and communication security. Software vulnerabilities can be identified more easily, network structures better protected, and cyber attacks repelled (Grimm, 2023, pp. 123–137). A model is trained with data describing the normal operation of an IT system. Once the model detects deviations due to changed data, it triggers an alarm and can also initiate automated actions (Schmidt, 2020, pp. 70–72). The rapid and automated detection of unusual data transfers allows for quicker identification and repulsion of attacks on IT systems. This increases the level of cyber security and strengthens communication security.

The use of Artificial Intelligence in the military already enhances the speed of action and decision-making, efficiency, and safety. Additionally, personnel and material resources are conserved, which can be used for other purposes in the future.

2 Future Use of Artificial Intelligence in the Military

The use of lethal autonomous weapon systems (LAWS), also known as 'killer robots,' revolutionizes warfare by enabling operations with unprecedented precision and speed. LAWS are weapons that can identify and attack targets without human intervention. They combine some of the previously mentioned applications and benefits of Artificial Intelligence.

They allow for more precise execution of operations and are not subject to human fatigue or distraction. At the same time, these systems can be used in areas and under conditions that are fundamentally dangerous to human life, such as contaminated areas, mined or extreme weather zones (U.S. Department of Defense, 2019, pp. 15–18; Cummings, 2017, pp. 21–24). Another benefit of LAWS is their ability to process large amounts of data in real time, allowing for quick adaptation to dynamic combat scenarios. They can also operate effectively in complex environments by using advanced sensors and algorithms, leading to greater efficiency and effectiveness in military operations.

Whether in the form of drones or robots, these systems are continuously evolving and will shape the battlefield of the future (Scharre, 2018, pp. 201–203).

Another developmental stage is swarm technology. Supported by Artificial Intelligence, the use of various autonomous devices such as drones, robots, or vehicles on land, water, or air can be synchronized through combat clouds. Synchronization enables these autonomous systems to act in a coordinated manner in a swarm (David & Nielsen, 2016, pp. 65–69).

Individual units communicate across dimensions and can act quickly and flexibly. They align their behavior with each other. There is no longer a need for central deployment coordination. Decentralized command control leads to greater robustness. The failure of individual elements does not result in the overall system's incapacity. The fact that swarm formations can operate independently of human operators leads to a reduction in the burden on forces and changes the battlefield. Swarm technologies on the battlefield affect all levels of military planning and operation. The degree of automation is increasing, and it is technologically possible to deploy fully automated systems.

3 Presentation of the Political Debate Status

While the battlefield of the future will include and be shaped by fully automated weapon systems, there is currently no regulation of the use of Artificial Intelligence in the military context.

International discussions and negotiations on a regulatory solution are stalled at the political level (Dahlmann et al., 2021). There is still no clarity on how far the use and utilization of Artificial Intelligence in weapon systems should be regulated and limited. Critical discussions mainly focus on fully autonomous offensive weapon systems that can make attack and kill decisions independently. Regulating the use of autonomous weapon systems is challenging because a clear distinction between "worrisome" and "non-worrisome" forms of autonomy is required.

3.1 United Nations

Since 2014, the United Nations (UN) has been intensively addressing the issue of autonomous weapon systems (AWS). The first warning about autonomy in weapon systems was issued by the UN Special Rapporteur on extrajudicial, summary, or arbitrary executions, Christof Heyns, to the UN Human Rights Council (Heyns, 2013, pp. 20–21). Consequently, an expert group with non-binding recommendations was established. Since 2017, the Group of Governmental Experts (GGE) has been discussing AWS within the framework of the United Nations Convention on Certain Conventional Weapons (CCW, Hoffberger-Pippan et al., 2022, p. 1).

However, there are significant disagreements within the GGE, particularly regarding the definition of AWS and the required level of human control over machines (Dahlmann et al., 2021, p. 1). Most states prefer the term 'human control,' while the USA prefers the concept of 'appropriate human judgment.' Another problem is the rivalry between major powers: China seeks to become an AI superpower, and its definition of AWS raises concerns that it may not adhere to regulations. Although Russia has stated that adequate human influence over AWS is essential, it opposes further regulations under international law beyond the existing framework.

UN Secretary-General Guterres has taken the position since 2018 that fully autonomous weapon systems that operate without human control or supervision and cannot be used in accordance with humanitarian international law are politically unacceptable and morally reprehensible. Therefore, they should be banned under international law (United Nations, 2018, pp. 67–71). In his *New Agenda for Peace 2023*, the Secretary-General reiterated this call. He recommended that states adopt a legally binding instrument to ban autonomous weapon systems that operate without human control or supervision by 2026 and regulate all other types of autonomous weapon systems (United Nations, 2023, p. 27).

3.2 NATO

NATO also plays an active role in the discussion about AI and autonomous weapon systems. In September 2020, Mircea Geoană called on NATO to invest in modern technologies like AI to remain competitive and defend the democratic order during

the online forum *HumanAIze* (Geoană, 2020). Geoană pointed out that the use of AI could not only improve the efficiency and responsiveness of NATO forces but also be crucial to remaining competitive against potential adversaries investing in these technologies. In October 2021, NATO defense ministers adopted a NATO AI Strategy, agreeing on a responsible use of AI (NATO, 2021, pp. 3–5). This strategy includes several key elements:

- Ethics and Accountability: NATO commits to an ethical and responsible use of AI to ensure that all AI systems are developed and used in accordance with international law, including humanitarian international law.
- Interoperability and Cooperation: The strategy promotes cooperation and information exchange among member states to ensure the interoperability of AI systems. This ensures that the various national AI developments are compatible and can work together effectively.
- Protection and Defense: NATO emphasizes the importance of protection against cyber threats and the defense of critical infrastructures through the use of advanced AI technologies. This includes measures to detect and repel cyber attacks that may target AI systems themselves.
- Innovation and Technological Leadership: The strategy supports continuous research and development in the field of AI. This includes investments in educational programs and the promotion of cooperation with industry and academic institutions.
- Transparency and Trust: NATO is committed to creating transparency regarding the use of AI systems to gain and maintain public and international community trust. This is enabled by publishing guidelines and reports on the use of AI in military contexts.

The strategic measures of NATO aim to proactively address the opportunities and challenges associated with the use of AI and autonomous systems and to continuously develop NATO's capabilities in this crucial area.

In February 2023, the NATO Data and Artificial Intelligence Review Board (DARB) began developing standards for the user-friendly and responsible use of Artificial Intelligence (NATO, 2023). The standards aim to help companies and institutions design future AI and data projects in accordance with international law and NATO's norms and values.

3.3 European Union

The European Union has also taken significant steps to regulate the use of Artificial Intelligence and AWS and ensure that these technologies are used responsibly and in accordance with ethical principles. These measures include both political demands and legislative initiatives.

In January 2021, members of the European Parliament (MEPs) called for a comprehensive ban on LAWS and other autonomous weapon systems in a resolution titled "Artificial intelligence: questions of interpretation and application of

international law provisions" (Artificial Intelligence, 2021). This demand reflects the deep concern of the parliamentarians that such systems, which can make lethal decisions without human intervention, pose significant ethical and security risks. The MEPs argued that the use of LAWS could undermine the fundamental principles of humanitarian international law and human rights. In this resolution, the MEPs emphasized the need for strict oversight and clear regulations to ensure that the development and use of AI and autonomous systems always remain under human control. They called on EU member states and the European Commission to engage in international negotiations to achieve a binding ban on such weapon systems.

In addition to the political demand for a ban, the 27 member states adopted uniform rules for the development, marketing, and use of AI within the EU on May 21, 2024. The Artificial Intelligence Act is based on a risk-based approach and sets specific requirements for AI systems depending on their risk to society. According to the AI Regulation (Council of the European Union, 2024, Article 2, Paragraph 3), military systems are, however, excluded from the scope of the regulation.

3.4 G7

At their meeting in Kyoto 2023, the G7 countries agreed on eleven principles for regulating AI to establish guidelines for handling the technology for the first time. These "International Guiding Principles for Organizations Developing Advanced AI Systems" (Kölling & Volkery, 2023) apply not only to development but also to science, civil society, and the private and public sectors.

The agreement was challenging as the G7 countries have different views on regulation. While the EU has already adopted the AI Regulation, which regulates transparency, risk management, and controls, the USA, the UK, and Japan are more skeptical of strict regulation. Therefore, the document allows for different approaches in various jurisdictions and remains vague to offer broad interpretive leeway.

However, the EU managed to push through some key points:

- Labeling of AI Content: Companies should develop methods to label AI-generated content, such as digital watermarks, to combat disinformation.
- Accountability: Organizations must be accountable for the consequences of their AI use, and there should be control mechanisms to ensure the effectiveness of voluntary rules.
- Risk Management: Companies should identify potential risks during product development and detect weaknesses after deployment and report them publicly.
- Security: Access to AI and data should be secured against external access, and research on risk mitigation and AI security should be prioritized.
- Human Rights: The document emphasizes the adherence to universal human rights and democratic values, stating that the use of AI that violates these is unacceptable.

Although the guidelines are non-binding, they are intended to serve as a guide (Scott, 2023).

3.5 France

France sees an intense global competition for the use of Artificial Intelligence, aiming not to fall behind its competitors. In military matters, AI is not an end in itself but can help the French armed forces fulfill their tasks, according to the AI strategy from the Ministry of the Armed Forces (Ministère des Armées, 2019, pp. 3–4).

However, commercial AI is considered only limitedly useful for the military as there are specific requirements for military use, both in terms of tasks to be performed, the type of data to be processed (infrared images, radar data), and performance and robustness requirements. In other words, the more specific it becomes militarily, the more the military must develop it themselves.

The intention is to use AI to achieve speed, better reconnaissance and target acquisition, and faster and more targeted actions. At the same time, compliance with international law is to be ensured. Operations are to be accelerated and optimized. AI-supported data and threat analyses and decision-making can protect soldiers by, for example, individualizing healthcare.

However, opposing AI could be capable of predicting the French course of action, thus removing the element of surprise. The paralysis of 'command capabilities' by neutralizing, deceiving, or distracting French technologies poses a danger. The French strategy also points out that AI is not very robust yet and can be manipulated. Additionally, there are problems with machine learning, such as bias and the urgent need for useful data (Ministère des Armées, 2019, pp. 6–7).

France sees a global AI arms race, with the USA and China as leading powers in this field. Behind these two superpowers, France sees the EU as a middle power, whose strict stance on legal and ethical issues can be both a strength and a weakness. A strength in developing and using norm-setting power, supported by many actors in the public and private sectors. The risk is seen in the possibility that this could lead to restrained research and business development policies or excessive regulation hindering investments.

3.6 Germany

In Germany, the use of Artificial Intelligence in weapon systems, despite extensive public discussions on AI in general (e.g., Chat GPT), is hardly addressed. While other countries like the USA, China, and France have already developed specialized military AI strategies, Germany lacks corresponding measures. On the one hand, Germany lacks sufficient expertise in this area; on the other hand, the topic has been relatively unattractive for German politics so far (Brink, 2024). The political will to prioritize the issue of 'AI in military applications' is missing in Germany. For years, there was neither political nor public interest in the further development and modernization of the German armed forces. The need for modern equipment for occasional deployments within international crisis management was considered

secondary. Although interest in better and modern equipment for the Bundeswehr has increased following the Russian attack on Ukraine, it does not go so far as to seriously and strategically address questions about the use of AI in the Bundeswehr and in weapon systems, their limits, and the determination of scenarios. The current deficits in basic equipment that need to be caught up are too great.

The Bundestag's AI Enquete Commission already called for a security policy guideline document for the military use of AI in the last legislative period, but this has not yet been implemented. A position paper published in 2023 by the AI & Defense working group, which includes representatives from the Bundeswehr, the Fraunhofer Society, and the University of the Bundeswehr Munich, emphasized the need for a national military AI strategy (Brink, 2023).

There is broad consensus within the scientific community that Germany urgently needs such a strategy as Berlin otherwise risks losing international ground and jeopardizing the Bundeswehr's interoperability in this field within NATO in the medium term. Despite individual considerations within the Bundeswehr, there is no comprehensive guideline regulating human–machine interaction (Brink, 2023).

The traffic light coalition rejects the use of lethal autonomous weapon systems, but also does not bring forward any regulatory proposals in this context.

3.7 USA

The United States places the highest priority on the research and development of Artificial Intelligence. US federal agencies were already urged by President Trump's American AI Initiative to prioritize any investments and support in the field of Artificial Intelligence (Reitmeier, 2020, pp. 10–11).

The guiding principles for the use and promotion of Artificial Intelligence are the ten principles for regulating AI. These principles cover the most important elements of the American value system: physical and economic security, civil liberties, human rights, and the protection of intellectual property.

Compliance with the principles is monitored, among other things, by the National Security Commission on Artificial Intelligence (NSCAI).

Research and use of Artificial Intelligence in the US military to date follow Directive 3000.09, adopted in 2012. The requirements and principles set out in it are intended to minimize the likelihood and consequences of failures in autonomous and semi-autonomous weapon systems that could lead to unintended deployments (Sayler, 2024). The requirements outlined in the directive include the following points:

1. Autonomous and semi-autonomous weapon systems should be designed so that commanders and operators can exercise appropriate human judgment over the use of force.

2. Persons who authorize, direct, or operate such systems must do so with due diligence and in accordance with the law of war, applicable treaties, safety regulations for weapon systems, and applicable rules of engagement.
3. The weapon system must have demonstrated adequate performance, capability, reliability, effectiveness, and suitability under realistic conditions (Sayler, 2024).

On the one hand, the US military clearly positions itself to keep humans 'in the loop.' However, two definitional loopholes are already built into Directive 3000.09: The term "appropriateness" invites broad interpretation as it is usually at the situational discretion of the users. At the same time, the Pentagon defines "human judgment over the use of force" not as manual human control of the entire weapon system. Rather, it refers to comprehensive human involvement in decisions about how, where, when, and why a weapon is used. The deployment of the weapon system can then proceed largely autonomously. To date, the USA has not ruled out the development and use of LAWS. Numerous high-ranking representatives of the US military have already pointed out that Washington reserves the right to develop, procure, and use LAWS if geopolitical adversaries of the USA decide to procure or use LAWS.

Washington is thus interested in responsible regulation of the use of AI in the military field but keeps enough maneuvering space open to deploy LAWS if necessary.

3.8 China

The People's Republic of China pursues an ambiguous strategy regarding LAWS. On the one hand, China is the only P5 member (P5 stands for permanent five and includes the five permanent representatives on the UN Security Council: USA, China, Russia, France, and the UK) advocating for a ban on LAWS. On the other hand, it tries to avoid stronger regulation through a very narrow definition of these systems.

The Chinese definition of LAWS includes five fundamental characteristics: lethality, autonomy, impossibility of termination, impact, and evolution (Boulanin et al., 2017, pp. 14–17). This definition contrasts sharply with the definitions of other contracting states, which focus on broader technological features, particularly on autonomous target functions.

Many critics argue that the Chinese definition sets an extremely high threshold for the types of technologies that would be subject to legal regulation. The focus on full autonomy could, for example, nullify the regulation of systems with a rather low degree of autonomy (Boulanin & Verbruggen, 2017, pp. 34–36).

Despite these criticisms, China emphasizes the need for clear rules to prevent the development of this particular weapon category. Beijing argues that the rapid technological progress could soon lead to weapons without human supervision, especially in high-tech countries like the USA (Kania, 2017, pp. 18–20).

In practice, however, China invests heavily in AI research and development and sees it as a central building block of the future military. Chinese military theorists see AI as a potential key to surpassing the USA as the world's strongest military force (Allen, 2019, pp. 5–7).

The focus of the People's Liberation Army's modernization efforts has so far been on mechanization and informatization. However, increasing emphasis is now placed on the intelligentization of warfare. This essentially means the militarization of the Fourth Industrial Revolution, i.e., using the Fourth Industrial Revolution for the development of new weapon systems based on Artificial Intelligence (State Council Information Office of the People's Republic of China, 2019, pp. 24–26).

3.9 Russia

Russia recognized the importance of AI both in general and for military purposes early on. Soviet military strategists already saw the potential of new technologies such as computer science and digitization in the 1970s, which could massively change warfare through more precise and controllable weapon systems. However, these efforts ended in the 1980s due to poor economic conditions.

In 2012, Putin called for increasing Russia's defense capability with the help of science, particularly through digitization. In October 2019, he approved a new National Strategy for the Development of AI until 2030 (Mewes, 2023). This strategy sets short-term goals until 2024 and medium-term goals until 2030. For the use of AI, seven principles were named, including the protection of human rights and freedoms, security, transparency, technological sovereignty, integrity of innovation cycles, reasonable frugality, and support for competition (Office of the President of the Russian Federation, 2019).

In December 2019, AI research received the status of a strategic program. The programs adopted in 2019 have a total budget of around $26 billion. Companies engaged in AI research have the right to tax relief. Additionally, six AI research centers are funded by the government.

In the fall of 2021, the Russian IT industry created an ethics code for Artificial Intelligence. Putin had already called for standards for the use of AI in 2019. Among other things, it was decided that AI should not make the final decision. At the same time, Russia rejects a legally binding GGE framework for LAWS and refers to the responsibility of individual states (Working Paper of the Russian Federation, 2022, pp. 2–3).

In August 2023, defense companies presented their latest developments and products at ARMY 2023 (Naval News, 2023). The focus was on unmanned aerial and ground vehicles. UAVs are to go into mass production, and progress is also to be made in the field of unmanned ground vehicles. There is no concrete information on this yet, much is currently being coordinated and will be implemented in the near future.

Despite these efforts, Russia currently lacks the capacity to keep up with the West or China in developing new systems (Bitzinger & Raska, 2022, p. 313).

Instead, the focus is on improving existing systems to compensate for the strategic advantages of rivals. The Russian military shows weaknesses in this area and is dependent on partners.

3.10 Iran

Iran hardly engages in regulating autonomy in weapon systems. Although the country is not a member of the CCW, it participated in the meetings on LAWS in 2016 and 2018–2019 (Human Rights Watch, 2020, pp. 23–25). According to Human Rights Watch, Iran has neither commented on the concerns raised by the removal of human control over the use of force nor supported proposals for negotiating a new international ban treaty (Human Rights Watch, 2020, pp. 23–25). At the 78th UN General Assembly in 2023, Iran abstained from voting on Resolution L.56 on autonomous weapon systems (Statement by Iran, 2023).

Meanwhile, Iran invests heavily in AI development and sees it as a key technology in the military field that can cover other weaknesses. Iranian military literature suggests that the army, air force, and the Revolutionary Guard Corps aim to gain an advantage by deploying AI-powered systems on the battlefield as soon as they become operational, no matter how rudimentary and unreliable they may be.

Despite difficulties in procuring hardware and processors, there is a large pool of qualified workers in the public and private sectors contributing to AI development in Iran. According to the Nature Index, Iran ranked 13th in the most AI publications between 2015 and 2019 (Wodecki, 2022). Tehran aims to use this pool of expertise to integrate AI into the country's drone fleet, air defense network, and command systems (Sheikh, 2023). There are speculations that Iran is considering AI-powered air defense systems capable of operating without human intervention.

Drones, however, are the military area in which Iran is most advanced. At the same time, Tehran has ambitions for armed remotely controlled ground robots and underwater vehicles.

4 What Needs To Be Done

The regulation of Artificial Intelligence has been stalled for over a decade. Although Western-influenced states share a common value foundation, concrete definitions and implementations are lacking. This disunity is evident in the debates about banning LAWS and defining the 'Human-in-the-Loop.' Additionally, it remains unclear whether the use of AI in the military should be regulated (as proposed by the EU) or whether the focus should be on the effects of AI use in the military context (as favored by the USA).

Geopolitical competitors of the West, particularly Moscow and Beijing, show little interest in serious regulation of AI. They fear that such regulation would put them at a disadvantage in a confrontation with Western-influenced states. This situation makes it difficult to create a globally uniform regulation that is necessary to effectively address the challenges and risks of AI in the military context.

A substantial, sustainable, and meaningful regulation of AI applications in the military must encompass the following four points:

- Broad International Participation: The regulation must not be limited to Western-influenced states. To be effective, it must be based on a broad foundation of signatory states. The circle of participating states should therefore be as wide as possible.
- Uniform Definitions: Clear and uniform definitions for automation in weapon systems, LAWS, and the integration of human operators are essential. These terms must be consistently understood and applied worldwide to ensure effective regulation.
- Clear Regulatory Focus: The focus of regulation must be clearly defined. It must be decided whether the use of AI in weapon systems or the effects of AI use in the military context should be regulated. A unified direction is crucial for effectively implementing regulatory measures.
- Transparent Review: The regulation must be transparent, and its compliance and implementation by the signatory states must be regularly reviewed. Only through continuous monitoring and adjustment can it be ensured that the regulations remain effective and keep pace with technological developments.

The importance of these points is underscored by current geopolitical tensions. While some Western-influenced states often emphasize that ethical and human rights aspects should be at the center of AI regulation, other states frequently pursue more pragmatic and less restrictive approaches. This leads to a fragmented regulatory environment that hinders global cooperation.

Effective regulation of AI in the military field thus requires not only consensus and cooperation among states with a consistent value foundation but also the integration of other states as partners. Only through a broad internationally coordinated regulation can the potential of AI in the military be used responsibly and the risks minimized.

5 SMART—Strategic Military AI Regulation Talks

Regulating Artificial Intelligence in military applications is one of the central issues for the future of any existing and future security order. The use of AI for military purposes is already a reality. The Russian war of aggression against Ukraine has significantly increased attention to the possible areas of application of AI on the battlefield—among other things, through the use of drones.

To achieve effective regulation, as many states as possible must be involved in the global process of regulating the use of AI for military purposes. It is important not to rely on any already established cooperation structure such as the EU, NATO, G7/G20, or ASEAN, as this would exclude numerous willing states from the outset.

Regulation under the auspices of the United Nations is also not a suitable solution as there is a risk that important points will be blocked or at least significantly watered down early on to reach compromises. Points from other thematic working groups under the United Nations umbrella could become negotiating material for regulating Artificial Intelligence in military applications. Any compromises would thus be pre-shaped and not solely based on the negotiation rounds for the use of Artificial Intelligence in military applications. Formats like the AI Safety Summit 2024 in London, with the concluding Bletchley Declaration, are much more promising than the UN level: 28 states and the EU came together at the Safety Summit in London to discuss regulatory mechanisms and possibilities.

The goal of a newly formed initiative should be an open process in which any state can participate that shares the fundamental values and ethical principles underlying the regulation of AI in the military.

To ensure the credibility of the goal, a clear condemnation of the unregulated spread and use of LAWS must be at the beginning of any legal framework. The principle that the final accountability for killing people in war always lies with human actors must be firmly anchored in future regulations (Koch et al., 2024; Spreen, 2023). Even with accountability chains arising from the cooperation or interaction of various organizational systems and levels, the responsibility always lies with the human. The definition of 'Human-in-the-Loop' can be based on the US model. The decision to use lies with a human commander or operator, while decisions and executions within an operation can be entirely AI-controlled. The deployment decision includes that the human commander defines the operational framework.

The newly formed initiative could be called Strategic Military AI Regulation Talks (SMART) to reflect the significance and scope of this endeavor.

Global regulation of AI in the military is important but difficult to achieve uniformly. Major powers mostly have little interest in binding themselves with a regulatory framework, risking falling behind in comparison with geopolitical competitors. Winning at least one major power for SMART is the prerequisite for the success of this new control regime. If oriented toward the US model, Washington would not have to make painful compromises and could act as a driving force for the SMART initiative.

Thus, the USA could demonstrate its willingness to participate in an international regulatory framework and encourage other key players in this field to commit to clear values regarding the use of AI in the military. This would allow the United States to present itself as a benevolent superpower and win over other medium and major powers. It should also be in the strategic interest of the United States to

subject the spread and use of LAWS to a legal framework, as is the case with nuclear weapons in the form of the NPT.

Every state participating in the initiative should have at least developed its own strategy for the use of AI in military applications in advance. The strategies should become the subject of negotiation within the initiative, enabling more targeted discussions and facilitating compromise formation.

In the context of the debate on LAWS, the idea that lethal autonomous weapon systems may not always be necessary to deter the use of autonomous systems still often comes up short. Opposing LAWS, for example, could also be rendered harmless through cyber operations. Cloud-based AI systems can be neutralized through kinetic attacks on the corresponding cloud server farms. An alternative to kinetic operations is the use of special forces trained and prepared in numerous countries for sabotage acts on server farms. Defense against self-learning autonomous weapon systems (third-level AI) can be disrupted by capabilities in Electronic Warfare (EW), such as blocking communication possibilities between individual units. However, future technological developments will make non-networked LAWS possible, reducing the importance of the cyber and EW components within the logic of defense and deterrence.

The technological advancements in Artificial Intelligence and its use are subject to disruptive dynamics. Therefore, SMART must be a dynamic international initiative with a dynamic regulatory framework. Unlike rigid frameworks like the Nuclear Non-Proliferation Treaty (NPT), periodic review conferences and evaluation formats should discuss the need for adjustments in the regulatory framework.

A key element of regulation should be ensuring transparency and accountability. International monitoring and inspection mechanisms must be established to oversee compliance with regulations. These mechanisms could be led by an independent international organization that regularly publishes reports and enforces sanctions for violations. Especially in staffing the independent international organization, appropriate representation of all states must be ensured, and the involvement of these must be guaranteed in a rotation principle.

The role of science and technology in SMART is also of great importance. Research institutions and technology companies should be involved in the regulatory process to ensure that the technical possibilities and limitations of AI systems are understood and considered. Cooperation between military and civilian research institutions could contribute to ensuring that ethical and security standards are adhered to define clear terms, deployment scenarios, and limits, SMART should use practical simulations and demonstrators that allow for targeted illustration of contentious deployment scenarios, making discussions more realistic.

Additionally, within SMART, investment in the education and training of military personnel should be made. This enables users of AI systems to understand the technological, ethical, and legal implications, which not only promotes compliance with regulations but also improves the operational effectiveness and safety of deployments. The international community faces the challenge of finding a balance

between leveraging the benefits of AI in the military and minimizing the risks. A comprehensive, inclusive, and transparent regulatory process as aimed for by SMART could make a crucial contribution to the responsible use of AI in future conflicts.

- Artificial Intelligence is already widely used in the military context. The possibilities are continually evolving, and the battlefield of the future will include and be shaped by fully automated weapon systems.
- Currently, there is no regulation of the use of Artificial Intelligence in the military context. However, it is one of the central issues for the future of any existing and future security order.
- For the regulatory process, a new initiative independent of the United Nations is needed: Strategic Military AI Regulation Talks (SMART). SMART must be a dynamic initiative with a dynamic regulatory framework. In annual review conferences and evaluation formats, the necessity of adjustments in the regulatory framework should be discussed, and technological developments should be continuously re-evaluated.
- To achieve effective regulation, as many states as possible must be involved in the global process of regulating the use of AI for military purposes. The regulation must cover the spread of LAWS and their deployment scenarios.
- Each state participating in the initiative should have at least developed its own strategy for the use of AI in military applications in advance, as the strategies will become the subject of negotiation within the initiative.
- To define clear terms, deployment scenarios, and limits, *SMART* should use practical simulations and demonstrators.

References

Allen, G. C. (2019). *Understanding China's AI Strategy: Clues to Chinese Strategic Thinking on Artificial Intelligence and National Security*. Center for a New American Security (CNAS). https://www.cnas.org/publications/reports/understanding-chinas-ai-strategy

Artificial Intelligence: Questions of Interpretation and Application of International Law. (2021, January 20). European Parliament. Texts adopted A9-0001/2021. https://www.europarl.europa.eu/doceo/document/TA-9-2021-0009_EN.html

Bitzinger, R., & Raska, M. (2022). Die militärische Modernisierung in China und Russland und die Vierte Industrielle Revolution. *Sirius*. https://doi.org/10.1515/sirius-2022-3006

Boulanin, V., & Verbruggen, M. (2017). *Mapping the Development of Autonomy in Weapon Systems*. International Peace Research Institute (SIPRI). https://www.sipri.org/publications/2017/policy-reports/mapping-development-autonomy-weapon-systems

Boulanin, V., Bruun, L., & Goussac, N. (2017). *Autonomous Weapon Systems and International Humanitarian Law: Identifying Limits and the Required Type and Degree of Human-Machine Interaction* (iPRAW Report). https://www.sipri.org/sites/default/files/2021-06/2106_aws_and_ihl_0.pdf

Brink, N. (2023, June 26). *Machine denkt, Mensch lenkt?* Internationale Politik. https://internationalepolitik.de/de/maschine-denkt-mensch-lenkt

Brink, N. (2024, February 14). *KI-Strategie? Fehlanzeige!* Internationale Politik. https://internationalepolitik.de/de/ki-strategie-fehlanzeige

Bundeswehr. (2019, November). *Künstliche Intelligenz in den Landstreitkräften—ein Positionspapier des Amts für Heeresentwicklung*. https://www.bundeswehr.de/resource/blob/156024/d6ac452e72f77f3cc071184ae34dbf0e/download-positionspapier-deutsche-version-data.pdf

BWI GmbH. (2022). *Künstliche Intelligenz: BWI entwickelt Lösungen für die Bundeswehr*. https://www.bwi.de/fileadmin/images/presse/20220124_Presseinformation_KI-BWI_entwickelt_Loesungen_fuer_die_Bundeswehr_V1.0.pdf

Council of the European Union. (2024, January 26). *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts* (8115/21). https://data.consilium.europa.eu/doc/document/ST-5662-2024-INIT/en/pdf

Cummings, M. L. (2017). *Artificial Intelligence and the Future of Warfare*. Chatham House, the Royal Institute of International Affairs. https://www.chathamhouse.org/sites/default/files/publications/research/2017-01-26-artificial-intelligence-future-warfare-cummings-final.pdf

Dahlmann, A., Hoffberger-Pippan, E., & Wachs, L. (2021, April). *Autonome Waffensysteme und menschliche Kontrolle. Konsens über das Konzept, Unklarheit über die Operationalisierung.* SWP-Aktuell/A 31. https://doi.org/10.18449/2021A31

David, R. A., & Nielsen, P. (2016). *Report of the Defense Science Board Summer Study on Autonomy*. Defense Science Board. Accession Number: AD1017790. https://apps.dtic.mil/sti/citations/AD1017790

Geoană, M. (2020, September 28). *Rede des stellvertretenden NATO-Generalsekretärs Mircea Geoană beim HumanAIze Forum*. https://www.nato.int/cps/en/natohq/opinions_178354.htm

Grimm, R. (2023). Künstliche Intelligenz (KI) zur Abwehr von Cyber-Angriffen und Cyber-Angriffe auf KI. *Journal of Cybersecurity and Artificial Intelligence*. https://doi.org/10.1007/s11623-023-1743-7

Heyns, C. (2013, April 9). *Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions.* United Nations Human Rights Council, A/HRC/23/47. https://www.refworld.org/reference/themreport/unhrc/2013/en/96228

Hoffberger-Pippan, E., Vohs, V., & Köhler, P. (2022, June). *Das Scheitern der VN-Expertengespräche zu Autonomen Waffensystemen*. Stiftung Wissenschaft und Politik. https://www.swp-berlin.org/publications/products/aktuell/2022A36_ExpertengespraecheAutonomeWaffensysteme.pdf

Human Rights Watch. (2020). *Stopping Killer Robots: Country Positions on Banning Fully Autonomous Weapons*. https://www.hrw.org/report/2020/08/10/stopping-killer-robots/country-positions-banning-fully-autonomous-weapons-and

Kania, E. B. (2017). *Battlefield Singularity: Artificial Intelligence, Military Revolution, and China's Future Military Power*. Center for a New American Security (CNAS). https://www.cnas.org/publications/reports/battlefield-singularity-artificial-intelligence-military-revolution-and-chinas-future-military-power

Koch, W., Spreen, D., Talves, K., Wagner, W., Lillemäe, E., Klaus, M., Viidalepp, A., Cooper, C. G., & Pekarev, J. (2024). On the Ethics of Employing Artificial Intelligent Automation in Military Operational Contexts. *IEEE Transactions on Technology and Society, 5*(2), 231–241. https://doi.org/10.1109/TTS.2024.3405309

Kölling, M., & Volkery, C. (2023, October 12). *G7 beschließt 11 Gebote für KI.* Handelsblatt. https://www.handelsblatt.com/technik/ki/kuenstliche-intelligenz-g7-beschliesst-elf-gebote-fuer-ki/29442462.html

Mewes, B. (2023, November 25). *Künstliche Intelligenz: Putin will Russland zur KI-Macht aufbauen.* heise. https://www.heise.de/news/Kuenstliche-Intelligenz-Unter-Putin-soll-Russland-zur-KI-Macht-werden-9539907.html

Ministère des Armées. (2019, September). *Rapport De La Task Force IA Septembre 2019*. https://www.defense.gouv.fr/sites/default/files/aid/20200108-NP-Rapport%20de%20la%20Task%20Force%20IA%20Septembre.pdf

NATO. (2021, October 22). *Summary of the NATO Artificial Intelligence Strategy*. https://www.nato.int/cps/en/natohq/official_texts_187617.htm

NATO. (2023, February 7). *NATO starts work on Artificial Intelligence certification standard*. https://www.nato.int/cps/en/natohq/news_211498.htm

Naval News. (2023, August 17). *Army 2023: KMZ aus Russland stellt neues unbemanntes Oberflächenfahrzeug vor*. https://www.armyrecognition.com/news/navy-news/2023/army-2023-kmz-from-russia-unveils-new-unmanned-surface-vehicle

Office of the President of the Russian Federation. (2019, October 10). *Decree of the President of the Russian Federation. National Strategy for the Development of Artificial Intelligence Over the Period Extending up to the Year 2030* (Trans.). https://cset.georgetown.edu/wp-content/uploads/Decree-of-the-President-of-the-Russian-Federation-on-the-Development-of-Artificial-Intelligence-in-the-Russian-Federation-.pdf

Reitmeier, G. (2020, October). *Lizenz zum Töten. Künstliche Intelligenz in den Waffensystemen und neue Herausforderungen für die Rüstungskontrolle*. Potsdam-Babelsberg: Friedrich-Naumann-Stiftung für die Freiheit. https://shop.freiheit.org/#!/Publikation/945

Sayler, K. M. (2024, February 1). *Defense Primer: U.S. Policy on Lethal Autonomous Weapon Systems*. Congressional Research Service. https://crsreports.congress.gov/product/pdf/IF/IF11150

Scharre, P. (2018). *Army of None: Autonomous Weapons and the Future of War*. W. W. Norton & Company.

Schmidt, T. (2020). Künstliche Intelligenz in der Cybersicherheit—zwischen Hype und Notwendigkeit. *Wirtschaftsinformatik & Management, 12*, 70–74. https://doi.org/10.1365/s35764-020-00244-4

Scott, M. (2023, September 7). *G7 Countries Commit to AI Code of Conduct*. Politico. https://www.politico.eu/article/g7-countries-commit-to-ai-code-of-conduct

Sheikh, A. (2023, August 28). *Irans Oberbefehlshaber beleuchtet KI-Integration für eine verstärkte Luftverteidigung*. Cryptopolitan.

Spreen, D. (2023). Lethal Autonomous Weapon Systems (LAWS). On the Ethics of Automation in the Military from the Perspective of Social Systems Theory. *Sõjateadlane (Estonian Journal of Military Studies), 21*, 10–40. https://doi.org/10.15157/st.vi21.24177

State Council Information Office of the People's Republic of China. (2019). *China's National Defense in the new era*. Foreign Languages Press.

Statement by Iran. (2023, November 1). *UNGA First Committee*. https://reachingcriticalwill.org/images/documents/Disarmament-fora/1com/1com23/eov/L56_Iran.pdf

United Nations. (2018). *Securing our common future—An Agenda for Disarmament. Office for Disarmament Affairs*. https://unoda-epub.s3.amazonaws.com/i/index.html?book=sg-disarmament-agenda.epub

United Nations. (2023, July). *Our Common Agenda Policy Brief 9—A New Agenda for Peace*. https://www.un.org/sites/un2.un.org/files/our-common-agenda-policy-brief-new-agenda-for-peace-en.pdf

U.S. Department of Defense–DoD. (2019). *Summary of the 2018 Department of Defense Artificial Intelligence Strategy. Harnessing AI to Advance Our Security and Prosperity*. https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF

Wodecki, B. (2022, February 1). *Iran Vies to Become Top 10 AI Nation by 2032*. AI Business. https://aibusiness.com/verticals/iran-vies-to-become-top-10-ai-nation-by-2032

Working Paper of the Russian Federation. (2022, July 18). *Application of International Law to Lethal Autonomous Weapon Systems (LAWS)*. https://documents.unoda.org/wp-content/uploads/2022/07/WP-Russian-Federation_EN.pdf

Acceptance Model of Artificially Intelligent Military Technologies in the Small Country Context

Kairi Talves, Priit Värno, and Eleri Lillemäe

Abstract This article examines the perceptions of AI among the military personnel of Estonia, a small country on the eastern edge of NATO, based on data collected during interviews in the armed forces. Due to its small size and geographical location, Estonia finds itself in a unique security situation, which requires the country to constantly adapt to changing threats. The article proposes a model for military technology acceptance in a small country framework, focusing on differences in perceptions of the usability of technology in a military context. The article also discusses trust-related issues of military technology. Results indicate that military personnel see autonomous technologies offering clear benefits by reducing risks to human lives. Ethical and legal concerns about responsibility and decision-making are important, but from a more practical perspective. The study also stresses the importance of user experience. At the same time, a lack of confidence in the decision-making capabilities of AI and the need for human control are hindering achieving the full potential of the systems. Our proposed model describes both military-specific and contextual factors, providing insights into the potential benefits and challenges of AI implementation in the context of a small country's military.

1 Introduction

The rapid development of artificially intelligent technologies is also increasing their usage in workplaces and daily life, leading to a greater necessity to understand the attitudes about them and the user's acceptance process. One of the core principles of successful innovation is that a new invention is accepted and taken into use. It means that a method, technology, or approach has moved from the experimental phase to the application: no longer a novelty but something normal and

K. Talves (✉) · P. Värno · E. Lillemäe
Estonian Military Academy, Tartu, Estonia
e-mail: kairi.talves@mil.ee; priit.varno@mil.ee; eleri.lillemae@mil.ee

K. Talves, D. Spreen (eds.), *Artificial Intelligence in Military Technology*,
Artificial Intelligence, Simulation and Society 192,
https://doi.org/10.1007/978-3-031-95578-5_6

institutionalized (Deutsch, 1985). However, the new technologies in the military require a change in multiple procedures, e.g., military doctrine, training, and capability planning. It is, therefore, a long-term process with multiple interactions, where technology and its impact should always be carefully considered.

Attitudes are important motivators that lead to the intention to use technology and actual usage of them (Ajzen, 1991). Most technology acceptance theories stress the importance of attitude–behavior relationships and seek the factors determining the success or failure of specific technology usage. To analyze intentions, it is important to understand the person's and the community's attitudes toward the behavior and also how the person values the community's standards (Fishbein & Ajzen, 1975). The commissioning and underutilization of new technologies often result in high costs, and therefore, including personal and organizational factors in analyzing technology acceptance is of great importance (Venkatesh & Hillol, 2008). Hence, we take the Technology Acceptance Model (TAM) framework as a theoretical basis to analyze how different factors influence attitudes toward and acceptance of new technologies in the military. Although TAM and its further adaptations are widely used to measure the acceptance of new technologies, it has found less use in military organizations. Here, validating and modeling the acceptance can offer valuable insights, taking into account the specificity of military organizations.

We analyze the acceptance of new technologies in the military through the example of unmanned ground systems (UGS). UGS is a field of technology that is largely under development and seeking to apply significant artificial intelligence capabilities. Exploring the attitudes toward such emerging technologies helps understand how defense forces implement innovation and what factors influence their utilization and acceptance. The article employs a small country case study: the study was carried out in Estonia, which, with its small army, needs to be flexible and fast-adapting but is particularly influenced by global technological and security challenges. We take a closer look at how such context influences attitudes toward new technologies.

Due to the novelty of the technology under review, we consider our research as exploratory to identify the main factors influencing the use of artificially intelligent technology. In the following sections, we provide an overview of the theoretical grounds of technology acceptance, the methodology used for this study, the results, and a discussion, which, based on the research findings, proposes a model of military technology acceptance in a small country framework.

2 Theoretical Background

One of the first models for researching acceptance of technology is the Technology Acceptance Model, developed by Fred D. Davis in the 1980s (Davis, 1985). The model is based on the theory of reasoned action, including attitudes and social norms toward the use of technology (Fishbein & Ajzen, 1975) and the theory of planned behavior, adding the factor of perceived behavioral control (Ajzen, 1991).

TAM aimed to create a better understanding of the user's acceptance process, thereby supporting the developers with a theoretical model for creating and applying new software and providing a method for testing user's acceptance of them. In its time, TAM was the first and most used tool for empirically testing end-user responses to new technological inventions.

Later, researchers developed TAM further into TAM2 by adding social and cognitive variables like social norms, the voluntariness of technology use, and the rise of quality output to understand better workplace-related factors in technology acceptance (Venkatesh & Davis, 2000). Both TAM and TAM2 include perceived usefulness (PU) and perceived ease of use (PEU) as the core factors affecting the attitudes and intentions of using the technology (Li, 2010). Perceived usefulness describes the extent to which the user believes the technology will improve his performance, and perceived ease of use the degree to which the person believes that a particular system is easy to use, without considerable effort (Davis, 1989). Even though TAM has been a widely used concept, it has some shortcomings. While focusing on individual user attitudes, TAM does not specify the external factors that have proved important in later studies. For example, technology-related factors like the faults in technologies (Masrom, 2007) or organizational factors like the decision-making process over the adoption of new systems (Bagozzi, 2007) play a significant role in the acceptance process. The comprehensive model should take into account both the personal and organizational factors that influence how new technologies are perceived. This is especially the case if we take the example of a military organization that seeks the best solutions for military capabilities. Military capability planning, by looking at the options of modernizing the forces and introducing new technologies, has to carefully consider the technical maturity and organizational absorbance of prospective new systems (Bistron & Piotrowski, 2021).

The adoption and use of new technologies in a military organization is costly. Besides purchase costs, the life-cycle costs—like maintenance and possible extra personnel costs—have to be considered. Although artificial intelligence is largely expected to diminish the burden of human soldiers, the use of high-tech systems is not entirely human-free. It may require qualified experts and engineers to uphold the systems. Another aspect crucial for the military is the reliability and dependability of the technology in use, which brings us to the issue of trust toward the technology. In different domains and communities, trust in technology has been described in different manners (Holden & Karsh, 2010). Research in military organizations has indicated that factors such as perceived risk and perceived control (Chamata & Winterton, 2018) influence the trust and are of great importance in affecting the attitude and intention to use new technologies. As unmanned systems and artificial intelligence are relatively new and unfamiliar, trust in them is subjacent because of the possible risk of unintended consequences. As an example of a recent study, soldiers' trust in a manned aircraft superseded the trust in an unmanned system when the danger to own troops is increasing, even though the unmanned platforms could carry more firepower (Macdonald & Schneider, 2019). Trust also depends on the users' technological self-efficacy and confidence in using the technology (Levy & Green, 2009). Questions of trust in technology, and the relationship

between risk and control lead to the discussion on the ethical aspects of artificially intelligent military technology (Koch et al., 2024; Spreen, 2023). Autonomous technologies on the battlefield will increase the efficiency of military operations but may also pose increased danger to human lives, thereby leading to an ethical question of how much control the soldier is willing to overhand to the machines (Galliott, 2018).

In addition to personal and organizational factors, the societal context plays a role in the uptake of novel solutions. Most countries see the commitment to technological achievements as a catalyst for economic growth and direct their citizens toward a technology-supportive mindset (Vu & Lim, 2022). However, the size of the country is heavily directing the implementation of new technologies. Small countries are particularly hampered by existing vulnerabilities: economic dependency on global fluctuations, narrow margins of error, lack of strategic capabilities, and lack of human resources—to name some of the most important ones (Briffa & Männik, 2023; Kenworthy et al., 2023). From a military point of view, smaller countries are not the primary definers of the global strategic threats, nor do they have enough resources to define the main technological imperatives in warfare (Bartmann, 2002). To overcome these shortcomings, small countries must adapt quickly and effectively to the environment of possible threats. Adaptability and cooperation are crucial not only in the need for military alliances but also in seeking the financial, political, and social support of other countries and international organizations (Bailes et al., 2016). Nevertheless, small countries are difficult to conquer and occupy when seriously committing to national defense and investing in advanced military technologies (Thorhallsson & Steinsson, 2017).

In this study, we seek answers to two research questions. Firstly, to assess the attitudes toward accepting specific emerging technology—how are the perceived usefulness and ease of use described in the context of adopting UGS in the military? Secondly, what are the prospects for using artificially intelligent technologies in the armed forces? The results will be discussed in a small country framework to assess its capacities and barriers to adopting new military technologies.

3 Methodology

3.1 Interviews and Data

Data was collected from July to September 2021 in Estonia under the framework of the European Union-funded project Integrated Modular Unmanned Ground Systems (iMUGS). One of the project's aims was to study the general public's and military personnel's attitudes about UGS and autonomous technologies. The research group carried out interviews with 18 current and previous officers and employees of the Estonian Defense Forces (EDF) and Estonian Defense League (EDL), among them 14 officers, two non-commissioned officers, and two civilians. The ranks of the interviewed persons ranged from second lieutenant to lieutenant general, and all of them had a higher education.

Most interviewees had no extensive background knowledge in robotics, AI, or other topics closely related to the development of autonomous technologies. However, to acquire multiple opinions about the technology, the interviewees were selected from the units that had experience in testing and using UGS and also among the higher rank officers of the EDF and EDL who have a deeper insight into current and future ambitions and goals of the EDF (commanders, heads of command).

Interviews were conducted in a face-to-face semi-structured interview format, with one exception, which was conducted via a video teleconference platform. The interviews consisted of 21 leading questions that covered general attitudes about the use of UGS, AI, and implementation of autonomy on the battlefield, its ethical and legal issues, as well as questions about challenges of innovation in the military. The questions were open-ended, and the interviewers encouraged the open and free-flowing discussion. All the interviews were voice recorded, except one recorded by taking notes. After the interviews, they were transcribed and marked with anonymous code names for each interviewee. Finally, the audio files were deleted. Interviews were conducted in Estonian, but the transcripts were translated into English, so the Estonian and English transcripts are available for all interviews (Wagner et al., 2022).

3.2 Data Analysis

We used a two-step qualitative content analysis (Mayring, 2022), following the steps of deductive and inductive category assignment regarding the types of attitudes related to autonomous technologies. First, we created deductive categories based on technology acceptance theories and identified the main categories related to the perceived usefulness and perceived ease of use of the UGS. Secondly, we used inductive category development to identify expressions related to the development and use of AI technologies from a wider perspective and to identify attitudes about the use of emerging technologies in the Estonian military. The coding was conducted by two authors with the support of the QCAmap platform (https://www.qcamap.org/). The coders used multiple iterations in creating categories and inter-coder verification (only the codes that both authors agreed upon were kept) to ensure transparency and reliability of coding.

4 Results

In the following section, we present the factors from interviews that reflect the acceptance and adoption of UGS and artificially intelligent technologies in the usefulness and ease of use aspects of military context. To present the variety of opinions, illustrate the analysis, and ‘give voice’ to the interviewed experts, we have added the quotations from the interviews.

4.1 Tactical Value

This perspective encompasses characteristics wherein interviewees reflected on how machines could outperform humans. Interviewees highlight that machines do not experience fear, they do not get tired, and machines do not need to be motivated like humans. Soldiers require rest and experience fatigue, whereas machines can operate continuously. In this context, UGS are not viewed as autonomous actors but as tools to automate processes and mitigate human errors.

> At the same time, an electronic system, as long as its batteries are charged, can monitor and observe 24/7, which is not something humans are very good at. [*Participant 17*]

Interviewees also contend that teaching machines is easier than teaching humans. While human learning is time-consuming, machine learning is inherently a simulation. Managing people is perceived as more challenging than overseeing machines that follow orders without question. While soldier morale can fluctuate and leaders have to deal with it, with machines, 'changing batteries' suffices.

> The machine does not run away from you, the machine probably does not feel fear or shame or, I do not know, whatever. So, in that respect, you set the task, and obviously, it is easier to control the machine because you do not have to motivate the machine why it has to do something. [*Participant 13*]

Therefore, interviewees foresee a potential shift in leadership with the increasing automation and use of unmanned vehicles. They anticipate a more technical approach to leadership, characterized by clearer and more limited instructions for machines to comprehend. Leadership is expected to adhere more closely to specified protocols and manuals.

It is also noted that as the machines do not feel the threat or fear like people, their decision-making process is faster and more effective than humans. They can take additional risks that people would not take. The interviewees emphasized that the role of technology should be to do things more easily, efficiently, and faster. The most significant effect of automation is seen in the increased speed of decision-making and opening fire.

> /.../ machines come to the battlefield, the main effect is speed and replacing fighting with human morale. [*Participant 18*]

Interviewees bring out that machines are more expendable. When leading soldiers, leaders must think about possible risks and their acts are more conservative to preserve human lives. Interviewees contend that unmanned ground vehicles can save lives, and in dangerous situations, machines can go first and make the first decisive move. Using machines can change the tactic of the unit—the unit can be more aggressive and take more risks.

> With this machine, you can kind of go poking around and see what happens, but you cannot really put a soldier out there and see what happens. Human lives are of a different type of value, after all. [*Participant 4*]

4.2 Cost–Benefit Ratio

Another consideration is cost-effectiveness in implementing new technology. Interviewees question whether the benefits of current technological solutions outweigh the costs associated with acquiring and maintaining them, favoring the development of conventional powers. Unmanned vehicles are perceived as a supportive tool, just one among various technological solutions.

> I think the art of war and warfare, at least in the foreseeable future, is still between people. And that autonomous systems are still a technical solution, another technical solution, as they have come and gone. [*Participant 2*]

While interviewees acknowledge the upside of automation in freeing people from tasks machines can perform, they repeatedly express skepticism about technology's current state and maturity, asserting that true artificial intelligence has not yet been achieved. The current technologies are deemed insufficiently advanced, and until autonomy reaches higher levels, machines do not genuinely free people from tasks. Even if machines alleviate some tasks, additional human operators are required, whose workforce needed to maintain processes might be even more expensive than the initial cost of the task.

> /.../ this artificial intelligence is still... it does not really exist. So far. It is a bit of a commercial expression, but we could be talking about artificial intelligence as such if the system were to have a kind of 'free will'. But no system has that, this artificial intellect is ultimately a program that behaves according to certain rules after all /.../. [*Participant 3*]

Discussing cost-effectiveness, respondents also brought in the perspective of Estonia as a small country in a geopolitically challenging situation. The consensus among interviewees was that limited resources constrain opportunities for adopting new technologies. Given the scarcity of resources, the perception was that autonomous technologies do not offer sufficient cost-effectiveness. Nonetheless, technology was recognized as essential in compensating for the small population, serving as an enabler.

4.3 Ethical Aspects

The interviewees used rather pragmatic viewpoint while discussing about ethical questions of autonomous technologies in warfare. They primarily focused on accountability issue by asking who is ultimately responsible for the actions of such technologies and the use of lethal force. The main legal and ethical problems were seen in cases where the machine itself would be given the decision-making authority. As long as humans make decisions, no ethical problems are seen. However, one consequence of introducing autonomous weapon systems to the battlefield was seen in increasing the role and number of legal advisers. They would monitor the situation, interpret it, and tell the operator whether they can 'press the button.'

Mostly, autonomous technology was seen as another tool that might extend the decision-making chain, but ultimately, the decision rests with humans. Therefore, the use of UGS was not perceived as an ethical problem since it was seen that these machines execute the will of people.

> It makes no difference whether I press the trigger myself or if I have taught something to do so at that moment; it is still my creation. [*Participant 8*]

However, it was mentioned that using such technology means that humans are further away from the battlefield, which changes how immediate the threat is perceived to be for human lives. In this sense, machines become buffers, allowing quicker action and greater risk-taking on the battlefield. This led to discussions on how proportional it is to use such technology against an enemy. For instance, some interviewees emphasize that deploying such systems in warfare could increase the violence on the battlefield.

> If you can send a robot to the battlefield with a certain degree of impunity, you will definitely be braver in making tactical decisions. At some point, that courage can turn into audacity. [*Participant 17*]

Overall, it was argued that if employing such technology is to save human lives, then its use is ethical. Nonetheless, the question of whether warfare itself can be deemed ethical was raised. The interviewees also addressed the issue of technological maturity, stating that discussing the ethics of such systems is currently unreasonable, as the systems are not yet mature enough for thorough evaluation. The use and development of autonomous technologies were deemed ethical for another reason—because the rest of the world also partakes in their development and deployment.

4.4 Experience with Technology

The feeling of being skilled in using technology is tightly connected with a profound experience of using it. Any new technology to be accepted needs multiple rounds of testing and getting acquainted with the specificities of the technology. As the UGS is relatively new for most interviewees, they stressed the importance of experience from many different angles, from the field practice for learning everyday handling of the technology to wider knowledge about the capabilities and potential of the technology in military operations.

> If you do not know the technology, it is complicated. If you are trained to use the technology, then things will be easy. /---/ So you have got to take it and play around with it, if you use it, you are going to have even more questions than you had before. You do not know anything to ask yourself before. If you use, you will have questions that can be answered. [*Participant 16*]

Besides getting to know the working principles of the technology, the trust in the technology is strongly related to experience. Different interviewees with and

without actual experience with such technology express how exposure to technology increases trust and not being exposed decreases or makes the person hesitant toward the possible outcomes. Without any actual contact, just by guessing about the capabilities, which many interviewees told to be the case of 'full autonomy' or 'true AI,' it is not possible to build up the usage options of the technology.

> I think that it is constant because, even based on the trials that we conducted with the ground systems, trust for the device is quick to come. It weighs over a ton, 1,200 kilos, and nobody, not once, was afraid to stand in front of it, like what if it did not stop after the operator removed his finger? They walk one meter away. There is no question of will I be overrun from behind. The trust is quick to come and I think that the same goes for artificial intelligence. [*Participant 8*]

> Given that, today, I have not seen a system like that, I would say that, currently, I would not trust it. If it could give me its opinion and show me the actual image and then I would confirm with a yes or no answer... Today, I have not seen such a system and I do not have faith in it. [*Participant 6*]

Growing trust and comfortability with the use of new technologies are also related to the evolving maturity of the complex technology. One of the interviewees used an example of the relative maturity of unmanned aerial systems compared to such technology in the ground domain. It shows that technology that is more mature and widely acknowledged is, in turn, considered trusted, and compliant to use.

> If you ask a soldier that, what if we have a ground battle without men, it probably will not fly, no one will understand it. But if you have an air battle without people, it does not seem so utopian. If an area will be developed further, it is taken more naturally as a process. [*Participant 10*]

Being familiar with technology is also seen as important in a wider development perspective. On the one hand, it helps to make informed decisions about developing and procuring the technology. To be an equitable partner for the defense industry and other parties developing technologies, the interviewees see the wider experience as a potential. Here, cooperation is expected to avoid hesitations in saying yes or no without adequately understanding the technology. On the other hand, experience is a source for new ways of thinking, such as materializing in use cases and operational scenarios in armed forces and sorting out the potential of the constantly evolving technology. However, a stepwise careful approach is considered appropriate, especially in terms of new technologies that inherently need multiple iterations of development to become reliable technology-wise and usable combat-wise.

> The most effective way is that the defense forces could then use these machines one way or the other and these use cases would perhaps come through practice /---/ hand it over and let them use it. As I like to say, small steps, small moves /---/ soldiers use it, problems come up, ideas start to emerge maybe where you could use it in that maneuver. [*Participant 3*]

> I am convinced, as I said at the beginning, that as automation is coming more and more, it is better to think for ourselves, even if we do not make big breakthroughs with our systems, but to go through these baby steps and figure out where the bottlenecks are. We already know that the machine is not yet smart enough to do everything by itself, but if at some point it becomes that smart, we should think about how we can best use these unmanned systems. [*Participant 17*]

4.5 Maturity of Technology

Maturity of autonomous systems is essential to ensure sufficient technological readiness to conduct military operations and to ensure that systems are operational on the battlefield. Quite many interviewees stress that AI-based unmanned ground systems need serious development to be operationally useful and practical to use on the battlefield. For example, one of the interviewees points out that current AI technologies are developed only by following principles of scaling and speeding up. However, a qualitative shift in technology development must occur to advance to the next level.

> Yes, it is accelerating, but accelerating only quantitatively. The fact that processors are going faster, that data volumes are getting bigger, that more data is coming, that more seemingly intellectual-like things can be done based on it – it still does not give us the human-like objects. For true artificial intelligence to emerge, a revolution is needed in this field. /---/ There must be a serious revolution in technology before we can start talking about these things. [*Participant 3*]

The lack of maturity of autonomy is considered an overall problem with the technology as such. It is seen as a clear hindrance in the civil sector with self-driving cars, which is even amplified in the military, where the terrains and operational conditions of the technology need more complex development.

> I am not sure if... I think this technology, no matter what robot we are talking about, I think this autonomy or self-control is so immature now. Even Tesla, who has probably gone the furthest with it today... You should be very smart and definitely on the asphalt... if you bring this thing to a terrain or nature where there is no asphalt and beautiful curbs, it will become many times more difficult. [*Participant 4*]

Throughout the interviews, the present technology is evaluated against the possible future, autonomous capabilities of the UGS. The variety and complexity of performance are embedded with the maturity of technology and its autonomous capabilities. One of the prevailing opinions is that autonomous systems have to replace soldiers, especially in routine or dangerous tasks. Until the technology has not developed far enough, it is unable to do that or is doing it to a limited extent. Here, ease of use is also considered an important factor. For most interviewees, the desirable autonomous technology includes three main characteristics: humans have to stay in the loop, technology has to be reliable and easy to use, and it has to be able to replace soldiers in certain tasks.

> Well, we have to go back to, well, why we need technology in the first place. Technology must help us do something easier, faster, more efficiently, better: we may use whatever epitaph. And, well, as I have formulated in my mind that in the Estonian Defense Forces, technology must enable us to replace a person where a machine can do the job. [*Participant 13*]

> It depends on the intelligence it has. If you can control it with simple commands, not that you have to control it with a remote control – like command it, rather than direct it, it would be quite useful. [*Participant 2*]

Some interviewees see the use of autonomous systems from a mission-centric perspective—it has to fulfill the needs of the specific task or maneuver being an incremental part of it. The maturity of autonomy should enable the combination of artificial intelligence with human intelligence to achieve the enhancement of human soldiers as an integrated force or as a fully integrated system or system of systems.

> It depends on the task I have, what I have to achieve. What is the composition of different means? We could also add some sensors, but sending the UGV to the battle is a bit like sending a tank into battle, there should be a bit of consideration and there should be the other systems behind it. [*Participant 15*]

> Our job, I think, is to take it into account and try not to teach them (machines) but to learn from them, we need to learn and develop together and find ways to deal with it all together /---/ we must learn to co-exist and create added value together. If you ask, will my thoughts be applicable for the next ten years? No, but whether we like it or not, we are on our way there. [*Participant 18*]

4.6 Reliability of Use

Reliability of use is often discussed in the same context as the maturity of technology, especially in connection with possible operational scenarios. However, it is also described as the overall principle of technology in the military domain—the technology has to be robust, safe, and reliable. From the user practice of current UGS technology, interviewees mostly bring out the logistical challenges that must be solved. It is stated that logistical problems are coming up in a very early phase of use, especially when moving vehicles long distances or with maintenance. These factors also induce cost–benefit comparisons with other types of (manned) technology from a technical and human resources requirements perspective.

> We have bounced a lot to the fact that it still needs to be transported by some other machine somewhere on the battlefield and there is the machine that is transporting it and there is the team that is doing it. And in fact the question arises very easily—if we already have some kind of SUV that can get this UGV there and if we already have people for that anyway, then why do we need this UGV at all. [*Participant 3*]

> If we knew that these were unable to, for a long time, I mean like a week, to keep up with people. If an additional chain of logistic units was added to a battlefield and their only assignment was to solve the issues of these machines then this is something to consider. [*Participant 17*]

In terms of artificial intelligence related issues, the most crucial aspects of reliability are robustness and trustfulness of AI, non-deceivability and resistance to adversary activities. Robustness on AI is most often connected with unpredictability. As rather non-familiar autonomous technology is coming to the battlefield, its acting is unpredictable for many. For example, low adaptability is named; compared to human soldiers, a certain level of rigidness is anticipated with the technology. Another issue is related to the trust in the system's reliability, the problems with

targeting, but also issues with technological breakdowns (non-compliance, unexpected errors, etc.) were mentioned.

> A soldier acts automatically. Now, if the circumstances changed and the device should adjust its behavior then the question arises: how quickly will technology be able to follow the next command or to reorient? Or if the person giving a command was absent, what will technology do, these are the two question marks, for me. [*Participant 16*]

> When it comes to weapon systems, I am a bit more skeptical. How do you make them specifically identify an adversarial combatant or adversarial technology? [*Participant 7*]

> If it included a weapon system and there was just another system failure, an order failed, for instance, at the moment when it should have fired a certain way, and the system that was able to make the "fire stop" command failed to operate or to direct the fire some other way and still made the shot in the initial direction, well, it may endanger our own people. [*Participant 14*]

From a technology resilience point of view, many interviewees bring out the vulnerability of systems to adversary attacks—be it harming the technology with cyber or electronic warfare means, but also outright takeovers and the exploitation of technology by hostile forces. Most of these concerns are rooted in the belief about the intrinsic vulnerability of the technology and the capabilities that can be attacked or harmed.

> And if we were to speak about the probable immediate future for these weapon systems then the given parameters, if we knew their given parameters, then we would be able to fool them. [*Participant 5*]

> I remember right now, there has been a talk about intelligence, for example, you can send it somewhere far ahead, closer to your enemy. But again, the problem comes. If I send this machine with radio stations on it, with encrypted technology on it, with some sensors and things we do not want to be got by the enemy on it... And it gets to know what technology we use, which cryptographic algorithms we use... [*Participant 3*]

5 Discussion

In the model shown in Fig. 1, we present the characteristics identified in the current study that influence the acceptance of new technologies. The factors influencing acceptance from a military perspective are depicted on the left as dark grey ovals, while the contextual factors, particularly acute in a small country context, are shown on the right as ovals with the light gray background.

From a *usefulness perspective*, a pragmatic and practical attitude toward artificially intelligent technologies prevails. One of the main advantages of autonomous systems is considered to be their ability to replace human soldiers. This is deemed necessary for conducting dangerous operations and taking greater risks in maneuvers, and it is also expected to reduce the training and commanding burden on human soldiers (tactical benefit). Usefulness is also tied to the cost–benefit ratio of the new technologies: whether they outweigh the direct and indirect costs of

replacing the old technology and whether they are capable enough to perform the tasks in the manner expected of them.

Ethical issues arise not so much from a moral point of view but mainly from a pragmatic (and legal) perspective and are related to who would be responsible for the actions of AI-based technologies if such machines were to have decision-making authority. It was generally believed that human beings will always remain responsible, either directly (by 'pushing the button') or indirectly (as the creator of the technology), and that cancels out the concern of responsibility.

It could also be expected that as autonomous technologies enable humans to be further away from direct contact with the battlefield, and therefore reduce the risk to human lives, this tunes down the threat perception and makes warfare more precise and economical. Interestingly, however, participants saw that less threat to human lives works the other way around and actually enables to take greater risks and increase the speed of decision-making against adversaries. As stated also in Huh Wong et al. (2020), "the greatly increased decision-action speeds associated with autonomous systems have the potential to escalate conflict much more quickly," which in turn raises the question about if and how the principle of proportionality is going to change in the future.

From an *ease-of-use perspective*, the importance of user experience of new technologies is brought out through the need to acquire practical skills to use them and increase competence to partner in developing and procuring new technologies. In the technological maturity area, the study revealed a considerable gap between existing technology readiness and its desired functionality and capabilities.

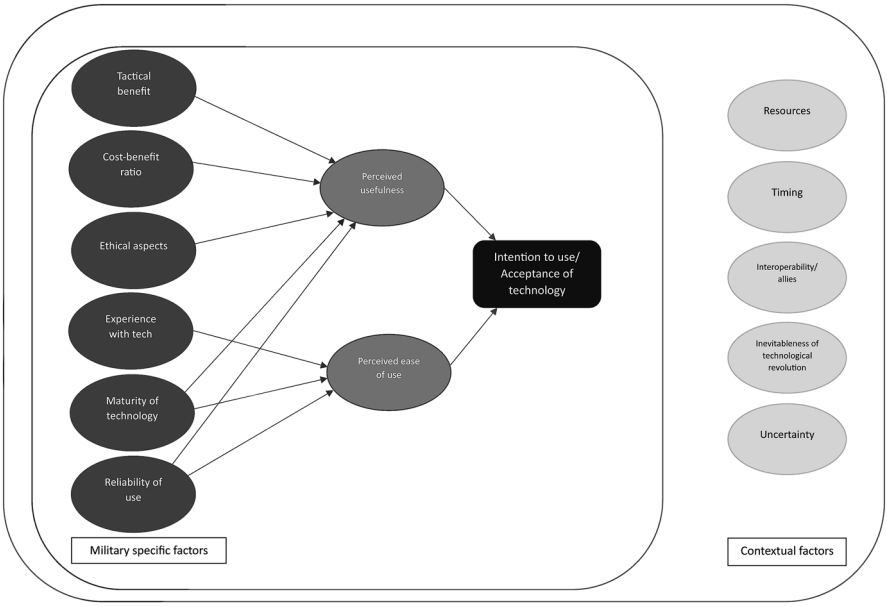

Fig. 1 Factors for technology acceptance (*source*: own visualization)

Interestingly, we found that the maturity of the technology is considered crucial from both, technology usefulness and ease-of-use perspective, ultimately influencing the acceptance of the technology.

Looking ahead, AI technologies are expected to be, above all, easy and convenient to use and integrate into human–machine and machine–machine functions. The latter is closely related to the reliability of use, which comes from general expectations of military technology: any technology used by the defense forces must be reliable, safe, and robust. The fallibility of technology is considered important—it is seen as presenting a number of vulnerabilities that can already be described today. However, there is also much discussion about the fact that constantly evolving new technologies bring along risks that cannot be foreseen today.

Respondents also reflected in their interviews on the perspective of Estonia as a small country, which refers to the importance of contextual factors in technology acceptance. For a small country, the main challenges in deploying new technologies are related to issues of resources, timing, interoperability, global technological push, and uncertainty. In a resource-constrained context, the question is which capabilities should be invested in, in which sectors, and what is the balance between "off-the-shelf" technology and innovation? The dilemma faced in implementing new technology is the trade-off between its costs and benefits. Advocates of emerging technologies understandably focus on their benefits. However, the costs are nontrivial and include those related to the uncertainties surrounding any emerging technology—will it work, at what cost, and when (James, 2013)?

The timing issue is even more complex—the rapid development of technology makes it challenging to decide on the maturity and optimality of the technology for deployment. The dependence of small states on large weapon systems and other technologies also adds an important compatibility issue. At the same time, an inevitable technological revolution is pushing the decisions about deploying new technologies. However, uncertainty has been said to be the main constraint on such decision processes involving innovative technology. Deciding, for instance, whether a new device is worth the investment needed to perfect it is largely a matter of reducing uncertainties to manageable dimensions. The very novelty of emerging technologies means that our expectations about them—how they will be used, who will use them, and why—are constructed around considerable uncertainty (Smith, 2020).

6 Conclusions

In conclusion, this study highlights the multifaceted factors influencing the acceptance of new technologies within the armed forces, particularly in the context of small countries. Our proposed model outlines both, the military-specific and contextual factors.

The military-specific factors include tactical benefit, cost–benefit ratio, ethical aspects, experience with technology, maturity of technology, and reliability of use.

From a military perspective, autonomous technologies offer clear benefits, such as reducing risks to human lives. Ethical and legal concerns about responsibility and decision-making authority are significant, but rather from a practical viewpoint. Interestingly, our study brings out that distancing humans from direct battlefield contact might increase risk-taking and violence, as the autonomous machines are acting as a buffer and are seen as more expendable. The study also emphasizes the importance of user experience, highlighting the need for practical skills and technological competence.

We argue that these aspects affect how useful and easy to use the technology is perceived to be, which in turn affects the intention to use and accept novel technologies in the military sphere. However, we also claim that contextual factors, such as resources, timing, interoperability, inevitability of technological revolution, and uncertainty, play a role in acceptance and influence the military-specific factors.

For small countries like Estonia, resource constraints, timing, and global technological pressures complicate the adoption of new technologies. Balancing immediate costs with long-term benefits, along with the uncertainties of emerging technologies, presents a persistent challenge. The rapid pace of technological advancement makes it even harder to decide when and what technology to adopt.

The study also shows that despite the perceived benefits, there is a significant gap between current technological readiness and the desired functionality, which affects overall acceptance. As the maturity of autonomous technologies is currently rather low and mostly requires an operator, the readiness to accept fully autonomous systems and see their possible benefits and challenges is yet hard to grasp. Therefore, acceptance on different levels of autonomy needs to be studied further alongside technological developments.

To validate the model and the weight of importance of the different factors proposed in the current study, a quantitative study could be carried out in the militaries of different countries to assess the drivers and contexts of adopting new technologies.

References

Ajzen, I. (1991). The Theory of Planned Behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Bagozzi, R. P. (2007). The Legacy of the Technology Acceptance Model and a Proposal for a Paradigm Shift. *Journal of the Association for Information Systems, 8*(4), Article 12. https://doi.org/10.17705/1jais.00122

Bailes, A. J., Thayer, B. A., & Thorhallsson, B. (2016). Alliance theory and alliance 'Shelter': the complexities of small state alliance behavior. *Third World Thematics: A TWQ Journal, 1*(1), 9–26. https://doi.org/10.1080/23802014.2016.1189806.

Bartmann, B. (2002). Meeting the Needs of Microstate Security. *The Round Table*, 361–374. https://doi.org/10.1080/00358530220000103 35

Bistron, M., & Piotrowski, Z. (2021). Artificial Intelligence Applications in Military Systems and their Influence on Sense of Security of Citizens. *Electronics, 10*(7), Article 871. https://doi.org/10.3390/electronics10070871

Briffa, H., & Männik, E. (2023, September 14–16). *Confronting the Future: Cultivating Small State Resilience through Strategic Foresight* [Workshop presentation]. Workshop: Europe in an Uncertain World: Openness, Resilience and Vulnerabilities of Small States and Middle Powers, Athens, Greece.

Chamata, J., & Winterton, J. (2018). A Conceptual Framework for the Acceptance of Drones. *The International Technology Management Review, 7*(1), 34–46. https://doi.org/10.2991/itmr.7.1.4

Davis, F. D. (1985). *A Technology Acceptance Model for Empirically Testing New End-User Information Systems: Theory and Results* [Doctoral dissertation, Massachusetts Institute of Technology, Sloan School of Management]. MIT. http://hdl.handle.net/1721.1/15192

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly, 13*(3), 319–339. https://doi.org/10.2307/249008

Deutsch, K. W. (1985). On Theory and Research in Innovation. In R. L. Merrit & A. J. Merrit (Eds.), *Innovation in the public sector* (pp. 17–35). Sage.

Fishbein, M., & Ajzen, I. (1975). *Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research*. Addison-Wesley.

Galliott, J. (2018). The Soldier's Tolerance for Autonomous Systems. *Paladyn, Journal of Behavioral Robotics, 9*(1), 124–136. https://doi.org/10.1515/pjbr-2018-0008

Holden, R. J., & Karsh, B.-T. (2010). The Technology Acceptance Model: Its Past and its Future in Health Care. *Journal of Biomedical Informatics, 43*(1), 159–172. https://doi.org/10.1016/j.jbi.2009.07.002

Huh Wong, Y., Yurchak, J., Button, R. W., Frank, A. B., Laird, B., Osoba, O. A., Steeb, R., Harris, B. N., & Joon Bae, S. (2020). *Deterrence in the Age of Thinking Machines*. RAND Corporation. https://doi.org/10.7249/RR2797

James, A. D. (2013). *Policy Brief: Emerging Technologies and Military Capability*. Nanyang Technological University, ETH Zürich. https://www.files.ethz.ch/isn/174574/Policy%20Brief-Emerging%20Technologies%20and%20Military%20Capability.pdf

Kenworthy, P., Kirby, P., & Vorisek, D. (2023, January 26). *Small States are the Canary in the Coal Mine for the Global Economy, and they are Struggling.* Brookings. https://www.brookings.edu/articles/small-states-canary-in-the-coal-mine-global-economy-are-struggling/

Koch, W., Spreen, D., Talves, K., Wagner, W., Lillemäe, E., Klaus, M., Viidalepp, A., Cooper, C. G., & Pekarev, J. (2024). On the Ethics of Employing Artificial Intelligent Automation in Military Operational Contexts. *IEEE Transactions on Technology and Society, 5*(2), 231–241. https://doi.org/10.1109/TTS.2024.3405309

Levy, Y., & Green, B. D. (2009). An Empirical Study of Computer Self-efficacy and the Technology Acceptance Model in the Military: A Case of a US Navy Combat Information System. *Journal of Organizational and End User Computing (JOEUC), 21*(3), 1–23. https://doi.org/10.4018/joeuc.2009070101

Li, L. (2010). *A Critical Review of Technology Acceptance Literature*. Grambling State University. http://swdsi.org/swdsi2010/SW2010_Preceedings/papers/PA104.pdf

Macdonald, J., & Schneider, J. (2019). Battlefield Responses to New Technologies: Views from the Ground on Unmanned Aircraft. *Security Studies*, 28(2), 216–249. https://doi.org/10.1080/09636412.2019.1551565

Masrom, M. (2007). Technology Acceptance Model and E-Learning. *Technology, 81*.

Mayring, P. (2022). *Qualitative Content Analysis: A Step-by-Step Guide*. Sage.

Smith, F. L. (2020). Quantum Technology Hype and National Security. *Security Dialogue, 51*(5), 499–516. https://doi.org/10.1177/0967010620904922

Spreen, D. (2023). Lethal Autonomous Weapon Systems (LAWS). On the Ethics of Automation in the Military from the Perspective of Social Systems Theory. *Sõjateadlane (Estonian Journal of Military Studies)*, (21), 10–41. https://doi.org/10.15157/st.vi21.24177

Thorhallsson, B., & Steinsson, S. (2017). Small State Foreign Policy. *Oxford Research Encyclopedia of Politics*. https://doi.org/10.1093/acrefore/9780190228637.013.484

Venkatesh, V., & Davis, F. D. (2000). A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Management Science, 46*(2), 186–204. https://doi.org/10.1287/mnsc.46.2.186.11926

Venkatesh, V., & Hillol, B. (2008). Technology Acceptance Model 3 and a Research Agenda on Interventions. *Decision Sciences, 39*(2), 273–315. https://doi.org/10.1111/j.1540-5915.2008.00192.x

Vu, H. T., & Lim, J. (2022). Effects of Country and Individual Factors on Public Acceptance of Artificial Intelligence and Robotics Technologies: A Multilevel SEM Analysis of 28-country Survey Data. *Behaviour & Information Technology, 41*(7), 1515–1528. https://doi.org/10.1080/0144929X.2021.1884288

Wagner, W., Talves, K., Viidalepp, A., Otsus, M., Pekarev, J., Lillemäe, E., Cooper, C., & Hellestveit, C. (2022). *Attitudes and Experience of Military Personnel with Unmanned Ground Systems: Research Report*. Estonian Military Academy.

AI and Military Conflict

Human Factor and Military Technology in Warfare: A Historical Perspective

Igor Kopõtin, Kaarel Piirimäe, and Arto Oll

Abstract The development of military organization has always gone hand in hand with the development of military technology. A great many inventions were, from the outset, connected with warfare or were used in warfare. Weapons systems created as a result of the development of military technology have had a major impact on warfare, especially in the twentieth century, when the Industrial Revolution finally bore fruit, expanding throughout Europe. The experience of military history shows that any military technological invention can only be effective if it is successfully integrated into the military organization and its use is thought through within a theoretical concept. The mere possession of military technology does not always produce results or win the war. In the final analysis, war is still a war between people, and the human factor will always be decisive, because the weapon, whatever it may be, will always remain a tool in human hands (assuming that humans will be able to maintain control of artificial intelligence in the future). This is the nature of war, as military history has demonstrated, but it is also the very nature of artificial intelligence that it makes the future essentially unpredictable.

Similarly to Francis Fukuyama (Herwitz, 2000, pp. 222–234), Martin van Creveld essentially declared the end of military history in 1991. In his work *The Transformation of War*, he made a prediction that after the end of the Cold War,

I. Kopõtin (✉)
Estonian Military Academy, Tartu, Estonia
e-mail: igor.kopotin@mil.ee

K. Piirimäe
Institute of History and Archaeology, University of Tartu, Tartu, Estonia
e-mail: kaarel.piirimae@ut.ee

A. Oll
Esonian Maritime Museum Foundation, Tallinn, Estonia
e-mail: arto.oll@meremuuseum.ee

K. Talves, D. Spreen (eds.), *Artificial Intelligence in Military Technology*,
Artificial Intelligence, Simulation and Society 192,
https://doi.org/10.1007/978-3-031-95578-5_7

there would no longer be conventional warfare. Instead, he believed that regular armies would decrease in numbers and focus more on performing policing functions in low-intensity conflicts. Accordingly, he believed that military technology would evolve, leading to an end in the production of conventional weapons as we know them. Furthermore, Creveld referred to weapons as expensive toys for generals and politicians and argued that conventional military technology had no military future at large because the emphasis was shifting toward hybrid warfare, where influence operations through information warfare and cyber warfare were expected to play a significant role (Creveld, 1991, pp. 206–210).

The Russo-Ukrainian war shows that Creveld, along with several Western and Russian military experts, has been deeply mistaken or at least rushed in dismissing conventional warfare. Exactly 100 years after the start of World War I, the Russo-Ukrainian conflict, which threatens to escalate globally, paradoxically demonstrates that despite technological progress in warfare, frontlines and trench warfare have reemerged and maneuvering on the battlefields is as difficult as hundred years ago (Ukraine's commander-in-chief, 2023). What are the reasons for this? Could the answer lie in the interaction between military technology and armed forces, which can be simplistically referred to as finding the right balance between weapons and humans?

On 25 October 1854, Lord Cardigan led a frontal attack by the Light Brigade of the British cavalry on the Russian artillery battery in the Battle of Balaclava (Adkin, 1997; Woodham-Smith, 1991). Immortalized in verse by Alfred, Lord Tennyson, this ill-fated assault resulted in disorder and heavy losses. This remarkable battle is interesting from several perspectives. On one hand, it was a clash between brave cavalrymen and military technology—lances against cannons. On the other hand, this desperate and hopeless attack was ultimately caused by what is known as human error—a misinterpreted order. Lastly, this battle took place during the Crimean War, which can be considered one of the first wars of the Industrial Era where Russia clearly suffered due to outdated military organization and technological inferiority compared to Western countries (Kagan, 2002, pp. 132–133). Let us now examine "Lord Cardigan's mistake" from these three angles—technology, human factor, and historical context.

1 Military Technology and Stalemate in Warfare

One of the main modern theories that helps to understand the influence of military technology on military history is the model known as Revolution in Military Affairs (RMA). According to this theory, one of the most important RMAs was the Industrial Revolution, which resulted in the invention and production of a large amount of weaponry with incredible firepower. The development could be seen as having culminated in the crisis of 1914 when a stalemate emerged, leading to a protracted positional war. The problem was that the firepower was so devastatingly strong that

the attacking side could not overcome it. According to the theory, the First World War is also called one of the RMAs in which the legacies and effects of the French Revolution and the Industrial Revolution combined to set a pattern for twentieth-century warfare (Murray & Knox, 2001, p. 6).

In this way, a continuous frontline was formed and maneuvering disappeared from the battlefield. Was advanced military technology solely to blame for this? Certainly not. The problem lay, among other things, in the lack of corresponding doctrinal concepts for the use of military technology (Hacker, 2005, p. 257). Essentially, major powers entered World War I with tactics reminiscent of the Napoleonic Era, seeking an opportunity to engage in one decisive battle. The problems that arose during World War I had actually been evident earlier, even in the Second Boer (1899–1902) and the Russo-Japanese War (1904–1905). However, the countries involved in these wars and those observing them failed to draw the correct conclusions from these experiences (Doughty, 2011, pp. 176–177).

The Mukden Battle (1905) is an illustrative example, in which more than half a million soldiers participated on both sides and the frontline stretched for 100 km and 60 km in depth. However, the Japanese attempted to encircle the Russian troops in that battle using the classical scheme of maneuver from the Battle of Cannae (216 BC) or the Battle of Sedan (1871). The increased firepower that favored defense, the emergence of mass armies, as well as insufficient mobility of advancing forces prevented the Japanese from completing their maneuver (Menning, 2002, pp. 204–205). One might think that if Schlieffen's Plan had been executed in 1914 by motorized troops, the Germans could have easily captured Paris. In fact, German troops had both railways and motorized transport companies to provide supplies. At the Battle of Marne, the German supply roads were so long that the nearest operational railway station for first German Army was 140 km behind the front, and in the motorized transport companies 60% of the cars were out of order due to intensive logistic activities (Herwig, 2016, p. 211). By fate alone, it was precisely the French who used railway network in Parisian suburbs and mobilized all civilian transport in Paris, and managed to concentrate their sixth army against the right flank of advancing Germans in the Battle of the Marne 1914, thus saving their position. The motorized and railway maneuver at the Battle of the Marne has even been referred to as a "revolution in transportation" (Doughty, 2011, pp. 170, 176). In general terms, it can be said that one problem leading to positional warfare in 1914 was inflexible military thinking and limited ability of armies as organizations to draw conclusions and apply them as innovations for achieving greater military effectiveness. The same can be said about the greatest confrontation between battleships in the Battle of Jutland 1916, where a decisive naval engagement yielded no desirable results. This was the consequence when historical memory was implemented into naval service culture and therefore defined as the nature of future warfare. The outcome of the battle even called into question the principles of conducting naval warfare and led a prominent naval thinker, Admiral Sir Herbert Richmond, to conclude that the battleship could be defeated by more versatile ships of inferior size (Baugh, 1993, p. 33). So, despite the rapid development of naval technology along with

complicated weapon systems, the navy entered the twentieth century with Nelsonian tactics in mind.

2 The Impact of Military Technology on Military Thinking

However, military technology in World War I not only benefited the defending side. The attacking forces constantly thought about the technical possibilities of breaking through a heavily fortified enemy defensive line. To achieve this, toxic chemicals were used, artillery was concentrated, and suppressive fire was developed, tanks and airplanes were employed, flamethrowers and hand grenades were utilized, as well as the development of small arms such as light machine guns aimed at increasing the firepower of the attacking side (Hacker, 2005, p. 258). Regarding the German army's spring offensive in 1918, it is worth noting the successful use of small unit tactics based on elite highly motivated assault troops, which demonstrated the value of soldiers' morale preparation (Addington, 1994, p. 166). On the one hand, it was a new tactic adapted to the conditions of trench warfare, but on the other, it relied on what might be called elite soldiers who stood out for their exceptional personal courage. Despite occasional success in breaking through enemy defenses due to technological innovation—for example, in battles like Cambrai or Second Battle of Marne or Brusilov Offensive—converting tactical success into operational advances proved elusive for advancing troops. The problem lay both in insufficient troop mobility and inadequate military-theoretical foundations for offensives, which prevented breakthrough forces from exploiting their successes.

One of the responses to the aforementioned problems was the mechanization of troops, as well as the development of military aviation after World War I. Some historians distinguish the interwar period separately as the so-called era of mechanization, others emphasize advances in military theory and operational art (Hacker, 2005, p. 256; Corum, 1992, pp. 122–125). Either way, it was a time not only of explosive development in military technology, but also of military thought, which sought to determine the nature of future wars and create a theoretical basis for the use of military technology, primarily to return maneuver to the battlefield. Despite being pioneers in tank warfare, with an impressive number of armored vehicles by the end of WWI, the Brits essentially abandoned further development of tank forces in the late 1920s. Major General J. F. C. Fuller, a proponent for mass utilization of tanks and a British military theorist, believed that aggressive and adventurous foreign policies pursued by Adolf Hitler contributed to fostering tank forces as an offensive armament. This explains Fuller's sympathy toward Hitler (Fuller, 1977, p. 244). Even as late as the 1930s, many generals in Germany did not believe in the success of tanks on the battlefield, and there was talk that effective anti-tank weapon would prepare tanks for the fate of the cavalry horse in the future warfare. On the other hand, it demonstrates the importance of political pressure in military innovation, which could overcome the inertia in the thinking of generals who tend to draw

on the experience of previous wars (Murray & Millet, 1998, pp. 24–43, 45–46). In any case, limited foreign policy goals of the British leadership probably contributed to the neglect of developing large-scale tank formations (Bond & Alexander, 1986, pp. 605–606, 610–612).

The experience of the Second World War shows that the military-theoretical justification for the use of military technologies became a decisive factor in achieving military success. For example, the Germans had fewer tanks than allied forces during the French campaign in 1940. Furthermore, German tanks were no better and significantly inferior to their Western counterparts in terms of tactical and technical characteristics. The success of German forces from 1939 to 1941 was primarily based on the organization of tank troops into large operational units—divisions and corps, and later into armies (Pöhlmann, 2016, pp. 210–211; Miksche, 1994, pp. 232–234; Przybyło, 2019, pp. 127–129; Ganz, 2016, pp. 3–4, 12). Thus, one can say that it was ultimately human factors that determined attitudes toward technological innovation. It is noteworthy that the modern American theorist William S. Lind believes that in contemporary maneuver warfare, commanders' decisiveness bordering on adventurism plays a significant role, as maneuver warfare does not have a fixed template and requires the application of mission command principles (Lind, 1985, pp. 7, 22–23). Undoubtedly, in Lind's arguments, one can recognize the eternal Clausewitzian war-theoretical debate—whether warfare is an art or a science. However, we can also ask: If they are art and science at the same time, how do they relate to each other?

To some extent, Germany's success in developing blitzkrieg theory can be compared to that of the Soviet Union's development of deep operations theory as doctrinal justification for using military technology to achieve its main objective of operational art—breakthrough through fortified defense and converting tactical success into operational ones. Interestingly enough, these two revolutionary theories in warfare were developed by countries that suffered defeat during WWI. In contrast, France—which emerged victorious from WWI—remained trapped by outdated but proven practices regarding engineering defense structures and heavy artillery usage. This resulted in significant investment being made toward fortifying France's borders by what is known as the Maginot Line, which stood in stark contrast to maneuver-based theories employed by both Germany and Soviet Russia. Therefore, it can be concluded that theoretical justifications for using tanks and tactical aviation directly supporting land forces were based on interpretations drawn from lessons learned during previous wars. These justifications proved more successful for those who ended up on the losing side.

In any case, military technological innovation is far from being military innovation. The point is that a military may commission military technological innovation but may not always correctly assess its impact on the battlefield and on war in general, which is what happened to the Western allied forces in the First World War and the interwar period. From this point of view, military innovation is more than a mere military technological invention, which in turn is only a part of it. In other words, "military innovations combine and utilize existing technologies and techniques in new ways," in which the ability of the armed forces as an organization to implement

change by creating a conceptual framework for the use of technology plays an important role (Horowitz & Pindyck, 2023, pp. 85–86, 88, 90–93).

3 Military Technology and War Success

To some extent, one might think that the use of innovation and military technology in the history of conventional warfare was primarily the choice of a weaker adversary seeking opportunities to defeat a numerically and qualitatively superior enemy (Mey, 1998, pp. 310, 313). A prime example in this regard would be the French Jeune École naval concept of the 1870–1880s, which advocated the use of innovative technology (torpedo boats, submarines, and underwater weaponry) to challenge a much stronger British navy in open seas. This was an original theoretical naval concept to attack British sea lines of communications with smaller warships that can also avoid fleet engagements with a more powerful battle fleet (Roksund, 2007). Eventually this strategical concept was neglected on the basis that technology could not provide an ocean-going torpedo boat.

One can recall the biblical example of David using a sling against the physically stronger Goliath or the ancient Greeks employing disciplined citizen-soldiers in phalanx formations against their Persian foes (Hassan & Brosius, 2017, pp. 207–276; Worthington, 2017, pp. 503–573, compare Delbrück, 2003). Military innovation became a key factor in the success of the Israeli army against its numerous Arab adversaries in several consecutive wars (Bitzinger, 2021). Innovation and military technology may also be one of the main directions for developing the defenses of countries bordering Russia, or Ukraine in the ongoing war, given its more limited material and human resources. On the other hand, the mass production and use of military technology is a prerogative of wealthy countries. This trend is especially noticeable during and after the Cold War. Nowadays, traditional military technology—tank and jets, which were innovative in the interwar period, have become very expensive and the loss of a few modern tanks or aircraft in a battle can affect the conducting of the whole operation or campaign. That is why war theorists who advocated the RMA theory, including van Creveld itself, have wrongly predicted the fate of the dinosaurs for traditional weapons, believing that they will be replaced by lighter and more high-tech tools and weapon systems (Creveld, 1991, pp. 208–209).

The mass production of weapons systems had always been expensive and linked to the country's economic capacity. It is worth remembering that arms race between the USA and the USSR eventually led to serious problems in the Soviet economy and may be linked to the collapse of the USSR. After the end of the Cold War, the US attempted to maintain its global supremacy by permanent inventing and using of advanced military technologies. Despite the fact that the adversaries of the US were attracted by its technologies and also used US invented weapons, the most effective methods against the US proved to be asymmetric warfare (O'Hanlon, 2000, pp. 2–4). Thus, the advanced military technologies applied by the US did not lead to decisive

victories, most strikingly in the war in Afghanistan. Equipped with the most advanced weapons and having excellent organization, the Western coalition suffered defeat, and the Afghan National Army they created ceased to exist within a matter of hours. Samples of the latest American weaponry became war trophies for Afghan insurgents. In the Iraq war, US forces and their allies proved effective against the Iraqi regular army, but encountered asymmetric resistance from insurgents, which has not yet been fully resolved there. It showed that a weapon, no matter how powerful and clever, is only a tool in the hands of the military, and that a doctrinaire and organizational integration of weapon systems and adaption to the changing face of the battle is needed to achieve strategic objectives of war. Equally paradoxical is the phenomenon where war is fought mostly with the mass use of the previous generation of weapon systems (Hacker, 2005, p. 262), and the most innovative military technology cannot be a Wunderwaffe to turn the tide of war.

Time and again, we see that the presence of military technology does not guarantee military success in and of itself. What is the mystery behind this? Some researchers believe that the main reason is once again a mismatch between strategy and actual conditions of war, because new weapons require new doctrine, tactics, and organization (Hacker, 2005, p. 273). Principles of conventional warfare simply do not work in asymmetric and counterinsurgency or non-linear warfare. A prime example of this is General William Westmoreland's attempt to gather enemy guerrillas in one place during the Vietnam War to defeat them in one decisive battle (Daddis, 2014, p. 94). However, the jungles of "Iron Triangle" in Mekong Delta or areas near Khe Sanh bore little resemblance to the fields and hills of Gettysburg. Thus, it is evident that also the Russian political-military leadership made mistakes when choosing their strategy toward Ukraine in February 2022 when they attempted to replicate operations like seizing Czechoslovakia in 1968 or Afghanistan in 1979. This would have worked under conditions where there was no determined resistance from the enemy; and it left woefully unresolved the problem of holding territory if a guerrilla war broke out.

Historical experience shows another important aspect in the use of military technologies. This is the effect of soldier's morale and risk tolerance, because, as Clausewitz has put it succinctly, "moral values cannot be ignored in war" (Clausewitz, 1976, p. 137). It was precisely the moral superiority of Afghan insurgents that gave them a crucial advantage over Afghan government forces and partially even over allied troops (Johnson, 2016, pp. 248, 250–257). It could be said that Western societies are sensitive to military casualties, which undoubtedly became one of the factors leading to the defeat of allies in Vietnam or later in Afghanistan. So, Herfried Münkler describes human societies as "post-heroic societies" (Münkler, 2007). Perhaps this phenomenon, also called "post-heroic warfare," began with the experience of World War I, after which there was talk of a "lost generation" who had experienced the horrors of trench warfare. It was this aspect, as well as personal combat experience, that prompted Sir Basil Liddell Hart to develop his theory known as "indirect approach." In order to avoid large human losses in future war armed conflicts still being effective in achieving war objectives, Liddell Hart proposed focusing on developing mechanized forces during peacetime

(Hart, 1994, pp. 231–232). In this way, the West from then on systematically tried to develop military technology to save the lives of its soldiers.

On the one hand, Western theorists were guided by humane feelings toward preserving the lives of their soldiers, but on the other hand, the number of casualties among the civilian population of the enemy increased. In general, this was associated with the development of a common understanding of total war, within which war ceased to be just a confrontation between armies but encompassed all populations in warring countries. Germany did not develop strategic aviation because the theory of Blitzkrieg itself excluded the possibility of transitioning to a protracted total war that Germany could not win due to its lack of necessary raw materials reserves. Nevertheless, after World War II, the West began paying more attention to developing military technologies in order to compensate for limits on manpower (Hacker, 2005, p. 273) and constantly decreasing morale among its soldiers. This process led to what can be considered as a permanent Revolution in Military Affairs in the United States and resulted in what can be considered the first high-tech war during Operation Desert Storm in 1991. As a result, the use of military technology became much more pervasive than during World War II or Cold War years.

In terms of air force utilization, Anglo-American theorists relied on the theory of Giulio Douhet, who saw aircraft as an opportunity to increase artillery range. Thus, an airplane became like a flying gun and an aviation unit like an artillery battery (Douhet, 2009, p. 200). An important distinction between Western theorists studying bomber aviation development and Soviet or German authors primarily lay in assigning strategic goals for aviation—breaking enemy will through massive bombing campaigns targeting industrial and political centers. One of the goals of strategic bombing was to break the morale of the population of the enemy country (Biddle, 2002, p. 245). In practice, however, the massive aerial bombardment of enemy territory did not produce the expected strategic results, neither in Germany during World War II, in Vietnam in the 1960s and 1970s, nor in the NATO campaign over Yugoslavia in 1999 (Gentile, 2000, pp. 191–194). It should also be remembered that massive bombings of Germany and Vietnam did not directly lead to the achievement of strategic goals in the war, but caused losses and resentment or even some consolidation of the enemy country's population. As we can see, Russian bombings of Ukraine with the aim of destroying industry and infrastructure, as well as intimidating the population, do not have a significant effect. At the same time, it is necessary to acknowledge the very successful actions of the Air Force against Yugoslavia, Iraq, and Libya in 1990s and 2000s (Gentile, 2000, pp. 191–194). However, as can be seen from the example of Ukraine, conventional wars are still fought for control over territories and land cannot be considered captured until an infantryman sets foot on it.

4 Military Technology by the Russian Way of War

It is noteworthy that the idea of RMA was partially borrowed by Western authors from Soviet military theory (Bukkvoll, 2011, p. 684; Metz & Kievit, 1995). Essentially, the entire course of military history in RMA is viewed through the prism of the influence of military technology, in the development of which Russia usually did not have a leading role. Moreover, the development of military theory in Russia before World War I was characterized by a clash between the so-called academic and national schools of military thought. One of the main questions addressed by Russian theorists at that time was precisely the relationship between weapons and soldiers, which they linked to the debate about what was the Russian "national art of war." While "academics" believed that principles of warfare were universal for all nations but their application differed based on national characteristics, representatives of the national school or "traditionalists" emphasized moral superiority of Russian soldiers over enemies. In search for a national idea in military science, "traditionalists" juxtaposed weapons with soldier morale and considered frontal attack with cold weapons (simplified as bayonet attack) as decisive to be enshrined in tactical doctrine (Steinberg, 2010, pp. 50–51; Pintner, 1986, p. 367; Baumann, 2002, p. 145). The Light Brigade's charge at Balaklava battle, which the British considered a mistake, was elevated by the Russians to the status of tactical doctrine, as a result of which, one can argue, Russia "successfully" lost both Russo-Japanese War and World War I. It should be noted that their opponents—"academics"—proposed instead the idea that weapons complemented soldiers, serving as means for achieving victory. Thus, relying on Western theorists' ideas, they placed emphasis on firepower and maneuver over frontal bayonet attacks.

One could say that one of the culminating moments in Russian military theory was the creation of the theory of deep operations, which essentially established a doctrinal concept for utilizing mass mechanization to achieve operational success (Kipp, 1990, pp. 106–110). It was an outstanding theory, hindered by weak organization and equipment within the army, insufficient training, and political repression against Soviet military personnel. In any case, this theory can be called the maneuver revolution in the RMA in response to the previously dominant firepower revolution (Kagan, 2010, p. 79). The concept of the deep operation, born out of the workings of Soviet military thought in the interwar period, as well as the invention of the art of operations, inspired US forces to change their conceptual perceptions in the 1980s and develop their effective AirLand Battle operational methods as demonstrated in the Gulf War (Bukkvoll, 2011, p. 684). But the Soviet concept of deep operations is also influencing the Ukrainian military leadership today as they seek a way out of the attritional positional war. However, they are greatly inhibited by the lack of sufficient numbers of modern armor, aircraft and even artillery projectiles. Same as in the Second World War, armor without close air support or artillery support is helpless against fortified infantry and artillery.

Overall, it can be said that the idea of operational art was developed by Russians from the perspective of a continental state's army, and the theory of deep operations

primarily focused on skillfully employing mechanized troops en masse. It was thus different from the ideas of Fuller and German enthusiasts, who imagined relatively small elite armies made up of mechanized forces overcoming the opposition of numerically larger but technologically less-advanced opponents. As is known, Soviet military theory had some influence on Western military science but is unlikely to be adopted by powers, such as the United States and Great Britain that have historically had a stronger emphasis on naval and aerial power. In any case, the notion of perceived moral superiority among Russian soldiers remains relevant in Russia even today as they attempt to compensate for their technological lag. The presence of vast human resources is evident even in the current Russo-Ukrainian war where there seem to be no significant difficulties for the Russian leaders or the population at large regarding accepting substantial human losses.

Currently, there is a so-called drone discussion going on—whether the use of unmanned weapon systems is the threshold of a new revolution in military affairs (Rossiter, 2023, pp. 253–255). To what extent can technology replace humans on the battlefield? Though it remains to be seen, the Russo-Ukrainian War may present another RMA. Although both sides for the most part rely on very conventional, even antiquated weaponry—Russia in particular making use of every item it can finds in its vast Cold War-era stores (Ukraine lacking such a luxury)—at the same time there is also a true arms race with both sides deploying vast numbers of unmanned aerial as well as land vehicles, which advance in complexity and ability almost by the hour. Latest advancements have seen AI detecting and targeting enemy in what appears to be a new generation of network-centric warfare. These systems are also developed and deployed by Israeli Defence Forces (IDF) in what is quite a different kind of warfare in Palestine but still seem to point toward a new quality in tactical methods. As to Ukraine, this is an example how a side with an inferiority in resources can compensate with smarter use of technology. Whereas Russia invests in innovation also, Ukraine seems to have the advantage of having a more decentralized armed forces where private initiative in encouraged, having the result that lessons are identified, learned, and adopted quicker than in a centralized system, such as the Russian armed forces. It remains to be seen if the organizational agility and adaptability will lead to significant success in the battle field, or the Ukrainian efforts will fizzle out as those of Imperial Germany in 1918 and have to succumb to a superiority of resources of the other side.

Nevertheless, it is worth recalling Clausewitz, who believed that "war is an act of human intercourse" and is "part of man's social existence" (Clausewitz, 1976, p. 149). It can be argued that where there are humans, mathematical logic does not always work. Eventually, the objective of war is to impose one's will on the other, until the adversary accepts defeat, and this is a matter of psychology as much as kinetic force (Clausewitz, 1976, p. 75). However, a war solely between robots, with no human casualties, would make it a fiction. Therefore, it is probably true what Professor Holger H. Mey wrote: "War without human life will remain an illusion" (Mey, 1998, p. 318). On the other hand, one can imagine a future army achieving maximum effectiveness, employing AI, robots, etc., and being able to annihilate all of humanity.

Of interest in this regard is British psychologist Norman F. Dixon's interpretation of Lord Cardigan's Light Brigade charge. In it he saw fundamental problems in the British army's organization and in the military training of officers, when incompetent nobles were placed in command of larger units. Ultimately, it was the human inability to orientate themselves in a rapidly changing tactical situation, and shortcomings in command and communication, that led to the disaster of the Light Brigade in the Battle of Balaclava in 1854 (Dixon, 2016, pp. 98–99). Even with advanced in warfighting systems deploying AI, the shortcomings of the human factor and of the specific features of military organizations will never be entirely overcome.

Returning to the operational impasse in the Russo-Ukrainian war, it should be noted that the war is being fought between adversaries who have emerged from the same Soviet military-theoretical school. In terms of military technology utilization, Russia's progress has turned out to be a facade, and its army still relies on mass deployment. To what extent it can innovate, having inherited from the Soviet period and preserved a top-down, authoritarian military culture, is a moot question. On the other hand, Ukraine's army lacks sufficient saturation with Western models of weaponry, and it is forced to make due with "weapons of the weak," such as the UAVs. Will it be sufficient to break the stalemate in what has become a protracted positional warfare is still to be seen. The solution may lie not in a new technology but in managing larger operational formations beyond brigades. This is operational art that neither Russia nor Ukraine has thus far mastered. Perhaps this is precisely because neither the West nor Russia (and Ukraine) anticipated the possibility of a major conventional war. In any case, merely relying on military technology alone will not ensure victory in warfare if there is no corresponding theoretical conceptualization of its application.

5 Conclusions

The development of military organization has always gone hand in hand with the development of military technology. The development of military technology can be characterized in terms of defensive and offensive: the sword and spear were opposed by the shield, just as the tank was opposed by anti-tank weapons. A great many inventions were, from the outset, connected with warfare or were used in warfare: the internal combustion engine, the telegraph and the telephone, or even the Internet. Weapons systems created as a result of the development of military technology have had a major impact on warfare, especially in the twentieth century, when the Industrial Revolution finally bore fruit, expanding throughout Europe. Military organizations that failed to take account of the specificities of military technology faced a problem: the other side's firepower was so massive that it did not allow for conducting of successful maneuvers. This problem culminated in the First World War in the trench warfare, which forced the belligerents to look for military technological solutions as well as to think of new operational concepts. The

countries that lost the war—Germany and Russia (the Soviet Union)—were the most successful in this area, having developed the relevant concepts in the light of modern military technology. These were so innovative that war theorists and practitioners return to them even today.

Every war gives a major impetus to the development of military technology. This was also the case in the Second World War, when weapons systems developed rapidly. Another wave of innovations in technology and tactics emerged when the West realized that it may be forced to fight a conventional war despite the existence of nuclear weapons, precisely because the catastrophic consequences of their use had their deployment unthinkable but may not deter an aggressive adversary launching a conventional war. This dilemma has not been solved even now.

With the collapse of the Soviet Union, many people thought there would be no more conventional war in the future, and that the armed forces would mostly be used for police functions. Consequently, the focus shifted to developing precision weapons and cyber warfare. Traditional weapon systems for conventional warfare became very expensive and their production began to decline. However, the outbreak of a full-scale war between Russia and Ukraine in 2022 overturned "the end of war" proposition, which war theorists had proclaimed, and a new war of position emerged, despite the availability of modern high-technology. In all likelihood, the problem again lies in the lack of a corresponding operational concept that would take account of the specific features arising from the development of new military technology.

The experience of military history shows that any military technological invention can only be effective if it is successfully integrated into the military organization and its use is thought through within a theoretical concept. The mere possession of military technology does not always produce results or win the war. In the final analysis, war is still a war between people, and the human factor will always be decisive, because the weapon, whatever it may be, will always remain a tool in human hands (assuming that humans will be able to maintain control of artificial intelligence in the future). This is the nature of war, as military history has demonstrated, but it is also the very nature of artificial intelligence that it makes the future essentially unpredictable.

- Military history shows that any military technological invention can only be effective if it is successfully integrated into the military organization.
- The mere possession of military technology does not always produce results or win the war.
- Human factor will always be decisive, because the weapon will always remain a tool in human hands.
- Nature of artificial intelligence makes the future warfare essentially unpredictable.

References

Addington, L. H. (1994). *The Pattern of War Since the Eighteenth Century* (2nd ed.). Indiana University Press.

Adkin, M. (1997). *The Charge: The Real Reason Why the Light Brigade Was Lost*. Leo Cooper.

Baugh, D. E. (1993). Admiral Sir Herbert Richmond and the Objects of Sea Power. In J. Goldrick, & J. B. Hattendorf (Eds.), *Mahan Is Not Enough: The Proceedings of a Conference on the Works of Sir Julian Corbett and Admiral Sir Herbert Richmond* (pp. 13–49). Naval War College Press.

Baumann, R. F. (2002). The Russian Army 1853–1881. In I. F. W. Kagan, & R. Higham (Eds.), *The Military History of Tsarist Russia* (pp. 137–150). Palgrave Macmillan. https://doi.org/10.1007/978-0-230-10822-6_8

Biddle, T. D. (2002). *Rhetoric and Reality in Air Warfare. The Evolution of British and American Ideas about Strategic Bombing, 1914–1945*. Princeton University Press.

Bitzinger, R. A. (2021). Military-technological Innovation in Small States: The Cases of Israel and Singapore. *Journal of Strategic Studies, 44*(6), 873–900. https://doi.org/10.1080/01402390.2021.1947252

Bond, B., & Alexander, M. (1986). Liddell Hart and De Gaulle: The Doctrines of Limited Liability and Mobile Defence. In P. Paret, G. A. Craig, & F. Gilbert (Eds.), *Makers of Modern Strategy from Machiavelli to the Nuclear Age* (pp. 598–623). Clarendon Press. https://doi.org/10.2307/j.ctv8xnhvw.24

Bukkvoll, T. (2011). Iron Cannot Fight—The Role of Technology in Current Russian Military Theory. *Journal of Strategic Studies, 34*(5), 681–706. https://doi.org/10.1080/01402390.2011.601094

Clausewitz, C. v. (1976). *On War* (M. Howard, & P. Paret, Ed. and Trans.). Princeton University Press (Original work published 1832).

Corum, J. S. (1992). *The Roots of Blitzkrieg: Hans von Seeckt and German Military Reform*. University Press of Kansas.

Creveld, M. (1991). *The Transformation of War*. The Free Press.

Daddis, G. A. (2014). *Westmoreland's War. Reassessing American Strategy in Vietnam*. Oxford University Press.

Delbrück, H. (2003). *Geschichte der Kriegskunst. Das Altertum. Von den Perserkriegen bis Caesar*. Nikol. (Original work published 1900).

Dixon, N. F. (2016). *On the Psychology of Military Incompetence*. Basic Books.

Doughty, R. A. (2011). Winning and Losing. France on the Marne and the Meuse. In M. S. Neiberg (Ed.), *Arms and the Man. Military History Essays in Honor of Dennis Showalter* (pp. 169–192). Brill. https://doi.org/10.1163/ej.9789004206687.i-275.28

Douhet, G. (2009). *The Command of the Air*. The University of Alabama.

Fuller, J. F. (1977). *The Conduct of War 1789-1961. A Study of the Impact of the French, Industrial and Russian Revolutions on War and Its Conduct*. Methuen.

Ganz, A. H. (2016). *Ghost Division. The 11th "Gespenster" Panzer Division and the German Armored Force in World War II*. Stackpole Books.

Gentile, G. P. (2000). *How Effective Is Strategic Bombing? Lessons Learned from World War II to Kosovo*. New York University Press.

Hacker, B. C. (2005). The Machines of War: Western Military Technology 1850–2000. *History and Technology, 21*(3), 255–300. https://doi.org/10.1080/07341510500198669

Hart, L. B. (1994). The Indirect Approach. In L. Freedman (Ed.), *War* (pp. 231–232). Oxford University Press.

Hassan, C., & Brosius, M. (2017). The Persian Wars, 492–450 BC. In M. Whitby, & H. Sidebottom (Eds.), *The Encyclopedia of Ancient Battles* (Vol. I, pp. 207–276). Wiley Blackwell. https://doi.org/10.1002/9781119099000.wbabat0100

Herwig, H. H. (2016). *Marne 1914. Eine Schlacht, die die Welt veränderte?* Ferdinand Schöningh.

Herwitz, D. (2000). Francis Fukuyama and the End of History. *South African Journal of Philosophy, 19*(3), 222–234. https://doi.org/10.4314/sajpem.v19i3.31317

Horowitz, M. C., & Pindyck, S. (2023). What is a military innovation and why it matters. *Journal of Strategic Studies, 46*(1), 85–114. https://doi.org/10.1080/01402390.2022.2038572

Johnson, R. (2016). The Taliban. In B. Heuser, & E. Shamir (Eds.), *Insurgencies and Counterinsurgencies. National Styles and Strategic Cultures* (pp. 246–266). Cambridge University Press. https://doi.org/10.1017/9781316471364.012

Kagan, F. W. (2002). Russias Small Wars 1805–1861. In F. W. Kagan, & R. Higham (Eds.), *The Military History of Tsarist Russia* (pp. 121–136). Palgrave Macmillan. https://doi.org/10.1007/978-0-230-10822-6_7

Kagan, F. W. (2010). The Rise and Fall of Soviet Operational Art, 1917–1941. In R. Higham, & F. W. Kagan (Eds.), The Military History of the Soviet Union (pp. 79–92). Palgrave Macmillan. https://doi.org/10.1057/9780230108219_6

Kipp, J. W. (1990). Mass, Mobility, and the Origins of Soviet Operational Art. In C. W. Reddel (Ed.), *Transformation in Russian and Soviet Military History: Proceedings of the Twelfth Military Symposium U. S. Air Force Academy 1986* (pp. 87–116). USAF Academy.

Lind, W. S. (1985). *Maneuver Warfare Handbook*. Westview Press.

Menning, B. W. (2002). Mukden to Tannenberg: Defeat to Defeat, 1905–1914. In F. W. Kagan, & R. Higham (Eds.), *The Military History of Tsarist Russia* (pp. 203–225). Palgrave Macmillan. https://doi.org/10.1007/978-0-230-10822-6_11

Metz, S., & Kievit, J. (1995, June 27). *Strategy and the Revolution in Military Affairs: From Theory to Policy*. Strategic Studies Institute, US Army War College. https://www.jstor.org/stable/resrep11727

Mey, H. H. (1998). The Revolution in Military Affairs: A German Perspective. *Comparative Strategy*, *17*(3), 309–319. https://doi.org/10.1080/01495939808403148

Miksche, F. O. (1994). Blitzkrieg. In L. Freedman (Ed.), *War* (pp. 232–234). Oxford University Press.

Münkler, H. (2007). Heroische und postheroische Gesellschaften. *Merkur, 61*(8/9), 742–752.

Murray, W., & Knox, M. (2001). Thinking about Revolutions in Warfare. In W. Murray, & M. Knox (Eds.), *The Dynamics of Military Revolution 1300–2050* (pp. 1–14). Cambridge University Press. https://doi.org/10.1017/CBO9780511817335.001

Murray, W., & Millet, A. R. (Eds.). (1998). *Military Innovation in the Interwar Period*. Cambridge University Press. https://doi.org/10.1017/CBO9780511601019

O'Hanlon, M. (2000). *Technological Change and the Future of Warfare*. Brookings Institution Press.

Pintner, W. (1986). Russian Military Thought: The Western Model and the Shadow of Suvorov. In P. Paret (Ed.), *Makers of Modern Strategy from Machiavelli to the Nuclear* Age (pp. 354–375). Princeton University Press. https://doi.org/10.2307/j.ctv8xnhvw.17

Pöhlmann, M. (2016). *Der Panzer und die Mechanisierung des Krieges. Eine deutsche Geschichte 1890 bis 1945*. Brill, Schöningh.

Przybyło, Ł. (2019). Building Military Doctrine based on History and Experience: 20th century examples from Germany, France, Israel and the US. *Estonian Yearbook of Military History, 9*, 114–150.

Roksund, A. (2007). *The Jeune École. The Strategy of the Weak*. Brill.

Rossiter, A. (2023). Military Technology and Revolutions in Warfare: Priming the Drone Debate. *Defense & Security Analysis, 39*(2), 253–255. https://doi.org/10.1080/14751798.2023.2178500.

Steinberg, J. W. (2010). *All the Tsar's Men: Russia's General Staff and the Fate of the Empire, 1898–1914*. Woodrow Wilson Center Press; Johns Hopkins University Press.

Ukraine's Commander-in-chief on the Breakthrough he needs to Beat Russia. General Valery Zaluzhny admits the war is at a stalemate. (2023, November, 1st). The Economist. https://www.economist.com/europe/2023/11/01/ukraines-commander-in-chief-on-the-breakthrough-he-needs-to-beat-russia

Woodham-Smith, C. (1991). *The Reason Why: The Story of the Fatal Charge of the Light Brigade*. Penguin Group.

Worthington, I. (2017). Campaigns of Alexander the Great, 336–323 BC. In M. Whitby, & H. Sidebottom (Eds.), *The Encyclopedia of Ancient Battles* (Vol. II, pp. 503–573). Wiley Blackwell. https://doi.org/10.1002/9781119099000.wbabat0230

On the Responsible Use of Artificially Intelligent Systems in Future Warfare

Wolfgang Koch, Jörg Vollmer, and Florian Keisinger

Abstract The augmentation of natural perception and volitive agency of humans through artificially intelligent technology necessitates a sober reflection. This applies above all to the use of Artificial Intelligence (AI) in the military domain. A transparent public discourse on such technologies needs to be conducted. With a view to the responsible use of artificially intelligent systems in future warfare, this chapter considers the interplay between science and technology and military requirements. Vice versa, understanding the potential of AI-based technologies will open up new operational options. Moreover, light is shed on legal and ethical considerations. A subjective assessment of the current security situation forms the starting point from which a military threat analysis emerges. A clarification of the role of AI in data and information fusion makes it clear why fusion engines need to accelerate the targeting cycle. Based on what is legally required and necessary for mission success, considerations of human-centric systems design form a central section. The Future Combat Air System serves as an example. Because military AI affects the way armed forces think and act, this chapter concludes with three recommendations. May it inspire a fruitful exchange between the different communities dealing with military operations and defense research.

W. Koch (✉) · J. Vollmer
Fraunhofer Institute for Communication, Information Processing and Ergonomics (Fraunhofer FKIE), Wachtberg, Germany
e-mail: wolfgang.koch@fkie.fraunhofer.de; joerg.vollmer@fkie.fraunhofer.de

F. Keisinger
Zentralverband der deutschen Seehafenbetriebe (ZDS), Hamburg, Germany
e-mail: florian.keisinger@zds-seehaefen.de

K. Talves, D. Spreen (eds.), *Artificial Intelligence in Military Technology*,
Artificial Intelligence, Simulation and Society 192,
https://doi.org/10.1007/978-3-031-95578-5_8

The augmentation of natural perception and volitive agency of human persons through artificially intelligent technology necessitates a sober reflection. Indeed, at the "sharp end of digitalization"[1] (Rieks, 2022), i.e., when utilizing Artificial Intelligence (AI) in the military domain, general challenges of AI-based machines become apparent as if seen in a spotlight. Hence, a transparent public discourse on the development of AI-based defense technologies is inevitable (Bossong et al., 2022).

While it seems necessary to consider military systems that are available on the market as a first step toward establishing deterrence against an aggressor, the allied nations' own plans, concepts, and their own will to fight in combat will define the required capabilities and identify capability gaps. Military planning therefore relies on advanced research on AI and comprehensive systems engineering to assess the strengths, weaknesses, opportunities, and risks of AI-based technologies from an operational perspective.

With regard to the responsible use of artificially intelligent systems in future warfare, this chapter considers the development of military AI technologies as an interplay of research push emphasizing the impact advancements in science or technology on innovation and concepts pull, where military requirements define, which technical capabilities are needed. Vice versa, understanding the potential of AI-based technologies will open up new operational options. Moreover, light is shed on legal and ethical considerations.

A naturally subjective assessment of the current security situation forms the starting point from which elements of a military threat analysis emerge. On this basis, a conceptual clarification of the role of AI in data and information fusion makes it clear why fusion engines need to accelerate the targeting cycle. Based on what is legally required and necessary for mission success, considerations of human-centric systems design form a central section. The Future Combat Air System serves as an example of their possible implementation. Regarding the influence of military AI on the way armed forces think and act, this chapter concludes with three recommendations. May it inspire a fruitful exchange between the different communities dealing with military operations and defense research.

1 An Overview of the Current Security Situation

Since February 24, 2022, at the latest, and more than thirty years after the end of the Cold War, Western societies are being forced to learn again, what a truly sustainable and precious good security is to achieve their common good.[2] Because without security, all other individual, social, cultural, economic, or even ecological goods

[1] We have translated quotes from German into English.

[2] "The political community […] exists for the Common Good: This is its full justification and meaning and the source of its specific and basic right to exist," summarizes a classical text a long tradition of political thought: "The Common Good embraces the sum total of all those conditions

remain unattainable. "In a world that turns everything into weapons, we cannot pretend to be just a soft power," stated Josep Borrell, High Representative of the European Union for Foreign Affairs and Security Policy, back in November 2021. Or to put it another way: "We like the world of Kant, but we will prepare ourselves to live in the world of Hobbes" (Gutschker, 2021).

Evidently, the US-American political scientist Francis Fukuyama was not right with his thesis of the "end of history" (Fukuyama, 1992), which heralds its own end with comprehensive world peace. Quite the opposite is the case. Part of this new and austere reality is the fact that armament activities are increasing around the globe (SIPRI Yearbook, 2024), with the focus not only on the hopefully rather symbolic pursuit of nuclear weapons, but on the use of latest AI-based technologies in combination with uncrewed platforms and weapon systems in all military domains.

The defense policy guidelines for the German armed forces, the *Bundeswehr*, dated November 2023, draw the following conclusions, to take an example: "Contemporary national and alliance defense" is the "core mission" of the *Bundeswehr*. However, this mission would not be able to withstand a war of attrition. With the "single set of forces," it must "cover the entire range of missions and tasks." The conclusion seems unavoidable that the defensive fight needs to be won already at the beginning of an aggression. Nevertheless, it will take time to build up the necessary industrial production lines to re-fill military stocks as an imperative first measure. Moreover, national key technologies relevant to Software Defined Defense (SDD) (Soare et al., 2023) must be developed quickly and made widely available in view of dramatically shortened innovation cycles as seen in the Ukraine war. For this reason, the German defense policy guidelines call for "the increasing expansion of methods and applications of artificial intelligence" and robotics (Bundesministerium der Verteidigung, 2023, p. 11).

In a digitalized world, there is no room for non-digital defense. Weapon systems must also use artificially intelligent automation now and in the future in order to be able to fulfill tactical and operational tasks effectively and precisely in the interests of the commanders. Already in training, but above all in operations, AI will accelerate military conflicts and change them altogether. The fundamental question is how soldiers can efficiently and responsibly work together with automated reconnaissance platforms and weapon systems that are no longer controlled remotely but also at least partly by AI. According to the German Federal Ministry of Defense, the importance of AI for the Bundeswehr lies therefore "not in the choice between human or artificial intelligence, but in an effective and scalable combination of human and artificial intelligence to ensure the best possible performance" (Bundesministerium der Verteidigung, 2019, p. 27). Besides ergonomic aspects, this statement also covers the legal and ethical dimensions of artificially intelligent weapon systems. According to these lines, the European Council has early demanded that "the allocation of functions between humans and AI systems should follow

---

of social life which enable individuals, families, and organizations to achieve complete and efficacious fulfilment" (Ecclesia Romana, 1975, p. 981).

human-centric design principles and leave meaningful opportunity for human choice. This means securing human oversight over work processes in AI systems" (High-Level Expert Group on AI, 2019, p. 14).

If NATO is "ready, willing, and able" to "defend every inch of allied territory," as the NATO Summit 2024 in Washington reaffirmed (Bundesministerium der Verteidigung, 2024), "contemporary national and alliance defense" includes credible deterrence and socially supported military capabilities to defend against an aggressor. This must not exclude operations on the territory of the aggressor.

A distinction is to be made between the strategic, operational, and tactical levels. At the strategic level, considering, for example, the use of nuclear weapons, technological aids for automating decision processes should be viewed with particular caution (Johnson, 2024, Chap. 4). Examples of AI-based technologies on the operational and tactical levels are systems for Intelligence, Surveillance, and Reconnaissance (ISR) and weapon systems for offensive military operations within a defensive overall strategy such as air combat systems or loitering munition. In this context, Multi-Domain Operations (MDO) will play a key role and require the synchronization of the effect achieved in several combat domains. Multi-Domain Operations are carried out simultaneously in all domains involved, but in such a way that their main effort can be shifted between the domains at will. Collaborative decision-making in the MDO context does not imply, however, that decision-makers from each domain have arbitrary access to the means of another domain as soon as they believe access is needed.

2 Some Elements of a Military Threat Analysis

A coherent assessment of the current military situation from a geostrategic and security policy perspective is expressly *not* the subject of this chapter. Nevertheless, presumed developments among potential adversaries are to be considered, i.e., primarily Russia. The potential theater of war between NATO member nations and Russia is likely to be North-Eastern Europe and Central Eastern Europe, where all military dimensions need to be considered.

Firstly, the question arises as to what options for action Russia still has after the Ukraine war or will have again in the foreseeable future. More particularly speaking, what forces, what military and technological capabilities will be at Russia's disposal, how will Russia adapt or change its military doctrine, how will its posture develop?

Secondly and more intimately related to topic of this chapter, the following question calls for an answer: What will Russia or any other potential adversary be capable of realizing with regard to artificially intelligent automation? There is no sufficient evidence of this aspect with a view of the ongoing Russian aggression toward Ukraine, which is only to some extent a relapse into WW I warfare. On the side of the defenders, platforms such as *Gepard*, *Leopard*, and field artillery as

examples, which have been discarded as supposedly obsolete, are unexpectedly relevant again, if operated according to the network-centric paradigm.

Since artificially intelligent automation and Electronic Warfare (EW) are essentially based on applied mathematics, an area where Russian scientists are traditionally strong, Russian capabilities should not be underrated despite the fact that numerous Russian scientists have left their home country. Since the NATO member states cannot materially afford any war of attrition, as is currently the case in Ukraine, rapid and effective counterattacks on the aggressor's own territory are necessary, for which artificially intelligent automation is required as an enabling technology within the interoperable framework of cooperating NATO forces in multiple domains.

Finally, Western countries should have at least a peripheral view of China, as NATO member states will be involved in the Indo-Pacific region in any kind of a US-led coalition of the Global West, to which certain contributions must be made at least as a sign of solidarity. It would therefore also be necessary to look at how Chinese military forces, capabilities, means, structures, and processes are developing and what role highly advanced Chinese AI developments are playing in this context.

In particular, the United States—Navy, Air Force, Marines—are consistently preparing for a geostrategic confrontation situation. Its allies should be able to contribute in their own particular way to crisis and conflict management and above all remain interoperable and compatible with allies. Technologically demanding challenges are more likely to arise in this direction than with regard to future Russian forces and capabilities. A look at Iran, North Korea, and other powers that are increasingly moving geopolitically into a camp led by China and Russia is necessary as well.

In general, a strictly threat-oriented analysis and security policy should be advocated, with a focus on Russia and its MDO capabilities, rather than following the established path of "capabilities-based planning" (Taliaferro et al., 2019) or "capabilities-based, threat-informed planning" (North Atlantic Treaty Organization, 2022), which apparently looks only at the existing capabilities offered to the Supreme Allied Commander Europe (SACEUR).

The next question to be answered would be in which direction main allies and partners are developing, above all the USA. Into which direction are these NATO member nations going technologically, how are they changing the overall posture of their armed forces? For the United States of America, the United Kingdom, France, or Israel, for example, artificially intelligent and automatically operating capabilities are first-order force multipliers, which will make the decisive difference between first-class and inferior armed forces in the future.

This applies even more to the interoperability of national armed forces with the armed forces of NATO partner countries, primarily Germany and the USA, and includes the need for uncrewed platforms, as well as swarms of them, and artificially intelligent, automated capabilities. The intention is to be able to deter a Russia that is politically, diplomatically, economically, militarily, and technologically supported by China to make war in Europe less likely, thereby strengthening security

and peace. The danger of war is reduced if the risk of war is significantly increased for the aggressor ("the problem of reducing the danger of war by increasing the risk of war," Luhmann, 1993, p. 105). In the medium term, however, it seems to be probable that US air and maritime power will be far less available in Europe. Moreover, modern national armed forces should be able to play a role in other military tasks such as crisis management, i.e., be interoperable and be able to contribute selected capabilities that are useful for the alliance within means and capabilities in their inventory. Provided these observations are valid, three conclusions seem plausible.

Firstly, the supposedly obsolete platforms such as tanks, artillery, anti-aircraft missiles must be considered, including spare parts and ammunition stocks to maintain adequate operational readiness and achieve credible endurance in a 30-day war. As the war in Ukraine shows, which cannot yet be conclusively assessed, networked digitalization and thus artificially intelligent automation is proving to be a key factor even in supposedly 'traditional' operations. In question is not 'heavy metal versus software,' but 'software-enabled heavy metal.' This seems to be dominant and not the fact of whether the main adversary, Russia, is technologically superior or not. It may be different with China, but China is not (yet?) Europe's main military adversary.

Secondarily, numerous Western countries seem to have an urgent need to invest step by step in the latest technologies in addition to major weapon platforms, spare parts, and ammunition—primarily in digital ISR and C2 capability, i.e., in technologies for achieving situational awareness, faster 'sensor to shooter' links than the enemy is capable to execute, decision support, etc.

Thirdly and on a broader scale than immediate defense capabilities require, one could consider whether state-of-the-art AI technology and automation developments could be used to achieve such superiority over Russia that it would be forced to abandon its imperial goals altogether. If this were the case, even if Russia were supported by China, it would be strategically worthwhile to invest massively in this direction.

Decisive for situational awareness is answering the question of 'what' in a military situation. Detection algorithms inform about the existence of relevant objects and phenomena, while classification algorithms provide information about their properties, i.e., their essence and intents. Important building blocks are furthermore algorithms that infer object interrelations. Finally, situation pictures have to indicate decision relevance, such as estimated threat levels and the status of own resources. Moreover, situation pictures must inform the users about their limitations and gaps as well as on the integrity of the situation picture, on the one hand, i.e., on the "known unknowns"[3] associated to it. On the other hand, it must correspond to the tasks, roles, and abilities of the decision-maker.

Besides their role in situational awareness and decision-making, algorithms also transform acts of will, i.e., decisions made, into partially or fully automated command sequences for controlling networking platforms, multifunctional sensors,

[3] According to the "Rumsfeld Matrix" (Krogerus & Tschäppeler, 2012, pp. 86–87).

communication links, and weapons systems. The question of 'why' to achieve an effect is crucial for the algorithm design. According to the four ways of answering to why questions, the intentions correspond to

1. The *final cause*, usually specified by performance parameters, characterizing the intended effect
2. The *effective cause* indicating by which concrete algorithms the effect is to be achieved
3. The *formal cause* answering the question, according to which predefined rules this should happen
4. The *material cause* indicating necessary resources to be used with their respective properties.

The close link between the formal and the final cause corresponds to the principle that military ends are to be achieved according to certain legally binding rules, e.g., the Rules of Engagement (RoE). The mission preparation phase corresponds to the link between material and formal cause. Finally, impact assessment determines the extent to which the final cause of the intended action has actually been achieved and is basic for further action.

Since the North Atlantic Treaty has created a community of values for 75 years,[4] ethical and legal criteria must be implemented as technical design principles that shape AI-based technologies from the outset. Foremost, strict adherence to International Humanitarian Law (IHL) and Rules of Engagement (RoE) must be supported also technically. Secondly, the precise operational success made possible by AI-based automation and intended by human decision-makers is a positive moral good in the consequentialist sense, where moral quality of an action depends only on how desirable its consequences are (Schroth, 2023). Finally, soldierly virtues realize the concept of the "citizen in uniform,"[5] as formulated by the German Armed Forces, to provide an example. Only humans can act responsibly, machines merely achieve effects.

Aspects of AI-based technologies that must be discussed on the tactical level are (Koch & Keisinger, 2024):

- Context: Controlling the space, duration, time, and conditions
- System: Functioning, capabilities, and limitations in given operational circumstances
- Environment: Situational awareness and understanding of the environment, proper training
- Reliability: Consequences of use and reliability as the likelihood of failure in realistic operational environments against adaptive adversaries

[4] "The Parties to this Treaty […] are determined to safeguard the freedom, common heritage and civilization of their peoples, founded on the principles of democracy, individual liberty and the rule of law" (North Atlantic Treaty Organization, 2023, Preamble).

[5] "As a citizen in uniform, the soldier sees himself both as a member of the armed forces and as part of society" (BMVg FüSK III 3, 2017, p. 3).

- Supervision: Human ability to intervene
- Accountability: Standards of authority and accountability of human operators, team-mates, and commanders
- Ethics: Preserve human dignity and agency, uphold moral responsibility in decisions to use force.

3 Artificially Intelligent Data Fusion in Defense

Comprehensive data and information fusion from all available sensors and non-sensor sources, both model-based and data-driven, in short: AI, as this notion is understood here, already plays a key role for allied defense. Without this powerful technology, there are no effective armed forces, which depend on information superiority and decision dominance on land, at sea, in the air and space, or in the cyber space, i.e., in all military domains.

AI-driven multiple source information fusion transforms massive data streams from a vast variety of sources into actionable information for optimized management of multifunctional sensor systems and sensor networks, communication links, and other resources as well as for Command and Control (C2) of weapon systems, including Electronic Warfare (EW), and stationary or moving platforms. Moreover, the resulting situational awareness and decision-making and automated executing capabilities are enablers of improved interoperable effectiveness of allies cooperating with each other in combined MDOs.

In view of these considerations, artificially intelligent information fusion for defense poses a timeless question:

> How to decide well in terms of military action according to what is recognized as true in terms of reliable situation pictures and insight into their deficiencies in the "fog of war," i.e., their "known unknowns" or even "unknown unknowns"? (Krogerus & Tschäppeler, 2012, p. 86)

Turned into systems engineering, this leads to three fundamental tasks:

1. Design information fusion and decision support in a way that human beings are not only mentally but also psychologically able to master each situation.
2. Identify technical design principles that facilitate the responsible use of artificially intelligent C2 and Manned-unManned Teaming (MuM-T).
3. Guarantee that human decision-makers in such support systems still have full superiority of information, decision-making, and execution of action.

"Of course all thought is art" observed the Prussian general and military philosopher Carl von Clausewitz (1770–1831). "The point where the logician draws the line, where the premises […] end […], is the point where art begins" (Clausewitz, 1976, p. 148). For this reason, digital ethics and a corresponding ethos and morality are essential soft skills for commanders and staff, but also for information fusion

engineers, to be built up systematically in parallel to operational and technical excellence.

Engineers do not need to execute military operations, as soldiers will not program systems for situational awareness and command & control. Both, engineers and soldiers, however, should be able to assess the strengths and weaknesses, risks and opportunities of AI-enabled operations and technologies. The associated operational and technical competence, as well as the digital morality required is teachable. It addresses key questions of the soldierly dignity and responsible systems design, which are aggravated using AI for defense, but not fundamentally new.

In the age of digitalized military operations, loops to OBSERVE, ORIENT, DECIDE, and ACT, according to the military strategist John Boyd (1927–1997) (Boyd, 1976), and then to ASSESS the effect of the actions taken (OODAA loops), are dramatically accelerating and thus to be executed "at machine speed" in a network-centric and collaborative way. Moreover, the pragmatic US-definition of AI as "the ability of machines to perform tasks" that "normally require human intelligence," also includes physical assistance systems such as AI-controlled exoskeletons or robots. For this reason, the immediate physical presence of human beings in a potentially lethal environment is becoming increasingly dispensable. "Even decades-old old technology can still be AI," such as aircraft autopilot, missile guidance, and signal processing systems (Allen, 2020).

Quite in line with the US use of it, the term AI does not only comprise machine or deep learning, e.g., but a whole world of data-driven and model-based algorithms, including approaches to Bayesian and Machine Learning, game theory, and adaptive resources management.[6] This "World of Algorithms," realized by the "art of computer programming" (Knuth, 1962–2022) and enabled by qualitatively and quantitatively appropriate testing and training data, running on distributed computing devices drives a data processing cycle that starts from elementary signals, measurements, and observer reports collected from multiple and heterogeneous sources.

4 New Engines for Accelerating OODAA Loops

Interoperability in all military domains does not mean imply direct access any means of certain military domain from any other domain. On the contrary, each domain must rather maintain its own competences and specific capabilities by developing them further in the sense of a common understanding of strategic, operational, and tactical planning. The German Army's concept of AI-enabled MDO is an example of a domain-specific sub-operation under the leadership of a domain leader. Sensors, effectors, and support services of different domains are to achieve spatial and temporal superiority under unified command, focused on operational objectives. The essential prerequisite of MDO is the end-to-end digitalization of all levels

[6] Just as an example among many others: Koch (2014).

and forces, which creates the preconditions for effect-oriented information superiority and decision dominance.

In future defense scenarios, crewed and uncrewed systems (UxS) form a comprehensively networked system of systems. Cooperating multiple-sensor, multiple-effector UxS protect the soldiers or assets and execute reconnaissance or combat missions with electronic or kinetic impact, whereas satellites, early warning, refueling, or transporting will be integrated. The core infrastructure needed are the so-called combat clouds, which fuses all required data and makes mission-relevant information available in real time and provides means for adaptive resources management.

The US definition of AI explicitly includes 'even decades-old AI,' such as aircraft autopilots, missile guidance, and signal processing systems. Though many AI technologies are in a sense 'old,' there have been technological breakthroughs that have greatly increased the diversity of applications in defense where AI is practical, powerful, and useful. Many recent achievements have been focused on Machine Learning (ML) and data-driven algorithms more generally. Such algorithms are closely related to mathematical statistics and encode knowledge that is automatically 'learned' from data in AI Models. Due to the extremely large number of numerical values that characterizes them, AI Models are no longer directly accessible to direct human understanding, i.e., are in a sense black boxes that may sometimes be turned into grey boxes by using methods from Explainable AI (XAI), perhaps exaggeratedly called so.

Algorithms for harvesting information from data fusion and collecting data via adaptive resources management belong to the methodological core of cognitive and volitive engines for Intelligence, Surveillance, and Reconnaissance (ISR), for C2, and MuM-T that assist the intelligent minds and autonomous wills of commanders and staffs. The concepts of mind and will to be assisted, and, therefore, of consciousness and autonomy, bring human beings as persons into view that are 'somebody' and not 'something.' Within this framework, new types of engines enhance the perceptive mind and the deliberate will of persons, who alone are capable to perceive intelligently and to act autonomously (Koch et al., 2024):

- *Cognitive engines* fuse massive streams of sensor, observer, context, and mission data for producing comprehensive situation pictures, the basis for conscious human cognition to plan, perceive, act, and assess effects appropriately.
- *Volitive engines* transform overall decisions of deliberate and responsible human volition into chains of automatically executed commands for data acquisition, sub-system control, and achieving effects on objects of interest.

Cognitive and volitive assistance provided by such machines will enable decision-makers to remain capable of acting in complex situations with spatially distributed, moving assets and on short time scales. In a sense, certain processes that underlie conscious perception and causal action and that were previously reserved for humans are transferred to machines on which they may be executed at enormously reduced processing time and scaled to enable massive processing at highly increased data rates. By this, they enable Human Performance Enhancement (HPE)

way beyond the natural human levels. Exceeding natural levels in this way is nothing unusual. Even the steam engine surpassed all natural physical capabilities.

Nevertheless, processes triggered by such cognitive and volitive engines are to be distinguished from natural intelligence and autonomy in the sense that they enhance the perceptive mind and the active will of persons, who alone perceive intelligently and act autonomously, understood as a moral right and the capability of persons to think for oneself and decide in a way that achieves a freely set effect. For this reason and in accordance with NATO's strategy on the use of military AI, to name an example, and a large community of defense scientists, the responsibility of human decision-makers is pivotal.

5 Human-Centric Design of Information Fusion

The importance of automation, to take an example, was recognized as early as in 1957, one year after the term 'AI' was coined at the Dartmouth Conference, when the conceptual architect of the German Armed Forces Wolf von Baudissin (1907–1993) wrote that, thanks to automation, "human intelligence and manpower will once again be able to be deployed in the area that is appropriate human beings: in monitoring and controlling the machines and in eliminating unforeseen malfunctions" (Baudissin, 1969, p. 174).

According to high-rank documents of the German Federal Ministry of Defense, to refer to the example that was previously mentioned, the importance of AI does not lie "in the choice between human or artificial intelligence, but in an effective and scalable combination of human and artificial intelligence to ensure the best possible performance" (Bundesministerium der Verteidigung, 2019, p. 27). Comprising the ergonomic as well as the ethical and legal dimensions of AI-based systems, this statement calls for ethically aligned AI-based systems engineering as a fundamental military requirement.

Ethical criteria can only become practicable if it is possible to translate them into technical design principles to be considered in the technology development from the outset. Particular care must be taken to what needs to be adhered to at any rate in a Kantian sense, i.e., international law or the rules of engagement, what is to be achieved, as mission success is also a moral good in a consequentialist sense, and the soldierly virtues in an Aristotelian sense constitute the concept of the 'citizen in uniform' and may comprehensively guarantee soldierly, and therefore human, dignity.

Consideration should be given to how the classic doctrine of virtue could be transformed into system-ergonomic design principles. It is evident that there are alternative conceptualizations for the design of AI-based systems. Nonetheless, the authors maintain that virtue ethics, a framework that has been extensively debated from Aristotle through Thomas Aquinas to contemporary scholars, continues to hold significant relevance in the current context. This ethical approach provides a foundation for evaluating the moral dimensions of AI system design, emphasizing

the interplay between virtuous military decision-makers with machine assistance for responsible actions.

As guiding principles, virtues can shape artificially intelligent machines in the sense of responsible judgment. The four Cardinal Virtues of European ethics are mentioned only as an indication:

1. *Prudence* entails conformity to reality (Pieper, 1996a). Accordingly, *Innere Führung* could, within the context of the German Armed Forces, be interpreted as education and self-cultivation toward the virtue of prudence, meaning the ability to objectively perceive the realities surrounding our actions and, depending on their nature and significance, let them decisively influence the appropriate action. It is essential to consider that due to various uncertainties, a certain degree of ambiguity always persists. This is especially true in the 'Fog of War.' Cognitive machines belong to this area.
2. *Justice* is facilitated by prudence in the sense that it considers the situation in which people act (Pieper, 1996b). Prudence substantiates the tangible potentiality to act justly. The essence of justice, however, resides in being the supreme and intrinsic manifestation of virtue itself: The virtuous individual is fundamentally just. Volitive machines, which also take into account the rule to be obeyed to in the sense of normative assistance, would be included here.
3. *Fortitude* is characterized by the readiness to accept wounds in the struggle for the realization of the good. This virtue is, thus, is directed toward the obstacles that arise in the realization of the good (Pieper, 1996c). Therefore, it is dependent on prudence. Fortitude without prudence is not truly fortitude. In this context, reflective assistance based on cognitive machines can provide support.
4. As its Latin origin implies, *temperance* aims at assembling a unified whole from diverse parts. Against this background, "to temper" and "temperance" also means "to realize order in oneself," especially in combat (Pieper, 1996d). Even in this environment, technical system assistance is conceivable.

Without a clearly defined conception of humanity that enables responsible use of technology, grounded in the concept of virtues, and without the embodiment of such an understanding through the education of military decision-makers, it is argued that the design of machine assistance aligned with virtues and aimed at supporting morally and ontologically sound decisions becomes unattainable.[7]

As the war in Eastern Europe or the attacks in the Gulf of Aden with severe impact on the global economy, show, artificially intelligent drone technology may serve as an example. Within this context, it must firstly be clarified whether the technical prerequisites for the responsible use of partially or fully automated reconnaissance and combat drones is feasible, i.e., its compatibility with soldierly dignity. The spectrum ranges from Remotely Piloted Air Systems (RPAS), in which the entire targeting cycle is completely under human control, via partially automated

[7] See Koch and Matter (2024) for literature dealing with the difficulties pertaining to the concrete translation of such virtues and a more sophisticated discussion of virtue ethics within the sphere of technology.

individual drones and fully automated swarms of drones to Loitering Munition (LM), which can wait for hours for a target to be detected and then engaged.

According to the Federal Ministry of Defense (Drucksache 20/10266, 2024), LM is defined as "a guided missile that searches for targets on the ground in an operational area the remains at different altitudes depending on the mission profile and engages a (moving) target with high precision after being released by an operator." As LM is intended for single use against enemy targets, it is not considered an unmanned aerial system (UAS), which is designed for repeated use, though mixed forms are possible.[8] LM increases the protection of own forces and means by separating the operating personnel from the system. In this way, a precise and scalable effect can be achieved in depth against key enemy capabilities. The longer duration of operation compared to other ammunition also enables more accurate target reconnaissance and more precise engagement after release by the operator. This shows in particular that LM is not currently a "fully autonomous lethal weapon system" in the use case presented by the German government. The continuous control of the loitering munition by an operator makes its use comparable to other effectors already in use.

However, technical developments and operational necessities must be anticipated. As soon as deployed LM exceeds a certain and foreseeable limit in number, continuous control of the individual LM can no longer be carried out by the operating personnel. In addition, electronic defense by the opposing side can disrupt the connection to the drone, so that more autonomy is necessary in order to be able to use the system type sensibly. This may result in a situation comparable to the deployment of a minefield. Since sea mines have long been equipped with detonation mechanisms that can be attributed to a certain kind of artificially intelligent automation, it would be obvious to transfer the corresponding provisions of international law to loitering munitions in a technology-agnostic sense (BMVg R I 3, 2018). The so-called fire-and-forget weapons with sensory seeker heads have been around for a long time and are in use. It would therefore be a perfectly legitimate question to ask whether these weapons should not be replaced by artificially intelligent and ethically aligned weapon systems that can be used responsibly until the final weapon effect released.

6 On the FCAS Ethical AI Demonstrator Project

In the spirit of these considerations and for the first time in Germany, an intellectual struggle over the technical implementation of ethical and legal principles accompanies a major air defense project from the outset. In the European Future Combat Air System (FCAS), manned jets of the latest generation are elements of a complex and comprehensively networked system of systems. Unmanned remote carriers protect

[8] Such as the German MAUS system (Frank, 2024).

the pilots as loyal wingmen and accompany them on reconnaissance and combat missions.

The FCAS Ethical AI Demonstrator (E-AID), based on exemplary scenarios discussed by the Luftwaffe, identifies ethically relevant requirements for FCAS systems engineering, focusing on individual functions to be executed within the OODA Loop. So far, OBSERVE and ORIENT have been examined with regard to critically reflected situational awareness. DECIDE and ACT relate more directly to military action. The scenarios are intended to provoke ethical dilemma situations that are to be examined from a consequentialist and virtue ethics perspective. The IHL, which can be rather 'cruel,' is to be kept 'deontologically' at any rate. The central question is how ethically acceptable action under extreme time pressure and masses of data can be technically supported. More concretely:

1. Ethically aligned system design must determine the situation picture with its limitations as reliably as possible. The use of artificially intelligent information fusion, which may be turned into a 'grey' box, is indispensable. Full 'explainability' seems to be an unfulfillable promise.
2. It must be checked automatically, which of the conceivable options for action are legally compliant, i.e., instantaneously. If, for whatever reason, military personnel DECIDE in a way, that does not comply with the law or the RoE, they must be informed of this in an appropriate manner.
3. Automated functions are to be provided that quickly calculate the probable consequences of the respective decision alternatives in the sense of a consequentialist evaluation of the ACT step and present them in an ergonomically comprehensible manner. This aspect is related to ASSESS.
4. Soldierly virtues are acquired, for example, in dealing with various forms of bias or grey boxes, by confronting military personnel with ethical dilemma situations in a digital twin in the run-up to a mission.
5. The interplay between consequentialist ASSESS and the exercise of soldierly virtue influences mission planning and personnel selection. The problem of self-protection would at least partly be eliminated by unmanned platforms. Dilemma situations between mission fulfillment and the protection of non-combatants remain.
6. Under certain circumstances, combat decisions must be made automatically. DECIDE on the use of such a system in operations and on its technical design in advance must be consciously made by humans—beyond the operator in the cockpit—and they must take responsibility for them. The operator then represents the 'human in the loop' by making a situation-dependent 'nevertheless not' decision.
7. Dilemmata remain even then. Consequentialist and virtue ethical considerations are not made during the operation, but by parameterizing the system in preparation for the operation. A situation-dependent 'nevertheless' of an operator must remain possible.

The considerations lead to the thesis that the technical prerequisites for the responsible use of partially and fully automated systems within the framework on

FCAS can be created. Moreover, this can be done in such a way that the risk to non-combatants and to soldiers deployed is minimized in accordance with the rules of engagement, or at least is considerably lower than when using alternative weapon systems.

However, this does not mean that technological development will 'naturally' lead to responsibly usable, artificially intelligent standoff weapon systems or that the quality of the decision-making basis for their use cannot be further improved. Even the development of ethically irresponsible AI-based technology is entirely possible and may even be pursued by opponents.

These considerations include the conception of well-thought-out rules of engagement that address the risks of these AI-based technologies, which permeate all technical system components from their very design principles and comply with international humanitarian law, ethical values, and the soldierly dignity. In accordance with the inherent nature of defense technology developments, the potential threat to own forces from hostile use of AI-based technologies needs to be countered. It is, thus, one of the tasks of defense research to counter this threat.

7 Responsibly Using Military AI?—A Necessity

Only natural intelligence can assess plausibility, develop understanding, and agency. "The uncontrolled pleasure in functioning, which today is almost synonymous with resignation to technical automatism, is no less alarming than the dashing, pre-technical feudal traditions because it suggests the unscrupulous, maximum use of power and force," Baudissin observed already in the 1950s. "This not only contradicts the maxims of a liberal order of a state, but also the realities of modern war, in which ethics and reason agree on minimizing the possible use of force" (Baudissin, 1969, p. 180).

These words ring true not only for shaping the soldierly ethos in the digital age. There is a more general need for a new 'enlightenment' in dealing with AI maturely, ethically, and intelligently, i.e., man's release from his self-imposed immaturity. "*Sapere aude*! Have courage to make use of your *own* understanding!" according to Immanuel Kant (1996, p. 17). Anthropocentrism in this sense underlines the ethical and legal dimensions of artificially intelligent automation, which characterize the use of AI in defense systems.

Military AI comprises, thus, more than just the aspect of a technical innovation. It influences the entire way armed forces think and act. Since we feel encouraged to assume that a broader consent within the international defense science community might be achieved, we are closing with three recommendations:

- Digital ethics and corresponding morals are part of the human competencies that need to be developed and expanded to develop and deploy AI-based defense technologies responsibly. Consideration should be given to leadership philosophies and personality development instruments, such as *Innere Führung*, proven

in the German Armed Forces,[9] as a guiding principle for the development of ethical competence and to encourage its systematic development with regard to AI in defense.

- In addition to the operational added value of military AI, ethical skills in dealing with such technologies and ethical acceptance in the eyes of the conscience of individual soldiers, but also in the eyes of the research, development, and defense planning communities, are essential characteristics of successful innovation. Only then, AI in military domain will become acceptable before the conscience of the individual soldiers, but also in the broader view of the Common Good of the society as such.
- In analogy to the oath of Hippocrates for physicians, who are just as obliged to take responsibility as soldiers and defense scientists, the oath, which was considered indispensable when the German Bundeswehr was founded, should be viewed with a fresh eye in the context of digitalization in defense.

References

Allen, G. (2020). *Understanding AI Technology*. US Department of Defense, Joint Artificial Intelligence Center. https://apps.dtic.mil/sti/pdfs/AD1099286.pdf

Baudissin, W. G. v. (1969). *Soldat für den Frieden. Entwürfe für eine zeitgemäße Bundeswehr*. Pieper.

Bernzen, E., Peddinghaus, D., & Sieger, R. (2016). Innere Führung—Leadership Culture in Camouflage. *Ethics and Armed Forces, 2016*(1), 46–50. https://d-nb.info/1104523671/34

BMVg FüSK III 3. (2017, November). *Innere Führung—Selbstverständnis und Führungskultur (Zentrale Dienstvorschrift, A-2600/1)*. Bundesministerium der Verteidigung https://www.bmvg.de/de/themen/verteidigung/innere-fuehrung/staatsbuerger-in-uniform

BMVg R I 3. (2018, March 22). *Humanitäres Völkerrecht in bewaffneten Konflikten (Zentrale Dienstvorschrift A-2141/1, Version 2)*. Bundesministerium der Verteidigung. https://www.bmvg.de/resource/blob/93612/7d6909421eacad4ddc7dcdfdf58d42ca/b-02-02-10-download-handbuch-humanitaeres-voelkerrecht-in-bewaffneten-konflikten-data.pdf

Bossong, B., Rieks, A., & Koch, W. (2022, February 02). *Künstliche Intelligenz für die Landesverteidigung*. Frankfurter Allgemeine Zeitung. https://www.faz.net/aktuell/wirtschaft/in-welchem-rahmen-ist-ki-sinnvoll-fuer-die-verteidigung-17765528.html

Boyd, J. R. (1976, September 03). *Destruction and creation*. U.S. Army Command and General Staff College. https://www.coljohnboyd.com/static/documents/1976-09-03__Boyd_John_R__Destruction_and_Creation.pdf

Bundesministerium der Verteidigung. (2019, October). *Erster Bericht zur Digitalen Transformation des Geschäftsbereichs des Bundesministeriums der Verteidigung*. https://www.bmvg.de/resource/blob/143248/7add8013a0617d0c6a8f4ff969dc0184/20191029-download-erster-digitalbericht-data.pdf

Bundesministerium der Verteidigung. (2023, November). *Verteidigungspolitische Richtlinien 2023*. https://www.bmvg.de/resource/blob/5701724/5ba8d8c460d931164c7b00f49994d41d/verteidigungspolitische-richtlinien-2023-data.pdf

[9] According to von Baudissin, the overall goal of *Innere Führung* is to reconcile the functional conditions of operational armed forces with the principles of a democratic constitutional state (Bernzen et al., 2016). See also the article by Peter Andreas Popp in this book.

Bundesministerium der Verteidigung. (2024, July 12). *NATO-Gipfel 2024 in Washington—Auf Deutschland ist Verlass: Pistorius in Washington.* https://www.bmvg.de/de/themen/dossiers/die-nato-staerke-und-dialog/nato-gipfel-2024-washington

Clausewitz, C. v. (1976). *On War* (M. Howard, & P. Paret, Ed. and Trans.). Princeton University Press (Original work published 1832).

Drucksache 20/10266. (2024, February 22). https://dserver.bundestag.de/btd/20/104/2010456.pdf

Ecclesia Romana. (1975). Pastoral Constitution Gaudium et Spes. In A. Flannery (Ed.), *Vatican Council II: The Conciliar and Post Conciliar Documents* (pp. 903–1014). Liturgical Press.

Frank, D. (2024, Oktober 04). *Deutsche Kampfdrohnen Maus für ukrainische Spezialeinheit KRAKEN*. cpmDefenceNetwork. https://defence-network.com/kampfdrohne-maus-ukraine-spezialeinheit-kraken

Fukuyama, F. (1992). *The End of History and the Last Man*. Free Press https://pages.ucsd.edu/~bslantchev/courses/pdf/Fukuyama%20-%20End%20of%20History.pdf

Gutschker, T. (2021, November 10). *Europa ist in Gefahr*. Frankfurter Allgemeine Zeitung. https://www.faz.net/aktuell/politik/ausland/josep-borrells-neues-konzept-fuer-die-eu-verteidigungspolitik-17627660.html

High-Level Expert Group on AI – AI HLEG. (2019, April 08). *Ethics Guidelines for Trustworthy AI.*. European Commission. https://doi.org/10.2759/346720

Johnson, J. (2024). *The AI Commander: Centaur Teaming, Command, and Ethical Dilemmas*. Oxford University Press.

Kant, I. (1996). An Answer to the Question: What is Enlightenment? In M. J. Gregor (Ed.), *Practical Philosophy. The Cambridge Edition of the Works of Immanuel Kant* (pp. 11–22). Cambridge University Press. (Original Work published 1784).

Knuth, D. E. (1962–2022). *The Art of Computer Programming* (Vol. 1–4). Addison Wesley Longman.

Koch, W. (2014). *Tracking and Sensor Data Fusion—Methodological Framework and Selected Applications*. Springer. https://doi.org/10.1007/978-3-642-39271-9

Koch, W., & Keisinger, F. (2024). How Can Responsible AI be Implemented? In J. M. Schraagen (Ed.), *Responsible Use of AI in Military Systems* (pp. 37–58). CRC Press. https://doi.org/10.1201/9781003410379-4

Koch, W., & Matter, N. (2024). Who Decides What to Do? On Ethics, Ethos, and Morals of AI in the Military Defence Domain. *Philosophy, Theology and the Sciences, 11*(2), 252–268. https://doi.org/10.1628/ptsc-2024-0018

Koch, W., Spreen, D., Talves, K., Wagner, W., Lillemäe, E., Klaus, M., Viidalepp, A., Cooper, C. G., & Pekarev, J. (2024). On the Ethics of Employing Artificial Intelligent Automation in Military Operational Contexts. *IEEE Transactions on Technology and Society, 5*(2), 231–241. https://doi.org/10.1109/TTS.2024.3405309

Krogerus, M., & Tschäppeler, R. (2012). *The decision book: Fifty models for strategic thinking* (J. Piening, Trans.). W. W. Norton.

Luhmann, N. (1993). *Risk. A Sociological Theory* (R. Barrett, Trans.). de Gruyter (Original work published 1991).

North Atlantic Treaty Organization. (2022, March 31). *NATO Defence Planning Process*. https://www.nato.int/cps/en/natohq/topics_49202.htm

North Atlantic Treaty Organization. (2023, October 19). *The North Atlantic Treaty. Washington D.C.–4* April 1949. https://www.nato.int/cps/en/natohq/official_texts_17120.htm?selectedLocale=en

Pieper, J. (1996a). Traktat über die Klugheit. In B. Wald (Ed.), *Schriften zur Philosophischen Anthropologie und Ethik: Das Menschenbild der Tugendlehre* (pp. 1–42). Meiner (Original work published 1937).

Pieper, J. (1996b). Über die Gerechtigkeit. In B. Wald (Ed.), *Schriften zur Philosophischen Anthropologie und Ethik: Das Menschenbild der Tugendlehre* (pp. 43–112). Meiner (Original work published 1953).

Pieper, J. (1996c). Vom Sinn der Tapferkeit. In B. Wald (Ed.), *Schriften zur Philosophischen Anthropologie und Ethik: Das Menschenbild der Tugendlehre* (pp. 113–136). Meiner (Original work published 1934).

Pieper, J. (1996d). Zucht und Maß. Über die vierte Kardinaltugend. In B. Wald (Ed.), *Schriften zur Philosophischen Anthropologie und Ethik: Das Menschenbild der Tugendlehre* (pp. 137–197). Meiner (Original work published 1939).

Rieks, A. (2022, February 21). *Digitalisierung ist die DNA der Luftwaffe* (Interview). Hardthöhen-Kurier. https://hardthoehenkurier.de/digitalisierung-ist-die-dna-der-luftwaffe

Schroth, J. (2023). Konsequentialistische Ethik. In C. Neuhäuser, M.-L. Raters, & R. Stoecker (Eds.), *Handbuch Angewandte Ethik* (pp. 37–43). J.B. Metzler.

SIPRI Yearbook. (2024). *Armaments, Disarmament and International Security—Summary*. Stockholm International Peace Research Institute. https://www.sipri.org/sites/default/files/2024-06/yb24_summary_en_2_1.pdf

Soare, S. R., Singh, P., & Nouwens, M. (2023, February 17). *Software-defined Defence: Algorithms at War (Research Paper)*. The International Institute for Strategic Studies (IISS). https://www.iiss.org/research-paper/2023/02/software-defined-defence/

Taliaferro, A. C., Gonzalez, L. M., Tillman, M., Ghosh, P., Clarke, P., & Hinkle, W. (2019, February 01). Introduction to Capability-based Planning (CBP) and its Comparison to Threatbased Planning. In Institute for Defense Analyses (Ed.), *Defense Governance and Management: Improving the Defense Management Capabilities of Foreign Defense Institutions a Guide to Capability-Based Planning (CBP)* (Research Report, IDA Document NS D-10369, pp. 1–4). http://www.jstor.org/stable/resrep22853.4

Artificial Intelligence (AI) and the Bundeswehr from the Perspective of Innere Fuehrung

Peter Andreas Popp

Abstract Innere Fuehrung (InFü) is based on the model of the responsible 'citizen in uniform.' It has been the intellectual foundation and leadership concept of the Bundeswehr since its founding. Sooner or later, however, the further development of artificial intelligence (AI) will pose challenges because technological progress threatens the final responsibility of humans. This article does not argue against the use of AI as such. Its purpose is to sensitize military and political decision-makers so that InFü does not become a naively derived and thus determined parameter of technological development in the future. The security policy environment is deliberately included. In this way, AI can become a great opportunity to ensure the operational readiness of the armed forces and, at the same time, to promote the maturity of citizens in uniform with their duty to always face up to their own ethical responsibility. The Bundeswehr, and especially its educational and training institutions, are facing their greatest test—not only with regard to the professional level of information and knowledge in scientific-technical and socio-political-cultural fields! Education and well-considered, value-oriented training are the order of the day across all ranks.

In 1989, that strangely distant epochal year from today's perspective, a book dealing with the fate of German cities right after 1945 appeared on the market bearing the evocative title *So much beginning never was* (Glaser, 1989). But what does architecture have to do with artificial intelligence (AI), the skeptical reader might ask, especially as we are not talking here about the application of computer-based technology, i.e., AI, in the contradictory fields of architecture aesthetics, commerce, functionality, and the art of engineering.[1]

[1] AI is not hardware, but extremely fast and self-perfecting software, whereby the boundary between hardware and software is blurred. So far relevant definitions of the European Parliament and the Fraunhofer Institute express this well. The EU Parliament's definition of AI is: "AI is the

P. A. Popp (✉)
Fürstenfeldbruck, Germany

K. Talves, D. Spreen (eds.), *Artificial Intelligence in Military Technology*,
Artificial Intelligence, Simulation and Society 192,
https://doi.org/10.1007/978-3-031-95578-5_9

Quite a lot due to the title of the book. The statement 'So much beginning never was' points to the elementary problem of technical innovation related to the associated development perspectives. Does AI signify a technical evolution, i.e., a sustainable change adapted to the requirements of the times with the prospect of concrete idealistic and material added value, or does it embody a revolution—in turn understood as an expression of fundamental change with the result that the world as we know it in its inadequacy and beauty is now completely turned upside down and therefore no longer recognizable?

As far as the military is concerned, the challenge is whether the selective use or even the general use of AI will not completely overturn all the military truths we have come to love and believe to be safe. For all the esteem in which Innere Fuehrung is held, there are elementary things at stake for it too; after all, it is based on the Enlightenment view of humanity. Heretically asked: Isn't Innere Fuehrung itself at risk sooner or later?

In any case, our descendants will know whether AI does not support Adorno and Horkheimer's (capitalism-critical) thesis of the *Dialectic of Enlightenment* in its very specific way (Horkheimer & Adorno, 2002).[2] But let us stay in the here and now: This challenge is increasingly arising as a result of the stages toward the development of actual AI, i.e., the ability of the machine to develop its own consciousness and creativity beyond programmed algorithms and then to act as an adaptive unit much faster and at the same time 'more intelligent' than humans 'accurately and without error,' because it is 'more perfect' and ad hoc self-optimizing in comparison to humans. Consciousness is understood here as critical self-reflection in combination with the capacity for suffering and empathy as well as the endeavor to answer the question of meaning (*Sinnfrage*).[3]

---

ability of a machine to display human-like capabilities such as reasoning, learning, planning and creativity. AI enables technical systems to perceive their environment, deal with what they perceive, solve problems and act to achieve a specific goal. The computer receives data—already prepared or gathered through its own sensors such as a camera—processes it and responds. AI systems are capable of adapting their behavior to a certain degree by analyzing the effects of previous actions and working autonomously." (European Parliament, 2023). The Fraunhofer Institute (Fraunhofer IPA, 2024) defines AI as follows: "Artificial Intelligence (AI) provides machines with capabilities comparable to intelligent human behavior. The generic term AI encompasses problem-solving methods, including logic and planning procedures, which would require human intelligence. Probably the best known and most researched and applied method is machine learning." En detail s. https://epthinktank.eu/tag/artificial-intelligence/. Cf. in this context recently Steiger (2023) and Vohl (2023).

[2] The culturally pessimistic work was written in American exile between 1942 and 1944 and mutated into a cult book of the student movement in 1968. The decisive point for our topic is that the thesis of the 'dialectic of enlightenment' sensitizes us to the ambivalences of enlightenment. In any case, maturity is and remains an elusive good!

[3] AI does not yet have 'consciousness.' The question is when and how quickly this could be the case. Here, it also depends on the definition of what is meant by 'consciousness.' After all, AI operates algorithmically, i.e., in the medium of causality, and consciousness operates in the medium of meaning (and have feelings!). The following authors are just a few examples: Jost Halfmann, Elena Esposito, and Niklas Luhmann (cf. Spreen, 2023). Is 'consciousness' a philosophical category or

Humans would then no longer be partners, but dependent objects of the machine. The machine would not become a human, but a quasi-god in all its omnipotence; at least a being/entity with god-like omnipotence that absolutely and inescapably determines human behavior. The energy dimension of AI (i.e., the considerable amount of energy required to operate AI and which may relativize tactical-operational options!) cannot be deepen as a topic here. Contrary to the intended extreme dynamization, it will very probably turn future warfare and defense into a static matter—comparable to the military logistics of the eighteenth century or the French fortification system with the culmination of the Maginot Line. Likewise, the path toward the military and civilian use of AI affects humans both as a species and as individuals. There is a danger of negative and positive idolatry dealing with AI, neither a cult nor a religion or a total remedy for solving all problems on earth. That is why AI deserves to be viewed soberly; to paraphrase Max Weber (1946, p. 128): With "passion and perspective."[4]

That would be extremely tragic in a military context. Soldiers, high-ranking military decision-makers and—in a democracy, the primacy of politics applies!—their political superiors must be aware of the fact that AI–at whatever stage of development!—touches on the question of human existence. So how autonomous is the human being? What options for action do they have? Will they degenerate into willless objects that will inevitably have to submit to their fate? And then with the following 'side effect,' summarized in the question: 'Can, no, must the military take decisions without ethical reflection in the future, free of dilemmas?'

Addressing this directly does not mean dealing with incidental bells and whistles. On the contrary: war and the military are human endeavors. And since AI has become of utmost significance in the world of the military, despite German foreign policy being geared toward dialogue and based on the ideal of 'foreign policy as global domestic policy,' this nation's security policy cannot close itself off to the topic of 'AI.' If political and military decision-makers, as well as soldiers in general, are interested in their high task, then it would be grossly negligent not to do so. It would be pure luxury to capitulate to the challenges of AI, i.e., to bury our heads in the sand.

However, there should be no illusions: In terms of security policy, we are still a long way from following a set of economic rules modeled on those of the European Union when it comes to the use of AI. In this context, it should be noted as a 'thesis in the political arena' that dealing with AI could prove whether the European Union is actually prepared to pursue a serious security policy. In view of the current global situation, it is to be feared that the containment of this anarchic uncontrolled growth will continue for some time to come, particularly at the military level, and hopefully not at the cost of bloody experience. A global rule-based handling of AI would be absolutely desirable. Unfortunately, this is currently in the realm of utopia. But

something that can be mathematized? Or is it even both thanks to AI? Appealing Miller (2019) as well as recently Patrick Krauss (2024).

[4] AI also embodies a factor in terms of the 'political culture' of a community. Cf. Maurer (2021).

utopias have the quality that they can—for better or for worse—represent the reality of tomorrow. In short, we currently do not know whether AI has the political ability to take humanity a step further toward 'sovereignty of the United Nations as a subject of international law' and 'foreign policy as world domestic policy.'[5]

It must therefore be clearly addressed: Innere Fuehrung is not only the value compass of the armed forces. It is an integral part of security policy, too. Since the coherence of interests and value orientation is very important to German security policy and the application of AI is ambivalently revealing both a deficit and a plus in terms of credible security policy, Innere Fuehrung cannot and must not close itself off to the topic of 'AI.' Initial signs are hopeful that this will not be the case. The currently challenging task and question for Innere Fuehrung is: 'Not whether it gets involved, but how and where it gets involved in the implementation and provision of AI applications?' In short, technicians should be sensitized to the fact that only a holistic approach will lead to success, whereby one is once again confronted with implementing Innere Fuehrung in an explanatory sense. AI and Innere Fuehrung, both, form a synthesis.

The task set is quite comparable to the initial constellation of the Bundeswehr. How could the future German soldier (from the perspective of the time) pursue a morally responsible and ethically orientated profession in view of the abyss of inhumanity, particularly in the Second World War, and the destructive quality of nuclear weapons? Incidentally, this is not just a German problem; the appreciation of international (war) law after 1945 is a legal proof of this. The German answer to this was Innere Fuehrung as a warrant for military service by conscience-led obedience. The solution, too, was to ensure that deterrence remained credible and that the soldiers of the Bundeswehr did not form a foreign body in a society that was on its way to democracy.

Innere Fuehrung was, is, and always will be controversial. The recently published *Innere Fuehrung Handbook* (Zentrum Innere Führung, 2023) should not provide any false security in this regard (especially as the topic of 'AI' has not yet been or could not yet be addressed here …).[6] The Bundeswehr should not immediately sound all the alarm bells with regard to 'questions to Innere Fuehrung.' Because this shows neither self-confidence nor sovereignty. How could this finding in the sense of a plea for the questionability of Innere Fuehrung be any different? Benjamin Franklin's wise advice should also apply here: 'Our critics are our friends, they help us.' Innere Fuehrung has been and will continue to be an explanatory piece in the field full of tension defined by value-orientated action, obedience guided by conscience, military efficiency and—not to forget!—the ideal of civility, for which the term 'citizen in uniform' (*Staatsbürger in Uniform*) stands. Innere Fuehrung would be devoid of controversy if the military, too, only had 'common sense,' based on

[5] Cf. the argumentation of peace and conflict researcher Carl Friedrich von Weizsäcker (1977) on the prevention of the nuclear arms race.

[6] This makes the handbook itself a contemporary document. In connection with this statement, I would like to thank First Lieutenant Thomas Tüchsen (Zentrum Innere Fuehrung) for the stimulating discussions on the technical possibilities of AI.

objective rational decisions, and ignoring the fact that human nature is based on emotion, weakness and vulnerability, inadequacy, and the pursuit of power.

German democracy is currently required to internalize this fact again and again in order to defend its liberal order. For some, this is a fall from the dreamland of political illusion, for some even a bitter political truth. To prevent the unrealistic and fervent discussion about the use of 'combat drones' from celebrating its happy resurrection, the following postulate should be made: If AI helps to ensure security against the enemies of freedom effectively and efficiently, while preserving the human values and norms of the German Constitution (*Grundgesetz*) and, consequently, that of Innere Fuehrung, then there should really be nothing left to say against the use of AI.

But things are not quite that simple, however fascinating the technical options offered by AI may seem. If Innere Fuehrung is to remain the 'business basis' of the armed forces, then the Bundeswehr must take an 'offensive' approach to AI. 'Offensively' means: not rejecting from the outset, but skeptically in the sense of the German meaning of the Greek word 'σκέπτομαι (skeptomai),' namely 'open-mindedly examining.' This is definitely a tightrope walk, which—as the scrutiny is carried out by people—requires a certain degree of joined-up thinking and sensitivity. The realization must be that the end can be thwarted by the choice of means. The realization must also be that AI is a particular challenge for humans, so that they themselves should not become mentally sluggish. AI offers enough opportunities for (self-)deception with a lethal outcome.

It would be presumptuous to believe that Innere Fuehrung could erect a barrier against AI as such. Don Quixote would then send his regards. Clearly and unequivocally stated: This is not the purpose of Innere Fuehrung. However, it should mentally and instrumentally serve to sensitize us to always carry out a 'test loop' with AI, which protects us from blind faith and fatal euphoria.[7] It would be just as presumptuous to categorize AI as a 'wonder weapon.' There is no such thing as a world (including a military one!) that would be free of any problems thanks to AI, any more than the Schlieffen Plan made the First World War an 'obstacle-free three-month tour.'

AI forces us to first consider who and what we are (= step 1), and then to consider (= step 2) how Innere Fuehrung and thus the personnel of the armed forces must be structured in order to improve Innere Fuehrung with the help of AI—for the benefit of the defense capability of Germany and its allies. This is an ambitious undertaking and should not be confused with trivial instructions for self-optimization. It is about all areas of Innere Fuehrung. The long-term perspective should not be lost out of sight: The victory of freedom over tyranny, which makes, for example—see the main design field of 'civic education' (*Politische Bildung*)—more than ever the

[7] More than ever, Wolf Graf von Baudissin deserves critical reflection and not iconization for the sake of Innere Fuehrung (Baudissin, 1969, p. 62–63). Here Baudissin refers to the importance of human judgement in the context of mechanization. However, his statement there that there were no "decision-making machines" is no longer correct. Even then, this was no longer true, as cybernetic research in the 1950s and 1960s proves.

objectively informed soldier and not the soldier mentally at the mercy of propaganda and disinformation an absolute necessity.

What does this mean in terms of the size of the task? Ultimately, it is about the combination of (not only historical) awareness and technical open-mindedness. Technology must not act as a magnet deviating the needle of the compass of values and, therefore, setting us on the wrong course. What does consciousness mean? It is the knowledge of what defines us, so that we follow a path that does not lead to a dead end. In concrete terms, this can be summarized as a thesis:

Innere Fuehrung forms the military framework for the Enlightenment's concept of ideas and humanity, which sees the autonomous, rational human being as a value in itself. This means that the soldier, as the guarantor of a constitutional security policy, is both the protagonist and instrument of political rule, but never its compliant object. This view of humanity explicitly demands obedience for reasons of conscience. However, the Enlightenment as the idealistic breeding ground of the industrial revolution has a double face. This fact can be expressed in a formula that should always be considered with a question mark: 'Is humanity not also lost precisely because effectiveness and efficiency-orientated action to achieve a goal are clearly and thus one-sidedly emphasized?'

If courses of action are viewed purely rationally, i.e., without any empathy, this can either lead to minimization of violence or to its maximization. This was already abundantly clear in the phase of the Industrial Revolution, for which the Enlightenment provided the starting signal and which in turn had a massive impact on the military and military technology. The mass deaths in two world wars illustrate this vividly against the backdrop that technological progress increases the death of people outside the classic battlefield. All the more, this applies to the technical development of the atomic bomb, which literally ended the Second World War in the Far East 'in one fell swoop.'

Unlike US President Woodrow Wilson, who justified the United States' entry into the Great War of 1914–18 by arguing that this would be the war to end all wars with the yet-to-be-created League of Nations, AI, really, confronts us with the following fact: a war led entirely by it without any legal *and* ethical rules will indeed lead to the end of history. The extent to which the conceptual tools of military theorist Carl von Clausewitz, in particular his understanding of "absolute war," are still applicable here requires deeper reflection. It remains to be seen whether AI supports the thesis of the future viability or limitations of Clausewitz's thinking. The fact is that for Clausewitz, absolute war means the end of politics. He endeavored to reactivate the primacy of politics. In this sense alone, war has a serving function. Nevertheless: A war waged absolutely with AI could well lead to the end of the world–at least the end of the world as we have known it up to now.

Doubts cannot be dismissed as to whether this perspective actually represents an authentic "brave new world" or a variation on Aldous Huxley's (1932) negative utopia of the same name.[8]

[8] Cf. the essayistic review by the same author (Huxley, 1958). Two years before (exactly: 13 July 1956), AI's birth hour took place at a conference of computer scientists and scholars at Dartmouth College (N.H., USA).

The idea that an authoritarian or totalitarian regime that successfully utilizes AI will triumph over our open form of society, known as democracy, and metaphysically mutate in the process by creating the illusion of paradise on earth, is anything but intoxicating. Precisely because the enemies of open society (see the current regime in Russia under Putin) are no strangers to a nihilistic understanding of politics. In terms of security policy and ethically with regard to the self-image of the profession of soldier, it is more necessary than ever to deal with the opportunities and abysses of AI; and this in a world that is at a crossroads between rule-based international relations and a shark tank. Do we really realize this? What's more, democracy is currently facing a crisis of meaning that runs counter to its self-confidence and thus also to its ability to defend itself as a democratic way of life, which ultimately means that life remains worth living.

Unfortunately, the failure of democracy is a realistic option. AI alone will not be able to absolutely save us from the crash. But it can help us to better master our present and future, provided that soldiers in particular, are not lacking in ethical reflection. The alternative, in contrast, will not be a pleasant one: Our world would be a combination of the utopias of George Orwell, Aldous Huxley, and Carl Schmitt. And this means: mankind will be divided by aggressive continental dominions, superficially led by totalitarian dictators, but in reality dominated by unfallible machines. Thus, people would always be kept in cozy immaturity justified by Thomas Hobbes' perception "homo homini lupus est" as the only decisive argument.[9] AI would then even have a pacifying effect due to the wolfish nature of humans! AI's ambivalences, i.e., its use or abuse, deliver us the proof of man's maturity (*Mündigkeit*) and capacity for autonomy (*Autonomiefähigkeit*), as long as we keep in mind that we, ourselves, are the masters of our own fate. Frankly spoken and confessing it by the German soldier's point of view: ultimately in this sense, Innere Fuehrung is the litmus test for AI's compatibility with humanity.

What is the bottom line? AI is a fact that cannot be banished like the 'genie in the bottle.' With its technical applications, the question of the unintended consequences of social behavior also arises in the world of the military. AI confronts all members of the Bundeswehr with a huge learning task: are our ethical values, our level of information and knowledge in scientific-technical and socio-political-cultural fields such that the German military can handle AI responsibly? More than ever, Innere Fuehrung is facing a huge challenge and therefore a new test.

- The Concept of Innere Führung as the intellectual basis and leadership concept (valid for the Bundeswehr from the outset!) on the one hand and artificial intel-

[9] Cf. from the geopolitical perspective Burbank and Cooper (2023) sticking to Orwell's (1949) division of the world in three superstates/empires: (1) Oceania (the two Americas, Australia and the former U.K. as its European outpost, (2) Euroasia, and (3) Eastasia). Which combination ever, Carl Schmitt's axiom of 'friend or foe thinking in politics' (*Freund-Feind-Denken in der Politik*) plays always the part of a godfather. In-depth about Schmitt s. Simms' (2023) reflections on 'Großraumpolitik' and Mueller's (2011) concise analysis of Schmitt's profound influences in politics; not to forget Mehring (2022).

ligence on the other is in a state of tension, as sooner or later ethical challenges will arise in the use of AI that call into question the ultimate responsibility of humans.

- Will the machine then also replace the 'human being in uniform' with ultimate responsibility for ethical decisions? And what does this mean for the Bundeswehr's value compass, which is expressed in the concept of Innere Führung?
- The article does not argue against the use of AI as such. It is intended to sensitize military and political decision-makers to the fact that, even in the future, the concept of Innere Führung will not be a naively derived and thus determined variable of technical development.
- This is done with a particular focus on the security policy environment. AI must represent an opportunity to ensure the operational readiness of the armed forces so that the maturity of the citizen in uniform with his duty to always face up to his ultimate ethical responsibility is maintained.
- AI presents all members of the Bundeswehr with a huge learning task. Do our ethical values and level of scientific, technical, and socio-political-cultural knowledge allow the German military to handle AI responsibly? The concept of Innere Führung therefore faces more challenges than ever before.

References

Baudissin, W. G. v. (1969). *Soldat für den Frieden. Entwürfe für eine zeitgemäße Bundeswehr*. Piper.

Burbank, J., & Cooper, F. (2023). *Post-Imperial Possibilities: Eurasia, Eurafrica, Afroasia*. Princeton University Press. https://doi.org/10.1515/9780691251509

European Parliament. (2023, June 20). *What is Artificial Intelligence and How is it Used?* https://www.europarl.europa.eu/topics/en/article/20200827STO85804/what-is-artificial-intelligence-and-how-is-it-used

Fraunhofer IPA. (2024). *Definitionen*. https://www.ipa.fraunhofer.de/en/about-us/guiding-themes/ai/definition.html

Glaser, H. (Ed.). (1989). *So viel Anfang war nie. Deutsche Städte 1945–1949*. Siedler.

Horkheimer, M., & Adorno, T. W. (2002). *Dialectic of Enlightenment. Philosophical Fragments* (E. Jephcott, Trans.). Stanford University Press (Original work published 1947).

Huxley, A. (1932). *Brave New World. A Novel*. Chatto & Windus.

Huxley, A. (1958). *Brave New World Revisited*. Harper & Brothers.

Krauss, P. (2024, April 08). An der Schnittstelle von Gehirn, Geist und Maschine. Wie die Künstliche Intelligenz mithilfe der Neurowissenschaften auf das nächste Level gelangen wird. *Frankfurter Allgemeine Zeitung, 82*, 18.

Maurer, A. (Ed.). (2021). *Mit Leidenschaft und Augenmaß. Zur Aktualität von Max Weber*. Campus.

Mehring, R. (2022). *Carl Schmitt, Aufstieg und Fall. Eine Biographie* (2nd ed.). C. H. Beck.

Miller, A. (2019). *The artist in the machine. The world of AI-powered creativity*. MIT Press.

Mueller, J.-W. (2011). *Ein gefährlicher Geist. Carl Schmitts Wirkung in Europa* (2nd ed.). Wissenschaftliche Buchgesellschaft.

Orwell, G. (1949). *Nineteen-eighty-four*. Secker & Warburg.

Simms, B. (2023). *Die Rückkehr des Großraums*. Duncker & Humblot.

Spreen, D. (2023). Lethal autonomous weapon systems (LAWS). On the Ethics of Automation in the Military from the Perspective of Social Systems Theory. *Sõjateadlane (Estonian Journal of Military Studies)*, (21), 10–40. https://doi.org/10.15157/st.vi21.24177

Steiger, D. (2023). Der Einsatz Künstlicher Intelligenz (KI) bei der Anwendung militärischer Gewalt im Völkerrecht. In E. Hoffberger-Pippan, R. Ladeck, & P. Ivankovics (Eds.), *Digitalisierung und Recht. Jahrbuch 2023* (pp. 13–52). Nomos. https://doi.org/10.5771/9783748917533-13

Vohl, V. (2023). Europäische Regulierungsoptionen Künstlicher Intelligenz im militärischen Kontext. In E. Hoffberger-Pippan, R. Ladeck, & P. Ivankovics (Eds.), *Digitalisierung und Recht. Jahrbuch 2023* (pp. 53–80). Nomos. https://doi.org/10.5771/9783748917533-53

Weber, M. (1946). Politics as a Vocation. In *From Max Weber: Essays in Sociology* (H. H. Gerth, & C. W. Mills, Trans./Eds., pp. 77–128). Oxford University Press.

Weizsäcker, C. F. v. (1977). *Wege aus der Gefahr. Eine Studie über Wirtschaft, Gesellschaft und Kriegsverhütung*. Carl Hanser.

Zentrum Innere Führung (Ed.). (2023). *Handbuch Innere Führung*. ZInFü. Abteilung Weiterentwicklung Innere Führung. https://www.bundeswehr.de/de/organisation/zentrum-innere-fuehrung/handbuch-innere-fuehrung

Autonomous Weapons Systems Under International Law

Stuart Casey-Maslen

Abstract Whether weapons that rely on artificial intelligence for their targeting and use of force are a danger or a potential benefit to humanity is sharply disputed. This chapter argues that autonomous weapons systems do not, per se, violate any branch of international law. But, as is the case with any weapons, they must comply with international humanitarian law in any armed conflict and with international human rights law at all times. In armed conflict, this requires that autonomous weapons systems be able to distinguish between military objectives and civilians and civilian objects and in any event must not cause excessive incidental civilian harm that is reasonably foreseeable to a human commander. The United Nations Secretary-General has called on States to conclude a treaty to prohibit lethal autonomous weapon systems that function without human control or oversight by 2026. This target will not be met.

1 Introduction

In an open letter published on the *Future of Life Institute* website in 2018, signatories declared their belief that artificial intelligence (AI) "has great potential to benefit humanity in many ways, and that the goal of the field should be to do so." But, they cautioned, "a military AI arms race is a bad idea, and should be prevented by a ban on offensive autonomous weapons beyond meaningful human control" (Future of Life Institute, 2018).[1] Much international discussion has centered on what this

[1] The Lethal Autonomous Weapons Pledge had more than 5000 signatories at the beginning of May 2024.

S. Casey-Maslen (✉)
Insitute of International and Comparative Law in Africa (ICLA), University of Pretoria, Pretoria, South Africa
e-mail: stuart.maslen@sciencespo.fr

K. Talves, D. Spreen (eds.), *Artificial Intelligence in Military Technology*,
Artificial Intelligence, Simulation and Society 192,
https://doi.org/10.1007/978-3-031-95578-5_10

notion of "meaningful human control" may entail in practice without ever defining it precisely. Geoffrey Hinton, the renowned computer scientist often referred to as one of the 'Godfathers' of AI, has spoken of the risk that AI models "may well develop the goal of taking control" (Siddiqui, 2023). Should they do so, an "existential threat is," he believes, "relatively imminent" (Brown, 2023). In contrast, *Meta*'s chief AI scientist (and fellow AI 'Godfather') Yann LeCun has dismissed concerns that the technology could threaten humanity as "preposterously ridiculous," terming it a "projection of human nature on machines" (Dean, 2023).

The development of AI may thus be a global good or the beginning of the end of human civilization, depending on who and what one believes. Unquestionably, however, as the United States (US) Department of Defense has affirmed, AI's military applications will transform warfare (US Department of Defense, 2012). Already, major powers are investing heavily in the military applications of AI to develop air, land, and sea-based autonomous weapons systems (Human Rights Watch and Harvard Law School International Human Rights Clinic, 2023, p. 4) as well as in multi-domain warfare, where AI is expected to control the cooperation of domains and weapons. The technology—and especially deep-learning AI systems based on neural network technology—has given rise to machines that can themselves *decide* to target and kill individuals and destroy objects, all without authorization to do so from a human being. This autonomy in the use of lethal force has significant international legal ramifications.[2]

The fact that an international campaign (Campaign to Stop Killer Robots, 2023)[3] exists to prohibit lethal autonomous weapons systems (LAWS) and that more than 50 States have called for a treaty to prohibit (or at least regulate) autonomous weapons systems[4] is evidence that they are not per se illegal. As part of his "New Agenda for Peace," United Nations (UN) Secretary-General Antonio Guterres has called on States to undertake and conclude the negotiation of a legally binding instrument to prohibit lethal autonomous weapon systems that function without human control or oversight by 2026 (UN Secretary-General, 2023, p. 27). This highly ambitious target aside, three branches of public international law are particularly relevant to the existing regulation of such weapons: the law on inter-State use of force (also

[2] That is despite the etymology of the word 'autonomy,' which comes from the ancient Greek meaning 'having its own laws.'

[3] The *Campaign to Stop Killer Robots* is a non-governmental coalition launched in 2013, which seeks to ensure human control in the use of force. The campaign calls for new international law on autonomy in weapons systems.

[4] The US Department of Defense (2023, § G.2) offers a useful definition of an autonomous weapons system in its Directive of January 2023, which reads as follows: "A weapon system that, once activated, can select and engage targets without further intervention by an operator. This includes, but is not limited to, operator-supervised autonomous weapon systems that are designed to allow operators to override operation of the weapon system, but can select and engage targets without further operator input after activation." The definition in the 2023 Directive is substantively unchanged from that set forth in the 2012 Directive it replaced, although it excludes from its application "autonomous or semi-autonomous cyberspace capabilities." (US Department of Defense, 2023, § 1.1 (b) (1)).

known by its Latin moniker, *jus ad bellum*); international humanitarian law (also called the law of armed conflict or the laws of war); and international human rights law. The regulation of each of these branches of international law is considered in turn.

2 Autonomous Weapons Systems Under the Law on Inter-State Use of Force

Jus ad bellum regulates when one State may use force against, or on the territory of, another State. The foundational rules of the contemporary law are found in the Charter of the United Nations (UN Charter, 1945). A general prohibition on inter-State use of force is set forth in Article 2, paragraph 4 of the UN Charter: "All Members shall refrain in their international relations from the threat or use of force against the territorial integrity or political independence of any state, or in any other manner inconsistent with the Purposes of the United Nations." This is a rule of customary international law (International Court of Justice, 1986, § 188), which binds each of the 197 States in the world today.

The deployment of autonomous weapons systems by any State must comply with this general rule in the same way as it must with any other weapon. That is so, whether the armed force the weapon in question imparts is kinetic (e.g., bombs, shells, and bullets) or non-kinetic (e.g., incendiary, chemical, biological, nuclear, or directed energy) (Casey-Maslen, 2014, p. XX). It is no defense to argue that an autonomous weapons system "went rogue" and attacked another State. Under international law, a State is responsible for all force applied unlawfully to any other State by any of its agents (all organs, entities, or individuals that are under its direction or effective control). This includes responsibility for an illegal use of force by an autonomous weapons system even if the relevant organ, person, or entity "exceeds its authority or contravenes instructions" (International Law Commission, 2001, Art. 7). In such a case, State responsibility encompasses the duty to make full reparation for the injury caused by any unlawful action (International Law Commission, 2001, Art. 31).

But, potentially, the consequences of unlawful inter-State use of force are far more serious than the duty to make good any damage, whether that be material or moral in nature. For Article 51 of the UN Charter (reflecting a long-standing rule of customary law applicable to all States)[5] accords a State that has been the victim of an "armed attack" the right to use force in self-defense against the perpetrator of that attack. That right is not unlimited—it must, as discussed below, be exercised in accordance with the principles of necessity and proportionality (International Court

[5] The wording of Article 51 of the UN Charter refers to the right of self-defense as being "inherent," which means that it already existed as a customary rule.

of Justice, 1986, § 194)—but it is still far-reaching, allowing significant death and destruction to be directed against the culpable State in certain circumstances.

But just as not every use of force from human-controlled weaponry amounts to an armed attack, so not every use of force delivered by an autonomous weapons system gives rise to the right of self-defense. The International Court of Justice declared in its judgment of the merits in the 1986 *Nicaragua* case that an armed attack is the "most grave" form of the use of force (International Court of Justice, 1986, § 191).[6] The Court has never, though, explained exactly where the threshold lies between a mere illegal use of force and an armed attack. In a later judgment on the use of force by the United States, the *Oil Platforms* case,[7] the Court did not "exclude the possibility" that the mining of a single military vessel might be sufficient to "bring into play" the inherent right of self-defense. That said, in the facts at hand, even if it had been proven that Iran were responsible for the sea mine that hit a US warship and a missile that was fired against the US-flagged oil tanker, the *Sea Isle City*, these incidents did "not seem to the Court to constitute an armed attack on the United States" (International Court of Justice, 2003, §§ 64, 72).

Hence, if one or more autonomous weapons systems belonging to one State delivered significant force to the territory, armed forces, or political leadership of another in breach of the general prohibition on inter-State use of force, the victim State would be entitled to exercise its right of self-defense in response. This would be the case if, in the prevailing circumstances, recourse to the UN Security Council and/or the International Court of Justice or other peaceful diplomatic action would be unlikely to put an end to an armed attack or if it were ongoing. This principle of necessity—entirely distinct from its interpretation in international humanitarian law—thus dictates *when* force may be used (Ruys, 2010, p. 98).

The principle of proportionality under *jus ad bellum* then regulates *how much* force may be used (Gray, 2018, p. 159). This is in part a quantum issue even though the application of the principle does not limit the response to equivalence in the choice of weapon or weapons platform or the number of munitions fired. More significant in the application of the principle is the purpose of the use of force in self-defense. Thus, ending and repelling the armed attack, which includes driving out an invader from one's territory (whether metropolitan or non-metropolitan), is legitimate; regime change in the State behind the armed attack and annihilation of its armed forces are not (Randelzhofer, 2002, § 42).

In addition, aside from State responsibility for an unlawful use of force *ad bellum*,[8] potentially the senior military commanders and political leaders of the State that has launched the armed attack using autonomous weapons systems may

[6] The decision by the States that negotiated the UN Charter to incorporate a term other than "force" was not accidental, reflecting the higher threshold of an armed attack.

[7] The *Oil Platforms* case concerned action that occurred during the Gulf War between Iraq and Iran in the 1980s.

[8] The Rome Statute stipulates that: "No provision in this Statute relating to individual criminal responsibility shall affect the responsibility of States under international law" (Rome Statute, 1998, Art 25, § 4).

face personal criminal liability.[9] Launching a war of aggression is an international crime that is punishable under customary international criminal law and, in certain circumstances, by the International Criminal Court (ICC). This is so, where the act of aggression, "by its character, gravity and scale, constitutes a manifest violation" of the UN Charter (Rome Statute, 1998, Art. 8 *bis*, § 1). Nothing in the Rome Statute that underpins the ICC and gives it jurisdiction over international crimes precludes criminal liability where aggression was achieved using autonomous weapons. It would, though, be necessary to prove that the accused was aware that he or she was launching a war of aggression. This indicates that an unexpected malfunction in the algorithm leading to an unplanned attack against another State by autonomous weapons systems might be used as a defense by a person charged with aggression as an international crime.

3 Autonomous Weapons Systems Under International Humanitarian Law

The States Parties to the Convention on Certain Conventional Weapons (CCW, 1980) have been discussing lethal autonomous weapons systems since 2014, beginning with an Informal Meeting of Experts on the subject (UN Office for Disarmament Affairs, 2024). While a definition of LAWS has not yet been agreed upon, High Contracting Parties to the CCW have unanimously acknowledged that international humanitarian law "continues to apply fully to all weapons systems, including the potential development and use of lethal autonomous weapons systems" (CCW High Contracting Parties, 2019, Principle (a)). At the close of the annual meetings of the CCW Group of Governmental Experts in 2023,[10] it was reiterated that international humanitarian law "continues to apply fully to the potential development and use of LAWS," with the States in question declaring that "weapons systems based on emerging technologies in the area of LAWS must not be used if they are incapable of being used in compliance with IHL" (CCW Group of Governmental Experts, 2023, § 21(a) and (b)).

Two years earlier, the Chair of the CCW Group of Governmental Experts had annexed a valuable summary to the report on the Group's discussions. Therein, it was stipulated that States

> should commit not to use, or to develop, produce, acquire, possess, deploy or transfer with a view towards use, any weapons system based on emerging technologies in the area of lethal autonomous weapons systems that can perform the critical functions of selecting and engaging to apply force against targets without further intervention by a human operator, if:

[9] In respect of the crime of aggression, the provisions of the relevant article apply "only to persons in a position effectively to exercise control over or to direct the political or military action of a State" (Rome Statute, 1998, Art 25, § 3 *bis*).

[10] The Group of Governmental Experts on autonomous weapons systems was first established under the CCW in 2017.

> (a) It is of a nature to cause superfluous injury or unnecessary suffering, or it is inherently indiscriminate;
>
> (b) Its autonomous functions are designed to be used to conduct attacks outside a responsible chain of human command and control;
>
> (c) The incidental loss of civilian life, injury to civilians, and damage to civilian objects expected to result from the use of the weapon to conduct attacks cannot be reasonably foreseen or are not fully understood by a human operator; or
>
> (d) It is otherwise incapable of being used in accordance with international humanitarian law. (CCW Group of Governmental Experts, 2022, Annex III, § 25)

Whether autonomous weapons systems can—or will ever in the future be able to—comply with the rules of international humanitarian law in armed conflict is sharply disputed—among States, between scientists, and in the writings of international lawyers. For instance, the overt political and legal objective of the United States is to maintain its ability to develop and use LAWS responsibly but without new legal restrictions, on the basis that international humanitarian law "provides a robust and coherent system of regulation for the use of *all* weapons, including weapons with autonomous functions" (US Permanent Mission in Geneva, 2018, emphasis in original). Others argue that while the law is certainly applicable to LAWS, its application in practice is unclear. Thus, in April 2023, Mirjana Spoljaric, the President of the International Committee of the Red Cross (ICRC), called for "political leadership" from States to "negotiate and adopt a legally binding instrument regulating autonomous weapon systems." New international law, the ICRC believes, is needed to bring clarity and to "uphold and strengthen legal protections" (Spoljaric, 2023). Naturally, it is complicated to regulate something that is constantly evolving, as States have seen in relation to cyber operations.

4 Compliance with IHL Principles and Rules

In truth, as is the case with any weapons system, the use of LAWS in any armed conflict is already restricted by customary and conventional principles and rules, in particular those obligating distinction and proportionality in attack (CCW Group of Governmental Experts, 2023, § 21(c)) in addition to the prohibition of means or methods of warfare of a nature to cause superfluous injury or unnecessary suffering (Australia et al., 2023, Annex I). The former rules primarily protect civilians in armed conflict while the latter rule primarily protects combatants.

The "superfluous injury" rule prohibits the use of weapons whose innate characteristics are such that they would inflict gratuitous harm, which is to say, injury or suffering beyond what is necessary to render a fighter *hors de combat* (putting them out of the fight where they may no longer be attacked).[11] Suffering denotes

[11] Thus, in its 1996 Advisory Opinion on the Legality of the Threat or Use of Nuclear Weapons, the ICJ defined unnecessary suffering as "a harm greater than that unavoidable to achieve legitimate military objectives" (International Court of Justice, 1996, § 78). Injury denotes identifiable wounds and other physical harm to limbs, organs, senses, or other parts of the body, including through

primarily pain, but also extends to encompass severe psychological distress (Dinstein, 2016, p. 74; ICRC, 2005e). Given that autonomous weapons systems are delivering the same weapons as those delivered by human action, there is no additional legal analysis that needs to be performed that is specific to an autonomous system. Of far greater controversy in the analysis of the application of international humanitarian law to those systems are the twin principles of distinction and proportionality in attack.

4.1 The Principle of Distinction

The "cardinal" principle of distinction is international humanitarian law's most fundamental rule governing the conduct of hostilities (International Court of Justice, 1996, § 78). This is the law's term for the fighting and associated activities such as reconnaissance. Integral to the principle is the obligation upon each party to an armed conflict to distinguish in their military operations between civilians, on the one hand, and combatants (or civilians directly participating in hostilities) on the other, and to target only the latter. This obligation is explicit in both of the 1977 Additional Protocols to the Geneva Conventions (Additional Protocol I, 1977, Art. 48. Art. 51(2) and (3); Additional Protocol II, 1977, Art. 13(2) and (3)).

The International Court of Justice has determined that force may be used "regardless of the weapons employed" (International Court of Justice, 1996, § 39). This is widely considered to be a rule of customary law (Schmitt, 2017, p. 328). Accordingly, there is no requirement that a weapon be used by a person in order to be regulated by international humanitarian law. The law defines "attacks" in broad terms as "acts of violence against the adversary, whether in offence or in defence" (Additional Protocol I, 1977, Art. 49(1)).

According to the principle of distinction, an attack is unlawful not only if it is targeted against one or more civilians or the civilian population more broadly, but also if it is not directed at a lawful military objective. With respect to persons, a military objective is a combatant (in international armed conflict) or a person participating directly in hostilities (in non-international armed conflict). With respect to objects, "military objectives are limited to those objects which by their nature, location, purpose or use make an effective contribution to military action and whose total or partial destruction, capture or neutralization, in the circumstances ruling at the time, offers a definite military advantage" (Additional Protocol I, 1977, Art. 52(2); ICRC, 2005c). Thus, contrary to the view in some quarters, targeting State infrastructure may be lawful where it is being, or is about to be used, for military purposes. That is still the case even if the object is also used by civilians or is operated also for their benefit. In such circumstances, however, there will be an

---

burns. Such injuries may render a person temporarily incapacitated or with a permanent disability, such as through loss of sight or hearing, or a soldier may become a single, double, triple or even quadruple amputee.

assessment of proportionality in attack under international humanitarian law to see if the impact on civilians is excessive, as discussed below. But it is important to bear in mind that this rule, as it applies in international humanitarian law, is different to the proportionality assessment under the law on inter-State use of force already described.

An attack that is not directed against any such military objective is indiscriminate and thus unlawful (Additional Protocol I, 1977, Art 51(4); ICRC, 2005d). Further prohibited under conventional and customary international humanitarian law are acts or threats of violence whose primary purpose is to terrorize the civilian population (Additional Protocol I, 1977, Art 51(2); Additional Protocol II, 1977, Art 13(2)). Accordingly, one of the draft articles proposed in the context of the CCW by a group of Western military powers stipulated that "autonomous weapon systems must not be designed to […] [t]arget civilians or civilian objects, or to spread terror among the civilian population" (Australia et al., 2023, Draft Art. 1, § 1(a)).

4.1.1 The Definitions of "Combatant" and "Civilian"

A civilian is generally defined under the law in the negative as anyone who is not a member of the armed forces (ICRC, 2005b). The term "combatant," which technically is used in treaty law only with respect to international armed conflict,[12] encompasses all members of the armed forces who are authorized to participate directly in hostilities (ICRC, 2005a).[13] The definition of a civilian in a situation of non-international armed conflict is, in part, contested. In particular, it is disputed whether a member of a non-State armed group is considered to have a status akin to that of combatant for the purpose of lawful targeting. The ICRC has claimed that a committed military member of a non-State armed group has a "continuous combat function" and may therefore be attacked at any time (Melzer, 2009, pp. 27–36). The better view is that this notion does not reflect the law. Thus, the UN Commission of Inquiry on the Protests in the Occupied Palestinian Territory noted in its 2019 report that "continuous combat function" does not appear anywhere in a treaty of international humanitarian law and duly concluded that the concept "remains unsettled when assessed as custom" (UN Commission of Inquiry on the Protests in the Occupied Palestinian Territory, 2019, §§ 104–106).

[12] This is a conflict involving two States either directly or by proxy including a belligerent occupation or bombardment of another State, even if that hostile military operation meets with no opposition.

[13] Non-combatant members of the armed forces are principally dedicated medical and religious personnel (Geneva Convention III, 1949, Art. 33). They are not considered as civilians but must similarly not be attacked unless and until they engage in hostile acts against the enemy.

4.1.2 The Notion of Direct Participation in Hostilities

It is accepted that in any armed conflict a civilian is protected from attack "unless and for such time as they take a direct part in hostilities" (Additional Protocol I, 1977, Art. 51(3); Additional Protocol II, 1977, Art. 13(3)). In its 2009 *Interpretive Guidance on the Notion of Direct Participation in Hostilities*, the ICRC identified three components: threshold of harm to the enemy, direct causation, and belligerent nexus. Broadening its earlier interpretation of the term as "acts of war which by their nature or purpose are likely to cause actual harm to the personnel and equipment of the enemy armed forces" (Sandoz et al., 1987, § 1944) the ICRC now claimed that participation in hostilities can be founded by an act which either does, or is likely:

- Adversely to affect the military operations of a party to an armed conflict
- Adversely to affect the military capacity of a party to an armed conflict
- To kill or injure civilians protected against direct attack or
- To destroy (or damage) civilian objects protected against direct attack (Melzer, 2009, p. 47)

4.1.3 The Application of the Rules to LAWS: A Targeting Challenge

Can LAWS comply with these fundamental yet complex and sometimes disputed rules in a situation of armed conflict? In their 2012 report, *Losing Humanity: The Case against Killer Robots*, Human Rights Watch and The International Human Rights Clinic at Harvard Law School argued that the rule of distinction "poses one of the greatest obstacles to fully autonomous weapons complying with international humanitarian law. Fully autonomous weapons would not have the ability to sense or interpret the difference between soldiers and civilians, especially in contemporary combat environments" (Human Rights Watch and Harvard Law School International Human Rights Clinic, 2012, p. 30). This claim is in effect asserting that every fully autonomous weapons system now and in the future is inherently indiscriminate with respect to the targeting of persons. This is a bold claim. Some dispute it entirely while others call for nuance. Greipl, for instance, suggests that while the AI systems' output does not classify the individual's status under IHL as such, it will inform the military decision-makers' IHL categorization (Greipl, 2024).

A leading academic lawyer, Michael Schmitt, criticizes Human Rights Watch's "apprehension" as being counterfactual on the basis that military technology has advanced well beyond simply being able to spot an individual or object. He writes:

> Modern sensors can, inter alia, assess the shape and size of objects, determine their speed, identify the type of propulsion being used, determine the material of which they are made, listen to the object and its environs, and intercept associated communications or other electronic emissions. They can also gather additional data on other objects or individuals in the area and, depending on the platform with which they are affiliated, monitor a potential target for extended periods in order to gather information that will enhance the reliability of identification and permit target engagement when the target is relatively isolated. (Schmitt, 2013, p. 11)

He suggests that software in autonomous weapon systems "is likely to be developed" that enables individuals to be visually identified, "thereby enabling precision during autonomous 'personality strikes' against specified persons." "These and related technological capabilities," he concludes, "auger against characterization of autonomous weapon systems as unlawful per se solely based on their autonomous nature" (Schmitt, 2013, p. 11).

But it is one thing to conduct "personality strikes" against specified persons using facial recognition software or through confirmed identification of the insignia of an armed force. It is quite another to be able to lawfully identify and then kill a civilian who is directly participating in hostilities. Certainly, the carriage of arms is not sufficient to mark a person out as a lawful target in a situation of armed conflict. To argue otherwise is to make many if not most civilians in, for instance, the United States or Yemen the object of attack in an armed conflict involving either State, rendering the principle of distinction almost meaningless. The extent to which an algorithm can determine whether any given individual is harming the enemy or its ability to conduct operations is questionable. Likewise, assessing whether that harm would be the direct result of the target's conduct is also a challenge for AI. Furthermore, whether the target is assisting one party to the conflict against another may be beyond its capabilities.

As an anecdotal example, in a talk on LAWS at Stanford University in 2019, Paul Scharre reflected on his experiences as a special operations reconnaissance team leader in the US Army. He recounted one mission that sent him to the Afghanistan-Pakistan border to look for Taliban fighters. There, his team spotted a little girl of perhaps no more than six years of age accompanying a herd of goats and circling the team's position. The team soon realized the animals were a ruse and the child was, in fact, reporting their location to Taliban officials by radio. Applicable international humanitarian law would potentially have allowed Scharre's team to shoot as she was probably participating directly in hostilities through her actions. But they felt that shooting the child was not an option: "It would have been the wrong thing to do," he said. If an AI warrior had been in his place, however, he believes the outcome would have been different and the little girl would have been targeted and quite possibly killed. Scharre said this raised a question: "What if you didn't have a human there, to interpret these [international humanitarian rules], to bring that whole set of human values to those decisions?" (De Witte, 2019).

4.2 The Principle of Proportionality in Attack

The principle of proportionality in attack only comes into play if the principle of distinction is already being complied with. In straightforward terms, the proportionality principle dictates that parties to any armed conflict are prohibited from launching attacks which "may be expected" to cause "excessive" civilian harm when compared to the "concrete and direct military advantage anticipated." The principle applies as a matter of customary IHL irrespective of whether the armed conflict is

classified as international or non-international. In treaty law, however, the rule is set forth only in the first 1977 Additional Protocol (Additional Protocol I, 1977, Art. 57(5)(b)). It is not included in the second Additional Protocol (Additional Protocol II, 1977).

There are six main elements to the proportionality rule. First, there must be an "attack," which must have been subject to some form of prior planning. Second, the attack must be directed at a lawful military objective, meaning that the principle of distinction is complied with. Third, in the requisite assessment, which is conducted at the planning stage, the standard of foreseeability of harm from the attack is relatively low: "which may be expected." This is an objective, not a subjective standard. Fourth, the expected harm must be to civilians (involving death or physical injury) or concern damage to civilian objects (or be a combination of such reasonably foreseen consequences). Fifth, the civilian harm foreseen by the attacking party must be compared to the anticipated "concrete and direct" military advantage. The final element is that the attack is unlawful under the proportionality rule where that foreseen civilian harm would be "excessive" in comparison to the expected military advantage (Bellal & Casey-Maslen, 2022, § 10.4). What is excessive is of course a vague and contested notion.

There is no obligation under international humanitarian law, as is sometimes suggested, to "*ensure* [i.e., as an outcome] that civilian losses are not disproportionate to the direct and concrete military advantage anticipated to result from the attack" (Duffy, 2015, p. 373 [added emphasis]). The fact that civilians have been killed in any attack certainly demands that an inquiry be conducted to assess whether the principle of proportionality in attack was complied with. But compliance with the principle is not judged after the attack, based on the number of civilians that were harmed. Rather, compliance is judged *ex ante*, looking back in time at the situation when the attack was launched. What did the commander know about the risk to civilians at that time, or what should he or she be reasonably expected to have known? And based on that knowledge, was the attack justified because of the military benefit it was likely to bring? Would a reasonable commander have come to a similar decision to proceed with the attack? (Bellal & Casey-Maslen, 2022, § 10.5.)

Herein lies the problem. How is an algorithm going to be able to apply the proportionality rule with the necessary nuance—and "humanity" (a fundamental principle of international humanitarian law)—that one expects from a reasonable human commander? Unless the programming excludes the possibility of targeting where there is a risk of incidental civilian harm (which is certainly desirable but not required by the law), there is an aggravated risk of disproportionate attacks. It is noteworthy in this regard that in the draft articles proposed by the United States and several of its military allies, the wording pertaining to proportionality is exceptionally permissive. It is stipulated that autonomous weapon systems must not be designed to "[c]onduct engagements that would *invariably* result in incidental loss of civilian life, injury to civilians, and damage to civilian objects excessive in relation to the concrete and direct military advantage anticipated." The subsequent savings clause whereby autonomous weapon systems "may only be developed such that their effects in attacks are capable of being anticipated and controlled as

required in the circumstances of their use, by the principles of distinction and proportionality" (Australia et al., 2023, Draft Art. 1(1)(b)) does not soften the serious nature of the potential design flaw.

4.3 The Duty to Take Precautions in Attack

The principles of distinction and proportionality in attack are underpinned by the duty to take precautions in attack. The requisite precautions involve, among others, taking all feasible measures to verify that the target is a lawful one and taking all feasible precautions in the choice of weapons and their method of use to avoid or at least minimize incidental loss of civilian life, injury to civilians, and damage to civilian objects (Additional Protocol I, 1977, Art. 57(2)(a)(i) and (ii)). Reflecting the principle of proportionality in attack, States are further obligated not to launch an attack which "may be expected" to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be "excessive" in relation to the concrete and direct military advantage anticipated (Additional Protocol I, 1977, Art. 57(2)(a)(iii)).

The choice of autonomous weapons systems as the means of warfare in any conflict or attack is itself subject to legal challenge under the duty to take precautions in attack. A failure to take the necessary precautions in doing so engages the responsibility of the State under international law. But it does not readily lead to personal criminal responsibility for a war crime. Nowhere in the Rome Statute or in customary international criminal law is a willful failure to take precautions in attack a war crime (ICRC, 2005f).

5 Autonomous Weapons Systems Under International Human Rights Law

It may also be the case that autonomous weapons systems may be deployed in peacetime, for instance against violent criminals or riotous crowds. In such a situation, international humanitarian law is not applicable to any use of force. Instead, the action is to be adjudged under international human rights law as interpreted by reference to the law of law enforcement (Casey-Maslen & Connolly, 2018, pp. 79–107).

The Human Rights Committee that oversees implementation of the 1966 International Covenant on Civil and Political Rights has affirmed that "such weapon systems should not be developed and put into operation, either in times of war or in times of peace, unless it has been established that their use conforms with article 6 [guaranteeing the right to life] and other relevant norms of international law" (Human Rights Committee, 2019, § 65). Two fundamental human rights are especially important in the assessment of legality: the right to life and the right to freedom from torture or other cruel, inhuman, or degrading treatment ("ill-treatment"

for short). In each case, the application of two law enforcement principles is required: necessity and proportionality. In both cases, their interpretation and content differ materially to their homonyms under international humanitarian law and *jus ad bellum*.

5.1 Necessity Under the Law of Law Enforcement

The principle of necessity under the law of law enforcement dictates that, in order to be lawful, any use of force in policing must be necessary in the prevailing circumstances in order to achieve a legitimate law enforcement purpose. The nature and amount of the use of force must be no more than the minimum necessary in the circumstances *as the officer honestly and reasonably believed them to be*. Furthermore, once the need for force to be used has passed (for instance, because a suspect has ceased to resist arrest) any additional use of force will be unlawful. In particular, any punitive violence will violate the prohibition of ill-treatment under international human rights law.

The 1990 UN Basic Principles on the Use of Force and Firearms stipulate that, in carrying out their duty, law enforcement officials "shall, as far as possible, apply non-violent means before resorting to the use of force and firearms" (UN Basic Principles, 1990, Principle 4). Such non-violent means include verbal persuasion, the presence and authority of a police officer, and positive body language (Palmiotto, 2013, p. 245). For, as the US Department of Justice has observed: "The ability of a police officer to bring calm to a situation is a core policing skill" (US Department of Justice, 2015, p. 26). It is exceptionally hard to see how an autonomous weapons system could meet this requirement.

5.2 Proportionality Under the Law of Law Enforcement

If any given use of force complies with the principle of necessity, it must then also comply with the principle of proportionality. Failure to comply with both principles means that the force used will have been unlawful. The principle of proportionality acts as a ceiling on lawful force, adjudging whether the force used was reasonable when considering the threat posed by a suspect or the harm sought to be avoided. Thus: "Whenever the lawful use of force and firearms is unavoidable, law enforcement officers shall […] act in proportion to the seriousness of the offence and legitimate objective to be achieved" (UN Basic Principles, 1990, Principle 5). But is potential harm to an autonomous law enforcement official to be considered in adjudging proportionality? Is it analogous to the harm that may be caused to a human officer?

The restrictions imposed by international law on the use of firearms in policing operations apply the *law enforcement* principles of necessity and proportionality.

The rules, which are of a customary law nature (Inter-American Court of Human Rights, 2006, § 69), are contained in Principle 9 of the 1990 UN Basic Principles on the Use of Force and Firearms. The default rule is that law enforcement officials "shall not use firearms against persons except in self-defense or defense of others against the imminent threat of death or serious injury" and "only when less extreme means are insufficient to achieve these objectives" (UN Basic Principles, 1990, Principle 9). The latter element of the rule evidently reflects the principle of necessity while the former component embodies the principle of proportionality (Rodley & Pollard, 2011, p. 499). Imminence is not defined in the 1990 Basic Principles, but according to the UN Special Rapporteur on extrajudicial, summary, or arbitrary executions, an imminent threat should be considered "a matter of seconds, not hours" (Heyns, 2014, § 59). Given the far faster reactions of an autonomous weapons system, is imminence not to be construed as a fraction of a second rendering its use of a firearm in any set of circumstances a very last of last resorts? This would restrict the risk of a deadly outcome.

5.3 The Duty to Take Precautions

Already reducing the risk of a fatal outcome downstream from decisions taken upstream in the planning of a law enforcement operation is the duty to take precautions. This duty, which is again to be interpreted in an entirely distinct manner from the homonymous international humanitarian law principle, applies before and when using force. This principle emanates directly from international human rights law rather than the law of law enforcement. Thus, in its landmark judgment in the case *McCann et al. against the United Kingdom* in 1995, the European Court of Human Rights buttressed the understanding of the principles of necessity and proportionality by imposing a duty on the authorities to both plan and control a law enforcement operation "so as to minimize, to the greatest extent possible, recourse to lethal force" (European Court of Human Rights, 1995, § 194). An autonomous weapons system could be considered a means of reducing the risk to human officers but could it also be effectively programmed so as minimize the risk of injury?

6 Concluding Remarks

Many oppose the notion that LAWS can ever hope to meet the "principles of humanity and the dictates of public conscience" in the words of treaty law applicable in armed conflict (Additional Protocol I, 1977, Art. 1(2)). At least for now, key States have agreed that human responsibility for decisions on the use of weapons in armed conflict must be retained given that accountability "cannot be transferred to machines." This, States declare, "should be considered across the entire life cycle of the weapons system" (CCW High Contracting Parties, 2019, Annex III, Principle (b)).

But what if it turns out that autonomous weapons systems prove to be more protective of human life, especially that of civilians. Would we still wish to prohibit their use? The criminal acts of Russian armed forces in the war in Ukraine are the latest in a long line of armed conflicts over the past nine decades where war crimes are the norm not the exception. Is the dignity of the victim of the war crime of murder perpetrated by a person truly greater than that of the victim of an autonomous weapons system who is saved or only lightly injured? Already Ukraine has seen the first use of a sensor-fuzed cluster munition. The SMArt 155 munition produced and exported by Germany[14] to Ukraine's armed forces appears to comply with the exceptions listed in the 2008 Convention on Cluster Munitions for such an autonomous weapons system, thereby excluding the weapon from the prohibitions in the treaty (CCM, 2008, Art 2(2)(c)).[15]

In law enforcement operations, the circumstances in which an autonomous weapons system could lawfully be deployed are highly circumscribed. This is especially the case if lethal force is envisaged or is a capability of such a system. Even if an autonomous system fires less-lethal munitions as a form of policing, the law's requirements of necessity and proportionality and the human rights duty of precaution mean that the risk of infringing the right to freedom from ill-treatment is significant. This does not, however, per se eliminate the possibility that an autonomous weapons system could one day be used as a licit law enforcement official dispensing force only where absolutely necessary and strictly proportionate.

In sum, autonomous weapons systems are likely to have an increasing role in warfighting and may be capable of lawful use in certain, limited circumstances. Humans are already dependent on and intertwined with AI systems across a wide range of military decision-making processes (Greipl, 2024). The pressing need for far better civilian protection in armed conflicts does not require that every single weapon that is fired has to be under 'meaningful human control,' a concept that has not been satisfactorily defined in law. But autonomous weapons systems should never be used for everyday policing. They simply do not have the human touch that is integral to effective, human rights-compliant law enforcement.

[14] *Gesellschaft für Intelligente WirkSysteme mbH*, which produces the munition, is a joint venture between *Rheinmetall* and *Diehl BG*. The SMArt 155 munition (Suchzünder Munition für die Artillerie 155: "sensor-fuzed munition for 155 mm artillery" in English) consists of a 47 kg artillery projectile incorporating two sensor-fuzed fire-and-forget submunitions (Defense Update, 2008).

[15] A munition that, in order to avoid indiscriminate area effects and the risks posed by unexploded submunitions, has all of the following characteristics: (i) Each munition contains fewer than ten explosive submunitions, (ii) each explosive submunition weighs more than four kilograms, (iii) each explosive submunition is designed to detect and engage a single target object, (iv) each explosive submunition is equipped with an electronic self-destruction mechanism, and (v) each explosive submunition is equipped with an electronic self-deactivating feature.

7 Key Findings

- Whether weapons that rely on artificial intelligence for their targeting and use of force are a danger or a potential benefit to humanity is sharply disputed.
- Autonomous weapons systems do not, per se, violate any branch of international law.
- But any such system must—as is the case with any weapons—comply with international humanitarian law in any armed conflict and with international human rights law at all times.
- In armed conflict, this requires that autonomous weapons systems be able to distinguish between military objectives and civilians and civilian objects. When targeted at a lawful military objective, they must not cause excessive incidental civilian harm that is reasonably foreseeable to a human commander.
- The use of autonomous weapons systems for policing operations will invariably violate the law of law enforcement and the rights to life and/or to freedom from ill-treatment.

References

Additional Protocol I. (1977). *Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts (Protocol I)*. Adopted at Geneva 8 June 1977, entered into force 7 December 1978.

Additional Protocol II. (1977). *Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of Non-international Armed Conflicts (Protocol II)*. Adopted at Geneva 8 June 1977, entered into force 7 December 1978.

Australia, Canada, Japan, Poland, Republic of Korea, United Kingdom, and United States. (2023, May 15). *Draft Articles on Autonomous Weapon Systems—Prohibitions and Other Regulatory Measures on the Basis of International Humanitarian Law. ("IHL").* Submitted by Australia, Canada, Japan, Poland, the Republic of Korea, the United Kingdom, and the United States. UN doc. CCW/GGE.1/2023/WP.4/Rev.2, Annex I: "Relevant Cardinal Principles and Requirements of International Humanitarian Law," Prohibited Weapons.

Bellal, A., & Casey-Maslen, S. (2022). *The Additional Protocols to the Geneva Conventions in Context*. Oxford University Press.

Brown, S. (2023, May 23). *Why Neural Net Pioneer Geoffrey Hinton is Sounding the Alarm on AI*. MIT Management Sloan School. https://mitsloan.mit.edu/ideas-made-to-matter/why-neural-net-pioneer-geoffrey-hinton-sounding-alarm-ai

Campaign to Stop Killer Robots. (2023). *About Us*. https://www.stopkillerrobots.org/about-us/

Casey-Maslen, S. (Ed.). (2014). *Weapons under International Human Rights Law*. Cambridge University Press.

Casey-Maslen, S., & Connolly, S. (2018). *Police Use of force Under International Law*. Cambridge University Press.

CCM. (2008). *Convention on Cluster Munitions*. Adopted at Dublin 30 May 2008, entered into force 1 August 2010.

CCW. (1980). *Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which may be Deemed to be Excessively Injurious or to have Indiscriminate Effects (with Protocols I, II and III)*. Adopted at Geneva 10 October 1980, entered into force 2 December 1983.

CCW Group of Governmental Experts. (2022, February 22). *Report of the 2021 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems*. UN doc. CCW/GGE.1/2021/3. https://meetings.unoda.org/meeting/64589/documents

CCW Group of Governmental Experts. (2023, May 24). *Report of the 2023 Session of the CCW Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems*. UN doc. CCW/GGE.1/2023/2. https://meetings.unoda.org/meeting/67246/documents

CCW High Contracting Parties. (2019, December 13). Guiding Principles on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems. In *Final Report: Meeting of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons which may be Deemed to be Excessively Injurious or to Have Indiscriminate Effects, Geneva, 13-15 November 2019*. UN doc. CCW/MSP/2019/9. https://digitallibrary.un.org/record/3856241

De Witte, M. (2019, May 01). *In Drell Lecture, Speaker Calls for Ethics and Humanity as Militaries Expand Autonomous Weaponry*. Stanford Report. https://news.stanford.edu/stories/2019/05/ethics-autonomous-weapons

Dean, G. (2023, June 15). *One of the "Godfathers" of AI Says Concerns the Technology Could Pose a Threat to Humanity are "Preposterously Ridiculous."* Business Insider. https://www.businessinsider.com/yann-lecun-artificial-intelligence-generative-ai-threaten-humanity-existential-risk-2023-6

Defense Update. (2008, August 10). *SMArt Sensor-Fuzed Ammunition for 155mm Guns*. https://defense-update.com/20080810_smart.html

Dinstein, Y. (2016). *The Conduct of Hostilities under the Law of International Armed Conflict* (3rd ed.). Cambridge University Press.

Duffy, H. (2015). *The "War on Terror" and the Framework of International Law* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139028585

European Court of Human Rights. (1995, September 27). *McCann et al.* v. *UK. Judgment (Grand Chamber)*.

Future of Life Institute. (2018, June 06). *Lethal Autonomous Weapons Pledge (Open Letter)*. https://futureoflife.org/open-letter/lethal-autonomous-weapons-pledge

Geneva Convention III. (1949). *Convention (III) Relative to the Treatment of Prisoners of War.* Adopted at Geneva 12 August 1949, entered into force 21 October 1950.

Gray, C. (2018). *International Law and the Use of Force* (4th ed.). Oxford University Press. https://doi.org/10.1093/law/9780198808411.001.0001

Greipl, A. R. (2024, June 14). *Artificial Intelligence Systems and Humans in Military Decision-making: Not Better or Worse but Better Together*. Articles of War, Lieber Institute. https://lieber.westpoint.edu/artificial-intelligence-systems-humans-military-decision-making-better-worse

Heyns, C. (2014, April 01). *Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions.* UN doc. A/HRC/26/36.

Human Rights Committee. (2019, September 03). *General Comment no. 36: Article 6: Right to Life.* UN doc. CCPR/C/GC/36.

Human Rights Watch, & Harvard Law School International Human Rights Clinic. (2012, November 19). *Losing Humanity: The Case against Killer Robots (Report)*. https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots

Human Rights Watch, & Harvard Law School International Human Rights Clinic. (2023, February 14). *Review of the 2023 US Policy on Autonomy in Weapons Systems*. https://www.hrw.org/news/2023/02/14/review-2023-us-policy-autonomy-weapons-systems

ICRC. (2005a). *Customary International Humanitarian Law Rule 3: Definition of Combatants*. International Humanitarian Law Databases. https://ihl-databases.icrc.org/en/customary-ihl/v1/rule3

ICRC. (2005b). *Customary International Humanitarian Law Rule 5: Definition of Civilians*. International Humanitarian Law Databases. https://ihl-databases.icrc.org/en/customary-ihl/v1/rule5

ICRC. (2005c). *Customary International Humanitarian Law Rule 8. Definition of Military Objective*. International Humanitarian Law Databases. https://ihl-databases.icrc.org/en/customary-ihl/v1/rule8

ICRC. (2005d). *Customary International Humanitarian Law Rule 12*: Definition of Indiscriminate Attacks. https://ihldatabases.icrc.org/en/customary-ihl/v1/rule12

ICRC. (2005e). *Customary International Humanitarian Law Rule 70: Weapons of a Nature to Cause Superfluous Injury or Unnecessary Suffering*. https://ihl-databases.icrc.org/en/customary-ihl/v1/rule70

ICRC. (2005f). *Customary International Humanitarian Law Rule 156: Definition of War Crimes*. https://ihl-databases.icrc.org/en/customary-ihl/v1/rule156

Inter-American Court of Human Rights. (2006, July 05). *Montero-Aranguren et al. (Detention Centre of Catia)* v. *Venezuela*. Judgment (Merits).

International Court of Justice. (1986, June 27). *Case concerning Military and Paramilitary Activities in and Against Nicaragua (Nicaragua v. United States). Judgment (Merits)*.

International Court of Justice. (1996, July 08). *Legality of the Threat or Use of Nuclear Weapons*. Advisory opinion. https://www.icj-cij.org/case/95

International Court of Justice. (2003, November 06). *Case Concerning Oil Platforms (Iran* v. *United States). Judgment (Merits)*.

International Law Commission. (2001). *Draft Articles on Responsibility of States for Internationally Wrongful Acts*. UN doc. A/56/10.

Melzer, N. (2009). *Interpretive Guidance on the Notion of Direct Participation in Hostilities*. ICRC. https://www.refworld.org/policy/legalguidance/icrc/2009/en/68382

Palmiotto, M. J. (2013). *Policing: Concepts, Strategies, and Current Issues in American Police Forces forces* (3rd ed.). Carolina Academic Press. https://www.ojp.gov/ncjrs/virtual-library/abstracts/policing-concepts-strategies-and-current-issues-american-police

Randelzhofer, A. (2002). Article 51. In B. Simma (Ed.), *The Charter of the United Nations: A Commentary* (Vol. 1, 2nd ed., pp. 1397–1428). Oxford University Press.

Rodley, N. S., & Pollard, M. (2011). *The Treatment of Prisoners under International Law* (3rd ed.). Oxford University Press.

Rome Statute. (1998). *Rome Statute of the International Criminal Court.* Adopted at Rome 17 July 1998, entered into force 1 July 2002.

Ruys, T. (2010). *"Armed Attack" and Article 51 of the UN Charter: Evolutions in Customary Law and Practice*. Cambridge University Press.

Sandoz, Y., Swinarski, C., & Zimmermann, B. (Eds.). (1987). *Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949*. ICRC, & Martinus Nijhoff Publishers. https://www.legal-tools.org/doc/6d222c/pdf

Schmitt, M. N. (2013, February 05). Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics. *Harvard National Security Journal, 4*, 1–37. https://harvardnsj.org/2013/02/05/autonomous-weapon-systems-and-international-humanitarian-law-a-reply-to-the-critics

Schmitt, M. N. (Ed.). (2017). *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/9781316822524

Siddiqui, T. (2023, June 29). *Risks of Artificial Intelligence must be Considered as the Technology Evolves: Geoffrey Hinton*. University of Toronto News. https://www.utoronto.ca/news/risks-artificial-intelligence-must-be-considered-technology-evolves-geoffrey-hinton

Spoljaric, M. (2023, April 26). *Risks from the Unconstrained Use of Autonomous Weapons in Armed Conflict are Stark* (Statement given by Mirjana Spoljaric, President of the ICRC, at the Luxembourg Autonomous Weapons Systems Conference). https://www.icrc.org/en/document/risks-unconstrained-use-autonomous-weapons-armed-conflict-are-stark

UN Basic Principles. (1990). *Basic Principles on the Use of Force and Firearms by Law Enforcement Officials*. Adopted at Havana 7 September 1990.

UN Charter. (1945). *Charter of the United Nations.* Adopted at San Francisco 26 June 1945, entered into force 24 October 1945.

UN Commission of Inquiry on the Protests in the Occupied Palestinian Territory. (2019, March 18). *Report of the Detailed Findings of the Independent International Commission of Inquiry on the Protests in the Occupied Palestinian Territory.* UN doc. A/HRC/40/CRP.2.

UN Office for Disarmament Affairs. (2024). *Timeline of LAWS in the CCW*.

UN Secretary-General. (2023, July 20). *Our Common Agenda: A New Agenda for Peace (Policy Brief, 9)*. https://reliefweb.int/report/world/our-common-agenda-policy-brief-9-new-agenda-peace-july-2023

US Department of Defense. (2012). *Autonomy in Weapon Systems. Directive 3000.09.*

US Department of Defense. (2023, January 25) *Autonomy in Weapon Systems. Directive 3000.09 (Updated)*. https://media.defense.gov/2023/Jan/25/2003149928/-1/-1/0/DOD-DIRECTIVE-3000.09-AUTONOMY-IN-WEAPON-SYSTEMS.pdf

US Department of Justice, Civil Rights Division. (2015, March 04). *Investigation of the Ferguson Police Department*. https://www.justice.gov/sites/default/files/opa/press-releases/attachments/2015/03/04/ferguson_police_department_report_1.pdf

US Permanent Mission in Geneva. (2018, August 27). *Statement to the Meeting of the Group of Governmental Experts to the CCW on Lethal Autonomous Weapons Systems (LAWS)*. https://geneva.usmission.gov/2018/08/27/meeting-of-the-group-of-governmental-experts-of-the-high-contracting-parties-to-the-ccw-on-lethal-autonomous-weapons-systems/

Challenges for Communication

Trustworthy System Design: A Human Factors Perspective

Sonia Sousa, Gabriela Beltrão, Iuliia Paramonova, and Debora C. Firmino de Souza

Abstract The increasing use of autonomous and intelligent machines for strategic decision-making has disrupted the way that humans, organizations, and societies relate to technology. This phenomenon has raised awareness of the potential gains and losses of using such disruptive technologies, and the need to maintain more ethical, responsible, and trustworthy systems. This article stresses the importance of adopting an analytical view to incorporate end users' trust experiences during different human-centered design (HCD) problem-solving phases. The aim is to facilitate and support the integration of humans and systems, because otherwise, the consequence could be the failure of AI. It builds from the Human–Computer Interaction (HCI) body of knowledge, highlighting the importance of studying users' trust behavior to calibrate human system integration and mitigate possible communication challenges in system design. The article argues for the need to use Human–Computer Trust Scale (HCTS) as a research lens to enhance users' trust in current human–machine interrelations (HMI). Its goal is to help system designers and engineers further reflect on the potential benefits of using human trust factors as a driving element to create more trustworthy HMI collaborations.

S. Sousa (✉) · G. Beltrão · I. Paramonova · D. C. Firmino de Souza
School of Digital Technologies, Tallinn University (TLU), Tallinn, Estonia
e-mail: scs@tlu.ee; gbeltrao@tlu.ee; gabriela.de_moraes_beltrao@tlu.ee; juparam@tlu.ee; deboracs@tlu.ee; debora.conceicao_firmino_de_souza@tlu.ee; https://www.tlu.ee/en/node/110740; https://www.tlu.ee/en/node/110720; https://www.tlu.ee/en/node/110443

K. Talves, D. Spreen (eds.), *Artificial Intelligence in Military Technology*,
Artificial Intelligence, Simulation and Society 192,
https://doi.org/10.1007/978-3-031-95578-5_11

1 Introduction

The increasing use of autonomous and intelligent machines for strategic decision-making has disrupted the way that humans, organizations, and societies relate to technology. This phenomenon has raised awareness of the potential gains and losses of using such disruptive technologies and the need to maintain more ethical, responsible, and trustworthy systems. This article stresses the importance of adopting an analytical view to incorporate end users' trust experiences during different human-centered design (HCD) problem-solving phases. It aims to facilitate and support human–system integration because otherwise, the consequence could be the failure of AI. It builds from the Human–Computer Interaction (HCI) body of knowledge, highlighting the importance of studying users' trust behavior to calibrate human–system integration and mitigate possible communication challenges in system design. The article argues for the need to consider using Human–Computer Trust Scale (HCTS) as a research lens to enhance users' trust in current Human–Machine Interrelations (HMI). Its goal is to help system designers and engineers further reflect on the potential benefits of using human trust factors as a driving element to create more trustworthy HMI collaborations.

Like humans, AI-enabled machines (or artificially intelligent automation) should be able to demonstrate that they are trustworthy, i.e., explain their inner operating and intentions, demonstrate their competency, and follow ethical principles. Moreover, the trust that an individual is willing to place in such systems (i.e., trust attitude) depends on this person's goals and the uncertainty and vulnerability characterizing human–machine interaction (Lee, J.-G. et al., 2015; Mayer et al., 1995). In other words, a person's belief in whether a machine is capable of fulfilling a task as expected (i.e., trustworthiness) depends on the characteristics of the task and the person's perception of AI (Dong, 2010). This question is essential, especially considering that human decision-makers in social roles are accountable for the behavior of AI systems (Spreen, 2023, pp. 29–30).

To address these concerns, we use the HCTS empirical model to integrate trust as a quality of use in system design and development processes (Gulati et al., 2019). The following paragraphs will provide a general scope of the state of the art, focusing on users' trust research and addressing the problem's relevance; our focus is on the socio-digital influence of trust and explaining how it frames new interrelationships between artificially intelligent automation (AI) and humans.

Our contribution goes beyond studying trust as a component of a machine: we examine what potential human factors of trust can hinder or sustain a system's user experience quality (Sousa et al., 2023). By briefly describing the importance of adopting an analytical view in facilitating actions and supporting human–system integration, we highlight that trust lies in a set of aspects that allows us to perceive something (person, system, or service) as trustworthy. We claim that users' trust is reflected through their predisposition to act or behave in a given context, and this is demonstrated by users' intentions to perform a particular action that is important to

them, "irrespective of the ability to monitor or control that other party" (Mayer et al., 1995).

We will also illustrate how to use the HCTS psychometric scale as a user research lens to measure a system's perceived trustworthiness according to international software standards. We will bring three examples, inviting IT service providers, system designers, and engineers to critique their design propositions and become more informed on the potential challenges of design in the light of communicating and mediating users' trust in AI-enabled machines.

2 The Socio-Digital Influence of Trust

As society becomes more dependent on intelligent machines that perform their daily routines, discussions are arising on the need to develop novel mechanisms for mitigating potential threats, followed by the need to develop their operating, and examine how they can affect our well-being and decisions (Zuboff, 2015). This need is associated with AI's dependency on data and its potential misuse and bias, and/or the unjustified trust that people have in its ability to perform the expected task.

Hence, our belief on whether a machine is capable of fulfilling a task, and the characteristics of the task can affect our willingness to trust (i.e., trust attitude) and/or depend on such systems to perform a particular important task, especially when the characteristics of the task performed by an AI-based system can pose a high risk on human beings or cause extensive uncertainty and vulnerability (Mayer et al., 1995; Lee D., 2015).

Another aspect related to the current need for novel user research mechanisms is related to the fact that these data-driven machines operate like black boxes: their outputs cannot be explained, or in some cases, humans' limitations in understanding a complex subject might make them underestimate the potential risks to society (Rossi, 2018).

Therefore, to mitigate potential misuse and promote ethical principles, we could build a more trustworthy AI (TAI) relationship as proposed by software developers, regulators, etc. (European Commission, 2020; Sousa et al., 2023; Future of Life Institute, 2024; Madiega, 2021).

This approach of building more TAI relationships between humans and machines is predominantly implemented on the governmental level. It emphasizes the ethical principles of enforcing trustworthy AI applications by providing various regulatory guidelines to help IT providers abide during system design, development, and deployment of AI. Those include, for example, the *Ethics Guidelines for Trustworthy AI* (Cannarsa, 2021) the *European Union (EU) AI Act* (Future of Life Institute, 2024; Madiega, 2021) and *The Assessment List for Trustworthy Artificial Intelligence* (ALTAI) (European Commission, 2020), and the *CapAI* (Floridi et al., 2022).

Gerald Wagner argues that "technical standardization" tends to be underestimated in social and human science approaches (Wagner, 1994, pp. 153–155). Therefore, to mitigate potential misuse or miscommunication practices and promote

ethical principles for trustworthy AI, we need to consider an interdisciplinary approach that integrates the legal, technical, and socio-technical aspects of the human–AI relationship. However, such interdisciplinary approach requires novel HCI frameworks and expertise that focus on developing technology that would benefit people.

This section will illustrate the need for a human-centered AI (HCAI) approach strategy to utilize the existing legal, technical, and regulatory measures for designing a trustworthy system, while highlighting the complexity of the problem (Sousa et al., 2021, 2023). First, we will summarize three common approaches to incorporating and/or promoting ethical principles for a trustworthy AI: legal, technical, and socio-technical. We suggest adopting a socio-technical framework to help identify, study, categorize, and evaluate people who place trust in those machines.

2.1 Regulatory Standards

The regulatory approach aims to promote the development of AI in benefit of society by integrating the capability approach into AI ethics and design. It focuses on human-centric values and the broad, long-term impact of AI on human well-being. This perspective encourages the creation of AI systems that support and expand human capabilities rather than undermine or replace them (Floridi et al., 2018, 2022). It aims to ensure that AI system design is lawful, ethical, and robust. These guidelines emphasize the need for AI to be human-centric and respect human rights and values, enhancing autonomy and societal well-being while being technically secure and trustworthy. These include human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity and non-discrimination, societal and environmental well-being, and accountability (European Commission, 2019).

The EU AI Act also reflects the EU's commitment to balancing technological advancement with ethical and societal values by addressing the risks associated with specific uses of AI. The act classifies AI systems according to the level of risk they pose, ranging from minimal risk to unacceptable risk. It outlines strict requirements for high-risk applications. These requirements focus on transparency, data quality, oversight, and the need for human oversight to mitigate risks. The Act proposes outright bans for AI systems whose risk level is deemed unacceptable, including AI that manipulates human behavior to circumvent users' free will, and systems that allow "social scoring" by governments. Also, the Act demands that these high-risk AI systems' manufacturer must prove they can follow the ethical guidelines for a trustworthy AI (Future of Life Institute, 2024; Madiega, 2021).

These new regulations and principles emphasize the human-centered AI (HCAI) approach. They also bring a set of legal requirements and tools to instruct IT providers to reflect on various facets of legal requirements to develop ethical and trustworthy systems. An example of such tool is the Assessment List for Trustworthy Artificial Intelligence (ALTAI), provided by the European Union. It is a

self-assessment, or auditing method, to assess whether an AI-based assistance and automation systems is lawful, ethical, and technically robust (Sousa et al., 2023; European Commission, 2020). Similarly, the audit-based method Z-Inspection (https://z-inspection.org) is a co-design, self-assessment, and auditing method to assess trustworthy AI in practice (Zicari et al., 2021).

2.2 Technical Standards

The technical approach bridges the division between technical standardization and regulatory needs as stated above, and artificial intelligence standards' concepts and terminology (e.g., ISO/IEC 22989).

There are also mechanisms to support the development and deployment of machine learning (ML) systems like the *IBM AI Fairness, Explainability, Adversarial Robustness 360 toolkit* (https://art360.res.ibm.com), the *Google What-If Tool* (https://pair-code.github.io/what-if-tool; Wexler et al., 2019), and *Microsoft Fairlearn* (https://fairlearn.org; Bird et al., 2020), and the *Google Cloud Explainable AI service* (https://cloud.google.com/explainable-ai). *Google Model Card Toolkit* (https://www.tensorflow.org/responsible_ai/model_card_toolkit/guide/concepts) is a machine learning library that adds context and transparency into the development and performance of a model.

These open-source tools offer the technical support necessary to guide developers (e.g., code guides, tutorials, toolkits) and address some technical aspects of deploying trustworthy AI systems, such as fairness, robustness, and bias (Bellamy et al., 2019; Wexler et al., 2019; Simonyan et al., 2013).

To complement the abovementioned legal and technical approaches in developing, monitoring, and assessing trustworthy AI-based assistance and automation systems, we also need to take into account the socio-digital factors that deem those AI machines trustworthy. However, in this regard, the complex nature of problem has hindered how practitioners address the role of human factors and promote more sustainable, human-centric, and interdisciplinary practices in deploying trustworthy AI systems from a user's perspective (Sousa et al., 2023; Bach et al., 2022; Schmager & Sousa, 2021; Glikson & Woolley, 2020).

Thus, we argue for the need to seek new mechanisms to minimize this gap between theory, legal, and technical requirements and industrial practices to avoid further breaches of trust, both real and perceived (Sousa et al., 2024).

2.3 Socio-Technical Approach

A socio-technical approach to building and ensuring users' trust expectations advocates for a Human-Centric approach to system design. As Zieglmeier and Lehene (2021) noted, interface design can directly influence software trustworthiness.

Given the multidimensional nature of human interactions, the rise of AI has sparked societal discussions beyond technical aspects (Bach et al., 2022). AI is now deeply embedded in ethical and legal debates, raising questions about its potential benefits and risks. These concerns are culturally rooted and significant in various contexts. Therefore, novel human-centric approaches to building trustworthy systems are emerging, critically examining the integration of current and future intelligent automation systems. One such debate explores novel ways to make users trust these future systems and avoid potentially risky misuse. However, the complex nature of this issue prevents a common understanding of how to address it. If trust is not carefully examined as a socio-technical quality, it can lead to further breaches of trust, both real and perceived.

This approach is concerned with the impact of AI on people's well-being. It assesses the implications and effectiveness of using AI-based tools to perform basic daily routines or activities (Glikson & Woolley, 2020; Gillespie et al., 2023; Bach et al., 2022). It is focused on associating the system's design qualities such as integrity, predictability, familiarity, usability, and personalization, communicated through the User Interface (UI) affordances such as benevolence, credibility, and psychological safety and security to the user, while considering social structure as the key element in incorporating these aspects into the AI system UI affordances (Norman, 1988; Sousa et al., 2016).

However, addressing these socio-ethical elements, technical and design features, and user-related characteristics can be complex. The system's design functions and UI affordances are subjective qualities and can be interpreted differently depending on the application's socio-digital context and the circumstances of the user (Sousa et al., 2014). Therefore, to combine those socio-technical aspects, AI providers and practitioners must be prepared to promote and design iterative software solutions that, through collaboration between interdisciplinary teams, encourage delivering high-quality software aligned with end users' needs and the IT provider's goals.

For addressing such complex problems, scientific literature provides two frameworks to guide HCI specialists through the three phases of problem-solving: exploration, conceptualization, and evaluation (Oulasvirta & Hornbæk, 2016). Those frameworks are the Human-Centered AI Framework (HCAI) (Shneiderman, 2020) and the Human-Centered Trust Framework (HCTFrame) (Sousa et al., 2024). Both aim to guide the development of AI systems that prioritize human needs and values. Both argue that usability, user experience, and abiding by current standards and regulations are central qualities that foster more intuitive, trustworthy, and accessible AI interfaces.

The HCAI Framework, for instance, emphasizes user control and autonomy, ensuring that AI enhances rather than replaces human abilities, focusing on more ethical considerations such as bias, fairness, and privacy (Shneiderman, 2020). The HCTFrame, for instance, emphasizes the need to use a socio-technical lens of analysis to guide the AI provider and designers through the different aspects of AI software development, such as socio-ethical, organizational, and technical dimensions, and provides an easy-to-use, hands-on methodology that guides designers in

systematically measuring how end users' perceived trustworthiness can influence trustworthy design decisions (Sousa et al., 2024).

Both recognize the complex nature of AI system design that extends beyond mere technical specifications to encompass human elements, social structures, and organizational contexts. Here, designers are encouraged to establish a solid understanding of the cooperation between humans and AI machines, and validate the extent to which their design propositions influence their behavior. Both provide different assessment and validation strategies for addressing human needs and values when considering the different socio-technical lenses of analysis, such as organization, user interface, software, and hardware (Appelbaum, 1997; Bostrom & Heinen, 1977).

Such know-how and skills demand a solid foundation for user-centered design (UCD) software development practices (Bach et al., 2022; Sousa et al., 2024). It also combines novel theories and research findings with a practice focusing on realistic project work that is focused on HCI issues, and seeks solutions and action processes in less deterministic and more complex environments (i.e., Wicked problems, Buchanan, 1992).

The following paragraph further illustrates the link between the socio-technical approach and the associated need to address trust with the notions of risk according to the Artificial Intelligence Standard and the EU Act principles (Future of Life Institute, 2024).

3 Toward Incorporating Trust into System Design

Risk, according to artificial intelligence, is defined as "A function of the probability of a given threat and the potential adverse consequences of that threat's occurrence" (International Organization for Standardization, 2021). The notion of trust here relates to the aspect that end users must be assured that the AI system is trustworthy despite the potential risk.

Both notions of risk perception and trust are crucial for the human–computer trust relationship (Gulati et al., 2019; Kohn et al., 2021) as they can help to calibrate and mediate end user interactions and foster human–machine collaborations, i.e., willingness and motivation to use the system (De Visser et al., 2020; Sousa et al., 2016; Gulati et al., 2017, 2018). It is also related to the EU's efforts to guarantee that the AI system can explain its inner operating and intentions, demonstrate its competency, and abide by ethical principles. This notion of risk relates to uncertainty and end users' fears of potential breaches of trust, both real and perceived effects, which can either hinder or sustain the quality of user experience (Sousa et al., 2023).

Therefore, linking the need for users to understand the risk to the system's trustworthiness is crucial to maintaining a good user experience quality and ensuring that AI enhances people's well-being. It is as important as safety and security requirements (Abeywickrama & Ramchurn, 2024). Such communication features can be facilitated through the system's design functions, UI affordances, and its

inherent technical and socio-design subjective elements (Glikson & Woolley, 2020; Sousa et al., 2024; Gulati et al., 2019; Bach et al., 2022; Fimberg & Sousa, 2020).

In other words, regardless of the variations in the application and design of systems, examining users' trust toward AI-based automated machines becomes a relevant system feature, similar to assessing the system's usability and accessibility qualities (Kohn et al., 2021; Sousa et al., 2024).

Trust here is reflected in a person's willingness to use a particular system (i.e., trust attitude) after examining that machine's capability (competence) to perform a specific task, irrespective of its associated risks (i.e., the uncertainty and vulnerability it faces) (Lee, J.-G. et al., 2015; Mayer et al., 1995). Furthermore, an individuals' belief in whether a machine is capable of fulfilling a task as expected (i.e., trustworthiness) is directly affected by the user's characteristics, socio-ethical understandings, and system features (Bach et al., 2022; Sousa et al., 2016; Sousa et al., 2021; Sousa et al., 2023).

This section argues for using the Human–Computer Trust Scale (HCTS) as a research lens to measure a system's perceived trustworthiness. First, it describes the use of the proposed psychometric instrument as an analytical tool to avoid interactions that would make users vulnerable and lead to further breaches of trust, both real and perceived. Second, it brings three examples of how this analytical tool can inform designers and engineers about potential design pitfalls that could hinder users' trust in the system.

The HCTS follows a rigorous theory and scale development and validation procedures built upon the Human–Computer Trust Model (HCTM) initially proposed by Sousa et al. (2016). This model's novelty is that it views technology as part of a socio-technical set of interactions where the trust characteristics influence humans and technical and organizational systems.

The model builds upon previous conceptualizations proposed by Madsen and Gregor (2000), as well as trust in technology models by McKnight and Chervany (2002).

The HCTS is a psychometric instrument that is statistically validated in various application contexts (e.g., eVoting, SIRI, e-school, Internet of things (IoT), Human–robot interaction) and cross-cultural environments. It evaluates users' perceived trust through three key constructs:

- Perceived risk, which refers to the user's subjective assessment of the likelihood of encountering negative consequences when using the system (McKnight & Chervany, 2002) based on the notions of willingness and honesty (Gulati et al., 2018).
- Competence, reflecting the user's belief in the system's ability to perform the intended tasks effectively.
- Benevolence captures the user's perception of the system's intentions and whether these align with the user's best interests.

The HCTS' general trust level is based on the average score across these indicators, providing a standardized measure of user's trust in a given technology (Sousa et al., 2024; Gulati et al., 2019). These scores allow comparison across different

socio-demographic groups, which is essential for understanding trust variations among diverse user and stakeholder profiles. Additionally, it allows the comparison of the scores of the constructs and their items, adding more depth to the evaluation.

Researchers are also encouraged to use additional quantitative and qualitative measures with HCTS. Supplementary items can help researchers uncover the relationship between trust and other aspects of interaction with technology (e.g., design, operation, regulation). Questions targeting these aspects can help identify the attributes that shape individuals' trust in a system, and offer valuable insights into what should be done regarding trust calibration.

As with other standardized measures, HCTS aims to identify points of attention in the human–technology relationship that should then be further investigated. Separately, it does not explain the reasons behind the obtained trust scores. But combining it with other instruments can provide more depth to the trust results at the cost of more resources.

The following examples describe the results of studies using the HCTS to contemplate different problem-solving HCD phases (Oulasvirta & Hornbæk, 2016).

3.1 Example 1: Measuring Users' Trust in a Facial Recognition System in Mozambique

The first example illustrates the pivotal role of the HCTS as a tool in exploring the factors that influence trust in facial recognition systems for law enforcement in Mozambique. In this instance, the HCTM guided the design of a mixed-method research, which included a survey to uncover the factors affecting trust in these systems and interviews to investigate the reasons behind it.

In the study, the HCTS was part of the survey, primarily used as a tool to enable the comparison of trust levels between different socio-demographic groups in the country and, consequently, point toward paths that should be further investigated. Based on the responses to the survey and interviews (Beltrão et al., 2023), the study identified particularities of trust in facial recognition systems in the context of the country: citizens have a generally optimistic view of technology's implementation, shaped mainly by their concerns about security that the system could help mitigate. However, the country's lack of infrastructure and possibly flawed principles of use were points of major concern.

In addition, the survey identified that females and people who identify as part of minority groups had a higher propensity to trust these systems. The interviews revealed that this happened due to a feeling of greater vulnerability, resulting in a willingness to accept the (hypothetical) implementation of a technology that could, in their view, be more impartial than humans and improve their current situation.

This study sheds light on crucial aspects that need attention in the development and implementation of a critical AI system. Adopting the users' perspective revealed which aspects of technology require attention to avoid misuse due to lack of trust or

excess trust (De Visser et al., 2020), as well as which groups may be more vulnerable to issues. These steps are essential for understanding trust in technology from a socio-technical perspective, i.e., considering the social and organizational aspects of its deployment (Baxter & Sommerville, 2011). Above all, it demonstrates the complexity of trust in technology in society and the importance of contextual factors for understanding perceived trustworthiness.

3.2 Example 2: Measuring Users' Trust Contemplating the HCI Problem-Solving Phases

The second example contemplates using the HCTS to explore the differences in the formation of trust in technology in distinct national cultures, more specifically, Brazil and Singapore. The study uses partial least squares structural equation modeling (PLS-SEM) to identify the role of the HCTS constructs in trust formation in the example of different countries. PLS-SEM is commonly used to assess the validity of instruments that measure endogenous constructs and, in general, to analyze complex relationships in data (Hair et al., 2011).

In this case, PLS-SEM was used to compare the performance of the HCTS in two distinct countries, Brazil and Singapore. This comparison offers grounds to identify how cultural differences significantly affect the development of trust in technology on a more general level.

The results revealed that although the scale effectively predicted perceived trustworthiness in both places, the effect of the constructs in the result differed. This underscores the profound impact of cultural differences on trust formation, a key finding for those interested in the socio-technical aspects of technology.

These results align with the literature on national cultural differences and how they are interrelated to individuals' propensity to trust on a collective level (Hwang & Lee, 2012; Hofstede, 2011). While these results do not aim to explain cultural differences in depth, they shed light on how some aspects of technology can affect perceived trustworthiness differently due to cultural factors.

3.3 Example 3: Design Trustworthy Socio-Digital Systems

The third example demonstrates how the HCTS can contemplate the HCI problem-solving phases—conceptualization and evaluation (Oulasvirta & Hornbæk, 2016).

It uses the HCTS to explore, elicit, and conceptualize different factors that influence users' trust perceptions and illustrate how the HCTS can guide design recommendations using two different system applications—cryptocurrency and Human–Robot Interactions (HRI).

The first application study focused on exploring and eliciting the factors that can influence user interactions with cryptocurrency exchange graphical user interface platforms (Paramonova et al., 2023). The second application study validated a set of human-robot concepts on how trustworthiness factors in autonomous vehicles can influence human–machine collaborations (Pilacinski et al., 2023).

The results of the first study highlight the link between system design qualities like usability, risk mitigation, and perceived trustworthiness. However, also user satisfaction, understandability, learnability, ease of use or reputation, reinsurance behaviors, data visualization, transparency, and accountability are also taken into account. It also points out the need to further examine the specifics of UI affordances like credibility and reliability (reputation, time, recommendations, public media) that influence user trustworthiness perception alongside a user characteristic, i.e., Level of Expertise (Paramonova et al., 2023).

The results of the second application study indicate that trustworthy factors like reinsurance behavior are manifested through human–robot UI affordances such as communication, collaboration, observation, comparison, and control. Here, end users' trust might be affected by system anthropomorphisms (human likeness) affordances such as pupil size, speed, predictability, awareness, and proximity, which may be detrimental to collaborative robots depending on the application purpose (Pilacinski et al., 2023).

4 Conclusion

As the technical advances in human–machine interrelations become more complex (e.g., human–AI interrelations), it becomes harder for a non-specialist (i.e., layman) to fully understand and assess the associated risks and consequences of relying on a particular interaction with AI-automated machines. Consequently, this highlights the need to incorporate users' trust in AI-based system design. However, trustworthy human–AI interrelations can be dynamic (i.e., evolve through time) and complex (i.e., context-dependent), making its design a wicked problem. In summary, to build successful human–AI interrelations, AI system designers should consider trust as a system quality, address its multidisciplinary nature, and be equipped to combine the different socio-technical lenses of analysis. This includes identifying the socio-technical constraints, defining the design intent, and validating it with end users (Sousa et al., 2024). It also includes awareness that due to the complex and dynamic nature of the problem (i.e., wicked problem solution), there might not be a single solution for all, and for some, some risk acceptance is advised (Spreen, 2023, pp. 31–33).

Nonetheless, this socio-technical understanding is crucial in identifying potential threats and benefits to allow designers to anticipate and prevent issues that could undermine users' trust in the AI system.

Therefore, being prepared to exploit the opportunities for the benefit of citizens and society while facing the challenges raised by potential developments and

deployments of artificial intelligence should be an essential aspect of software development. Ultimately, this also implies preparing IT professionals with the necessary know-how to examine the influence and implications of users' trust behavior in the quality of the system appropriation and adoption.

The results above demonstrate the importance of using HCTS to further analyze and validate the design propositions during the three HCI problem-solving phases, as users' trust predisposition might vary depending on the application context (system use) and users' needs and goals (interactions) (Sousa et al., 2014; Fimberg & Sousa, 2020; Sousa & Bates, 2021).

Acknowledgments This study was partly funded by the Trust and Influence Program under Grant 21IOE051; European Office of Aerospace Research and Development; and U.S. Air Force Office of Scientific Research.

References

Abeywickrama, D. B., & Ramchurn, S. D. (2024). Engineering Responsible and Explainable Models in Human-Agent Collectives. *Applied Artificial Intelligence, 38*(1), 2282834. https://doi.org/10.1080/08839514.2023.2282834

Appelbaum, S. H. (1997). Socio-technical Systems Theory: An Intervention Strategy for Organizational Development. *Management Decision, 35*(6), 452–463. https://doi.org/10.1108/00251749710173823

Bach, T. A., Khan, A., Hallock, H., Beltrão, G., & Sousa, S. (2022). A Systematic Literature Review of User Trust in AI-enabled Systems: An HCI Perspective. *International Journal of Human–Computer Interaction, 40*(5), 1251–1266. https://doi.org/10.1080/10447318.2022.2138826

Baxter, G., & Sommerville, I. (2011). Socio-technical Systems: From Design Methods to Systems Engineering. *Interacting with Computers, 23*(1), 4–17. https://doi.org/10.1016/j.intcom.2010.07.003

Bellamy, R. K. E., Dey, K., Hind, M., Hoffman, S. C., Houde, S., Kannan, K., Lohia, P., Martino, J., Mehta, S., Mojsilovic, A., et al. (2019). AI Fairness 360: An Extensible Toolkit for Detecting, Understanding, and Mitigating Unwanted Algorithmic Bias. *IBM Journal of Research and Development, 63*(4–5), 4:1–4:15. https://doi.org/10.1147/JRD.2019.2942287

Beltrão, G., Sousa, S., & Lamas, D. (2023). Unmasking Trust: Examining Users' Perspectives of Facial Recognition Systems in Mozambique. In N. Jere, O. Isafiade, A. Ogunyemi, O. Anya, A. Sakpere, & D. S. Jat (Eds.), *AfriCHI'23: Proceedings of the 4th African Human Computer Interaction Conference* (pp. 38–43). Association for Computing Machinery. https://doi.org/10.1145/3628096.3628746

Bird, S., Dudík, M., Edgar, R., Horn, B., Lutz, R., Milan, V., Sameki, M., Wallach, H. M., & Walker, K. (2020, May). *Fairlearn: A Toolkit for Assessing and Improving Fairness in AI.* Microsoft, Technical Report MSR-TR-2020-32. https://www.microsoft.com/en-us/research/publication/fairlearn-a-toolkit-for-assessing-and-improving-fairness-in-ai

Bostrom, R. P., & Heinen, J. S. (1977). MIS Problems and Failures: A Socio-technical Perspective, Part II: The Application of Socio-technical Theory. *MIS Quarterly, 1*(4), 11–28. https://doi.org/10.2307/249019

Buchanan, R. (1992). Wicked Problems in Design Thinking. *Design Issues, 8*(2), 5–21. https://doi.org/10.2307/1511637

Cannarsa, M. (2021). Ethics Guidelines for Trustworthy AI. In L. A. DiMatteo, A. Janssen, P. Ortolani, F. de Elizalde, M. Cannarsa, & M. Durovic (Eds.), *The Cambridge Handbook*

of Lawyering in the Digital Age (pp. 283–297). Cambridge University Press. https://doi.org/10.1017/9781108936040.022

De Visser, E. J., Peeters, M. M., Jung, M. F., Kohn, S., Shaw, T. H., Pak, R., & Neerincx, M. A. (2020). Towards a Theory of Longitudinal Trust Calibration in Human–Robot Teams. *International Journal of Social Robotics, 12*(2), 459–478. https://doi.org/10.1007/s12369-019-00596-x

Dong, Y. (2010). The Role of Trust in Social Life. In Z. Yan (Ed.), *Trust Modeling and Management in Digital Environments: From Social Concept to System Development* (pp. 421–440). IGI Global. https://doi.org/10.4018/978-1-61520-682-7.ch017

European Commission. (2019, April 08). *Ethics Guidelines for Trustworthy AI (Report/Study)*. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

European Commission. (2020, July 17). *Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self-assessment (Report/Study)*. https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment

Fimberg, K., & Sousa, S. (2020). The Impact of Website Design on Users' Trust Perceptions. In E. Markopoulos, R. S. Goonetilleke, A. G. Ho, & Y. Luximon (Eds.), *Advances in Creativity, Innovation, Entrepreneurship and Communication of Design. Proceedings of the AHFE 2020 Virtual Conferences on Creativity, Innovation and Entrepreneurship, and Human Factors in Communication of Design, July 16–20, 2020, USA. Advances in Intelligent Systems and Computing*. *Advances in Intelligent Systems and Computing, 1218* (pp. 267–274). Springer. https://doi.org/10.1007/978-3-030-51626-0_34

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People – An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines, 28*, 689–707. https://doi.org/10.1007/s11023-018-9482-5

Floridi, L., Holweg, M., Taddeo, M., Amaya Silva, J., Mokander, J., & Wen, Y. (2022, March 23). capAI—A Procedure for Conducting Conformity Assessment of AI Systems in Line with the EU Artificial Intelligence Act. *SSRN*. https://doi.org/10.2139/ssrn.4064091

Future of Life Institute. (2024). *The EU Artificial Intelligence Act. Up-to-date Developments and Analyses of the EU AI Act*. https://artificialintelligenceact.eu

Gillespie, N., Anesa, M., Lizzio-Wilson, M., Chapman, C., Healy, K., & Hornsey, M. (2023). How do Sector Level Factors Influence Trust Violations in Not-for-profit Organizations? A Multilevel Model. *Journal of Business Ethics*. https://doi.org/10.1007/s10551-023-05429-6

Glikson, E., & Woolley, A. W. (2020). Human Trust in Artificial Intelligence: Review of Empirical Research. *Academy of Management Annals, 14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Gulati, S., Sousa, S., & Lamas, D. (2017). Modelling Trust: An Empirical Assessment. In *16th IFIP TC 13 International Conference on Human-Computer Interaction (INTERACT 2017)* (Vol. 10516, pp. 40–61). Springer. https://doi.org/10.1007/978-3-319-68059-0_3

Gulati, S., Sousa, S., & Lamas, D. (2018). Modelling Trust in Human-like Technologies. In *IndiaHCI'18: Proceedings of the 9th Indian conference on human-computer interaction*, pp. 1–10. https://doi.org/10.1145/3297121.3297124

Gulati, S., Sousa, S., & Lamas, D. (2019). Design, Development and Evaluation of a Human-Computer Trust Scale. *Behaviour & Information Technology, 38*(10), 1004–1015. https://doi.org/10.1080/0144929X.2019.1656779

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a Silver Bullet. *Journal of Marketing Theory and Practice, 19*(2), 139–152. https://doi.org/10.2753/MTP1069-6679190202

Hofstede, G. (2011). Dimensionalizing Cultures: The Hofstede Model in Context. *Online Readings in Psychology and Culture, 2*(1), 2307–0919. https://doi.org/10.9707/2307-0919.1014

Hwang, Y., & Lee, K. C. (2012). Investigating the Moderating Role of Uncertainty Avoidance Cultural Values on Multidimensional Online Trust. *Information & Management, 49*(3–4), 171–176. https://doi.org/10.1016/j.im.2012.02.003

International Organization for Standardization. (2021). *Information Technology–Artificial Intelligence–Overview of Trustworthiness in Artificial Intelligence Information Technology (ISO/IEC Standard No. TR 24028:2020)* (ISO/IEC Standard No. TR 24028:2020). https://www.iso.org/standard/77608.html

Kohn, S. C., De Visser, E. J., Wiese, E., Lee, Y.-C., & Shaw, T. H. (2021). Measurement of Trust in Automation: A Narrative Review and Reference Guide. *Frontiers in Psychology, 12*, 604977. https://doi.org/10.3389/fpsyg.2021.604977

Lee, D., Moon, J., Kim, Y. J., & Mun, Y. Y. (2015). Antecedents and Consequences of Mobile Phone Usability: Linking Simplicity and Interactivity to Satisfaction, Trust, and Brand Loyalty. *Information & Management, 52*(3), 295–304. https://doi.org/10.1016/j.im.2014.12.001

Lee, J.-G., Kim, K. J., Lee, S., & Shin, D.-H. (2015). Can autonomous Vehicles be Safe and Trustworthy? Effects of Appearance and Autonomy of Unmanned Driving Systems. *International Journal of Human-Computer Interaction, 31*(10), 682–691. https://doi.org/10.1080/10447318.2015.1070547

Madiega, T. (2021). *Artificial Intelligence Act (Briefing, EU Legislation in Progress)*. European Parliament: EPRS European Parliamentary Research Service. https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf

Madsen, M., & Gregor, S. (2000). Measuring Human-Computer Trust. In *11th Australasian Conference on Information Systems*, Vol. 53, pp. 6–8.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model of Organizational Trust trust. *The Academy of Management Review, 20*(3), 709–734. https://doi.org/10.2307/258792

McKnight, N. D., & Chervany, N. L. (2002). Trust and distrust definitions: One bite at a time. In R. Falcone, M. P. Singh, & Y. Tan (Eds.), *Trust in Cyber-societies: Integrating the Human and Artificial Perspectives (LNCS)* (Vol. 2246, pp. 27–54). Springer. https://doi.org/10.1007/3-540-45547-7_3

Norman, D. A. (1988). *The Psychology of Everyday Things*. Basic books.

Oulasvirta, A., & Hornbæk, K. (2016). HCI Research as Problem-Solving. In *CHI'16: Proceedings of the 2016 CHI Conference on Human factors in Computing Systems*, pp. 4956–4967. https://doi.org/10.1145/2858036.2858283

Paramonova, I., Sousa, S., & Lamas, D. (2023). Heuristics to Design Trustworthy Technologies: Study Design and Current Progress. In J. A. Nocera, M. K. Lárusdóttir, H. Petrie, A. Piccinno, & M. Winckler (Eds.), *Human-computer Interaction—INTERACT 2023. 19th IFIP TC13 International Conference, York, UK, August 28–September 1, 2023, Proceedings, Part IV IFIP Conference on Human-computer Interaction* (pp. 491–495). Springer. https://doi.org/10.1007/978-3-031-42293-5_60

Pilacinski, A., Pinto, A., Oliveira, S., Araujo, E., Carvalho, C., Silva, P. A., Matias, R., Menezes, P., & Sousa, S. (2023, January 27). The Robot Eyes Don't Have It. The Presence of Eye Gaze on Collaborative Robots Yields Marginally Higher Human Trust but Lower Performance. The Presence of Eye Gaze on Collaborative Robots Yields Marginally Higher Human Trust but Lower Performance. *SSRN*. https://doi.org/10.2139/ssrn.4339644

Rossi, F. (2018). Building Trust in Artificial Intelligence. *Journal of International Affairs, 72*(1), 127–134. https://www.jstor.org/stable/26588348

Schmager, S., & Sousa, S. (2021). A Toolkit to Enable the Design of Trustworthy AI. In C. Stephanidis, M. Kurosu, J. Y. C. Chen, G. Fragomeni, N. Streitz, S. Konomi, H. Degen, & S. Ntoa (Eds.), *HCI International 2021 - Late Breaking Papers: Multimodality, eXtended Reality, and Artificial Intelligence. 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings* (pp. 536–555). Springer. https://doi.org/10.1007/978-3-030-90963-5_41

Shneiderman, B. (2020). Human-centered Artificial Intelligence: Reliable, Safe & Trustworthy. *International Journal of Human–Computer Interaction, 36*(6), 495–504. https://doi.org/10.1080/10447318.2020.1741118

Simonyan, K., Vedaldi, A., & Zisserman, A. (2013). Deep inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. *arXiv preprint*, arXiv:1312.6034. https://doi.org/10.48550/arXiv.1312.6034

Sousa, S., & Bates, N. (2021). Factors Influencing Content Credibility in Facebook's News Feed: Inside View on the United Kingdom (UK) Post-Brexit. *Human-Intelligent Systems Integration, 3*, 69–78. https://doi.org/10.1007/s42454-021-00029-z

Sousa, S., Smorgun, I., Lamas, D., & Arakelyan, A. (2014). A Design Space for Trust-Enabling Interaction Design. In *MIDI'14: Proceedings of the 2014 Multimedia, Interaction, Design and Innovation International Conference On Multimedia, Interaction, Design and Innovation* (pp. 1–8). Association for Computing Machinery. https://doi.org/10.1145/2643572.2643578

Sousa, S., Lamas, D., & Dias, P. (2016). Value Creation through Trust in Technological-mediated Social Participation. *Technology, Innovation and Education, 2*(1–10), 5. https://doi.org/10.1186/s40660-016-0011-7

Sousa, S., Cravino, J., Lamas, D., & Martins, P. (2021). Confiança e tecnologia: práticas, conceitos e ferramentas. *RISTI – Revista Ibérica de Sistemas e Tecnologias de Informação, 10*(E45), 146–164.

Sousa, S., Cravino, J., & Martins, P. (2023). Challenges and Trends in User Trust Discourse in AI Popularity. *Multimodal Technologies and Interaction, 7*(2), 13. https://doi.org/10.3390/mti7020013

Sousa, S., Lamas, D., Cravino, J., & Martins, P. (2024). Human-centered Trustworthy Framework: A Human–computer Interaction Perspective. *Computer, 57*(3), 46–58. https://doi.ieeecomputersociety.org/10.1109/MC.2023.3287563

Spreen, D. (2023). Lethal autonomous weapon systems (LAWS). On the Ethics of Automation in the Military from the Perspective of Social Systems Theory. *Sõjateadlane (Estonian Journal of Military Studies)*, (21), 10–40. https://doi.org/10.15157/st.vi21.24177

Wagner, G. (1994). Vertrauen in Technik. *Zeitschrift für Soziologie, 23*(2), 145–157. https://doi.org/10.1515/zfsoz-1994-0205

Wexler, J., Pushkarna, M., Bolukbasi, T., Wattenberg, M., Viegas, F., & Wilson, J. (2019). The What-If Tool: Interactive Probing of Machine Learning Models. *IEEE Transactions on Visualization and Computer Graphics, 26*(1), 56–65. https://doi.org/10.1109/TVCG.2019.2934619

Zicari, R. V., Brodersen, J., Brusseau, J., Düdder, B., Eichhorn, T., Ivanov, T., Kararigas, G., Kringen, P., McCullough, M., Möslein, F., Mushtaq, N., Roig, G., Stürtz, N., Tolle, K., Tithi, J. J., van Halem, I., & Westerlund, M. (2021). Z-inspection®: A Process to Assess Trustworthy AI. *IEEE Transactions on Technology and Society, 2*(2), 83–97. https://doi.org/10.1109/TTS.2021.3066209

Zieglmeier, V., & Lehene, A. M. (2021). Designing Trustworthy User Interfaces. In *OzCHI'21: Proceedings of the 33rd Australian Conference on Human-Computer Interaction*, pp. 182–189. https://doi.org/10.1145/3520495.3520525

Zuboff, S. (2015). Big Other: Surveillance Capitalism and the Prospects of an Information Civilization. *Journal of Information Technology, 30*(1), 75–89. https://doi.org/10.1057/jit.2015.5

Challenges of Communicating Military Innovation

Natascha Zowislo-Grünewald and Franz Beitzinger

Abstract Military innovations are a special case of technological developments that affect societal subfields or society as a whole. They can have unintended consequences on other areas and are therefore viewed critically. The erosion of optimism about progress in modern societies has led to the fact that technical innovations are no longer seen uncritically and must legitimize themselves. The creation of consensus on risks is a crucial goal of innovation communication. Legitimizing scientific-technical developments cannot rely solely on scientific arguments, but must also consider the credibility of communication and the trustworthiness of the actors involved. Military innovations have a chance of acceptance if the logics of distribution of benefits and risks are seen as compatible with each other. This is particularly relevant for autonomous systems and AI. The risk perception of military innovations is strongly determined by uncertainties and subjective factors. Innovation communication must take into account the social, meaning-giving context and anticipate the expectations of the addressees. In the internal relationship, this means offering internal stakeholders a solution to the perceived goal-means discrepancies. Finally, innovation communication is not only a form of change communication but also a communication about the contexts in which innovations are created and used.

Communicating military innovation poses a significant challenge for organizational communication. In this essay, we will explore the difficulties that public relations and public affairs face in this field, as well as potential solutions.

Firstly, we will delve into the reasons why innovation generally represents a communications problem, followed by an analysis of the particular communication obstacles encountered within this realm. We will then apply the concept of

N. Zowislo-Grünewald (✉) · F. Beitzinger
Institute for Organizational Communication, University of the Bundeswehr Munich (UniBwM), Neubiberg, Germany
e-mail: natascha.zowislo@unibw.de; franz.beitzinger@unibw.de;
https://strategic-communication-management.de

K. Talves, D. Spreen (eds.), *Artificial Intelligence in Military Technology*,
Artificial Intelligence, Simulation and Society 192,
https://doi.org/10.1007/978-3-031-95578-5_12

innovation to the military context, considering the criteria for social acceptance of innovations and the unique challenges that military innovations present. The evaluation of associated risks is crucial to understanding why innovation is a communications problem, and we will explore how technological advancements can impact society as a whole. Furthermore, we will discuss how technological innovations encompass both a material, technical context and a social, semantic dimension, highlighting the complex nature of the communications problem. This complexity requires a nuanced approach to solving the issue, which we will explore in detail. Finally, we will examine the practical consequences of military organizations communicating about military innovation and how effective communication can foster greater acceptance and adoption of new technologies.

1 Innovation: A Communications Problem per se

An innovation is not merely a 'new invention' or something new; rather, it involves a dynamic process that requires acceptance and spreading to be considered an innovation. This concept can be traced back to Gabriel Tarde's sociological theory of change (1895), which posits that innovation is driven by processes of imitation. In line with this idea, Joseph Schumpeter distinguishes between mere invention and innovation, with the latter referring to "new combinations" that prevail in the market (Schumpeter, 1987, pp. 100–101, 128–129). The process of spreading innovation is commonly depicted in a scheme that includes the stages of invention, imitation, and diffusion, highlighting the importance of market success in determining whether an innovation has truly established itself.

An innovation is therefore not merely something new, but a novelty that gains acceptance and spreads: "To constitute an innovation, any or all of these must occur on a relatively large scale. Innovation is action and the results of action lead to repeated and widespread action" (Deutsch, 1985, p. 20). In essence, an innovation attains prevalence within a market, i.e., in public opinion, due to its perceived improvement. The location of the communication challenge in this context warrants examination. When viewed from the standpoint of an innovator seeking propagation, the predicament often resides in the sluggishness of the diffusion process for their innovation, prompting a desire for expedited promotion. Nevertheless, a closer examination reveals that the issue runs deeper, as innovation can have unintended and undesirable consequences on other fields, known as external effects in economics. These costs arise from the adaptation of innovation by other market participants who do not benefit from the general benefits of the innovation.

When the net benefits derived from an innovation are clearly much higher at the economic or societal level than the individual costs they cause, innovation can legitimately claim some basis for its existence. However, this is only possible if both the concept of progress itself is appreciated by society and the kind of 'progress' brought about by the specific innovation is able to gain widespread consensus within society.

In the field of economics, the notion that enhanced value and well-being stem from economic advancement is often embraced as a hallmark of a broader 'consensus on progress.' Yet, the complexity of modern societies transcends mere economic considerations. Consequently, consensus on progress outside the economic realm is less prevalent. However, it is crucial not to conflate this observation with the palpable decline in confidence regarding the continued proliferation and fortification of the Western liberal democratic societal model. This model faces mounting challenges from emerging ideological competitors, adding new dimensions to the ongoing discourse on societal evolution. The Russian attack on Ukraine is just one of several factors that calls into question this civilizational and political optimism about progress (Reckwitz, 2022). This decline in confidence in the Western societal model however is embedded in the erosion of the much more powerful scientific-technical optimism concerning progress that has characterized Western societies since the Age of Enlightenment.

As an indicator of this erosion, one might consider the changing reputation of the engineering profession. In the Western societies during the 1960s, engineers were still held in high esteem as guarantors of growth, prosperity, and peace. However, in recent years, they have been subject to increasing criticism in the discourse on technical topics, viewed as a threat to social cohesion and even natural foundations of life (Kunze, 2021). This shift in perception has led to a general suspicion of technical progress. This in turn is closely intertwined with the emergence of ecological thinking during this period, yet it represents a distinct development rather than a mere extension of ecological ideologies. Instead, it signifies a broader shift in societal attitudes and norms.

This shift is insightfully captured by Ulrich Beck's work on the "risk society" (1992), where he describes how technical progress is no longer met with unconditional welcomes and a "better future" is no longer equated with it. For Beck, the development of industrial society has led to the emergence of societal risks that call into question the ideological foundations of industrial society, including the general consensus on progress (Beck, 1992, pp. 200–203). The erosion of these innovation-friendly conditions—the disappearance of a universal optimism concerning progress—has transformed industrial society into a risk society. Such a society is confronted with risks it has created itself through technical and scientific progress. Thus, modern societies become reflexive, self-referential (Beck, 1992, pp. 19–20). In this 'postmodern' society, the modernization process becomes an issue, leading to a fundamental critique of civilization and a profound need for legitimacy.

According to Beck, such risks, which can arise from the use of new technologies, have also affected previous societies. However, in an industrial society, the logic of wealth production and distribution dominates; it does not expect a conflict with the distribution of risks. In contrast, a risk society is characterized by conflicts between the logics of distributing wealth and the risks that come with it (Beck, 1992, pp. 153–154), delegitimizing any progress perceived as having negative external effects. This leads to a politicization of risks and puts pressure on innovation-related actions to legitimize the risks being taken. Technical innovations therefore become a communications problem.

Beck stresses the socially constructed nature of risks, which leads him to argue that risk determinations are shaped by both—interest and fact (Beck, 1992, pp. 28–32). This perspective highlights the importance of values in risk assessments, resulting in diverse and often competing risk definitions. In a risk society, where information and science play a crucial role, different interpretations of risks fight for dominance in some sort of definitional battles (Beck, 1992, pp. 46–50).

Building on Ulrich Beck's insights, it becomes apparent that engaging in a discourse centered on risk is imperative when addressing innovations. The legitimacy of embracing risks holds particular significance, particularly within the realm of technologies. The center of gravity thus pertains to the perceived risks linked with innovations and changes. Attaining consensus on these risks stands as a pivotal objective in the communication of innovation. However, scientific arguments alone are insufficient in shaping risk perceptions. As Beck aptly notes, due to the pervasive scientification of contemporary society, science has paradoxically lost its monopoly on truth. Instead, competing scientific (and non-scientific) arguments vie for interpretative power, and the outcome of sense-making struggles is shaped by the level of credibility attributed to those advocating for them (Beck, 1992, pp. 163–169). Consequently, in the realm of innovation communications, the pertinent question arises: Which elements of belief structures can effectively be linked through communication?

2 Communicating Innovation: Why and How?

In this essay, innovations are conceptualized as either novel inventions or improvements that garner adoption due to their inherent quality, consequently achieving widespread acceptance. Such advancements may emerge from the reconfiguration of existing elements or from entirely original concepts. Contrary to common misconceptions, the sociological and economic underpinnings of innovation differentiate it from mere novelty or invention. The determination of whether something qualifies as an innovation is retrospective, contingent upon its ability to generate significant benefits. However, the assessment of benefit is inherently subjective, with potential users and adopters wielding the power to accept or disregard the innovation. Consequently, the central figure in the realm of innovation is not the originator of the novel idea, but rather the individual who discerns its potential and cultivates conditions conducive to its adoption, as posited by Joseph Schumpeter (1987) in his theory of innovation. This individual embodies the archetype of the entrepreneurial spirit.

The transformative potential of an innovation, be it a discovery, recombination, or original creation, hinges not on its intrinsic characteristics alone, but rather on the capacity of a creative entrepreneur to perceive its commercial viability. In economic terms, it is the demand side that ultimately adjudicates the worth of an innovation, based on a subjective evaluation of its added value. Whether a genuine opportunity for innovation truly exists becomes a secondary concern. Convincing potential users

or adopters of an innovation's significance becomes paramount; hence, innovation can be conceived as a process of sense-making (Svetlova, 2008).

The process of asserting meaning to something new is what transforms it into an innovation. This occurs within the social and cultural space, where established norms, values, traditions, routines, motives, interests, and other factors shape the interpretation of novelty. The competition among different patterns of interpretation within this discursive environment can lead to the successful assignment of meaning and the creation of a basis for the diffusion of innovation. Conversely, if the ascription of meaning is unsuccessful or lost, the innovation is likely to fail.

The findings of diffusion research unequivocally highlight the critical role played by user demand in the dissemination of ideas and innovation. According to Everett M. Rogers (2003), the adoption process can be delineated into distinct stages. Initially, individuals become aware of the existence of an innovation, followed by a process of conviction regarding its utility or benefit. This is accompanied by a decision to adopt or reject the innovation, and if adopted, it is integrated into one's daily life and structures. If the benefit is confirmed, the innovation will be retained; otherwise, it will be discarded (Rogers, 2003, pp. 169–218). This suggests that sense-making processes may not always follow a linear progression; instead, previously attributed meanings can be revoked, indicating that potential innovations can lose their significance.

Rogers highlights the communicative nature of the adoption process, emphasizing that diffusion "is the process by which an innovation is communicated through certain channels over time among the members of a social system" (Rogers, 2003, p. 5). This suggests that each stage of the adoption process involves communication, and innovation communications can impact the discourse in which sense-making (or denial) processes occur for potential innovations. As mentioned earlier, the relevance issue concerns which 'belief components'—such as convictions, opinions, attitudes—innovation communications can connect to.

To address this issue, the relative context of the innovation in question must be established. This essay has thus far focused on innovation in general, but the challenges of innovation communication in military contexts are of particular interest.

3 Innovation and the Military

At its core, the difference between military and civilian innovations is not as distinct as one might think, particularly in the realm of process innovations that must be implemented within and through an organization. This holds true for both military and civilian organizations, such as private businesses. In both cases, an innovation can be a product or process modification that must be implemented within the organization. A prime example of this is the implementation of logistics in the U.S. military during World War II. Its usefulness quickly became apparent to participants, leading to rapid adoption. Admiral Ernest J. King, commander of the U.S. Navy's Atlantic Fleet during World War II, famously remarked: "I don't know what the hell

this 'logistics' is that [Army Chief of Staff George C.] Marshall is always talking about, but I want some of it." This underscores that process innovations in military contexts are comparable to those encountered in civilian settings.

In the case of product innovations, the situation is somewhat different, as in the military context these are usually related to weapon technologies. The adoption of product innovations in military contexts is not solely dependent on the acceptance of those involved in the armed forces, but also on social acceptance of these means. This is particularly relevant when considering the potential destructive power of weapons technology or the consequences of its use. Comparing the acceptance of such technologies with that of large technologies in the civilian sector can provide valuable insights.

In principle, the acceptance of controversial technologies such as genetic engineering or nuclear power is assessed by weighing their benefits and risks. Fundamental opposition to the use of techniques or technologies is very rare. The German society, for example, has generally exhibited a high level of technology acceptance across various domains, including everyday consumer products and workplace tools (Hennen, 2002, pp. 2–3). However, larger-scale technologies that directly impact the immediate living environment, like those in the neighborhood or posing potential danger to individuals, are more difficult to accept. These decisions are typically made by the political field through interactions between the state, businesses, and the public. Typically, the decision to implement such technology, e.g., in the form of the construction of large-scale technical facilities, is made by the political field through an interplay between the state, businesses, and the public. This often leads to controversies that not only relate to the technology, but also to various political, social, or normative models. According to Michael Ortiz, controversies surrounding these technologies often involve concerns about loss of control, doubts about the ability of politics and experts to manage them effectively, ethical dimensions, and questions about the distribution of costs and benefits (Ortiz, 2021, p. 10).

Empirical findings on German citizens' acceptance of technology align well with corresponding research from other Western countries, making this example suitable for generalization. One crucial aspect of the acceptance of large-scale or risky technologies, such as weaponization technologies, is the perception of controllability of associated risks. When the balance between benefits and risks (including potential harm and control) is viewed positively, technology is readily accepted. This can be illustrated by comparing the German society's views on genetic engineering in agriculture and food production versus its use in medicine. In the former case, perceived risks to human health and the environment outweigh supposed benefits, while in the latter case, e.g., to treat genetic diseases, the balance tilts decidedly in favor of acceptance (Hennen, 2002, pp. 51–62).

Extended to the acceptance of military or military technology innovations, this means that the benefit conferred by such technologies must clearly outweigh any potential risks. These include ethical risks, such as conflicts with ethically based norms and worldviews, costs risks, risks associated with unintended effects, and the risk of losing control over the technology, such as conflicts escalating beyond control or the technology posing a threat to one's own civilian population. One area of

technology where these impacts are clearly visible in public discourse is autonomous systems, encompassing the broader thematic realm of artificial intelligence.

Military (technical) innovations stand a favorable chance of garnering acceptance and imbued significance when the alignment of benefit and risk distribution is perceived as congruent. As articulated by Ulrich Beck, this congruence entails a prevailing consensus on the notion of progress. This assertion holds particularly true when such technological advancements demonstrably bolster security measures in credible manners while concurrently safeguarding against the risk of unintended consequences, specifically the potential for deployment against one's own civilian populace.

4 Context Is King

The perception of military innovations' risks is largely shaped by uncertainties, rather than established knowledge. Subjective factors significantly influence the assessment of risks related to these innovations. These subjective evaluations are influenced by personal perspectives on the technologies or innovations themselves, as well as the actors associated with them. The evaluation process is value-laden and based on both knowledgeable and unknowledgeable assessments of the objectively verifiable risks and opportunities presented by the innovation or technology. Moreover, the credibility or trustworthiness of the actors involved in the innovation process can also impact the meta-level evaluation of context.

When knowledge or ignorance significantly influences the consensus on risks related to innovations, it may seem natural to assume that knowledge dissemination could address the communications challenge. However, this assumption is not well-founded. Subjectively, one's perception of the actors associated with technological or innovative developments (such as politics, science, business, or military) is incorporated into the evaluation of 'knowledge.' Trustworthiness and credibility are also considerations at this level. Moreover, the credibility of communication at the substantive level is also a crucial factor. As noted by Ulrich Beck (1992, pp. 163–169), scientific arguments often compete with one another in the discourse on risk, challenging the dominance of scientific truth in modern society. Ultimately, contradictory scientific arguments often contest for validity in risk discourses, and the credibility of both the argument and counterargument determines the outcome of these interpretive battles.

Moreover, scientific knowledge is not an independent entity, but rather a social construct that is shaped by its embedding in specific contexts. As such, it is only relative to the social circumstances in which it is produced. Bruno Latour's anthropological investigations at the Salk Institute in California (Latour & Woolgar, 1986) have highlighted the constructive and social nature of scientific knowledge, which is influenced by factors such as laboratory conditions, publication obligations, and scientific exchange. In this sense, scientific facts are not discovered, but rather constructed through a social process. While social construction does not imply that

scientific knowledge is arbitrary or detached from reality, it emphasizes that the nature of scientific truth is the result of scientific work rather than an inherent property of the natural world. As Latour and Woolgar note, "we do not wish to say that facts do not exist nor that there is no such thing as reality. In this simple sense our position is not relativist. Our point is that 'out-there-ness' is the consequence of scientific work rather than its cause" (Latour & Woolgar, 1986, pp. 180–182).

Latour challenges here the notion of a dichotomy between technology and human beings. For him, technology is inextricably linked to humanity; in other words, technology cannot exist without humans, nor can humans exist without technology. Additionally, both humans and technology form networks with technical artifacts. In essence, Latour's Actor-Network-Theory posits that social phenomena arise from the complex web of relationships between human actors and non-human actors (including objects, natural phenomena, structures, etc.), which are both materially and meaningfully (semiotically) interconnected.

To illustrate his point, Latour uses the example of a person using a gun. He contrasts two perspectives regarding this scenario: one that views a gun as an extension of the human actor, as a simple tool, and another that sees the gun as an autonomous force that transforms the user into a killing subject. However, Latour argues that both perspectives oversimplify the situation, as they fail to account for the synthesis of human and technical components involved in the act of using a gun. Instead, Latour posits that when a person uses a gun, something new is created: an actor–device network comprising both human and technical elements. This hybrid entity, or actant, is shaped by both human intentions and technological capabilities and gives rise to novel forms of agency and action. In the example provided, this could be described as a citizen-weapon or weapon-citizen. According to Latour, it is the actant—this third entity created by the synthesis of human and technical components—who acts. As a result of this hybridization, the person wielding the gun undergoes a transformation as well: they become a different person, and the gun gains a new quality, manifesting as a killing instrument rather than a tool for defense or protection. In other words, the brave citizen becomes a criminal, and the gun itself becomes an instrument of killing (Latour, 2000, pp. 213–216).

In addition to the technical aspects of scientific inquiry, Latour highlights the crucial role of such networks of human actors and their relationships with the material world in shaping scientific knowledge and technological innovation. The strict separation of subject and object or mind and nature is an oversimplification of the complex dynamic interactions that occur during scientific work (Latour, 2000, pp. 177–185). Instead, scientific knowledge and innovation arise from a web of human and material interactions that co-constitute one another. This means that facts are not fixed or objective, but rather are created and negotiated through the ongoing dialogue between researchers, their tools and techniques, and the objects of their study. Moreover, technology and innovation are not just the result of technical expertise, but also reflect the social and cultural contexts in which they are developed. The material and semantic aspects of scientific work are intertwined, such as new meanings and interpretations arise from the ongoing interaction between researchers and their objects of study. Therefore, it is not possible to fully

understand scientific knowledge or technological innovation without considering both the technical and social dimensions of these processes.

In his work, Latour seeks to reconcile (radical) constructivist and (radical) materialist perspectives on technology and innovation. Nevertheless, it is worthwhile to adopt a more social constructivist perspective in order to specifically highlight the role of contexts in the genesis of technologies or in the emergence of innovations. By examining the invention of the pneumatic bicycle tire, Trevor J. Pinch and Wiebe E. Bijker (1984) provide here a compelling example of how technological change or the diffusion of innovations is shaped by the social and cultural contexts in which it occurs. This case study highlights the role of human actors and their relationships with the material world in shaping scientific knowledge and technological innovation. Moreover, this perspective emphasizes that technological change is not solely determined by technical considerations, but also reflects the negotiation and interpretation of meaning among various stakeholders.

In the initial stages of this constructivist narrative, John Boyd Dunlop invented the pneumatic tire to mitigate the issue of vibrations encountered during bicycling, particularly on the prevalent penny-farthings of the era. The pneumatic tire marked a notable departure from the rigid, unpadded tires that prevailed at the time. Despite its innovative design, widespread adoption of the pneumatic tire did not materialize immediately. This can be attributed, in part, to the constructivist perspective, which posits that technological artifacts are inherently susceptible to varied interpretations and possess a degree of 'semantic malleability.'

While Dunlop conceived the pneumatic tire with specific objectives aimed at enhancing bicycling experience, the social actors engaged in its utilization—namely: bicyclists—formulated their own understandings of the technology, thereby imbuing the pneumatic tire with divergent socially constructed attributes. In their analysis of this social construction of technology, Bijker (1997, pp. 75–77) characterizes penny-farthings as emblematic of a "macho" biking culture, embodied by riders exuding a corresponding attitude. Bicycling, therefore, was predominantly perceived as a competitive pursuit, with comfort-enhancing technologies such as the pneumatic tire deemed antithetical to the quest for adrenaline-fueled thrills and the gain of distinction. Consequently, the pneumatic tire failed to gain significant traction in the marketplace, its potential hindered by the prevailing social constructs within the bicycling community.

The pneumatic tire's fortunes changed when it was reinterpreted and given new, similarly socially constructed properties. This shift occurred when bicycle races were won handily by those using such pneumatic tires, leading to a change in the perception of the technology. The comfort artifact of the pneumatic tire was thus reinterpreted into a functional, rational, sporty one, making it interesting again for the target group (Pinch & Bijker, 1984, pp. 427–429).

Different contexts can significantly impact the perception of technology and innovation, as every perception is subjective. The acceptance of new technology is often the result of social construction processes, in which the technical properties of the innovation play a role but are not the only factor. Contexts can be designed, interpreted, and reinterpreted, highlighting the importance of considering the social and cultural context in which technologies are developed and used.

5 Inside Organizations: Innovation as Deviant Behavior

The practical consequence of this is that innovation communications cannot be limited to simply conveying 'information.' Instead, they must take into account the social, meaningful context in which technologies are developed and used. This entails anticipating the expectations of one's audience, considering a range of factors such as strategic, tactical, long-term, short-term, economic, environmental, ethical factors, as well as blind spots and cognitive schemata. However, whose costs and benefits are at stake here, and for whom do innovations represent opportunities or risks?

In the context of achieving consensus regarding risks, it is important to emphasize a broad societal perspective. However, what are the stakes for society at large in terms of military innovation? The central aspects for the acceptance of such technologies have already been identified above: Military innovations must fulfill their purpose of ensuring security while maintaining a favorable balance of benefits and costs. Moreover, they must remain controllable to prevent harm to one's own civilian population. Anticipating the fear of loss of control is especially important in the context of autonomous systems and artificial intelligence, as these technologies can pose significant risks if not properly managed.

Both external and internal acceptance are crucial for fostering innovation, highlighting the dual importance of considering both dimensions. Organizations must address the concerns of internal stakeholders when integrating novel changes, as innovation communications inherently involve some form of change management (Zowislo-Grünewald & Beitzinger, 2021, pp. 110–117, 129–152). Similarly, successful implementation of military innovation within and through military organizations depends on acceptance from both internal and external parties. Without sufficient acceptance, resistance to change can arise, potentially disrupting the innovation process and causing failures.

Innovation fundamentally represents a departure from existing paradigms. Drawing from Robert K. Merton's anomie theory (Merton, 1938), innovation is conceptualized as a form of action aimed at resolving discrepancies between goals and means. Here, discrepancy denotes the inability of acting agents to attain desired objectives using available resources. Merton's framework, originally applied to social action and societal goals, can be extrapolated to the domain of innovation implementation. In this context, organizational actors embrace innovation as it offers improved means of achieving objectives compared to existing methodologies.

If one assumes, in the manner of a thought experiment, that innovative action viewed as deviant behavior is an expression of a goal-means discrepancy, Merton's (1938) approach can be used to identify further forms of deviant behavior that an organization would have to expect from its members when they implement this innovation. These types of actions also represent the great challenges of any change management process.

Innovation as action is just one way in which actors can respond to a goal-means discrepancy. Another adaptation response is conformist action, where actors exclusively set goals that can be achieved with the resources at hand. In the context of military innovation, this could lead to an acceptance of a lower level of military capability while still avoiding innovation. Another adaptation response is ritualism,

or over-conformity, where actors use whatever resources are already available, regardless of their suitability for achieving their goals. Translated back into the military innovations' context, this would involve rejecting new technologies because one does not want to give up on what already exists.

Another action type is retreat, where the perceived goal-means discrepancy results in a kind of abandonment (or passive rejection) of all goals and means. This could also be described as 'quiet quitting' of the organization's members or employees. Translated back into the present context, this means giving up entirely on trying to achieve the organization's goals. Rebellion is the opposite of this action type, where the acting agent responds to the perceived goal-means disparity by actively rejecting both the goals and means. In the context of military innovation, this could be described as active opposition to the innovation with the aim of creating something entirely new, but different from the proposed innovation.

Any organizational change management strategy must consider the various actions of actors within the organization who may be skeptical, resistant, or oppositional to the change. These actions can range from slowing down the change process to actively working against it. To successfully implement innovations within and throughout the organization, effective communication is crucial. Internally, this involves addressing perceived goal-means discrepancies by offering a solution that aligns with the possibilities of innovation. This complements the perspective of external innovation communication, which involves communicating the benefits and costs of innovation in all relevant contexts (e.g., security policy, economic, ethical) while emphasizing a solution to the issue of controllability of technological innovation. By telling a compelling story that highlights the subjective and socially unquestionable positive cost–benefit ratio of the innovation, both internal and external audiences can be convinced to willingly adopt and spread the innovation.

6 Conclusion

- Military innovations inherently pose a communications problem.
- Addressing this communications challenge goes beyond mere dissemination of information; it necessitates an understanding of the social and semiotic context influenced by the innovation.
- Contexts should be comprehended thematically, emphasizing subjective associations linked to each aspect touched by the innovation.
- Basic communicative patterns can facilitate both external and internal innovation communications:
  - Externally, societal acceptance of military innovations hinges on aligning the distribution of benefits and risks/costs while maintaining technology's controllability and manageability.
  - Internally, innovation communications transform into change communications, aiming to resolve goal-means discrepancies within the organization affected by the change.

References

Beck, U. (1992). *Risk Society. Towards a New Modernity* (M. Ritter, Trans.). Sage.

Bijker, W. E. (1997). *Of Bicycles, Bakelites, and Bulbs: Toward a Theory of Sociotechnical Change*. MIT Press.

Deutsch, K. W. (1985). On Theory and Research in Innovation. In R. L. Merrit, & A. J. Merrit (Eds.), *Innovation in the Public Sector* (pp. 17–35). Sage.

Hennen, L. (2002). Monitoring Technikakzeptanz und Kontroversen über Technik. Positive Veränderung des Meinungsklimas – konstante Einstellungsmuster. Ergebnisse einer repräsentativen Umfrage des TAB zur Einstellung der deutschen Bevölkerung zur Technik. In *Dritter Sachstandsbericht (TAB Arbeitsbericht Nr. 83)*. TAB (Büro für Technikfolgenabschätzung beim Deutschen Bundestag).

Kunze, R.-U. (2021). Krise des Fortschrittsoptimismus. In A. Grunwald, & R. Hillerbrand (Eds.), *HandbuchTechnikethik* (pp.71–75).J.B.Metzler. https://doi.org/10.1007/978-3-476-04901-8_13

Latour, B. (2000). *Die Hoffnung der Pandora: Untersuchungen zur Wirklichkeit der Wissenschaft*. Suhrkamp.

Latour, B., & Woolgar, S. (1986). *Laboratory Life. The Construction of Scientific Facts*. Princeton University Press.

Merton, R. K. (1938). Social Structure and Anomie. *American Sociological Review, 3*(5), 672–682. https://doi.org/10.2307/2084686

Ortiz, M. (2021). Kontrollverlust und Technologieakzeptanz in der (digitalen) Transformation. In *Akzeptanz- und Gestaltungsfaktoren eines heuristischen Modells*. Springer VS. https://doi.org/10.1007/978-3-658-35697-2

Pinch, T. J., & Bijker, W. E. (1984). The Social Construction of Facts and Artefacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit each Other. *Social Studies of Science, 14*(3), 399–441. https://www.jstor.org/stable/285355

Reckwitz, A. (2022, April 11). *Der erschütterte Fortschritts-Optimismus.* Deutschland Archiv. https://www.bpb.de/507282

Rogers, E. M. (2003). *Diffusion of Innovations* (5th ed.). Free Press.

Schumpeter, J. A. (1987). *Theorie der wirtschaftlichen Entwicklung. Eine Untersuchung über Unternehmergewinn, Kapital, Kredit, Zins und den Konjunkturzyklus* (7th ed.). Duncker & Humblot.

Svetlova, E. (2008). Innovation als soziale Sinnstiftung. In P. Seele (Ed.), *Philosophie des Neuen* (pp. 166–179). Wissenschaftliche Buchgesellschaft.

Tarde, G. (1895). *Les lois de l'imitation* (2nd ed., reprint 1993 ed.). Kimé Éditeur.

Zowislo-Grünewald, N., & Beitzinger, F. (2021). *Lehrbuch Strategisches Kommunikationsmanagement* (2nd ed.). LIT-Verlag.

Hybrid Warfare and the Defense of Discourse

Natascha Zowislo-Grünewald

Abstract Hybrid warfare in the twenty-first century is characterized by a multitude of terms that often remain undefined. German military documents do not use the term strategic communication uniformly and do not define it either. In contrast, the US military understands Strategic Communication as a holistic approach that uses various means to support American interests. The British military defines Strategic Communication as a means to promote national interests. The definition of the term narrative varies between documents. While the US military does not define it, it is defined in British documents as a story that provides an emotional justification for a political goal. NATO defines the term information environment as an environment consisting of individuals, organizations, and systems that collect, process, and disseminate information. The study shows that definitions of Strategic Communication, narrative, and information environment vary between documents. This can lead to misunderstandings and difficulties in analyzing and evaluating data. The integration of AI into Strategic Communication is also an important aspect that is not adequately considered in the documents.

1 Introduction

Informational Actions, Hybrid Warfare, Cyber Warfare, Information Operations, Information Space, Asymmetrical Warfare—the terminology used to describe modern warfare in the twenty-first century is manifold.

N. Zowislo-Grünewald (✉)
Institute for Organizational Communication, University of the Bundeswehr Munich (UniBwM), Neubiberg, Germany
e-mail: natascha.zowislo@unibw.de; https://strategic-communication-management.de

K. Talves, D. Spreen (eds.), *Artificial Intelligence in Military Technology*,
Artificial Intelligence, Simulation and Society 192,
https://doi.org/10.1007/978-3-031-95578-5_13

But what does it actually mean? Among Western military experts, a debate emerged following Mark Galeotti's initial analysis of Gerasimov's essay The Value of Science Is in the Foresight (Gerasimov, 2016). This debate centered on whether Gerasimov had indeed outlined the foundations of a new military doctrine. The discussion focused on the assessment that Russia, one year after the publication of the article in the Military-Industrial Courier in 2013, conducted an information warfare during the annexation of Crimea. Alternatively, it was suggested that Gerasimov had merely made a Russo-centric observation and classification of Western conflict management during the Arab Spring (Galeotti, 2018).

Two reasons could be responsible for this interpretative uncertainty:

(1) Gerasimov's terminology remains largely undefined. While he uses terms such as "informational actions," he emphasizes that there is only a superficial understanding of asymmetrical forms and means. He suggests that military science must create a comprehensive theory to address this issue, highlighting the importance of the work and research of the Academy of Military Sciences (Gerasimov, 2016, p. 25). From a systems-theoretical perspective, this exacerbates the contingency problem on the recipient's side, as even the sender acknowledges that things could be different (Esposito, 2013, pp. 49–60).
(2) In the Western world, the terminology associated with modern warfare and civilian digitalization seems to suffer from a "hyperinflation of terms coupled with a paucity of definitions" (Engesser, 2008)—which is to be expected in the context of digitization characterized by the 'advantage of practice.' The existence and impact of social media, grassroots and astroturfing movements, artificial intelligence, and self-learning algorithms are tangible for societies and individuals. However, describing their nature and modes of operation remains a subject of ongoing research.

    Once again, this results in an amplification of the contingency problem on the recipient's side as they must also contend with their own uncertainty (and the fact that things could be different).

To address these challenges outlined above, this paper evaluates official documents from the defense policy context relevant to organizational hierarchy regarding their use of specific terms. These terms are fundamental for understanding hybrid warfare according to communication research and are evaluated in descending order as follows:

- Strategic Communication(s)
- Strategic Narrative
- Discourse
- Information Environment
- Target Audience
- Stakeholder

Two objectives are pursued through this evaluation: Firstly, the aim is to gain insight into the state of affairs regarding the topic from military documents exemplifying 'the West.' The focus here is on obtaining an overview of whether and to what

extent the theoretical foundation for real phenomena of hybrid warfare has already progressed. This will help uncover potential research needs. Secondly, by comparing the available material, we attempt to identify conceptual differences within various approaches to Strategic Communication and provide indications for unification and generalization within the framework of NATO's alliance system.

Moreover, with regard to the functioning and learning methods of generative AI and Large Language Models (LLMs), which are generally considered instruments for automatically organizing and analyzing otherwise unmanageable data volumes in modern Strategic Communication, an overview of their possible learning fundamentals is essential for assessing output, outcome, and outflow (Watson & Noble, 2007).

Preliminarily, it should be noted that the lack or imprecision of conceptualizing Strategic Communication and its central concepts regarding the documented aggressiveness—especially in recent times—of (Russian) disinformation campaigns using AI-based tools is perplexing. There seems to be little perceptible sense of a forward-looking conceptual grasp of these terms outlining possible scope for action within the information space, let alone their embedding with regard to insufficient conception related to KI relevance. In short: Strategic Communication remains an underappreciated appendage of strategic and operational defense; thus, AI is unable to unfold its potential contribution in this area. Instead, it is only given the opportunity to act as a security threat—because hybrid aggressors make extensive use of AI.

2 Document Situation German Armed Forces (Bundeswehr; BMVg-WS; KK-KD; KK-AR; TF-AR)

The research revealed a multifaceted and highly decentralized document landscape regarding the terms under investigation within the German Armed Forces, which are publicly accessible but not 'published.' Therefore, quotation references are omitted here.

A BMVg workshop from 2022 (BMVg-WS) takes precedence in this hierarchy due to its focus on societal resilience as a central aspect of national security strategy following the 'change of times' marked by events such as the Ukraine conflict and hybrid threat situations (Fleischer, 2022).

The section on *Strategic Communication* is summarized as follows (all translations by the authors):

> Communication should be understood more strongly as an instrument for security policy at the strategic level. The change in strategic culture required due to 'change of times' (*Zeitenwende*) must also reach people's minds. Communication is a central element here. In the war of narratives, security-political communication is also a means of defense. It involves coordinated external communication among partners and allies. Germany's leader-

> ship role in Europe needs to be communicated even more strongly than before. This includes not only communicating results but also explaining underlying processes better.[1]

This section formulated as an intention statement lacks further context or concretization regarding its implementation at a later stage. Instead, there are additional documents of younger vintage associated with specific areas that address the topic.

As these are not organized hierarchically or a hierarchical assignment does not seem clearly possible, they are examined comparatively in their entirety:

(1) *Das Kommunikationskonzept der Reserve—Konzeptionelle Dokumenten landschaft* (The Communication Concept of the Reserve—Conceptual Document Landscape) K-3105/2 (KK-KD) dated June 10, 2021 (BMVg FüSK III 4, 2021).
(2) *Krisenkommunikation im Organisationsbereich AIN—Allgemeine Regelungen* (Crisis Communication in Organisational Area AIN—General Regulations) C1-600/0-7000 (KK-AR) dated October 20, 2023 (BAAINBw PIZ AIN, 2023).
(3) *Truppenführung—Allgemeine Regelungen* (Troop Leadership—General Regulations) C1-160/0-1001 (TF-AR) dated July 1st, 2022 (KdoH Op G5 Grds BTF TrFü and OpLaSK, 2022).

Regarding the term *strategic communication* there are no similarities between these three documents which is striking given their publication dates and current validity.

However, all investigated papers use the German equivalent of 'communication strategy' (*Kommunikationsstrategie*). The KK-KD defines this as follows:

> Communication Strategy
>
> A communication strategy comprises all medium or long-term measures by means of which an organization conveys its core messages in a targeted manner to dialogue groups outside so that it can achieve self set communicative goals. Communication strategies may include individual instruments, several combined ones.

As expected, the term *communication strategy* is, thus, hierarchically subsumed under Strategic Communications as a measure for communicating certain contents tailored to target audiences and objectives.

Document KK-AR features the following section regarding this concept:

> Constant updating on situation development; consultation about further procedure (communication strategy); if necessary:
>
> – establish language rules (model text)
> – write press/media release
> – formulate statements
> – select channels (target group analysis)
> – FAQs, contributions in Internet / Intranet
> – Press conference.

[1] The excerpts from the Bundeswehr documents have been translated into English for better understanding.

Here again communication strategies are understood as a distinct measure to be taken during crisis situations. A classification of this action into an overarching concept of Strategic Communications is missing.

The TF-AR defines the term as follows:

> Communication strategy: 9007. Influencing perception, attitude, will, and behavior in the InfoU can save bloodshed. Information and communication are central components for planning and conducting land operations. TrFhr comprehensively considers InfoU as a basis of targeted action. The orientation is given by means of narrative or information-communication strategies derived according to levels.

Unlike the documents KK-KD und KK-AR, TF-AR describes communications strategy in terms very close to Strategic Communication but fails to mention necessary adjustments for planning and structuring both concepts. While it states that this "orientation" should be provided through a "narrative," such narrative is situational (and thus variable) relying on factors like InfoU as basis and being derived by TrFhr according to levels.

The specific design of the communications strategy therefore depends not so much upon an overarching goal in terms of Strategic Communication but rather upon situation, person, and level-dependent derivation.

The term *strategic narrative* is only discussed within TF AR:

> 9024 Digital Direct Communication (DigDirKom) influences social networks/media Dynamics are influenced to promote dissemination of certain contents messages narratives favorable for own OpCon in order to influence respective target audiences.

What exactly constitutes a narrative remains unclear. It appears that according to the document it is some kind of 'quasi-message' or communication content, only being applied within digital direct communication.

Only KK-KD mentions *discourse*:

> 316 Through targeted communications civil leaders from economy public service and science political mandate holders employees as well as outstanding personalities shall be interested in security policy topics They are important multipliers to strengthen the discourse on defense politics tasks aims of Bundeswehr and its Reserve. Furthering these contents is an excellent task for InfoDVag.

What exactly constitutes a discourse remains unclear. The very general use suggests that it could have been replaced by words like 'discussion' or 'debate.'

Only one document (TF-AR) mentions *information space/information environment*:

> 2008 Methods of hybrid influence and warfare (hybrid warfare) comprise wide range civil military means instruments to achieve political goals already below threshold level conventional war They especially target InfoU undermining capabilities credibility legitimacy organs state as well societal cohesion Special challenge is represented by targeted influencing public opinion opponent.

Information space/environment (infoU) is not defined. Nonetheless, the effects of Strategic Communication are described ex negativo (from an enemy's perspective). Factors like credibility and legitimacy of the state and its organs are mentioned in this respect, as the term *social cohesion* is used.

Later on it says:

> 2016. The four dimensions of military action are Land, Air and Space, Sea, and the Cyber and Information Space (CIR). 9001. The CIR as a military dimension includes the virtual, physical, and cognitive sphere consisting of the Cyber Space (CR), the Electromagnetic Environment (EMU), and the Information Environment (InfoU).

At this point, the term is categorized, whereby it can be concluded from paragraph 2016 that the term information environment is not fundamentally assigned to all military dimensions. However, the previous fundamental classification of conflicts as "hybrid" means that this bridge is built indirectly.

Finally, TF-AR concludes with the following recommendation:

> 9023. In principle, CIR operations can be carried out in the InfoU with a wide variety of media.

Impact in the information environment is achieved with the help of 'various' media, whereby these are not further differentiated and a bridge to the other dimensions of military action (see paragraph 2016) again only takes place indirectly.

The term *target audience* (*Zielgruppe*) can be found in two of the three documents analyzed. KK-AR defines it like this:

> 421. In principle, a distinction is made between external and internal target audiences, but not in the sense of differentiating the messages, only in terms of language and frequency. All target audiences must be informed about relevant events in the crisis as required.

Target audiences are therefore not differentiated on the basis of their stakes or communicative goals, but in terms of the language to be used and the frequency of communication. However, no mention is made of the criteria according to which the style and frequency of communication should be varied without formulating communicative goals or analyzing target audiences. This is particularly interesting as it is stated below:

> 425. Crisis communication is considered appropriate for the target group if the target group
>
> – is reached via the chosen communication medium,
> – can be motivated to take up the message,
> – understands the message correctly, and
> – reacts to the message accordingly.

TF-AR describes the term target audience only in combination with "tactical direct communication":

> 9012. Commanders may have direct tactical communication staff at their disposal. They also advise and support the troops at a lower tactical level and have a direct and short-term effect on the behavior of defined target audiences by conducting talks and loudspeaker calls. Examples include the effect on trapped enemies with the aim of giving up the fight or directing population groups.

In this sense, target audience communication means direct communication. There is no explicit differentiation of messages according to target groups, and the concept refers primarily to "lower tactical levels."

Only KK-AR uses the term *stakeholder*:

> 401. Crisis communication describes all forms of communication with stakeholders (employees, citizens, media, other authorities/administrations, politicians, companies, associations and other advocacies) in order to resolve a crisis in the interests of the Bundeswehr and to have a positive impact on the perception of the crisis by third parties.

It is worth mentioning in this context that KK-AR uses both the term target audience and the term stakeholder without defining or differentiating between them precisely. However, the term stakeholder precedes the chapter on crisis communication in the document (see quoted section above).

3 Document Situation United States Armed Forces (US-DoM)

This examination is based on the *Department of Defense Dictionary of Military and Associated Terms* from 2017, the official reference work of the US armed forces, which is published directly by the US Department of Defense and is universally valid (Joint Staff, J-7, 2017):

> The DOD Dictionary of Military and Associated Terms (DOD Dictionary) sets forth standard US military and associated terminology to encompass the joint activity of the Armed Forces of the United States. These military and associated terms, together with their definitions, constitute approved Department of Defense (DOD) terminology for general use by all DOD components.

The dictionary also has a clear objective:

> This publication supplements standard English-language dictionaries and standardizes military and associated terminology to improve communication and mutual understanding within DOD with other US Government departments and agencies and among the United States and its allies.

It can therefore not be ruled out that certain subdivisions use different definitions or that other documents define the terms to be examined more precisely. However, the highly hierarchical structure of the Dictionary of Military and Associated Terms (here: US-DoM) and its topicality allow it to be used as a reference document.

The term *strategic communications* is defined by US-DoM:

> Focused United States Government efforts to understand and engage key audiences to create, strengthen, or preserve conditions favorable for the advancement of United States Government interests, policies, and objectives through the use of coordinated programs, plans, themes, messages, and products synchronized with the actions of all instruments of national power.

Strategic Communications is thus understood by the US armed forces as a holistic approach that uses a wide variety of means to support the implementation of American interests. It is important to note that no restrictive or guiding values ('within a democratic framework,' etc.) are associated with this.

The terms *strategic narrative* and *discourse* are not defined or mentioned by the US-DoM.

US-DoM defines the term *information environment* as follows:

> The aggregate of individuals, organizations, and systems that collect, process, disseminate, or act on information.

This very universal and broad definition follows on from the previous definition of Strategic Communications in its general validity, thus encompassing all possible state and non-state actors as well as allies, enemies, and 'uninvolved third parties' alike.

The term *target audience*, on the other hand, is clearly and generally defined by US-DoM:

> An individual or group selected for influence.

However, the US definition of target audience only becomes fully understandable in connection with the definition of the term *stakeholder*.

> In public affairs, an individual or group that is directly impacted by military operations, actions, and/or outcomes, and whose interests positively or negatively motivate them toward action.

While target audiences are selected (and therefore obviously have no choice as to whether and how they are 'affected' by Strategic Communication), this choice does not apply to stakeholder groups that actively interact with the organization through their own interests. It is also worth mentioning the restriction to the political sphere by limiting it to the area of public affairs.

This restriction/definition of the term pair target audience ←→ stakeholder makes sense from the perspective of the communicating organization (United States Government). For the US government, these seem to be exclusively stakeholders from its 'peer group,' which in turn can be made up of other political actors. Non-governmental, non-political actors, on the other hand, are considered by US-DoM to be passive target audiences.

4 Document Situation British Armed Forces (BAF-JDN)

The *Joint Doctrine Note 2/19—Defence Strategic Communication: an Approach to Formulating and Executing Strategy* (UK Ministry of Defence, 2019, here: BAF-JDN) forms the basis of the following investigation with regard to the British armed forces. Analogous to the structure of the US armed forces, the BAF-JDN is also a

top-down directive, the contents of which were authorized by the Chief of Staff and are considered binding. Here, too, it cannot be ruled out that subordinate documents exist which differentiate the terms examined in more detail—and thus possibly in a slightly modified form. However, due to the doctrinal nature of the document, it cannot be assumed that its contents are presented in a contradictory manner at lower levels or within the armed forces.

The term *strategic communications* is defined by BAF-JDN as follows:

> Advancing national interests by using Defence as a means of communication to influence the attitudes, beliefs and behaviours of audiences.

What is particularly striking is the general similarity of the definition to that in the US document US-DoM, while at the same time differing in a few crucial details. The British armed forces define the objective of Strategic Communication as "advancing national interests," while the US version speaks decidedly of "government interests." In addition, Strategic Communication appears here as an instrument of defense.

The definition of the term *strategic narrative* follows on from this:

> A strategic narrative is a story designed to provide an emotive justification for a policy goal, and in many cases how that goal is to be realised and the moral authority for doing so. The policy is the desire to move to a future state or maintain a current state (the ends). The strategy is the plan to get to that state (employing ways and means to achieve the ends). The narrative provides the 'why?' It provides a justification for action, and a justification can be very subjective.

The BAF-JDN is the only document of the national armed forces examined that contains a clear definition of the term narrative. A narrative is defined as a motive for action, whereby the psychological and legal dimensions of the term are apparently addressed equally (Hessick, 2006).

The term *information environment* is used 14 times throughout the document, but without defining it. Its use indicates that BAF-JDN derives the Information Environment from "common usage" (Röttger & Vedres, 2020). The importance of analyzing the information environment (IEA) as a fundamental instrument of Strategic Communication is pointed out. It is made up of "physical and virtual dimensions" and overlaps with the definition from TF-AR at this point.

The term *target audience* is not defined in BAF-JDN, but is mentioned 98 times. It becomes clear that a common usage is also assumed here. However, it also becomes clear in several places that the British definition of the term differs in particular from the Bundeswehr documents examined (KK-AR; TF-AR). Communication with the target audience—borrowed from private sector approaches—is always geared toward "how our product, service or brand can solve the problems that target audiences are facing" (BAF-JDN, p. 31).

The BAF-JDN does not define or use the term *stakeholder* or the term *discourse*.

5 Document Situation NATO (NATO-DB)

Finally, the terms are examined with regard to their definition in NATO. Therefore, the *NATOTerm—Official NATO Terminology Database* (here: NATO-DB) was used, which contains an official and therefore binding lexical definition of various terms and their use within the NATO context (NATO and NATO Standardization Office, n.d.).

> In the NATO military context, the integration of communication capabilities and information staff function with other military activities, in order to understand and shape the information environment, in support of NATO strategic aims and objectives.

The definition of the term *strategic communication* plays a hybrid role in the NATO context: on the one hand, a version of the term geared toward military operations ("with other military activities") can be recognized, which is certainly based on the understanding of the term "Kommunikationsstrategie" from the Bundeswehr documents KK-AR and TF-AR. On the other hand, there is also a classification in the overall strategy, which refers beyond the operational (and thus purpose-related) context of military operations ("in support of NATO strategic aims and objectives")—whereby at the same time the restriction must be made that the strategic objectives of a military alliance such as NATO are probably always defined in a more or less military context.

> A spoken or written account of events and information arranged in a logical sequence to influence the behaviour of a target audience.

In contrast to the definition in BAF-JDN, the NATO definition of the term *strategic narrative* does not mention the term story and also avoids attributing it as possibly "very subjective." Nevertheless, the NATO definition can be seen as similar in content to BAF-DJN: the story is restricted to the "spoken or written account of events and information" and the subjective perception to a "logical sequence"; finally, the purposeful meaning of a narrative from BAF-DJN ("It provides a justification for action") is replaced by a goal-oriented meaning ("influence the behavior of a target audience").

The term *discourse* is neither defined nor used in the NATO-DB.

> An environment comprised of the information itself, the individuals, organizations and systems that receive, process and convey the information, and the cognitive, virtual and physical space in which this occurs.

Just like US-DoM, NATO-DB defines the term *information environment* in such a way that it can also be applied outside the military context. In summary, the Information Environment can be described here as the environment and its perception or '(perceived) environment.' However, it remains unclear exactly how this 'environment' is defined—especially since it also includes cognitive and virtual environments.

> An individual or group selected for influence or attack by means of psychological operations.

Just like US-DoM, NATO-DB distinguishes the term *target audience* from the term stakeholder in that a target audience acts exclusively passively ("selected for influence"). Internal and external stakeholders—in the military context: friend and foe—are addressed equally.

> An individual, group or entity who can affect or is affected by the attainment of the end state.

As already indicated, the term *stakeholder* also has an active role in NATO-DB. In contrast to target audience, a stakeholder can actively influence the achievement of objectives. In contrast to target audience, stakeholders also include "entities," although it is not clear what this could mean. The specific meaning of the term "end state" also remains unclear.

6 Summary and Conclusion

After examining high-level documents from both leading national armed forces within NATO and the NATO terminology lexicon itself, it becomes clear that the terms examined in the context of Strategic Communication are sometimes defined completely differently, sometimes similarly and sometimes not at all.

There is fundamental disagreement, particularly with regard to how the subject area of Strategic Communication should be defined. While it is clear from the German Bundeswehr documents examined that strategic communication as a term is not used at all, but is only understood in its instrumental form as a communication strategy, both the British and American understanding of the term extends far beyond the military context. NATO is probably trying to build a bridge here, which understands Strategic Communication, on the one hand, as a means of achieving overarching, long-term goals, and on the other hand embeds this goal achievement in a military-operational context.

This divergence becomes important with regard to the definition of hybrid warfare and the understanding of when exactly we should speak of war in the twenty-first century. If one follows Gerasimov's view, nations in the age of hybrid warfare enter into a constant and 'perpetual' state of war, which in turn expands the conceptual understanding of Strategic Communication as a means of military operations, as a constant war that affects all social classes and also completely redefines the military operation itself.

The *narrative* as a superordinate means of standardizing or unifying the meaning of Strategic Communication is largely unknown within the documents examined, with only BAF-JDN and NATO-DB attempting to classify it—although this differs greatly in terms of content. While BAF-JDN clearly emphasizes the subjective meaning of the narrative, in NATO-DB this seems to have a more or purely cognitive effect, depending on the understanding of the term "logical (sequence)."

The overarching perception of "reality as context" (Blumenberg, 1969, p. 21) and the consequences arising from this are therefore only approximated by the definition in BAF-JDN. Only here is the recipient's 'acceptance' of reality extended by

a dimension of understanding that can also be seen as a question of 'why?' Answering this question does not directly lead to an increased acceptance of communicated content on the part of the recipient. Rather, the embedding of individual communicative measures in a larger causal context enables the recipient to take a position on (communicative) content in the first place.

Only KK-KD takes into account the concept of discourse as a delimitation of content, in which certain meaningful narratives can have an effect. The scalability of the term narrative—also and especially with regard to AI-supported automation—is complicated without a corresponding embedding framework.

A divergent understanding within the field of Strategic Communication is particularly evident in the example of the term pair target audience ←→stakeholder. While the German Bundeswehr documents examined use the terms either with the same meaning (KK-AR) or colloquially (KK-KD and TF-AR), the latter also appears to apply to BAF-JDN, which avoids the term stakeholder altogether and summarizes all potential message recipients as target audience.

In contrast, the US-DoM and NATO-DB differentiate the term pair qualitatively: while a target audience is passively 'exposed' to Strategic Communication, stakeholders actively shape Strategic Communication. Since Large Language Models (LLM) learn by processing large amounts of data (Luber, 2023), the 'lack of definition' mentioned in the introduction means that the outcome of these learnings may be unknown under the given circumstances—especially since hardly any reference is made to the different possible uses of the terms. What is meant by the terms target audience and stakeholder is therefore hidden behind the veil of human, intentional consciousness, which could make targeted learning difficult for an LLM and targeted evaluation difficult for an analyst.

This is because every AI-supported evaluation is followed by a human classification, which in turn must rely on the AI to distinguish between terms such as stakeholder, target audience, or narrative. It makes a decisive difference whether an American analyst is confronted with target audiences (and immediately assigns them a passive role based on their understanding of the term) or one with a different understanding of what a target audience is. It also makes a difference whether a generative AI takes into account the meaning and purpose of a narrative (explain the 'why?') when analyzing communications—or not.

Finally, this 'presupposed' understanding becomes clear with the term information environment. If the term is defined, it will apparently refer to the inherent and ascribed meaning of objects in an (undefined) environment. However, how this meaning is to be uncovered, differentiated, and analyzed remains largely unclear (cf. BAF-JDN, which speaks of an "information environment analysis"). The fundamental distinction between the 'true' and the 'attributed' meaning of information has long preoccupied epistemology and ontology, although no clear solution has yet been found (Husserl, 1971, pp. 7–8).

Overall, this results in the following situation (Table 1):

With regard to an LLM and generative AI in general, the problem of the learning basis therefore arises here: it is solely down to human perception which environmental factors are perceived as such and how one environment is differentiated from another environment.

Table 1 Definition of the examined terms (in the first column, the coloring indicates whether a uniform definition exists across all examined organizations; columns 2–5 indicate: ☐ no definition, ▭ vague/ambiguous definition, ▬ defined)

	Bundeswehr	USAF	BAF	NATO
(Strategic) Communication(s)	Defined only as "Kommunikationsstrategie" (Communication Strategy)	Protection of interests of the "US Government"	Protection of "National Interests"	Rather military; Achievement of "strategic aims and objectives"
(Strategic) Narrative	Definition unclear, rather to be understood as "message"	n/a	"Justification for action"	Limited in content ("spoken or written"), "to influence the behaviour of a target audience"
Discourse	Common usage: "debate," "discussion"	n/a	n/a	n/a
Information Environment	Purely military, not clear	Generalized, "everyone" who does "something" with information	Common usage	Common usage "perceived environment"
Target Audience	Defined as "direct communication" for "lower tactical levels"	Is influenced (passive)	Common usage	Is influenced (passive)
Stakeholder	Used – not defined	Influences (active)	n/a	Influences (active)

In particular, since the information environment also refers to the digital sphere which is spaceless due to its nature (and therefore able to connect to other environments), this makes it difficult (if not impossible) for generative AI to learn with the help of empirical values. For example, does Russian influence with the help of mass media (Russia Today) or social media (Facebook, X, etc.) belong to a specific environment? If this is the case, the environment of the environment (what world view prevails among the target audiences, what circumstances lead to a certain narrative potentially being followed?) must be included in the consideration—and how can these be delimited in terms of machine learning?

But even for the legitimization of the discourse 'Strategic Communication,' which takes place via a narrative—to take up two concepts here, which are partly taken up by the documents examined—a frayed definition is at best unfavorable, at worst harmful: because a narrative explains the 'why?' of an action, is a motive for action and guiding principle—and only functions for the recipient as a meaningfully understood connection, as a 'context.' This, in turn, seems difficult to achieve on the assumption that the terms used are understood as differently as the documents suggest.

Overall, the study revealed that certain fundamentally important definitions and a uniform understanding of the term are missing or are defined so differently that it is likely to be extremely difficult to master the large amounts of data that are available in the communicative sphere in the age of hybrid warfare and that must be systematically recognized and evaluated in order to counter bot networks, deep fakes, and propaganda and their effects.

The fact that AI, its possibilities, and also the way it works (especially as LLM) are not mentioned in the documents is particularly worrying. This can be partly attributed to the time lag between the release of ChatGPT (v3.5) in 2022, which

appeared to be revolutionary, and the documents examined (from 2017), but also seems 'incomprehensible,' especially with regard to the ongoing update of NATO-DB, for example, or possibly due to the fact that theoretical foundations (and the space of possibilities that opens up as a result) have not been sufficiently defined.

(Generative) AI is expected to make quantum leaps in the handling of large amounts of data. However, these expectations can hardly be fulfilled if the correct prerequisites are not provided. In the worst-case scenario, we receive the most elaborate results from generative AI—and yet we do not know which structures and mechanisms were used to produce them. Valery Gerasimov clearly and aptly formulated the key to solving this problem, despite all the vagueness and imprecision, in the title of his essay: "The Value of Science is in the Foresight." It is the task of scientific research to lend practical value to theoretical considerations through anticipation—just as it is to provide practical phenomena with as universal a theoretical foundation as possible in order to be able to make further deductions and drive innovation.

Without embedding the above-mentioned terms in the context of AI, the concepts cannot develop any relevance in the digital space. In other words, neither the analysis of the information space nor the strategic derivations for the creation of relevant communication strategies as a contribution to hybrid warfare—as a guarantee for the 'mental strength' of one's own soldiers as well as the population in the home country and country of deployment—can take place if the concept of Strategic Communication and, moreover, the relevance of AI for Strategic Communication remain (in some cases completely) underexposed.

- The use of fundamental terms on a national level within the NATO military alliance is not unified: misinterpretations or different interpretations are therefore most probable, knowledge transfer is hindered, an overarching analysis seems impossible.
- NATO itself does not define fundamental terms of organizational communication, which also hinders structured work and/or analysis.
- LLMs (and generative KI) could therefore learn wrongly and their evaluations could be understood wrongly.
- Fundamental structuring and unified definitions have to be able to re-translate practical outcome into a theoretical framework.

References

BAAINBw PIZ AIN. (2023, October 20). *Krisenkommunikation im Organisationsbereich AIN. Allgemeine Regelungen (C1-600/0-7000, Version 1, here as KK-AR)*. Bundeswehr.

Blumenberg, H. (1969). Wirklichkeitsbegriff und Möglichkeit des Romans. In H. R. Jauß (Ed.), *Nachahmung und Illusion. Kolloquium Gießen 1963 Vorlagen und Verhandlungen* (pp. 9–27). Wilhelm Fink.

BMVg FüSK III 4. (2021, June 10). *Kommunikationskonzept der Reserve. Konzeptionelle Dokumentenlandschaft (K-3105/2, here as KK-KD)*. Bundesministerium der Verteidigung.

Engesser, S. (2008). Partizipativer Journalismus: Eine Begriffsanalyse. In A. Zerfaß, M. Welker, & J. Schmidt (Eds.), *Kommunikation, Partizipation und Wirkungen im Social Web, Bd. 2: Strategien und Anwendungen. Perspektiven für Wirtschaft, Politik und Publizistik* (pp. 47–71). Herbert von Halem.

Esposito, E. (2013). Doppelte Kontingenz. In D. Horster (Ed.), *Niklas Luhmann – Soziale Systeme* (pp. 49–60). Akademie Verlag.

Fleischer, J. (2022). *Nationale Sicherheitsstrategie: Gesellschaftliche Resilienz im Fokus*. https://www.bmvg.de/de/aktuelles/nationale-sicherheitsstrategie-gesellschaftliche-resilienz-fokus-5514816

Galeotti, M. (2018, March 05). *I'm Sorry for Creating the Gerasimov Doctrine.* Foreign Policy. https://foreignpolicy.com/2018/03/05/im-sorry-for-creating-the-gerasimov-doctrine

Gerasimov, V. (2016). The Value of Science is in the Foresight: New Challenges Demand Rethinking the Forms and Methods of Carrying Out Combat Operations. *Military Review, 96*(1), 23–29. https://www.armyupress.army.mil/Journals/Military-Review/English-Edition-Archives/January-February-2016

Hessick, C. B. (2006). Motive's Role in Criminal Punishment. *Southern California Law Review, 80*, 89–151. https://ssrn.com/abstract=921111

Husserl, E. (1971). *Philosophie als strenge Wissenschaft*. Vittorio Klostermann.

Joint Staff, J-7. (2017, March). *DOD Dictionary of Military and Associated Terms (Here as US-DoM)*. https://www.tradoc.army.mil/wp-content/uploads/2020/10/AD1029823-DOD-Dictionary-of-Military-and-Associated-Terms-2017.pdf

KdoH Op G5 Grds BTF TrFü, & OpLaSK. (2022, July 01). *Truppenführung. Allgemeine Regelungen (C1-160/0-1001, Version 2.1, here as TF-AR)*. Bundeswehr.

Luber, S. (2023, October 05). *Was ist ein Large Language Model?* BigData Insider. https://www.bigdata-insider.de/was-ist-ein-large-language-model-a-d735d93bbc24d3c4091de8ce25aa36e8

NATO, & NATO Standardization Office. (n.d.). *NATOTerm – Home. Official NATO Terminology Database (here as NATO-DB)*. https://nso.nato.int/natoterm/content/nato/pages/home.html

Röttger, P., & Vedres, B. (2020). *The Information Environment and its Effects on Individuals and Groups. An Interdisciplinary Literature Review*. https://royalsociety.org/-/media/policy/projects/online-information-environment/oie-the-information-environment.pdf

UK Ministry of Defence. (2019, April). *Joint Doctrine Note 2/19. Defence Strategic Communication: An Approach to Formulating and Executing Strategy* (*Here as BAF-JDN)*. https://assets.publishing.service.gov.uk/media/5ce7fc2fe5274a4873de09e5/20190523-dcdc_doctrine_uk_Defence_Stratregic_Communication_jdn_2_19.pdf

Watson, T., & Noble, P. (2007). *Evaluating Public Relations. A Best Practice Guide to Public Relations Planning, Research and Evaluation* (2nd ed.) Kogan Page.

Holistic Bowtie Model of AI-Based Technology in Defense Systems

How to Think, Explore, and Implement Efficient and Meaningful Chains of Abilities, Control, Responsibility, and Accountability Between Humans, AI, Organization, and the Environment?

Frank Ole Flemisch, Dierk Spreen, Marie-Pierre Pacaux-Lemoine, Benjamin J. Knox, Kairi Talves, and John Christopher Brill

Abstract As AI creates a strong technological push on our societies and defense systems, increasing requirements to counteract hybrid warfare pull the demands for new methods to think defense in a systemic and interconnected way. The holistic bowtie model offers a method for understanding the connections between our defense systems and surrounding organizations and societies on an interconnected planet. It is derived from the class of bowtie models already in use for decades in hazard and accident analysis and is inspired by and a potential answer to the Chinese principle of 天下 Tianxia. The holistic bowtie model focuses on critical situations and is gradually expanded to include other system levels, their causal influences, and control loops. The article starts with the historical example of the spears of Schoeningen, introduces the concept step by step, puts it into the military context of modern (cog-

F. O. Flemisch (✉)
Fraunhofer Institute for Communication, Information Processing and Ergonomics (Fraunhofer FKIE), Wachtberg, Germany
e-mail: frank.flemisch@fkie.fraunhofer.de

D. Spreen
Institute for Organizational Communication, University of the Bundeswehr Munich (UniBwM), Neubiberg, Germany

Department of Business and Economics, Berlin School of Economics and Law Berlin (HWR), Berlin, Germany
e-mail: dierk.spreen@unibw.de; https://strategic-communication-management.de

M.-P. Pacaux-Lemoine
Departement of Human and Life Sciences, Université Polytechnique Hauts-de-France, Campus du Mont Houy, Valenciennes, France
e-mail: marie-pierre.lemoine@uphf.fr

K. Talves, D. Spreen (eds.), *Artificial Intelligence in Military Technology*,
Artificial Intelligence, Simulation and Society 192,
https://doi.org/10.1007/978-3-031-95578-5_14

nitive) warfare, discusses the role of cognition, emotions, truth, sensemaking, trust and mistrust, and finally agility as essential building blocks in a strong alliance.

1 Introduction

As the current war in the Ukraine shows, new automation- and AI-based technologies are applying pressure to adapt and instigating change in our defense systems. There are many indications that the skillful networking of humans, artificial intelligence (AI), organizations, and the environment can have a decisive impact on the modern battlefield, but also on the prelude to armed conflict between societies and states.

How can we think about and apply the multiple, potentially disruptive connections between AI, people, technologies, organizations, and the environment in a way that enables individuals, policymakers, and organizations to make appropriate decisions, act efficiently, and, when necessary, join forces to successfully defend against physical, virtual, or cognitive threats? Providing good models about thinking and acting in a strongly interconnected, AI-supported world can be an important contributor to credible deterrence.

In a multi-layered discussion process within NATO over several years—originally on questions of the controllability of AI and cognitive warfare—a number of insights into the character of the new challenges emerged. In this context, a class of models was developed that can be used to describe some of the challenges and solutions. Before the chapter goes into the details, a summary of the central insights:

(a) The challenges and complexities with AI are only partly of a technical nature, but primarily of a systemic nature, i.e., they lie in the combination of people, technology, culture, organizations, societies, and the environment.
(b) Our defense systems are interwoven and layered in many ways, starting with individual people and extending to weapon systems, teams, systems-of-systems, organizations, and societies.
(c) Only some of the challenges can be solved by looking at the individual layers in isolation. Another, rather underexposed part can only be understood and solved through a connected approach.

B. J. Knox
Topasbygget, Gjøvik, Norwegian University of Science and Technology (NTNU), Gjøvik, Norway
e-mail: benjamin.j.knox@ntnu.no

K. Talves
Estonian Military Academy, Tartu, Estonia
e-mail: kairi.talves@mil.ee

J. C. Brill
Air Force Research Laboratory, Wright-Patterson Air Force Base, Dayton, OH, USA
e-mail: john.brill.2@us.af.mil; https://www.randallroberts.com/obituaries/lt-col-john-brill

(d) Intelligence as a connected concept and an essential part of defense capabilities requires not only an understanding of the physical processes, but an integrated understanding of cognitive, organizational, cultural, ethical, moral, and perhaps also spiritual processes (e.g., Ferguson, 2023).
(e) Intelligent automation in defense systems is primarily generated in combination of actors, e.g., from people, technology, and organizations in interaction with their environment.
(f) The chains of capabilities, control, responsibility, and accountability play a key role in this.
(g) On the one hand, these chains must be efficient in order to be able to effectively *deter* aggressors such as Russia or system rivals such as China. Or, if military conflicts or wars do occur, to *win* them.
(h) These chains must be designed in such a meaningful way that they correspond to the value system of our democracies and are supported by the majority of our society and politics. One example of this is the concept of *Meaningful Human Control* (MHC) via automation or AI-based systems, as it is explored and applied by NATO.
(i) The speed of (re)design is obviously so high that it requires not just thinking about new defense systems, but also about new methods for system exploration and development, which integrate holistic thinking and MHC right away.

To address these insights, we propose holistic bowtie diagrams as one of several thinking tools for describing problems and solutions. This is a combination of the bowtie diagrams originally developed in the oil industry for hazard analysis with control and regulation diagrams from cybernetics.

Our article introduces the problem of holistic thinking in connection with AI and automation in defense, derives the model step by step, and illustrates it using the example of automated or AI-based systems in defense. The connection between ability, responsibility, control, and accountability is briefly outlined and placed in context. Additional concepts such as cognition, emotions, the paradox of truth, sense and nonsense, trust, and mistrust are briefly described and categorized before essential feedback loops are described holistically. As time and speed are crucial aspects in military conflicts, a discussion of agility and the concept of agile exploration concludes the chapter.

2 The Schöningen Spears

To understand and shape possible futures, especially in turbulent phases of world history, it might be worth taking a look back at the past, in particular at the history of defense technologies and methods. This approach is inspired by the method of Synectics, originated in the 1950s at the *Arthur D. Little Invention Design Unit* (e.g., Gordon, 1961). In a synectic approach, analogies with something not yet familiar are used to gain new insight into something familiar. Here, we assume that the technology of AI is more familiar to most readers, and we are using an example

of historic spears, which might not be too familiar to most readers, to gain insight into the nature of the integration between human and technology. It is important to note here that the unfamiliar examples do not have to be completely historically precise in the same way as historic scientists would describe them, but enough evidence that they can serve as an analogy for the more familiar target domain.

So, to gain insight into the nature of the human-AI challenge, let us take a look at an older example of a defense technology.

The Schöningen spears, for example (https://www.dieschoeningerspeere.de/en), which were found in a lignite mine in 1994, are the oldest wooden hunting weapons dating back over 270,000 years. Accumulations of horse bones with human scrape marks were also discovered near the hunting weapons, indicating that the spearmen—probably *Homo heidelbergensis*, a precursor of *Homo sapiens*—systematically hunted and rounded up wild horses. Hunting weapons such as these are also suspected of having contributed to the extinction of a number of animal species.

The Schöningen spears are an early example of weapon technology that certainly raised the effectiveness of the 'operation' and was very likely a game changer in the competition and wars of that time, analogous to how AI might be a game changer in defense systems of today. Already the Schöningen spears show how important it is to think beyond the pure technology of the weapon. The spears were not only optimized for purely technical functionality but were also adapted in shape and size to the user and bearer. This makes them an early example of including human factors in the design and engineering process (e.g., Milks et al., 2019), obviously with a much shorter distance between designer, implementer, and user compared to today's systems (Fig. 1).

It is believed that the organization of the hunt and the connection between humans, technology, organization, and environment played an important role, which we now refer to as Human Systems Integration. These wooden artifacts also demonstrate that the use of technology played an early role in human evolution. Presumably, this use for the purpose of hunting went hand in hand with social organization. According to the sociologist Heinrich Popitz, the reason for this is the "range of

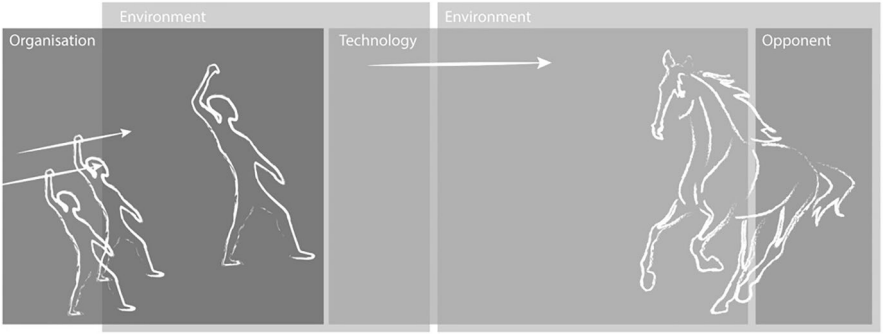

Fig. 1 Schoeningen spears as an early example for a human systems integration challenge of humans, technology, organization, and environment (*source*: Wasser & Flemisch, 2020)

capabilities" of the human hand (Popitz, 1995, p. 60). This range stems from its release of instinctive and reflex reactions, which in turn is associated with "learning ability and openness to action" (Pohlmann, 2000, p. 59). The performance diversity of the human hand (and thus of human beings as a biological species) includes, in particular, signaling functions (hand as a pointing organ), relationship functions (social contact and communication), and object contact, including the use of tools. The function of object contact can be further broken down into "touching–grasping–shaping–beating–throwing" (Popitz, 1995, p. 62), whereby the modes of beating and throwing, in particular, refer to the "suitability of the hand for guiding and handling objects," which includes the possibility of "violent alteration of objects, up to and including killing" (Popitz, 1995, p. 65). "Hand, brain, and eye" work together like an "organic-technical control loop" in which "manufacturing intelligence" typically becomes important (Popitz, 1995, p. 74). "In principle, this is the same circle of information and action that guides a computer-controlled anti-aircraft gun" (Popitz, 1995, p. 71). Popitz assumes that this control loop not only plays a role in the social evolution of human beings, but also in their biological developmental history. He writes: "If one considers the functioning of the technical-organic control loop, it becomes highly plausible to assume that experiences of technical action have contributed significantly to the biological formation of human beings" (Popitz, 1995, p. 72). But as "it has often not been possible to pass on acquired skills and knowledge," technological progress has also been rather slow (Popitz, 1995, p. 84). Popitz points out that the control loops of human system integration—the interaction between technology, humans, and communication—could play a role in the processes of biological and social evolution of human beings.[1]

Marean (2015) describes how, many millennia after the Schöningen spears, *Homo sapiens* set out from Africa to conquer the world with its weapon technology and, above all, with an "ultimate weapon": the ability to cooperate. Tomasello (2014) investigates the connection between co-operative hunting and fighting techniques and the development of language and cognition. Apparently, our tools are not without repercussions, but in turn influence our ability to think and act, which is sometimes described as "co-evolution" (Suhr, 2019). In general, assumptions about interactions between tool use and technology development and the development of social and communicative complexity are not uncommon in the sociology of technology (Popitz, 1995, pp. 78–138; Halfmann, 1996). Tool development and usage may even play a role in the "cognitive revolution" (Harari, 2015, pp. 22–28), which is said to have been a major breakthrough for *Homo sapiens* (Harari, 2015, pp. 23–24). It is noteworthy here that in sociology, it is assumed that social evolution is decoupled from biological evolution, so that "technological evolution can

[1] The use of technologies for the purpose of hunting (beating and throwing due to Popitz) is likely to have already taken place among the gatherers and hunters as an organized communicative interaction, although group cohesion and role differentiation may still have been rather weak compared to complex and differentiated societies (agriculture and the following). The attribution of the role of hunting to males and the implicit overemphasis on this role and the "struggle for existence" as an evolutionary factor is meanwhile viewed critically (Grupe et al., 2012, p. 57; Suhr, 2018, p. 13).

only be described as part of social evolution" (Halfmann, 1996, p. 98). However, this does not rule out the possibility that social evolution (which include the development of technology) an affect the biological constitution of human beings.

The Schöningen spears are an example of the fact that it is not enough to understand the physical principles of action, but that cognitive performance and cooperation are also important when competing with an intelligent opponent. Arthur describes key patterns of thought and action in which humans have increasingly understood natural phenomena and combined them to their advantage as "systems, interconnected, interacting, and mutually balanced" (Arthur, 2011, p. 38). Even today, segmentally differentiated indigenous people (Luhmann, 1984, p. 576) demonstrate hunting techniques that utilize a sophisticated cognitive influence, e.g., with the help of fire and smoke, before the weapon actually takes effect (Bird et al., 2005). Only recently has it been scientifically understood that animal cognition and hunting techniques interact in ways that resemble forms previously known only for *Homo* species (Wooster et al., 2024).

3 From Hunting to the Art of War and the Holistic Approach

With regard to the art of war, which may have developed in close interaction with the art of hunting (Eis, 1961; Popitz, 2017, pp. 125–126), it was understood early on in both Eastern and Western thought that cognitive factors play an essential role. Cognitive superiority can decide the future not only of an individual soldier or group, but of an entire army or society, even if the physical equipment is otherwise equivalent: "If you know the enemy and know yourself, your victory will not stand in doubt; if you know Heaven and know Earth, you may make your victory complete" (Tzu & Giles, 1910, pp. 112–113). It can be assumed that by heaven and earth, Sun Tzu means the nature of the landscape and terrain and the weather, but that he was also thinking well beyond this to the moral and spiritual realm. Many years later, Carl von Clausewitz wrote in *On War*—the book was published in 1832—that battle "is a trial of moral and physical forces through the medium of the latter. Naturally moral strength must not be excluded, for psychological forces exert a decisive influence on the elements involved in war" (Clausewitz, 1976, p. 127). It is noteworthy that Clausewitz aims at a relationship of interaction between mental and material factors (Clausewitz, 1976, p. 137). In this respect, one could say that he attempts a holistic approach.

Another 100 years later, intensive thought is being given by thinkers to when a war is "just" (e.g., Moseley, n.d.), and/or whether an advantage could be achieved with unlimited warfare that is unrestrained by any moral shackles (e.g., as "unrestricted warfare," Liang & Xiangsui, 2015). 'Unrestricted' refers to 'total,' as in 'total war.' The discourse on 'total war' followed the First World War and became a pattern of legitimation during National Socialist rule. The obvious problematic nature of such attempts to 'liberate' warfare and the use of force from moral, ethical, and humanitarian constraints makes it clear that thinking about the interplay of new technological possibilities and their constraints by law, ethics, and morality is of

considerable importance in modern defense and warfare. It also becomes increasingly clear that to understand this interplay, models, and perspectives beyond single weapons systems are necessary. It is important to note here that these more holistic models are NOT an alternative to traditional models, but complementary. We need good models on all levels of detail, from the single chemical molecule, screw, bolt, line of code, up to individual technologies, technical systems, human-machine systems, system-of-systems, organizations, societies toward the big system Earth and Space, in which we are all embedded.

A couple of recent concepts already point into this holistic direction: In response to these historical experiences and the Russian war of aggression in the Ukraine, the idea of "total defense," i.e., the networking of all civilian and military efforts for integrated deterrence and defense against an enemy, is being considered (Wither, 2020). In response to Russia's hybrid warfare, cognitive components of warfare are being intensively discussed as "cognitive warfare" which includes weapons, armies, but also organizations and societies (Masakowski & Blatny, 2023; Cluzel, 2021). In this context, the use and controllability of AI in defense is a major topic (van Diggelen & Draper, 2025). Controllability is understood here as meaningful human control (MHC) exercised by organizations and their human actors over technological or human-machine systems.[2]

The discussion in NATO on cognitive warfare and meaningful human control was the breeding ground for the holistic bowtie model. In this venture it became increasingly clear that NATOs system rival China is historically already thinking holistically for many millenniums as well, not only in terms of "unrestricted warfare" (Liang & Xiangsui, 2015), but even more so in societal and political terms: 天下 Tianxia is a Chinese term usually translated as "all under heaven" and refers not only to all creatures and/or the entire world, but also to a societal, cultural, and political concept to balance these interconnected systems in a way that they can flourish and be stable for a long time.

Tianxia is far from being just a historic concept. It is intensely used in the political debate in China (e.g., Tsang & Cheung, 2024), and it is increasingly adapted to modern times: Tingyang Zhao describes Tianxia not only historically, but also from a system perspective. A key point about Tianxia is that thinking about a (societal and political) system is expanded to such an extent that the distinction between the system and its environment no longer makes sense. Zhao describes that in this perspective, everything is considered inside, nothing is outside (Zhao, 2021). The world as a whole should be the starting point of an analysis to draft a global political order that transforms outer aspects to inner aspects. Zhao also describes the "gene of cooperation" as an essential part of Tianxia. This means that "co-existence precedes existence," i.e., cooperation seems to be more important to human beings than their own existence.

It is important to note here that for readers exclusively trained in western philosophy, ethics, and thinking, there are many hurdles which might provoke us to

[2] For an overview on this concept, see Mecacci et al. (2024).

reject the concept of Tianxia right away, giving away the chance to learn valuable insights from this concept. One stumbling block might be that the historic Tianxia had an emperor in the very center, who balances the different spheres between heaven and earth. This historic concept might spark the fear of democratic societies, having struggled for centuries to overcome or neutralize emperors, that a global Tianxia might become a global order under the rule of a new Chinese emperor-like leader. Even if Zhao (2021) makes it explicitly clear that Tianxia can also be applied to a multipolar political system, it could be vital for Western societies to take the tendencies of the PRC (Peoples Republic of China) to centralize even more under one powerful leader very seriously.

Another hurdle for Western thinkers trained in systems thinking, e.g., with Luhmann in mind, is the distinction between inside and outside, between system and environment. However, Zhao is aware of this challenge.

Whether we like Tianxia or not, we have to acknowledge that Tianxia is an ontologically rich thinking system, deeply rooted in history and at least from a Chinese perspective a viable thinking system to describe future global affairs. Reading Zhao cautiously and optimistically, the modern Tianxia also stresses the concept of co-existence and compatibility, and thus might open up a thinking and disputing space of mutual co-existence despite today's conflicts, e.g., about Taiwan. And even with a pessimistic view that the rivalry between western societies and China might increase and develop toward military conflicts, an understanding, and an answer of Western (and of global) thinkers to Tianxia is even more important.

Against the background of such contexts and experiences, the holistic bowtie diagram was gradually developed and trialed. It is not so deeply rooted into history and Eastern philosophy as Tianxia, but scientifically rooted into system thinking and cybernetics. The holistic bowtie model provides a stable bridge between the small and the large, and a controlled way to combine a perspective of systems and their environment with a perspective of nested systems, where larger meta-systems provide the environment for the smaller, embedded systems.

Thinking about the holistic bowtie model began with the realization that the above examples show, on the one hand, the advantage of cognitive 'thinking beyond' until everything can be thought of holistically 'all under heaven.' On the other hand, the examples also show that it still depends on the details, on the individual soldier, the individual team, and then on the connection of these systems with the higher-level systems such as the army, society, or alliances. How can the 'all under heaven' on the one side, and concrete weapon systems on the other side, how can global and local systems be thought of together in such a way that the connections that determine victory or defeat can be analyzed and designed in such a way that defeats are prevented and victories made possible? The bowtie model has several advantages in this context: Firstly, it enables an integrated analysis of trajectories across different system levels, i.e., it provides a holistic perspective without ignoring differences. Second, human–technology interactions can be systematically taken into account in the various system layers. Thirdly, it focusses on decisive risks and opportunities, dangers, and decision-relevant dilemma situations. Fourthly, it should be easy to

understand (plausibility) and fifthly, it opens up the design space to be operationalized technically as AI-based decision support.

4 Holistic Bowtie Model as a Macroscope

Theoretically speaking, the starting point for the bowtie diagram was hazard analysis, which has been used to analyze accidents in the oil industry since the 1970s. The main features of the original bowtie are (Figs. 2 and 3):

- Critical events are in the center. The mathematical, systemic theory behind it is the bifurcation theory, first mentioned by Henri Poincaré (1885), then generalized and described, for example, by Strogatz (1994). The theory focuses on the non-linear aspects of spatio-temporal causality, in which influence and the strength of influence are not distributed equally or linearly across space and time, but are concentrated in individual areas in such a way that these areas have significantly more influence than other areas.
- Preliminary aspects emphasize the temporal and/or causal relationship with the critical event ('before bang').
- Follow-up aspects clarify measures after the critical event ('after bang').

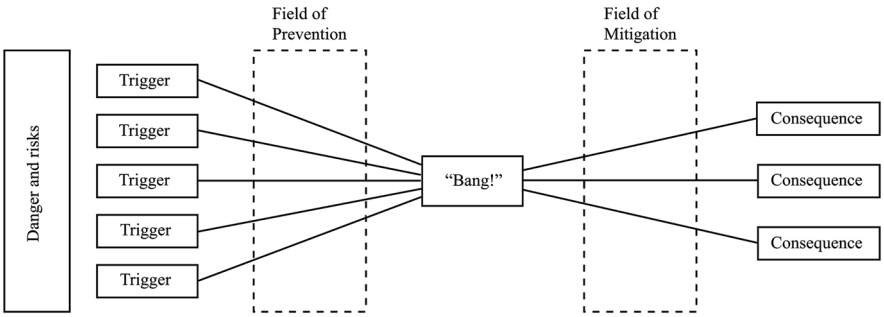

Fig. 2 Simplified bowtie diagram from the hazard analysis. When viewed from above, it looks like a butterfly (*source*: Rausand, 2011, p. 120, Fig. 5.3, own post editing)

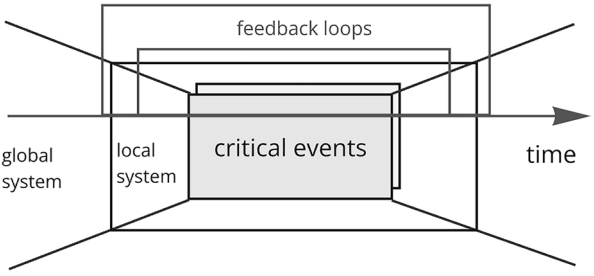

Fig. 3 Holistic bowtie model, generic example (*source*: Fraunhofer FKIE)

On the other hand, extensions to the holistic bowtie diagram were inspired by systems theory, e.g., systems engineering (SE, e.g., Haberfellner et al., 2019) or human systems integration (HSI, e.g., Booher, 2003):

- Systems are abstracted into subsystems and relationships.
- System layers are clearly identified and organized from the local to the global.
- Temporal and causal relationships are drawn as directed edges, similar to the feedback diagram of cybernetics (controller diagram).

By stringently connecting the local with the global, the holistic bowtie diagram combines the idea of the macroscope (Odum, 1983) with the synectics (e.g., Gordon, 1961). Metaphorically, the model can be compared with the perceptual characteristics of human vision: Human optical perception recognizes different resolution ranges between foveal and peripheral vision, which are assembled in the brain to form a coherent image of reality. What is obvious in the eye continues even more massively in our cognition, as models such as those by Wickens (2013) describe with the moderator "attention."

A further link to the bowtie model is the theory of bifurcation, which, based on chaos theory example of the butterfly (e.g., Lorenz, 1995, p. 134), describes that the relationships in the world are not linear, but that there are areas that have a greater influence than others. The critical events in hazard analysis often have a much greater influence on the fate of a company or a biosphere than other events. The main aim of bifurcation research is to describe the bifurcation points in such a way that they can be influenced, ideally controlled.

Figure 4 (upper part) shows a possible basic form of the holistic bowtie model, in which critical situations, events, and use cases are at the center, surrounded by layers of increasingly larger systems. This system is embedded in a system-of-systems, in organizations, which in turn are embedded in societies, which in turn are embedded in a system 'planet earth,' which is exposed to influences from outer space. In the context of this modeling, the choice of layers is a degree of freedom for the analyst to clarify the involvement of actors (humans and technologies), as well as their cooperation regarding their common goal and the constraints of the environment (Pacaux-Lemoine & Flemisch, 2019).

The model at the center of the example in Fig. 4 (lower part) is a dilemma model that can describe the cooperation between humans and automation or AI. Based on the model of human decision making by Boyd (1996), the model is expanded toward human-AI teaming, and the cooperation is described via two OODA loops, i.e., as an interlocking of observation (Observe), orientation (Orient), decision (Decide), and execution (Act). The dilemma situation is modeled in such a way that at the time of the decision on action or non-action it is not yet known whether it is a correct action or non-action or a non-correct action or non-action.

An example of the application of the dilemma model to AI technology in defense is a decision support system, e.g., for the defense of a military facility at an access gate. Based on deep neural networks, such a system could make recommendations about stopping or not stopping, fighting or not fighting, or other measures in the event of an approaching, as yet unknown, vehicle, or, depending on authorization,

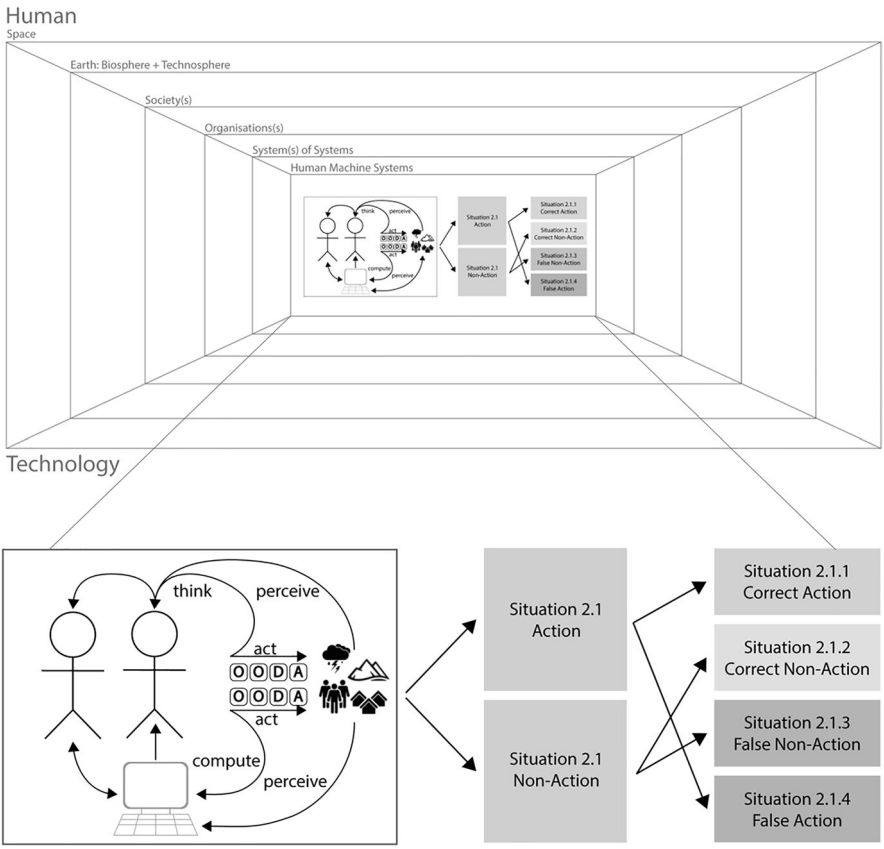

Fig. 4 Basic form of a holistic bowtie model, here with a dilemma model at the center (*source*: Fraunhofer FKIE)

even trigger them itself in the event of an immediate, strong, and well-identified threat (Flemisch et al., 2024). Another example of such AI-influenced dilemma situations are air defense systems, which already engage air targets fully automatically in certain situations (Flemisch & Nitsch, 2023).

5 Holistic Bowtie Model Applied to Defense

As Flemisch et al. (2012) already did for civilian systems, for military systems van Diggelen and Draper (2025) identify control as an essential system quality, which is now structured as effective and meaningful human control based on de Sio and van den Hoven (2018).

On this basis, Flemisch et al. (2024) outline the relationship between abilities, control, responsibility, and accountability: Based on an acceptable workload, good situational awareness, calibrated trust, and an ethical vision agreed with society, capabilities are built up and authorized for autonomous operativity. This process gives social agents, i.e., humans and organizations, decision-making responsibility. With sufficient abilities, authorization, and responsibility, control over situations is achieved, which in turn entails accountability of human actors who can be held accountable in the event of failure. This accountability is the result of a social negotiation process. Crucial to these chains of abilities, control, responsibility, and accountability are the necessary double and multiple ties: Whenever possible, control should be executed with sufficient ability and autonomy, authorization, and responsibility, and only then should it also trigger accountability. However, users should be able to trust in the technology and have sufficient information and competences to be held accountable. This in turn sets a framework for the use of technological developments in practice (Koch et al., 2024; Spreen, 2023). Situations should be avoided in which operators are nominally responsible but cannot exercise control at all due to insufficient skills or autonomy, as is also described as the "unsafe valley of automation" (Flemisch et al., 2017) or the "moral crumple zone" (Elish, 2019) when interacting with automation. The risk of delegating responsibility to machines (Krämer, 1992) in everyday, practical application without clarifying accountability in the organizations in which the machines are embedded must be avoided (Fig. 5).

The interlocking of these chains of ability, authority, control, responsibility, and accountability can also be shown across the different layers of the defense system. Thus, defense capabilities are always built incrementally by societies, organizations, human-technology systems, and individuals, starting from resources in the environment, before they can then be used in a defense situation. Decision-making authority also cascades from the global to the local and back again. This connection may seem trivial from the perspective of an individual who is already securely integrated in a hierarchical system, but it is essential for analyzing larger contexts, e.g., critical incidents, especially in conflicts between democracies and non-democratic actors. One example of this is 'false flag' operations or the deployment of irregular troops, for whose deployment state responsibility is then denied.

Fig. 5 Impact chains from moderators and central concepts via control to impact and accountability (*source*: Fraunhofer FKIE)

The example of trust is particularly critical here: if we analyze the chains more closely, we see that many decisions are made in the fog and under time pressure, relying on the actions of other actors. This trust is negotiated bi-directionally between local and global actors (e.g., calibrated trust according to Lee & See, 2004). If AI becomes one of the key cognitive actors in defense systems, trust or mistrust in AI will have a significant influence on its use.

Flemisch et al. (2024) demonstrate the application of the holistic bowtie diagram in the context of cognitive warfare. The system layers from the local to the global are connected via transversal layers, in this case a physical layer and a cognitive layer based on it. Cognition is already conceived here as connected cognition between humans and technology (e.g., AI), as has been conceptualized, for example, as a "joint cognitive system" (Hollnagel & Woods, 1983).

Cognitive warfare is primarily aimed at the cognitive layer, i.e., the combined cognition of humans and machines. To this end, cognitive warfare can attack and corrupt the chains of abilities, control, responsibility, and accountability. Trust in specific system actors can be shaken to such an extent that cooperation is severely disrupted.

6 From the Unipolar to the Bipolar Holistic Bowtie Model

The models presented so far already show a number of interrelationships, but completely ignore the fact that more than one party is involved in wars. To understand the interaction between system or even war opponents, the model is extended to include several parties.

Figure 6 shows a bipolar version of the bowtie diagram. The decisive factor in the modeled representation is the appropriate choice of differences on the one hand, but also of connections in order to work out the essential ones from the complexity of possible interactions on the other hand. It is important not to 'drown' in unimportant connections and at the same time to overlook crucial connections (complexity reduction).

7 Which Scientific Constructs Are Important and Sufficiently Connective to Address Holistic Defense?

A suitable selection of scientific constructs can make a decisive contribution to adequately mapping the complexity of issues without oversimplifying and thus overlooking crucial issues on the one hand, and without overloading the applicability of the model with too much complexity on the other. In an intensive discussion process on cognitive warfare (Flemisch et al., 2024), the following constructs were identified, which are also transferable to our discussion on AI.

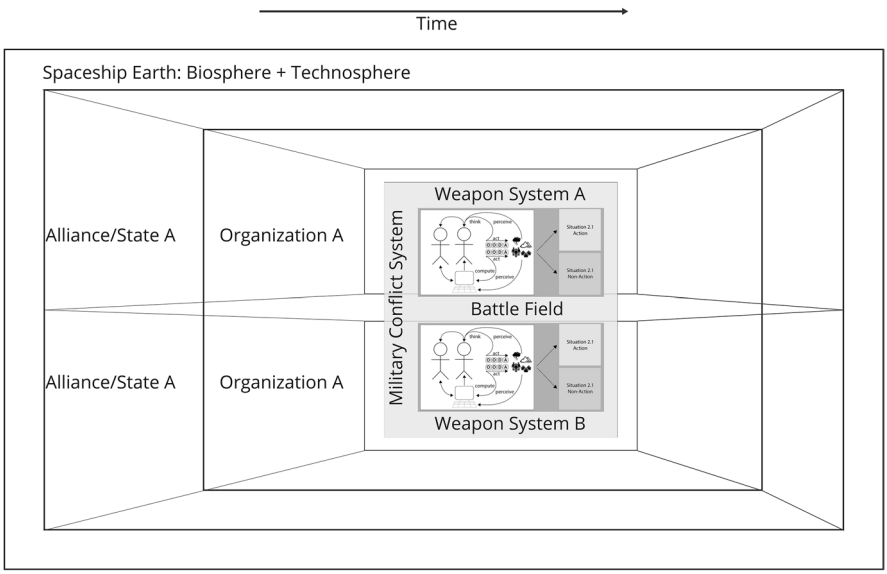

Fig. 6 Bi- and multipolar holistic bowtie model (*source*: Fraunhofer FKIE)

7.1 Cognition

Cognition is the process that produces intelligence. There are a number of valuable models and theories about cognition that are particularly useful for describing the intelligent, i.e., 'cognitive' interaction between people, organizations, technologies, and the environment. Thus, a strong branch of cognitive psychology has emerged for human cognition. The concept of cognition has not only been applied to computers, but the disciplines of human and machine cognition have been closely interwoven from early on, e.g., at the 1956 Symposium on Information Theory. One of the best-known examples is the work of Norbert Wiener (1988), who describes cybernetics as the control of machines or animals and already targets cognitive processes in this book, which has first been published in 1950. In his cybernetic concept, Wiener extends the cognitive idea to entire societies and their feedback loops. In 1959 Stafford Beer applied cognition to organizations, which is repeatedly taken up, e.g., as cognitive organization (Haun, 2016).

Hollnagel and Woods (1983) built on an early work by Jens Rasmussen and developed the concept of the joint cognitive system consisting of human and machine cognition, which can be analyzed and designed through cognitive systems engineering. Further landmarks in the understanding of cognition were to follow: in 1996, Hutchins describes, with the example of navigation tasks, the combination of individual cognition in cultural interaction with the environment, which enables a more sophisticated cognition with a significantly higher performance than

individual cognition. Hoc and Lemoine describe an actual cognitive approach to human–machine cooperation (Hoc & Lemoine, 1998; Hoc, 2001).

In sociological systems theory, cognition is closely linked to learning. A system (human or social) can react to disappointments in expectations either in a learning or normative way. In the first case, it learns to change its expectations "in order to match the disappointing reality" (Baraldi et al., 2021, p. 97). In the second case, it tries "to cling to the expectation despite the disappointing reality" (Baraldi et al., 2021, p. 97), i.e., it reacts in a normative way.

Based on Lovelock's GAIA hypothesis, Sam Palmisano, at that time Chairman of IBM, formulated a Smarter Planet Initiative in 2008 and called for a "smart decade" in which systems worldwide are to be networked and made cognitively efficient through the large-scale application of computer-based control loops. Buckminster Fuller (1969) developed a similar idea in his *Operating Manual for Spaceship Earth*. In 2022, based on Lovelock's GAIA hypothesis, astrophysicist Adam Frank speculated about intelligence on a planetary level, which is to be created through the further development of the technosphere and a synergetic connection with the biosphere in the form of connected cognitive processes (Valich, 2022).

With the symposium *Cognitive Warfare: The Future of Cognitive Dominance*, the cognitive paradigm was also used for defense in 2021. The need for cognitive superiority is emphasized. In 2022, Frank Flemisch warned in a NATO specialist meeting on cognitive warfare in Kjeller that striving for cognitive dominance could set off spiraling arms races between the USA and China that would be difficult to control, and instead proposed a goal of "cognitive eye level."

The discussion about cognitive superiority versus cognitive eye level remains closely intertwined with the discussion about the technological singularity, described by Ray Kurzweil (2005) as the point at which machine intelligence develops so quickly that it outstrips human intelligence. Kurzweil expects technology, biology, and intelligence to merge into machine intelligence (Kurzweil, 2005, p. 9, 358). The idea of a singularity is widely discussed, e.g., in Nick Bostrom's book *Superintelligence* (Bostrom, 2014). In the first decade of the idea, it was still largely ridiculed due to the cognitive weakness of many people's perception of exponential relationships (exponential growth bias), but since the breakthroughs in generative networks at the latest, the possibility has been taken seriously. The Covid-19 pandemic made a positive contribution to this, making the exponential growth described by Kurzweil and Bostrom understandable to larger sections of the population. Wallach and Allen (2008) take up this discussion about exponential growth and describe AI in their book on *Moral Machines* as a "fire" whose spread should be systematically controlled.

The idea of a cognitive eye level is contrasted with considerations of a "hyperwar" (Husain et al., 2018). The option of highly accelerated warfare is being discussed, in which an opponent is overwhelmed through the massive use of AI and autonomous capabilities. At the same time, cognitive speed raises major problems of controllability.

7.2 Emotions

Structuring the AI discussion as human and machine cognition could give the impression that purely rational arguments play a predominant role. However, it is known from cognitive research that feelings and sensations have a significant influence on decision-making and that these interact between the cerebrum and the rest of the body, as described by Damasio (1994) as embodiment and somatic markers. Research is now also being conducted in the technical field under the label 'Embodied AI' into the extent to which AI can also be embodied (Zia, 2024). While part of the research community sees emotions and cognition as complementary in a dualistic tradition (e.g., Draguhn, 2013), other parts point out that emotions are an integral part of cognitive processes (Foolen et al., 2012). Since emotions are obviously an integral part of public discourse and political and/or military decision-making (e.g., NATO ACT, 2024), we argue for an integrated, systemic view of emotions as part of cognitive processes, in which all layers from the individual to groups, organizations, and societies interact and resonate, as e.g., Rosa (2016) argues.

7.3 The Paradox of Truth

The discussion about emotions as part of cognitive processes has already shown that there is by no means a uniform use of this construct in the community, but rather different currents, some of which tend to stick to older approaches, while others explore newer approaches in the hope of pragmatic success and increasing recognition. The same applies to the construct of truth: one part of the community sees truth as something absolute to which research should refer. The picture is quite similar with regard to the problem of truth. For example, Einstein (1930, p. 193) described "goodness, beauty, and truth" as essential ideals that determine his thoughts and actions. Later, Heinz von Foerster provoked with the formulation "Truth is […] the invention of a liar" (Foerster & Pörksen, 2022, pp. 29–30), which leads to a paradox (Molter, 2017, p. 13). This discussion has led to a systematic investigation of how and why individuals, organizations, and societies identify something as truth (e.g., Glasersfeld, 1997). For example, Luhmann sees truth as the medium of the (communicative) scientific system and insists on the separation of system and environment: "No communication ever communicates the world. Communication does not communicate the world, it divides it into that which it communicates and that which it does not communicate" (Luhmann, 1990, p. 27). 'Consistency' between 'environment' and 'communication' is therefore an internal system model. System and environment do not come together—regardless of the type of system involved. If this were the case, the system would no longer be a system, and the environment would no longer be an environment. Ironically, Tianxi wants to overcome this paradox.

From the perspective of today's Western social systems theory, such an 'overcoming' is not possible.[3]

7.4 Sensemaking, Plausibility, and Accuracy

Weick et al. (2005) described the process of "sensemaking" that people, organizations, and societies go through in order to agree on something like a common understanding of truth. Crucial parts of this are:

- Retrospection back in time,
- representation of people through dialogues and narratives, including the social activities involved in exchanging narratives, and
- extracting clues from the context to select relevant information.

An important statement by Weick (1995) is that people seem to favor the plausibility of explanations over accuracy. This trade-off seems to be particularly decisive in the combination of human and machine cognition: "In an equivocal, postmodern world, infused with the politics of interpretation and conflicting interests and inhabited by people with multiple shifting identities, an obsession with accuracy seems fruitless, and not of much practical help, either" (Laroche, 1996).[4]

7.5 Trust and Mistrust

Establishing a meaningful resonance with the world and somehow creating a 'truth' and 'meaning' seems to be at the center of cognitive and emotional processes. An essential construct here especially when thinking and acting in the context of crises and war appears to be the trust that people, machines, organizations, or social communication contexts have in other actors or in information. From a systemic perspective, trust is the internal assessment of a quality, i.e., it is subjective per se. A key aspect of this is that there is always a gap between the complexity of future possibilities and the internal expectation, which trust in turn bridges but does not close. Luhmann describes trust as a mechanism in organizations or societies for reducing social complexity: trust reduces the "complexity of the future world" (Luhmann, 1979, p. 20). "[N]o decisive grounds can be offered for trusting; trust always extrapolates from the available evidence; it is [...] a blending of knowledge and ignorance. [...] Trust remains a risky undertaking" (Luhmann, 1979, pp. 26).

[3] Rather, this could lead to a pathological love of the world or such thing, to cite Sigmund Freud: "A love that does not discriminate seems to me to forfeit a part of its own value, by doing an injustice to its object" (Freud, 1962, p. 49).

[4] Translated from French.

In ergonomics and human systems integration, the construct of trust was originally used primarily to generate the highest possible trust of humans in the machine as well. It was not until Lee and See (2004) that a concept of calibrated trust was described for humans and automation, which should be neither too high nor too low, and which can be gradually transferred to human and technical cognition across all layers of a holistic system view.

Especially in military systems with their time-critical decision-making systems under risk, trust plays a decisive role in thinking and acting. Trust is obviously built up slowly, but can be quickly disrupted or destroyed, which further emphasizes the attractiveness of 'mistrust-building' measures, e.g., through false flag operations. This is possible because trust and mistrust are functionally equivalent (Luhmann, 1979, p. 71), i.e., these concepts answer to similar problems, but here with opposite directions.

8 Systemic Feedback Loops in Defense Systems

All of the individual constructs described above have a substantial effect on defense capability and are already being researched in their own right. But there is a major gap in research and action with regard to the interaction of these factors as feedback loops or processes: Cognitive incl. Emotional processes generate, based on authorization and capabilities, truths, meaning, trust or distrust, and actions based on these, for which in turn people and organizations must take responsibility. These processes have always been vulnerable to adversarial influence and are now being massively changed by the use of artificial intelligence (AI).

This need for a systemic view as feedback loops is already indicated by two other megatrends, and not just in defense: In addition to the ongoing mega-trend toward artificial intelligence, there is a strong trend toward the interlinking of different areas of life under the label 'hybrid war.' Gerasimov's article *The Value of Science Is in the Foresight*, which describes new forms and methods of warfare as the "use of political, diplomatic, economic, and other nonmilitary measures in combination with the use of military forces," is often seen as a landmark for this development (Gerasimov, 2016, p. 25). Against the backdrop of the annexation of Crimea in violation of international law and the Russian attack on the Ukraine, Gerasimov's article is read as a blueprint for a fake news war against Western democracies with the intention of eroding trust in the political system. Whatever the case, the issue of trust is currently at the center of strategic communication (Zowislo-Grünewald, 2025). The interweaving of different areas of life is playing an increasingly important role in defense—an insight that is not new and can already be found at one of the founders of Inner Leadership (Innere Führung), Wolf Graf von Baudissin, when he considers "stable, balanced conditions" to be an essential prerequisite for effectively countering "subversive warfare," because then there are "neither worthwhile targets nor the necessary support from the surrounding population" (Baudissin, 1969, p. 67).

The holistic bowtie diagram can now be used to precisely describe, operationalize, and in the long term automate the interdependencies of different system layers in their interactions, which could be used to develop a manageable tool to counter hybrid threats and risks in good time. Of course, it must be taken into account that the actions and communication of one's own side can very well give rise to the erosion of trust and mistrust. It would be too short-sighted to blame a hybrid opponent for everything that goes wrong. But every opponent is happy to take advantage of steep slopes.

Another trend that is self-reinforcing due to technological advances such as networking and computing power is the trend toward acceleration, which is described, for example, under the label "hyperwar" (Husain et al., 2018). Although acceleration and speed are a recurring motif in defense and warfare, the acceleration effects of technical mediators such as the internet or AI must be taken so seriously that our thought models must take them into account at a strategic point. In particular, if acceleration is to be achieved by shifting parts of the process to the machine, the chains of authorization, capabilities, trust, meaning, control, and responsibility (Fig. 7) must be specifically designed to prevent the chains from breaking. This could lead to a 'transhuman' constellation, which is very problematic, because accountability (*Verantwortlichkeit*) can no longer be attributed to human actors in social roles, thus risking anomie, i.e., a state of missing or weak social norms, rules, and order (Spreen, 2023, p. 25).

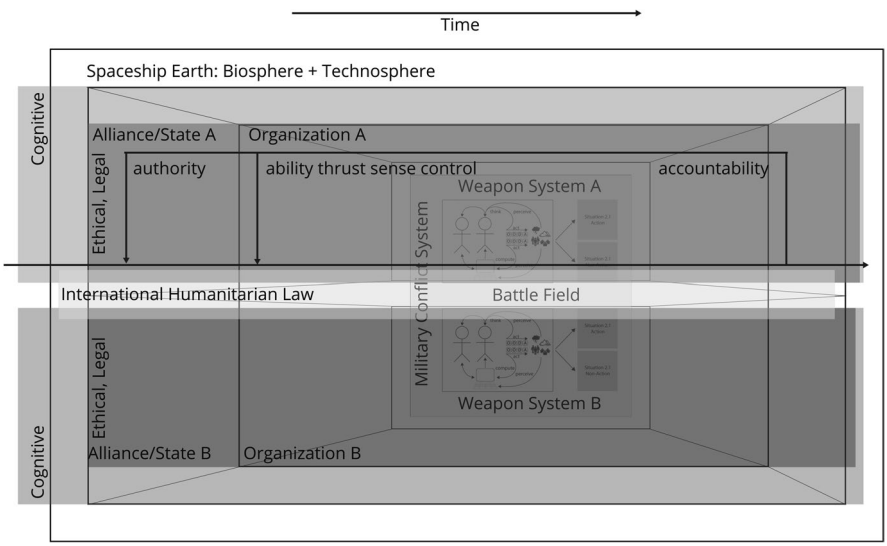

Fig. 7 Bipolar holistic bowtie diagram of cognitive warfare (*source*: Flemisch et al., 2024, Fraunhofer FKIE)

9 Outlook: Agile Exploration and Development of these Chains Toward Holistic Defense in a Strong Alliance

The acceleration described by Rosa (2016) or Husain et al. (2018) not only affects direct defense or warfare, but also the development and adaptation of capability, control, and responsibility chains. Based on historical examples, Flemisch et al. (2022) describe a concept of agile exploration of defense systems that precedes series development in order to identify, develop, and deploy the most promising combinations from the diverse combinations of defense material. A speeding up of exploration and development chains can currently also be observed in the Ukraine and will spread from there to other political defense systems. AI in conjunction with other technologies such as networking, ubiquitous computing, VR and, with a time delay, quantum computing will probably result in new technological pushes that could put social systems (especially politics, the economy, social security, and the military) under massive pressure. The combination of this often exponentially growing technology appears to be a particular challenge:

> As technology continues to exponentially accelerate, the interactions between various subsets of exponential technology will create opportunities to slingshot past the already breakneck speed of accelerating change in ways that are even stranger and more difficult to predict than the path of any individual exponential technology. (Ganz, 2016)

"Breakneck speed" could indicate significant risks or dangers here, for example if hypersonic weapons in conjunction with nuclear warheads and AI significantly destabilize the security situation and the political system reacts too slowly, suppression mechanisms prevail or the interrelationships between the various system layers are ignored. Thinking tools such as the holistic bowtie diagram can help to better understand the interrelationships and act accordingly, but they are only one piece of the puzzle on the way to a resilient security concept or perhaps even peace.

Although wars are always a possibility due to human beings' "ability to harm" and "exposedness to harm" (Popitz, 2017, p. 26), the aim should be to make violent conflict and war unattractive and thus less likely. A key issue is to drive up the costs for any violent aggressors. One should keep in mind what Luhmann formulated from a systemic perspective: "reducing the *danger* of war by increasing the *risk* of war" for potential aggressors (Luhmann, 1993, p. 105, footnote 9, emphasis in original). A strong, defensive alliance which can share costs and counterbalance any military or systemic aggression on a global level is the best respond to the threat of war.

- In the context of disaster analysis, the holistic bowtie model enables an integrated analysis of trajectories across different system levels, i.e., it offers a holistic perspective without ignoring differences.
- The strength of this model is that it keeps an eye on different system levels and their interactions.
- In relation to AI, particular attention is paid to the interlinking of abilities, control, responsibility, and accountability.
- Cognition, emotion, sensemaking, and trust are important influencing factors.

References

Arthur, W. B. (2011). *The nature of technology. What it is and how it evolves*. (paperback ed.). Free Press.

Baraldi, C., Corsi, G., & Esposito, E. (2021). *Unlocking Luhmann. A keyword introduction to systems theory* (K. Walker, Trans.). Transcript. https://doi.org/10.14361/9783839456743

Baudissin, W. G. V. (1969). *Soldat für den Frieden. Entwürfe für eine zeitgemäße Bundeswehr*. (P. V. Schubert, Eds.). Piper.

Bird, D. W., Bird, R. B., & Parker, C. H. (2005). Aboriginal burning regimes and hunting strategies in Australia's Western Desert. *Human Ecology, 33*, 443–464. https://doi.org/10.1007/s10745-005-5155-0

Booher, H. (Ed.). (2003). *Handbook of human systems integration*. Wiley.

Bostrom, N. (2014). *Superintelligence. Paths, dangers, strategies*. Oxford University Press.

Boyd, J. R. (1996). *The essence of winning and losing* (Unpublished briefing).

Clausewitz, C. V. (1976). *On war* (M. Howard, & P. Paret, Ed. and Trans.). Princeton University Press (Original work published 1832).

Cluzel, F. D. (2021). *Cognitive warfare, a Battle for the brain* (STO-MP-AVT-211). NATO Science and Technology Organization. https://www.sto.nato.int/publications/STO%20Meeting%20Proceedings/STO-MP-HFM-334/$MP-HFM-334-KN3.pdf.

Damasio, A. (1994). *Descartes' error. Emotion, reason and the human brain*. Avon.

Draguhn, A. (2013). Das Verhältnis von Emotion und Kognition aus Sicht der Hirnforschung. In S. Höfling & F. Tretter (Eds.), *Homo neurobiologicus. Ist der Mensch nur sein Gehirn?* (pp. 51–57). Hanns-Seidel-Stiftung e.V. https://www.hss.de/publikationen/homo-neurobiologicus-pub109

Einstein, A. (1930). What I Belive. Living philosophies XIII. *Forum and Century, 84*(4), 193–194.

Eis, G. (1961). Die Stellung der Jagd im mittelalterlichen System der Wissenschaften. *Zeitschrift für Jagdwissenschaft, 7*, 25–28. https://doi.org/10.1007/BF01956291

Elish, M. C. (2019). Moral crumple zones: Cautionary Tales in human-robot interaction. *Engaging Science, Technology, and Society, 5*, 40–60. https://doi.org/10.17351/ests2019.260

Ferguson, M. (2023). Toward a clinical neurospirituality. *Biological Psychiatry, 93*(9), S2. https://doi.org/10.1016/j.biopsych.2023.02.025

Flemisch, F., & Nitsch, V. (2023). Kooperative Systeme und hybride Intelligenz Plädoyer für ganzheitliche Mensch-Maschine-Integration. In N. Lammert & W. Koch (Eds.), *Bundeswehr der Zukunft. Verantwortung und Künstliche Intelligenz* (pp. 237–250). Konrad-Adenauer-Stiftung.

Flemisch, F., Heesen, M., Hesse, T., Kelsch, J., Schieben, A., & Beller, J. (2012). Towards a dynamic balance between humans and automation: Authority, ability, responsibility, and control in shared and cooperative control situations. *Cognition, Technology & Work, 14*, 3–18. https://doi.org/10.1007/s10111-011-0191-6

Flemisch, F., Altendorf, E., Canpolat, Y., Weßel, G., Baltzer, M., Lopez, D., Herzberger, N. D., Voß, G. M. I., Schwalm, M., & Schutte, P. (2017). Uncanny and unsafe valley of assistance and automation: First sketch and application to vehicle automation. In C. M. Schlick, S. Duckwitz, F. Flemisch, M. Frenz, S. Kuz, A. Mertens, & S. Mütze-Niewöhner (Eds.), *Advances in ergonomic design of systems, products and processes. Proceedings of the annual meeting of GfA 2016* (pp. 319–334). Springer. https://doi.org/10.1007/978-3-662-53305-5_23

Flemisch, F., Preutenborbeck, M., Kehl, C., Grünwald, C., Wasser, J., Baltzer, M., & Dahlmann, A. (2022). Human systems exploration for ideation and innovation in potentially disruptive defense and security systems. In G. Adlakha-Hutcheon & A. J. Masys (Eds.), *Disruption, ideation and innovation for defence and security. Advanced sciences and technologies for security applications* (pp. 79–117). Springer. https://doi.org/10.1007/978-3-031-06636-8_5

Flemisch, F., Spreen, D., Saariluoma, P., Gitanjali A.-H., Knox, B. J., Talves, K., & Brill, J. C. (2024). A holistic bowtie model of cognitive warfare: An interplay of technology, humans, societies and environment. In The NATO Science and Technology Organization

(Ed.), *Mitigating and responding to cognitive warfare. HFM-361 symposium held on 13–14 November, 2023 in Madrid, Spain*. STO-Meeting Proceedings Paper (STO-MP-HFM-361). NATO Science and Technology Organization. https://www.sto.nato.int/publications/STO%20Meeting%20Proceedings/STO-MP-HFM-361/MP-HFM-361-P04.pdf

Foerster, H. V., & Pörksen, B. (2022). *Wahrheit ist die Erfindung eines Lügners. Gespräche für Skeptiker* (13th ed.). Carl-Auer.

Foolen, A., Lüdtke, U., & Schwarz-Friesel, M. (2012). Kognition und emotion. In O. Braun & U. Lüdtke (Eds.), *Sprache und Kommunikation – Behinderung, Bildung und Partizipation* (pp. 213–229). Kohlhammer.

Freud, S. (1962). *Civilization and its discontents* (J. Strachey, Trans.). W. W. Norton & Company (Original work published 1930).

Fuller, R. B. (1969). *Operating manual for spaceship earth*. Simon and Schuster.

Ganz, J. (2016, September 29). No technology thrives alone: Progress is all about convergence. *singularityhub*. https://singularityhub.com/2016/09/29/no-technology-thrives-alone-progress-is-all-about-convergence

Gerasimov, V. (2016). The value of science is in the foresight: New challenges demand rethinking the forms and methods of carrying out combat operations (R. Coalson, Trans.). *Military Review*, *96*(1), 23–29. https://www.armyupress.army.mil/Journals/Military-Review/English-Edition-Archives/January-February-2016

Glasersfeld, E. V. (1997). *Radikaler Konstruktivismus. Ideen, Ergebnisse, Probleme*. Suhrkamp.

Gordon, W. J. J. (1961). *Synectics. The development of creative capacity*. Harper & Row.

Grupe, G., Christiansen, K., Schröder, I., & Wittwer-Backofen, U. (2012). *Anthropologie. Einführendes Lehrbuch* (2nd ed.). Springer. https://doi.org/10.1007/978-3-642-25153-5

Haberfellner, R., Weck, O. D., Fricke, E., & Vössner, S. (2019). *Systems engineering: Fundamentals and applications*. Birkhäuser, Springer Nature. https://doi.org/10.1007/978-3-030-13431-0

Halfmann, J. (1996). *Die gesellschaftliche "Natur" der Technik. Eine Einführung in die soziologische Theorie der Technik*. Leske + Budrich.

Harari, Y. N. (2015). *Sapiens. A brief history of humankind*. (imprint). Vintage Books.

Haun, M. (2016). *Cognitive Organisation. Prozessuale und funktionale Gestaltung von Unternehmen*. Springer Vieweg. https://doi.org/10.1007/978-3-662-52952-2

Hoc, J. M. (2001). Towards a cognitive approach to human-machine cooperation in dynamic situations. *International Journal of Human Computer Studies, 54*(4), 509–540. https://doi.org/10.1006/ijhc.2000.0454

Hoc, J. M., & Lemoine, M. P. (1998). Cognitive evaluation of human-human and human-machine cooperation modes in air traffic control. *International Journal of Aviation Psychology, 8*(1), 1–32. https://doi.org/10.1207/s15327108ijap0801_1

Hollnagel, E., & Woods, D. D. (1983). Cognitive systems engineering: New wine in new bottles. *International Journal of Man-Machine Studies, 18*(6), 583–600. https://doi.org/10.1016/S0020-7373(83)80034-0

Husain, A., Allen, J., Work, R., Cole, A., Scharre, P., Porter, B., Anderson, W., & Townsend, J. (2018). *Hyperwar. Conflict and competition in the AI century*. SparkCognition Press.

Koch, W., Spreen, D., Talves, K., Wagner, W., Lillemäe, E., Klaus, M., Viidalepp, A., Cooper, C. G., & Pekarev, J. (2024). On the ethics of employing artificial intelligent automation in military operational contexts. *IEEE Transactions on Technology and Society, 5*(2), 231–241. https://doi.org/10.1109/TTS.2024.3405309

Krämer, S. (1992). Symbolische Maschinen, Computer und der Verlust des Ethischen im geistigen Tun. In W. Coy, F. Nake, J.-M. Pflüger, A. Rolf, J. Seetzen, D. Siefkes, & R. Stransfeld (Eds.), *Sichtweisen der Informatik. Theorie der Informatik* (pp. 335–341). Vieweg. https://doi.org/10.1007/978-3-322-84926-7_23

Kurzweil, R. (2005). *The singularity is near. When humans transcend biology*. Viking Penguin.

Laroche, H. (1996). Karl E. Weick (1995), Sensemaking in organizations, Sage, Thousand Oaks, Californie (Recension). *Sociologie du travail, 38*(2), 225–232. https://doi.org/10.3406/sotra.1996.2274

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Liang, Q., & Xiangsui, W. (2015). *Unrestricted warfare: China's master plan to destroy America*. Echo Point Books & Media. (Original work published 1999).
Lorenz, E. N. (1995). *The essence of chaos*. U. Washington Press.
Luhmann, N. (1979). *Trust and power. Two works* (G. Poggi, Intro.). John Wiley & Sons.
Luhmann, N. (1984). *Soziale Systeme. Grundriß einer allgemeinen Theorie*. Suhrkamp.
Luhmann, N. (1990). *Die Wissenschaft der Gesellschaft*. Suhrkamp.
Luhmann, N. (1993). *Risk: A sociological theory* (R. Barrett, Trans.). de Gruyter.
Marean, C. W. (2015). The most invasive species of all. *Scientific American, 313*(2), 32–39. https://www.jstor.org/stable/26046104
Masakowski, Y. R., & Blatny, J. M. (Eds.). (2023). *Mitigating and responding to cognitive warfare*. STO Technical Report (STO-TR-HFM-ET-356). NATO Science and Technology Organization. https://www.sto.nato.int/publications/STO%20Technical%20Reports/STO-TR-HFM-ET-356/$$TR-HFM-ET-356-ALL.pdf
Mecacci, G., Amoroso, D., Siebert, L. C., Abbink, D., Hoven, J. V. D., & Sio, F. S. D. (Eds.). (2024). Research handbook on meaningful human control of artificial intelligence systems. Edward Elgar.
Milks, A., Parker, D., & Pope, M. (2019). External ballistics of Pleistocene hand-thrown spears: Experimental performance data and implications for human evolution. *Scientific Reports, 9*., Article 820. https://doi.org/10.1038/s41598-018-37904-w
Molter, H. (2017). *Wenn Wahrheit die Erfindung eines Lügners ist, dann ist Heinz von Förster ein Lügner*. https://systemagazin.com/wp-content/uploads/2018/07/Wahrheit-12.pdf
Moseley, A. (n.d.). *Just war theory*. Internet Encyclopedia of Philosophy (IEP). https://iep.utm.edu/justwar
NATO ACT. (2024). *Cognitive warfare*. NATO Allied Command Transformation. https://www.act.nato.int/activities/cognitive-warfare
Odum, E. P. (1983). *Basic ecology*. CBS College Publishing.
Pacaux-Lemoine, M.-P., & Flemisch, F. (2019). Layers of shared and cooperative control, assistance, and automation. *Cognition, Technology & Work, 21*, 579–591. https://doi.org/10.1007/s10111-018-0537-4
Pohlmann, F. (2000). *Die soziale Geburt des Menschen. Einführung in die Sozialpsychologie und Anthropologie der frühen Kindheit*. Beltz.
Poincaré, H. (1885). L'Équilibre d'une masse fluide animée d'un mouvement de rotation. *Acta Mathematica, 7*, 259–380. https://doi.org/10.1007/BF02402204
Popitz, H. (1995). *Der Aufbruch zur Artifiziellen Gesellschaft. Zur Anthropologie der Technik*. J. C. B. Mohr.
Popitz, H. (2017). *Phenomena of power. Authority, domination, and violence* (G. Poggi, Trans., A. Göttlich, & J. Dreher, Eds.). Columbia University Press.
Rausand, M. (2011). *Risk assessment. Theory, methods, and applications*. John Wiley & Sons. https://doi.org/10.1002/9781118281116
Rosa, H. (2016). *Resonanz. Eine Soziologie der Weltbeziehung*. Suhrkamp.
Sio, F. S. D., & Hoven, J. V. D. (2018). Meaningful human control over autonomous systems: A philosophical account. *Frontiers in Robotics and AI*, 5. https://doi.org/10.3389/frobt.2018.00015
Spreen, D. (2023). Lethal autonomous weapon systems (LAWS). On the ethics of automation in the military from the perspective of social systems theory. *Sõjateadlane (Estonian Journal of Military Studies)*, (21), 10–40. https://doi.org/10.15157/st.vi21.24177
Strogatz, S. H. (1994). *Nonlinear dynamics and chaos. With applications to physics, biology, chemistry, and engineering*. Perseus Books.
Suhr, D. (2018). *Das Mosaik der Menschwerdung. Vom aufrechten Gang zur Eroberung der Erde: Humanevolution im Überblick*. Springer. https://doi.org/10.1007/978-3-662-56830-9
Suhr, D. (2019). Ko-Evolution von Mensch und Technik. Bio- und technikphilosophische Perspektiven. In A. F. Koch, S. Kruse, & P. Labudde (Eds.), *Zur Bedeutung der technischen Bildung in Fächerverbünden. Multiperspektivische und interdisziplinäre Beiträge aus Europa* (pp. 159–172). Springer. https://doi.org/10.1007/978-3-658-25623-4_12
Tomasello, M. (2014). *A natural history of human thinking*. Harvard University Press.
Tsang, S., & Cheung, O. (2024). *The political thought of Xi Jinping*. Oxford University Press. https://doi.org/10.1093/oso/9780197689363.002.0003

Tzu, S., & Giles, L. (1910). *Sun Tzu on the art of war. The Oldest military treatise in the world* (L. Giles, Trans.). Luzac (Original work published around 500 BC).

Valich, L. (2022, February 16). *Can a planet have a mind of its own?* University of Rochester–News Center. https://www.rochester.edu/newscenter/planetary-intelligence-evolution-thought-experiment-510542

van Diggelen, J. V., & Draper, M. (2025). *Final report of RTG-HFM-330 "Meaningful human control over ai based systems"*. NATO STO (In press).

Wallach, W., & Allen, C. (2008). *Moral machines. Teaching robots right from wrong*. Oxford University Press.

Wasser, J., & Flemisch, F. (2020). In F. Flemisch, *Lecture on balanced human systems integration*. RWTH Aachen University (Lecture series).

Weick, K. E. (1995). *Sensemaking in organizations*. Sage.

Weick, K. E., Sutcliffe, K. M., & Obstfeld, D. (2005). Organizing and the process of sensemaking. *Organization Science, 16*(4), 409–421. https://doi.org/10.1287/orsc.1050.0133

Wickens, C. D. (2013). Attention. In J. D. Lee & A. Kirlik (Eds.), *The Oxford handbook of cognitive engineering* (pp. 36–56). Oxford University Press.

Wiener, N. (1988). *The human use of human beings: Cybernetics and society*. Da Capo Press (Original work published 1950 and revised in 1954).

Wither, J. K. (2020). Back to the future? Nordic Total Defence concepts. *Defence Studies, 20*(1), 61–81. https://doi.org/10.1080/14702436.2020.1718498

Wooster, E. I. F., Gaynor, K. M., Carthey, A. J. R., Wallach, A. D., Stanton, L. A., Ramp, D., & Lundgren, E. J. (2024). Animal cognition and culture mediate predator–prey interactions. *Trends in Ecology & Evolution, 39*(1), 52–64. https://doi.org/10.1016/j.tree.2023.09.012

Zhao, T. (2021). *All under heaven. The Tianxia system for a possible world order*. (J. E. Harroff, Trans.). University of California Press. https://doi.org/10.1525/9780520974210

Zia, T. (2024, November 15). *Advancing embodied AI: How Meta is bringing human-like touch and dexterity to AI*. Unite.ai. https://www.unite.ai/advancing-embodied-ai-how-meta-is-bringing-human-like-touch-and-dexterity-to-ai/

Zowislo-Grünewald, N. (2025). Hybrid warfare and the defense of discourse. In K. Talves & D. Spreen (Eds.), *Artificial intelligence in military technology: Sociological, cultural and ethical perspectives* (pp. 211–225). Springer Nature. https://doi.org/10.1007/978-3-031-95578-5_13