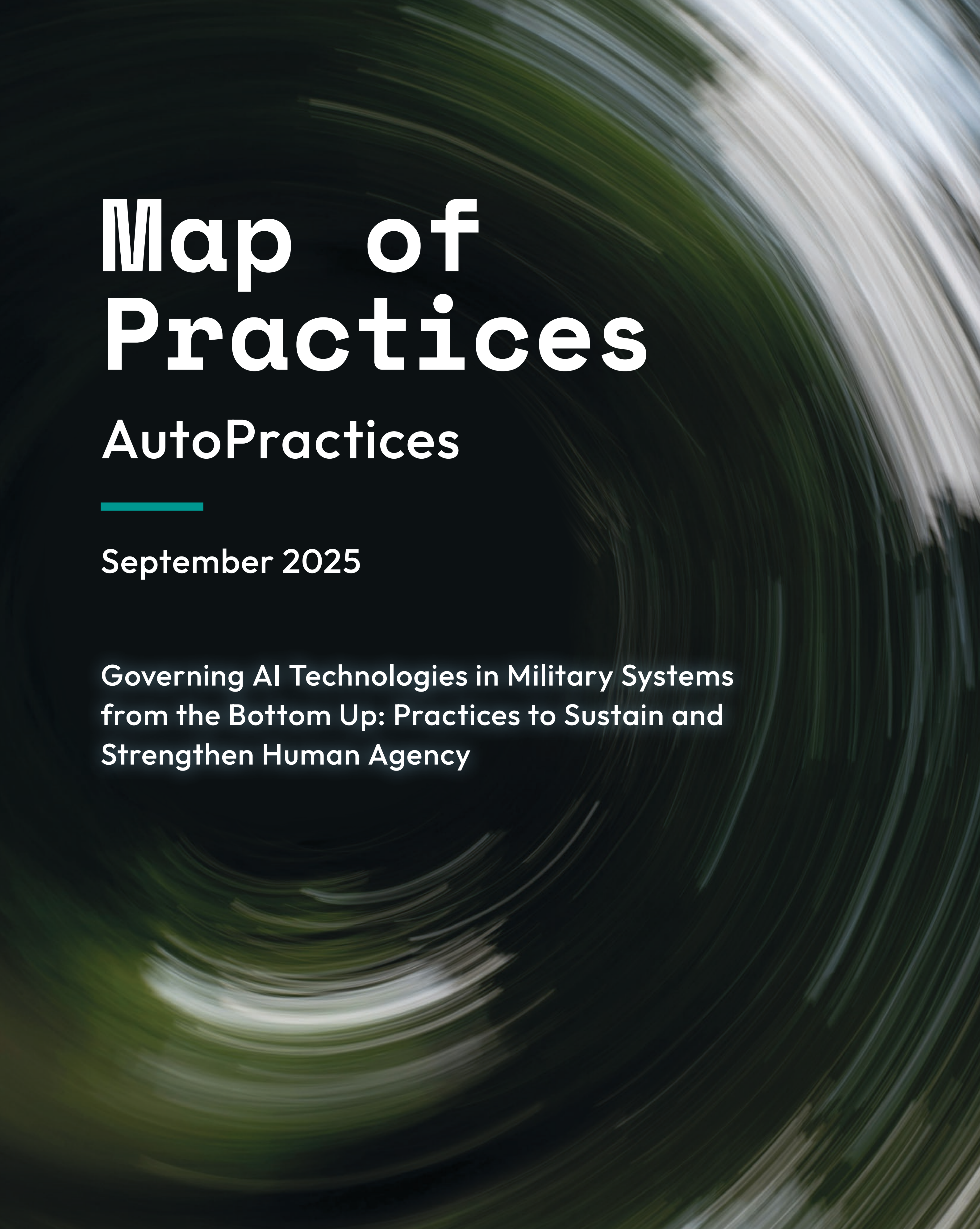


Map of Practices

AutoPractices

September 2025

Governing AI Technologies in Military Systems from the Bottom Up: Practices to Sustain and Strengthen Human Agency

CENTER FOR WAR STUDIES

sipri

AutoNorms

SDU

About the Center for War Studies

The Center for War Studies (CWS), established at the University of Southern Denmark (SDU) in 2012, brings together academics from political science, law, history, and cultural studies to contribute to major debates on the past, present, and future of war, as well as its impact on societies. We strive for interdisciplinary research that is relevant to policymakers and the society at large. We aim to contribute to ongoing debates on war and peace by illuminating their multiple dimensions. Through research excellence and societal relevance, we advance the understanding of the fundamental issue of war and peace. For more information about the CWS and its researchers, see https://www.sdu.dk/en/cws.

Funding

AutoPractices is a Proof of Concept project funded by the European Research Council (grant number 101156237). It runs from June 2024 to December 2025. It is based on the project "Weaponised Artificial Intelligence, Norms and Order" (AutoNorms), which has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 852123). For more information about AutoNorms and AutoPractices, see www.autonorms.eu.

Acknowledgements

The AutoPractices team is grateful to all stakeholders who participated in this research.

The AutoPractices team

Principal Investigator

Prof. Ingvild Bode, Professor of International Politics and Director of the Center for War Studies, University of Southern Denmark (SDU)

Research Team

- Dr. Anna Nadibaidze, Postdoctoral Researcher, SDU
- Dr. Alexander Blanchard, Senior Researcher, Stockholm International Peace Research Institute
- Shimona Mohan, in her personal capacity
- Ariel Conn, in her personal capacity
- Dr. Hendrik Huelss, Assistant Professor, SDU
- Dr. Qiaochu Zhang, Max Weber Postdoctoral Fellow, European University Institute
- Dr. Guangyu Qiao-Franco, Assistant Professor, Radboud University Nijmegen
- Dr. Tom Watts, in his personal capacity

This map of practices was prepared by Anna Nadibaidze and Ingvild Bode, with input from the AutoPractices team and the stakeholders.

How to cite this publication

The AutoPractices Project. (2025). *Map of Practices* (Odense: Center for War Studies).

Published by the Center for War Studies
University of Southern Denmark, Campusvej 55, Odense M 5230, Denmark

Cover photo by **Jerry Chen** on Unsplash (**unsplash.com/@jerry__chen__**)

Contents

Introduction to AutoPractices

As of July 2025, there are no international legally binding regulations specific to the development and deployment of artificial intelligence (AI) technologies in the military domain. Groups of states have agreed on sets of non-legally binding principles guiding the military uses of AI, whether in autonomous weapon systems (AWS)[1] or other applications.[2] However, top-down, state-led approaches to global governance in this area continue to face challenges such as different regulatory positions, competing interests, and diverging visions of the role of AI technologies in warfare and in society more broadly.[3]

As the European Research Council-funded project "Weaponised AI, Norms, and Order" (**AutoNorms**) has found, current practices in the design, training personnel for, and use of AI technologies in military systems have the potential to lead to a reduced exercise of human agency in use-of-force decision-making. A reduced form of human agency in military targeting raises ethical, legal, security, and operational concerns which are insufficiently addressed by sets of broad and often ambiguous principles featuring in current top-down frameworks.[4]

Considering this global challenge, the purpose of the European Research Council-funded project "Governing AI Technologies in Military Systems from the Bottom Up: Practices to Sustain and Strengthen Human Agency" (**AutoPractices**) is to initiate and accompany a process of social innovation to govern AI technologies in military systems from the bottom up. The AutoPractices project aims to co-create a set of 'best practices' in the form of a practical toolkit to sustain and strengthen the exercise of human agency when developing and using military systems integrating AI and autonomous technologies. The operational toolkit will be co-created together with stakeholders who represent different professional backgrounds and geographies.

This document outlines **a map of practices** which represents a step on the way towards the final toolkit. It is therefore meant as a transitional document.

The AutoPractices project runs from June 2024 until December 2025, and the operational toolkit is to be finalised by December 2025.

[1] The 11 guiding principles adopted by the United Nations Group of Governmental Experts on emerging technologies in the area of Lethal Autonomous Weapon Systems (UN GGE on LAWS) in 2019.

[2] The Responsible AI in the Military Domain (REAIM) Summits Call to Action (2023) and Blueprint for Action (2024); the United States Political Declaration on Responsible Military Use of AI and Autonomy (2023).

[3] Ingvild Bode, *Emerging Norms around Military Applications of AI: The Case of Human Control*, GC REAIM Expert Policy Note Series (The Hague: The Hague Centre for Strategic Studies, May 2025), https://hcss.nl/wp-content/uploads/2025/05/Bode-1.pdf; Ingvild Bode et al., "Prospects for the Global Governance of Autonomous Weapons: Comparing Chinese, Russian, and US Practices," *Ethics and Information Technology* 25, no. 5 (2023): 1–15, https://doi.org/10.1007/s10676-023-09678-x; Anna Nadibaidze, "Governance of AI in the Military Domain: International Law, Norms, and Ways Forward," in *Oxford Intersections: AI in Society*, ed. Dov Greenbaum (Oxford: Oxford University Press, 2025), https://doi.org/10.1093/9780198945215.003.0102.

[4] Ingvild Bode and Tom Watts, *Meaning-Less Human Control: Lessons from Air Defence Systems on Meaningful Human Control for the Debate on AWS* (Oxford & Odense: Drone Wars UK & Center for War Studies, 2021), https://dronewars.net/2021/02/19/meaning-less-human-control-lessons-from-air-defence-systems-for-lethal-autonomous-weapons/; Ingvild Bode and Tom Watts, *Loitering Munitions and Unpredictability: Autonomy in Weapon Systems and Challenges to Human Control* (Odense & London: Center for War Studies & Royal Holloway Centre for International Security, 2023), https://www.autonorms.eu/loitering-munitions-and-unpredictability-autonomy-in-weapon-systems-and-challenges-to-human-control/; Anna Nadibaidze, Ingvild Bode, and Qiaochu Zhang, *AI in Military Decision-Support Systems: A Review of Developments and Debates* (Odense: Center for War Studies, 2024), https://www.autonorms.eu/ai-in-military-decision-support-systems-a-review-of-developments-and-debates/.

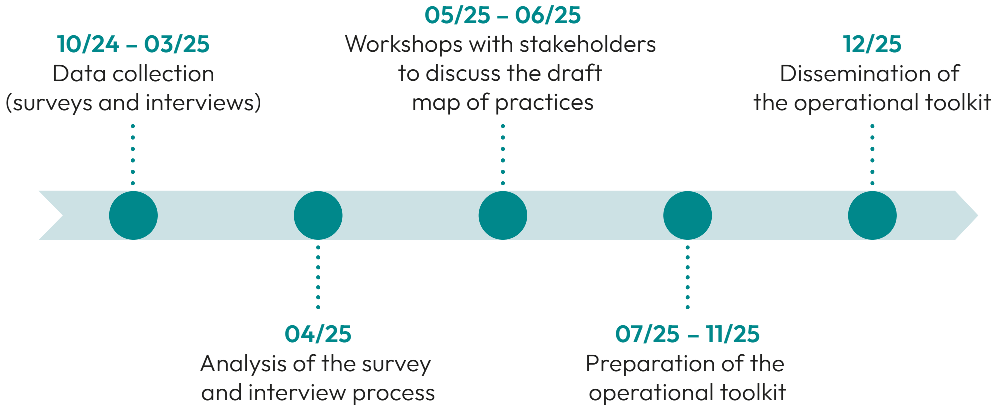
Data collection

This map of practices is based on data collected from stakeholders in two ways:

1. An **online survey questionnaire** completed by stakeholders who were willing to and gave their consent to participate. The survey was conducted via the SurveyMonkey platform (see the appendix for the full questionnaire).
   - **a** The questions appeared in the same order for all respondents.
   - **b** All questions were open-ended.
   - **c** All questions were optional.
2. One-on-one interviews conducted by a member of the AutoPractices team with stakeholders who gave their consent to participate.

The final operational toolkit will be based on 1) the survey responses; 2) the interview responses; and 3) the discussions held during two workshops (one virtual on 26 May 2025 and one in-person on 17 June 2025).

Figure 1. The main steps in the AutoPractices project process


10/24 – 03/25
Data collection (surveys and interviews)
04/25
Analysis of the survey and interview process
05/25 – 06/25
Workshops with stakeholders to discuss the draft map of practices
07/25 – 11/25
Preparation of the operational toolkit
12/25
Dissemination of the operational toolkit

Overview of stakeholders

Participation in this research is entirely voluntary and consent based. Stakeholders have been selected based on their knowledge and expertise of the integration of autonomous and AI technologies in military systems, as well as AI technologies more broadly.

To ensure interdisciplinarity and diversity of perspectives, stakeholders involved in this project include legal experts, military personnel (former and current), civil society representatives, academics, and researchers with different disciplinary backgrounds (humanities as well as social and natural sciences), as well as industry experts. All stakeholders have been invited to contribute **in their personal capacity**. Their views do not necessarily represent their states or institutions.

As of 1 July 2025, the project involves 47 stakeholders. They include political-ethical; legal; military; technical; and civil society experts representing all continents (except Antarctica). The AutoPractices team has aimed to secure a diverse stakeholder representation across these groups and geographical contexts. We refrain from displaying descriptive statistics about how many stakeholders per our own categorisation have been involved. This is to respect the fact that our categorisation may not align with how the stakeholders would categorise themselves. The research ethics committees at the University of Southern Denmark and the European Research Council have reviewed the AutoPractices project.

Terminology

Systems

In the AutoPractices survey or interview questions, we did not define ‘systems’ exclusively as either weapon systems, decision-support systems, or other types of systems integrating AI. This was to allow stakeholders to comment based on their background and expertise, no matter what type of systems they are most knowledgeable about. Some stakeholders explicitly mentioned AI-based (autonomous) weapon systems, others named decision-support systems, while others did not specify which systems they meant.

Practices and activities

In the AutoPractices survey or interview questions, we did not define ‘practices’ or ‘activities’ to allow stakeholders to interpret these terms according to their background and expertise. In the context of the AutoPractices project, we define practices as organised, linked, patterned activities performed by (groups of) people.[5] Activities are simply performances of a certain action by a person or group of people.

Agency

For the purposes of the AutoPractices project, the exercise of human agency in the context of interacting with AI systems is defined as:

> *The capacity to 1) understand and reasonably foresee a system’s functions and effects in a relevant context; and 2) to deliberate and decide upon suitable actions in a timely manner; and 3) to act in a way that can impact the use of the system.*

This definition is based on a literature review conducted by the AutoPractices team, which revealed the following common elements in the exercise of human agency:

- knowledge of the context (situational awareness) and of the system (its technical characteristics, capabilities, limitations),
- based on this knowledge, the ability to foresee how the system would function and its potential effects in the context of use,
- the ability to reflect/deliberate upon the effects of this system’s use, and following this deliberation, the ability to decide on an action/response and act upon this response (the ability to intervene in a timely manner), and
- the ability to engage in actions that make an impact/change in the world.

[5] Theodore R. Schatzki, “A Primer on Practices,” in *Practice-Based Education: Perspectives and Strategies*, by Joy Higgs et al. (Rotterdam: Sense Publishers, 2012), 13, https://doi.org/10.1007/978-94-6209-128-3_2.

The lifecycle of AI systems

The AutoPractices project takes as a foundation the model of the AI lifecycle[6] proposed by the IEEE Standards Association Research Group on Issues of Autonomy and AI in Defense Systems.[7] The work of the Research Group builds on other lifecycle frameworks that have been used in relation to military applications of AI,[8] but is more fine-grained.

To date, the IEEE lifecycle framework is the most comprehensive to have been developed with the specific challenges related to the military context in mind, while combining elements of frameworks from civilian domains. It is also the result of a joint, interdisciplinary exercise involving a group of experts with technical, political, ethical, legal, and military backgrounds.

A comprehensive lifecycle framework allows considering both 1) a micro perspective of practices of various groups of humans involved at each stage, and 2) a macro perspective of challenges for the exercise of human agency across the different stages.

The framework presents a granular way of thinking about the lifecycle with multiple points of human involvement and opportunities to exercise agency. In this model, the lifecycle of a military system integrating AI technologies includes the following 9 stages (see Figure 2):

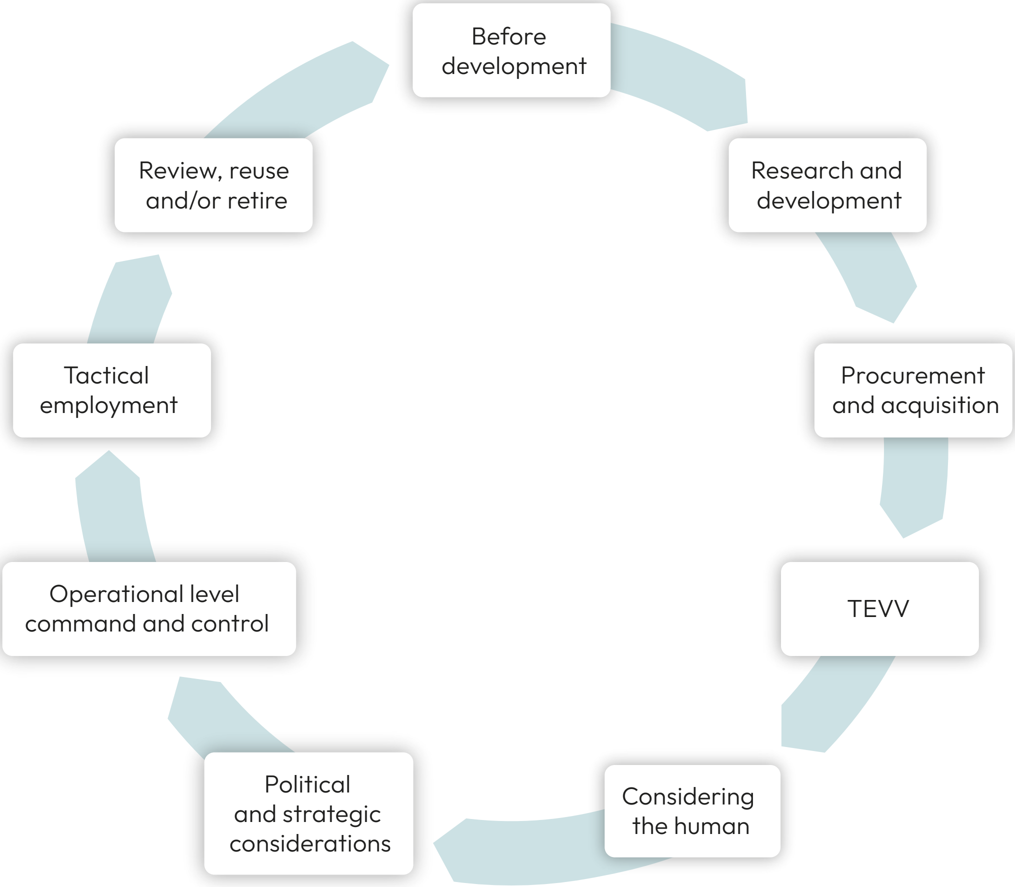
1) before AI system development
2) research and development
3) procurement and acquisition
4) Test, Evaluation, Validation and Verification (TEVV)
5) considering the human: education, training, and human-system integration
6) political and strategic considerations
7) operational level command and control
8) tactical employment
9) review, reuse and/or retire.

[6] We recognise that some stakeholders may not consider the term 'lifecycle' appropriate in the context of warfare. We have chosen to use this term in the context of AutoPractices because the 'lifecycle' is commonly used in technical literature describing the development, use, and post-use review of AI systems.

[7] IEEE SA Research Group on Issues of Autonomy and AI in Defense Systems, *A Framework for Human Decision-Making through the Lifecycle of Autonomous and Intelligent Systems in Defense Applications* (New York, NY: IEEE SA, 2024), https://ieeexplore.ieee.org/document/10707139.

[8] Merel Ekelhof and Giacomo Persi Paoli, *The Human Element in Decisions about the Use of Force* (Geneva: United Nations Institute for Disarmament Research, 2020), https://unidir.org/wp-content/uploads/2023/05/UNIDIR_Iceberg_SinglePages_web.pdf.

Figure 2. Lifecycle framework for systems integrating AI and autonomous technologies in the military domain.[9]


Before development
Research and development
Procurement and acquisition
TEVV
Considering the human
Political and strategic considerations
Operational level command and control
Tactical employment
Review, reuse and/or retire

Moreover, the IEEE Research Group highlights five activities that are ongoing across the nine stages of the lifecycle:

1) evaluation of legal, ethical, and policy concerns
2) responsibility, accountability, and knowledge transfers
3) considering the human: training, education and human-system integration
4) TEVV, monitoring, hardware system or software updates and interoperability, maintenance
5) risk assessment.

[9] Based on IEEE, *A Framework for Human Decision-Making through the Lifecycle of Autonomous and Intelligent Systems in Defense Applications*.

Methodology

Step 1 Identification of preliminary themes

Before the data collection, the AutoPractices team identified some preliminary themes that we expected to see in stakeholders' responses, based on a review of the literature. These themes did not dictate our analysis of the data. Rather, they were useful in starting us off in our structuring of the data analysis.
The preliminary themes were:

- Maintaining human responsibility and accountability
- Restrictions on the use of AI systems (spatial, temporal, context, type of targets, etc.)
- Ensuring that AI systems do not replace humans in critical targeting tasks
- Conducting appropriate levels of testing, auditing, and reviews
- Ensuring transparency via documentation and monitoring
- Training and education measures for the personnel involved
- Compliance with legal frameworks such as international humanitarian law (IHL)
- Understanding of the systems and technologies, predicting how they will work
- Appropriate levels of 'trust' or 'justified confidence' in the systems' outputs

Step 2 Data collection

The AutoPractices team shared the link to the online questionnaire (via the SurveyMonkey platform) with stakeholders via email. After stakeholders completed the survey, we exported the responses onto OneDrive and separated the responses from names of stakeholders to ensure anonymity. The names of stakeholders were only used for the purpose of sending invitations for follow-up interviews and workshops.

We also sent interview invitations via email. Interviews were recorded and transcribed with the help of Microsoft Teams in-built automated transcription tool and re-checked by a member of the AutoPractices team. Following the transcription, all recordings containing personal information were deleted.

Step 3 Data analysis and coding

First, we read through all the survey responses and interview transcripts several times to get a broad picture of the data.

Second, we created separate Microsoft Word files with responses for each survey question and files with each interview transcript. In these files, we included a series of initial themes. Some of them matched the preliminary themes, but others did not.

Third, using the software NVivo, we engaged in another round of coding with the objective of grouping the themes under more general categories that would allow clustering the data into sets of practices. We reviewed the themes and revised the initial lists of themes.

The analysis followed a qualitative approach. Given that all survey and interview questions were open-ended, the focus of the analysis was on common themes, patterns and clusters of key practices highlighted by stakeholders.

Step 4 Thematic analysis

Continuing in NVivo, we grouped the smaller themes into broader clusters that form the basis of this map of practices. These themes are listed in no particular order:

1. AI systems not replacing humans
2. Option for human intervention
3. End-user involvement
4. Education and training
5. Political and policy considerations
6. Testing and evaluation
7. Ensuring human accountability and responsibility
8. Risk assessment frameworks

Notes

Some phrases and sentences were coded into multiple themes. Practices might be similar or listed under several themes. The themes are therefore not mutually exclusive.

Moreover, not all stages of the lifecycle have been discussed by all stakeholders. Therefore, not all themes contain practices at every stage of the lifecycle.

The language used in the map of practices attempts to stay as close as possible to the language used by stakeholders in their survey and interview responses.

Map of practices

Theme 1 AI systems should not replace human decision-making, especially on the use of force

AI systems should be used in ways to support, augment, or enhance humans' abilities. AI systems should neither replace human personnel nor constrain humans' options for actions, for example by reducing human decisions to a veto or nominal approval.

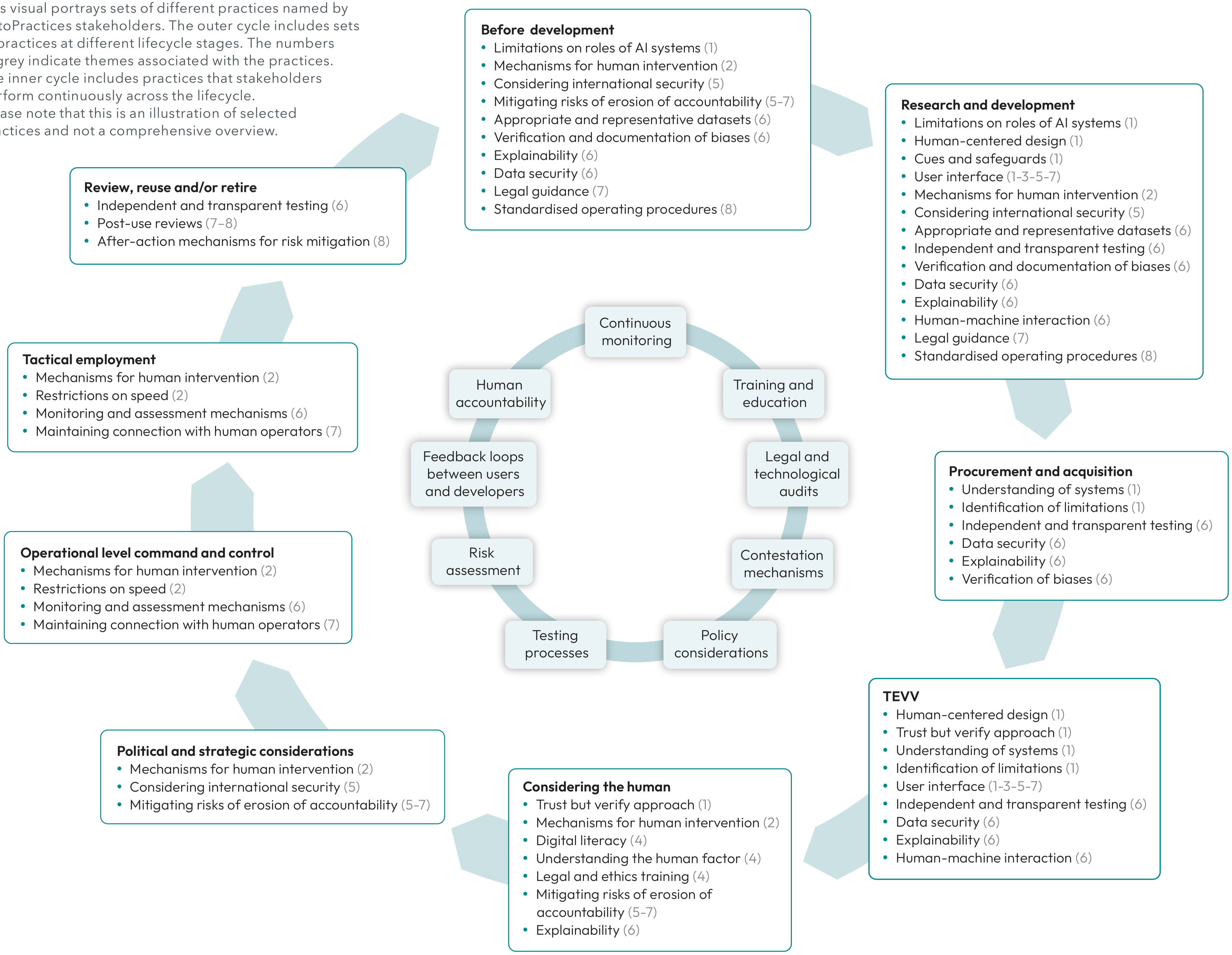
Practices applied **across the lifecycle** include:

- Setting defined objectives to achieve via the use of AI systems and ensuring these objectives align with political, strategic, ethical, and legal considerations. These considerations should be reflected in operational frameworks such as targeting doctrines and rules of engagement, as well as "standardised operating procedures or fragmentary orders" (S#27).
- Continuously monitoring systems to ensure that they function as intended via both legal and technological audits.
- Continuously training different actors involved, especially the users, to interact with AI systems in ways where the systems are employed as tools.
- Setting limitations on the roles/tasks performed by AI systems.
- Adopting a "trust but verify" approach where human decision-makers have a sufficient level of confidence in AI systems while remaining "the final arbiters in decision processes" (S#5). In other words, ensuring a balance between humans' sufficient understanding of the systems, and over-trust that would limit the exercise of human agency.
- Cross-validating decisions taken by humans at different stages to ensure that the objectives of using AI systems are respected and achieved.

Practices applied **at specific lifecycle stages** include:

- At stage 2 (R&D), developing a user interface that integrates delimitations between the roles and responsibilities attributed to both humans and AI systems. The interface should present the information in an appropriate manner. It should also be properly tested at stage 4 (TEVV).
- At stage 2 (R&D), ensuring ways for users to understand when the systems are not working as planned. Design and development should follow a human-centred design approach and built-in cues, safeguards, and standards that limit the autonomous functions of AI systems.
- At stages 3-4 (procurement and acquisition, TEVV), identifying already known and potential limitations of the systems and having a clear understanding of anticipated contexts and circumstances of use.
- At stage 9 (post-use), either adapting, revising, or potentially retiring the systems depending on how they perform at the use stage, especially in relation to whether this performance matches the political, strategic, legal, and ethical frameworks set earlier.

This visual portrays sets of different practices named by AutoPractices stakeholders. The outer cycle includes sets of practices at different lifecycle stages. The numbers in grey indicate themes associated with the practices. The inner cycle includes practices that stakeholders perform continuously across the lifecycle. Please note that this is an illustration of selected practices and not a comprehensive overview.



Before development
- Limitations on roles of AI systems (1)
- Mechanisms for human intervention (2)
- Considering international security (5)
- Mitigating risks of erosion of accountability (5-7)
- Appropriate and representative datasets (6)
- Verification and documentation of biases (6)
- Explainability (6)
- Data security (6)
- Legal guidance (7)
- Standardised operating procedures (8)

Research and development
- Limitations on roles of AI systems (1)
- Human-centered design (1)
- Cues and safeguards (1)
- User interface (1-3-5-7)
- Mechanisms for human intervention (2)
- Considering international security (5)
- Appropriate and representative datasets (6)
- Independent and transparent testing (6)
- Verification and documentation of biases (6)
- Data security (6)
- Explainability (6)
- Human-machine interaction (6)
- Legal guidance (7)
- Standardised operating procedures (8)

Procurement and acquisition
- Understanding of systems (1)
- Identification of limitations (1)
- Independent and transparent testing (6)
- Data security (6)
- Explainability (6)
- Verification of biases (6)

TEVV
- Human-centered design (1)
- Trust but verify approach (1)
- Understanding of systems (1)
- Identification of limitations (1)
- User interface (1-3-5-7)
- Independent and transparent testing (6)
- Data security (6)
- Explainability (6)
- Human-machine interaction (6)

Considering the human
- Trust but verify approach (1)
- Mechanisms for human intervention (2)
- Digital literacy (4)
- Understanding the human factor (4)
- Legal and ethics training (4)
- Mitigating risks of erosion of accountability (5-7)
- Explainability (6)

Political and strategic considerations
- Mechanisms for human intervention (2)
- Considering international security (5)
- Mitigating risks of erosion of accountability (5-7)

Operational level command and control
- Mechanisms for human intervention (2)
- Restrictions on speed (2)
- Monitoring and assessment mechanisms (6)
- Maintaining connection with human operators (7)

Tactical employment
- Mechanisms for human intervention (2)
- Restrictions on speed (2)
- Monitoring and assessment mechanisms (6)
- Maintaining connection with human operators (7)

Review, reuse and/or retire
- Independent and transparent testing (6)
- Post-use reviews (7–8)
- After-action mechanisms for risk mitigation (8)

Theme 2 Humans must have a possibility to intervene across the lifecycle

Ensuring human agency means including mechanisms for human actors to engage in critical reflections, question the output of AI systems, and intervene in the development and use of AI systems, if needed (depending on the context of use).

Practices applied **across the lifecycle** include:

- Implementing contestation mechanisms in the design of AI systems. Humans should have the opportunity to cross-check the outputs of AI systems with other sources of data and intelligence. This cross-checking exercise aids humans to critically assess and challenge the AI systems' outputs.
- Continuously training human actors, especially operators, on using those contestation mechanisms, especially in terms of taking the time to engage in critical reflections (political, legal, and ethical deliberations) when they understand that the systems are not performing as expected, and subsequently either slowing down, manually overriding, or stopping/deactivating the systems, if needed.
- Adopting restrictions on the speed to ensure the time that is needed for humans to assess and potentially intervene/override the AI systems, while allowing a rapid process when the situation is "absolutely time critical (e.g. defence against incoming fast-moving threat)" (S#23).

Practices applied **at specific lifecycle stages** include:

- At stages 1-2 (before development, R&D), including mechanisms allowing humans to cross-check and challenge outputs of AI systems in the design of AI systems.
- At stage 3 (procurement and acquisition), guaranteeing critical assessments and potential for interventions.
- At stage 5 (considering the human), training humans on using mechanisms designed in stages 1-2, including the scenarios where an 'off switch' might be used. Guidance and protocols on using the systems should account for the time necessary to assess AI systems' outputs.
- At stages 7-8 (operational, tactical), including an option to react and recall systems (especially weapon systems), as well as reducing the tempo by adopting "tactical patience" (S#33).
- At stage 9 (post-use), ensuring appropriate understanding of how systems worked during the employment stage, which would allow humans to decide on whether to make modifications or potentially not reusing the system.

Theme 3 Feedback loops between end-users and developers should feature across the lifecycle, including after the deployment of AI systems, and should reinforce human agency

End-users such as operators and commanders should be constantly involved in the design of AI systems (stages 1-2), even after the systems have been deployed.

Practices applied across the lifecycle include:

- Implementing feedback loops and consultations between users (operators) and developers, allowing them to develop a common understanding of the context of use.
- Giving the opportunity for key stakeholders including political and strategic decision-makers, developers, lawyers, ethicists and operators to provide input across the lifecycle.
- Holding ongoing discussions on ethical and legal aspects among end-users and between end-users and developers to "share experiences and perception of problems" (S#7).
- Developing processes and contracts that enable users to maintain control over critical design decisions in AI systems, including clauses where users could "ask for modifications if concept drift reduces meaningful human control" for no additional cost, rather than being subject to "vendor lock-in" on systems that no longer operate as they were designed (S#36).
- Establishing communication channels to transmit knowledge, concerns or risks across the chain of command.
- Conducting "user consultation and testing with as broad a representation as possible" (S#18).
- Fostering an organisational culture of "honest exchange and feedback" that allows "reporting issues, mistakes and errors without restraint" (S#38).

Practices applied at specific lifecycle stages include:

- At stage 2 (R&D), involving users in the design of the system's interface.
- At stage 4 (TEVV), involving users in the testing of AI systems and giving them the opportunity to provide input. Stakeholders such as lawyers, ethicists and those who can evaluate social implications should also be involved in TEVV.

Theme 4 Training and education measures should be implemented across the lifecycle

Training and education should not be restricted to users or operators of systems. These measures should apply to actors such as political and strategic decision-makers, researchers, developers, engineers, and technical staff. Moreover, training should be continuous and constantly updated: it is not an "one-and-done event" (S#25).

Concrete measures for training and education include scenario exercises, simulations, wargames, mock drills, and experiments. They must be based on realistic, but diverse, conditions of use, including unexpected situations and edge cases.

There are three broad areas of training measures to ensure the exercise of human agency:

1. **Technical education (digital literacy)**
   - Continuously educating humans involved, especially users and operators, on the evolving technical aspects of AI systems so that they understand the technologies involved, how they function, and importantly, their limitations. Being able to reasonably predict and foresee a system's behaviour is a key part of exercising human agency.
   - Educating operators on how to critically assess the output of the system, intervene, or potentially stop the system as well as how to document the issue after the employment stage, or in other words, "quickly cease use and report issues when AI systems do not operate as expected" (S#6).
2. **Psychological education (understanding the human factors)**
   - Raising the involved humans' awareness about how they make decisions both without and with technologies. Personnel should know about various aspects of interaction with AI systems such as automation biases and cognitive biases, assumptions made by humans and integrated into AI systems, and risks of de-skilling, among others.
   - Considering the diversity of humans involved (in terms of gender, educational background, etc.) when studying human-machine interaction. Education and training should also about being prepared to use the whole socio-technical infrastructure involved, not just one AI system.
3. **Legal/ethics training**
   - Training humans involved about applicable legal frameworks, especially international humanitarian law, civilian harm mitigation, as well as military ethics, ethical considerations, and evaluation of broader societal impacts.

Theme 5 Humans must continuously engage in political and policy considerations during the development and use of AI systems

Prior to the development of AI systems, and throughout the lifecycle, relevant policymakers need to ask critical questions about the purpose of the systems, the intended uses, and whether these systems are politically, legally and ethically appropriate for the context of use.

Practices applied across the lifecycle include:

- Fostering organisational cultures that promote exchange, feedback and communication that allow engaging in those reflections across the lifecycle.
- Delimiting who is responsible for activation, use, setting parameters, setting mission goals, and terminating the systems.
- Setting specific instructions for all actors to match the political and policy considerations.
- Setting limitations or restrictions, whether geographical, spatial, or not targeting humans, for example, based on evaluations of risks for humans affected by the use of AI systems: "consideration should extend to indirect, long term and reverberating effects" such as "human costs beyond physical effects" or the "impact on the natural environment" (S#33).

- Framing human-machine interaction within the relevant organisational cultures as well as the complexities of how humans interact with each other and how machines interact with each other within those structures: "it is therefore not just a question of human-machine interaction, but also of human-human-machine interaction" and of "machine-machine interaction in the context of human decision-making processes and their embedding in organisational cultures" (S#14).
- Incorporating concerns surrounding international security, including the proliferation of technologies, escalation, lowering the threshold for the use of force, or reinforcement of some narratives about AI, for example that AI inevitably increases efficiency and precision.
- Mitigating the risks of erosion of culture of accountability and how decisions can cumulatively affect targeting processes and ultimately also targeting doctrines.

Practices applied **at specific lifecycle stages** include:

- At stages 2-3 (R&D, procurement and acquisition), integrating any restrictions on spatial, geographical boundaries or types of targets into an appropriate interface: "this is only possible if these measures are conceived of beforehand" (S#27), at the early stages of the lifecycle.
- At stage 9 (post-use), setting post-use legal reviews to evaluate whether political but also strategic, legal, and ethical considerations have been met.

Theme 6 AI systems should undergo extensive and ongoing testing procedures

Testing, evaluation, validation and verification should be an ongoing set of practices throughout the lifecycle.

Practices applied **across the lifecycle** include:

- Involving an appropriate dataset and recording limitations or biases within the training data via clear, transparent documentation.
- Conducting independent and transparent testing processes that include decision traceability.
- Including input and feedback from end-users to match the planned context of use.
- Recognising that updates of systems or adaptations of training models may cause problems. Potential updates and adaptation need to be incorporated into measures and risk assessment frameworks.
- Implementing regular reviews with metrics assessing the performance and risks in accordance with international legal obligations, especially international humanitarian law and the principles of distinction, proportionality, and precaution.
- Adding parameters that would ensure that, if systems do not meet the requirements, new rounds of testing would be required. Similarly, imposing conditions on how long the system could be used without needing to be reviewed or re-tested.
- Ensuring transparent access to the training data and parameters, especially for machine learning systems, in a way that allows humans to "obtain explanations on the causal link between AI inputs and outputs, and the functioning of algorithmic processes" (S#33).

- Minimising errors that could lead to unforeseen outcomes, for example via technical audits and failsafe mechanisms, while recognising that malfunctions cannot be fully eliminated.
- Including recording methods and monitoring processes such as mechanisms that ensure transparency and facilitate access by relevant parties.
- Ensuring data security.
- Educating human personnel about the safety risks in case systems are hacked or jammed.
- Striving for representative datasets and documenting issues with data such as biases to create records for auditing.
- Testing systems in ambiguous contexts and running through how actors would behave.

Practices applied **at specific lifecycle stages** include:

- At stages 1–3 (before development, R&D, procurement and acquisition), involving human supervisors who can verify the biases in the systems or the data.
- At stages 1–5 (before development, R&D, procurement and acquisition, TEVV, considering the human), ensuring understandability and predictability of AI systems. If this requirement is not fulfilled at the design and testing levels, acquisition should be prevented.
- At stage 2 (R&D), applying measures "against potential AI-induced harm" and integrating security measures to ensure a "security-by-design approach during R&D" (S#6).[10]
- At early stages such as stages 2–3 (R&D, procurement and acquisition), detecting technical malfunctions or uncertainties because "if flaws aren't caught early, those flaws become baked into the system, undermining human decision-making down the line" (S#25).
- At stage 4 (TEVV), considering human-machine interaction in testing, not only technical characteristics but also how humans interact with the AI systems. Testing should sufficiently match the situation of employment (stages 7-8, operational and tactical). Many aspects of testing will therefore depend on the planned contexts of use.
- At stages 7–8 (operational and tactical), implementing real-time battlefield monitoring and assessment mechanisms to detect malfunctions and recall the system if needed.

Theme 7 Practices across the lifecycle need to ensure human accountability and responsibility

Ensuring human agency means keeping track of accountable and responsible humans throughout the stages of the lifecycle, given that only human agents/natural persons can hold legal accountability for violations of international humanitarian law (I#1).

Practices applied **across the lifecycle** include:

- Clearly distributing roles among the human actors involved at each stage.
- Applying mechanisms tracing decisions to specific actors to ensure transparency. They should enable human operators to have access to data or information that may help the operators understand what led to this output or recommendation.

[10] Zhang Ling gave consent to be acknowledged by name for direct quotes from S#6.

- Maintaining consistent records and documentation to ensure transparency of how responsible individuals or teams use tools for certain functions. At the same time, this should not involve extensive surveillance, but rather "the security of a team built of trust and excellent communication" (S#8).
- Allocating enough resources into building these mechanisms and planning this resource allocation in advance, at the early stages of the lifecycle.
- Developing a concept of operations and match these concepts with the computational components so that, at the employment stage, "responsible people can have relied on the practices that went before them" (I#1).
- Adopting a verification regime.

Practices applied **at specific lifecycle stages** include:

- At stages 1-2 (before development, R&D), incorporating legal guidance (especially on international humanitarian law) and defining accountability measures since the beginning.
- At stage 2 (R&D), implementing safeguards or measures to mitigate risks, for example, via agile co-design (S#8).
- At stages 7-8 (operational and tactical), ensuring a continuous connection between the systems and human operators.
- At stage 9 (post-use), after the use of AI systems and in case of violations of international humanitarian law or other legal frameworks, engaging in a process to apply liability and ensure "judicial agency" (S#14) via "strong after-action mechanisms" (S#27).

Theme 8 The use of AI systems should follow appropriate and detailed risk assessment frameworks

Actors should implement frameworks to assess risks of using AI systems in the military domain. The type of framework, however, may depend on the context and the systems.

Practices applied **across the lifecycle** include:

- Considering the differences and distinctive features of types of conflicts when assessing the risks and necessary frameworks/measures.
- Considering the contexts of use, as "specific frameworks will be necessary for high-risk/impact use cases, such as AI-enabled weapons (including AWS), AI in decision support (especially related to use of force), AI in cyber, AI in information operations, among others… this is where the most impactful work could and should be done, especially given all of these use cases are current concerns and in operational use" (S#23).
- Engaging in exercises together with other states or organisations to identify best practices and benchmarks for AI systems to meet objectives while assessing risks.
- Adopting appropriate (and continuously updated) cybersecurity measures because in case of a cyber-attack, "the performance of the system can utterly change from one day to the next" (S#8).[11]
- Ensuring that AI systems are interpretable, and that the parameters and weights are used according to broader legal, ethical and strategic considerations.
- When it comes to weapon systems, improving traceability and transparency by marking hardware and recording data about operators and tasks.

[11] Joanna Bryson gave consent to be acknowledged by name for direct quotes from S#8.

- Classifying risks in a tier system (e.g., untenable, high-level, mid-level and low-level), as well as deciding on safeguards to prevent and mitigate different sets of risks.
- Classifying how the AI systems relate to the use of force (especially relevant for AI in decision-support systems).

Practices applied **at specific lifecycle stages** include:

- At stages 1-2 (before development, R&D), adopting standardised operating procedures (SOPs) across the lifecycle, especially at earlier stages when speed is not such a prominent issue.
- At stages 2-3 (R&D, procurement and acquisition), approaching and building an awareness of risks at the earliest phases of the lifecycle as well as adopting a framework to minimise risks during stage 4 (TEVV).
- At stages 1-4 (before development, R&D, procurement and acquisition, TEVV), adopting safeguards for online learning.
- At stage 9 (post-use), including risk assessments in the form of technological audits or legal reviews conducted after the use of the systems. After-action mechanisms can ensure legal accountability as well as reveal technical malfunctions and vulnerabilities.

Risk assessment frameworks that stakeholders consider helpful and important, although these frameworks might currently not be applied to the military domain, include:

- The European Union's AI Act
- The US National Institute of Standards and Technology (NIST)
- The International Organization for Standardization (ISO), e.g., ISO/IEC 27001, ISO/IEC 23894
- Civil aviation industry standards and the International Civil Aviation Organization (ICAO)
- IEEE SA Framework or other IEEE standards, e.g., IEEE 7007-2001
- The International Atomic Energy Agency (IAEA)
- The Chemical Weapons Convention
- The Organization for Economic Cooperation and Development (OECD) guidelines on AI ethics
- The NATO AI Strategy
- The NATO Responsible AI toolkit
- The US Responsible AI toolkit
- The REAIM Summit's Blueprint for Action
- The Australian Voluntary AI Safety Standard
- The Trusted Autonomous Systems (Australia) Responsible AI For Defence Toolkit (Consultation)
- The UNIDIR taxonomy of risks
- The UK Ministry of Defence Dependable Artificial Intelligence (AI) in Defence Directive (JSP 936 V1.1)
- Responsible AI principles
- Legal reviews of advanced cyber capabilities
- The 3D (design, development, deployment) framework
- Checklists or flowcharts
- National or international confidence building measures
- Self-regulation schemes

Conclusion and overview

This map of practices is the foundation for AutoPractices' work towards the main objective of the project: co-creating, with the stakeholders involved, a practical toolkit to sustain and strengthen the exercise of human agency in the use of military AI systems.

The map highlights the diversity of activities that actors, including political and strategic decision-makers, developers, engineers, commanders, operators, lawyers, ethicists, and others can perform to contribute to the exercise of human agency **both across the lifecycle of AI systems and at specific stages.**

The practices mentioned across the eight themes identified above can be grouped into three broad categories (see Table 1, based on terminology in S#3).

First, there are **technical practices** that relate to aspects such as hardware, software, data, and cybersecurity.

Second, there are **policy practices** that relate to establishing operational norms and constraints (including, but not limited to, legal) on how AI systems should be developed and used.

Third, there are **procedural practices** that include adapting organisational cultures, concepts of operations, rules of engagement, special instructions to mitigate risks specifically associated with uses of AI systems.

These sets of practices contribute to ensuring the exercise of human agency in human-machine interaction within the military context, especially in decision-making on the use of force.

The AutoPractices operational toolkit will expand further on some of these practices and their contribution to the exercise of human agency in the military domain.

Table 1. General overview of practices

Technical practices	Policy practices	Procedural practices
Human-machine interaction		
Hardware Software Learning techniques Cybersecurity measures Data security Testing and evaluation Design Monitoring	Restrictions on use of systems Legal reviews Technological audits Feedback loops and communication mechanisms	Organisational culture Concept of operations Rules of engagement Special instructions Education and training

Appendix – Survey questionnaire

Dear Participant,

Thank you for agreeing to take part in this survey. The results of this questionnaire will feed into the project "Governing AI Technologies in Military Systems from the Bottom Up: Practices to Sustain and Strengthen Human Agency" (AutoPractices), funded by the European Research Council (Proof of Concept grant no. 101156237).

The purpose of the AutoPractices project is to initiate and accompany a process of social innovation to govern autonomous and AI technologies in the military domain. The project will co-create a set of best practices with stakeholders in the form of a practical toolkit to sustain and strengthen human agency and accountability for the use of AI systems in the military. To co-create this practical toolkit, we approach stakeholders across diverse professional backgrounds and geographies. You have been selected as a research participant based on your knowledge and expertise on AI systems, including in the military domain.

Participation is voluntary, and you are free to decline to answer some questions and can leave the data collection activity at any time without giving a reason. The research team will keep your data strictly confidential and anonymised.

This questionnaire focuses on your understanding (based on your respective area of expertise) of how human agency should be exercised throughout the lifecycle of systems integrating AI technologies.

For the purposes of this survey, the exercise of human agency in the context of interacting with an AI system is defined as: The capacity to 1) understand and reasonably foresee a system's functions and effects in a relevant context; and 2) to deliberate and decide upon suitable actions in a timely manner; and 3) to act in a way that can impact the use of the system.

The lifecycle of a military system integrating AI includes the following stages:

1) before AI system development; 2) research & development; 3) procurement & acquisition; 4) Test, Evaluation, Validation and Verification (TEVV); 5) education & training; 6) political and strategic considerations; 7) operational level command and control; 8) tactical employment; and 9) review, reuse and/or retire.

Questions

Q1. What do you think ensuring the exercise of human agency across the lifecycle of a military system integrating AI means? Please respond in max. 1-2 sentences.

Q2. Based on your area of expertise, what are the most important concerns to address at various stages of a system's lifecycle (listed above) to ensure the exercise of human agency? Please name your top 3 concerns.

Q3. What do you consider as key activities contributing to ensuring the exercise of human agency when it comes to human-system integration/human-machine interaction throughout an AI system's lifecycle? Please name your top 3 activities.

Q4. What types of activities would be detrimental to the exercise of human agency across an AI system's lifecycle? Please name your top 3 activities that you find detrimental.

Q5. What do you consider key activities to ensure responsibility and accountability throughout an AI system's lifecycle? Please name your top 3 activities.

Q6. What are some of the key activities needed to ensure the exercise of human agency in Testing, Evaluation, Validation and Verification (TEVV), monitoring, and maintenance? Please name your top 3 activities.

Q7. Based on your area of expertise, do you consider certain training and education measures essential to ensure the exercise of human agency across a system's lifecycle? If yes, which ones? Please name a maximum of 3 measures.

Q8. How should risks related to systems integrating AI technologies be assessed? Are there any existing frameworks which you would consider useful or important?

List of references

Bode, Ingvild, Hendrik Huelss, Anna Nadibaidze, Guangyu Qiao-Franco, and Tom Watts. 2023. "Prospects for the Global Governance of Autonomous Weapons: Comparing Chinese, Russian, and US Practices." *Ethics and Information Technology* 25 (5): 1–15. https://doi.org/10.1007/s10676-023-09678-x.

Bode, Ingvild, and Tom Watts. 2021. *Meaning-Less Human Control: Lessons from Air Defence Systems on Meaningful Human Control for the Debate on AWS*. Oxford & Odense: Drone Wars UK & Center for War Studies. https://dronewars.net/2021/02/19/meaning-less-human-control-lessons-from-air-defence-systems-for-lethal-autonomous-weapons/.

Bode, Ingvild, and Tom Watts. 2023. *Loitering Munitions and Unpredictability: Autonomy in Weapon Systems and Challenges to Human Control*. Odense & London: Center for War Studies & Royal Holloway Centre for International Security. https://www.autonorms.eu/loitering-munitions-and-unpredictability-autonomy-in-weapon-systems-and-challenges-to-human-control/.

Ekelhof, Merel, and Giacomo Persi Paoli. 2020. *The Human Element in Decisions about the Use of Force*. Geneva: United Nations Institute for Disarmament Research. https://unidir.org/wp-content/uploads/2023/05/UNIDIR__Iceberg__SinglePages__web.pdf.

IEEE SA Research Group on Issues of Autonomy and AI in Defense Systems. 2024. *A Framework for Human Decision-Making through the Lifecycle of Autonomous and Intelligent Systems in Defense Applications*. New York, NY: IEEE SA. https://ieeexplore.ieee.org/document/10707139.

Nadibaidze, Anna. 2025. "Governance of AI in the Military Domain: International Law, Norms, and Ways Forward." In *Oxford Intersections: AI in Society*, edited by Dov Greenbaum. Oxford: Oxford University Press. https://doi.org/10.1093/978 0198945215.003.0102.

Nadibaidze, Anna, Ingvild Bode, and Qiaochu Zhang. 2024. *AI in Military Decision Support Systems: A Review of Developments and Debates*. Odense: Center for War Studies. https://www.autonorms.eu/ai-in-military-decision-support-systems-a-review-of-developments-and-debates/.

Schatzki, Theodore R. 2012. "A Primer on Practices." In *Practice-Based Education: Perspectives and Strategies*, by Joy Higgs, Ronald Barnett, Stephen Billett, Maggie Hutchings, and Franziska Trede, 13–26. Rotterdam: Sense Publishers. https://doi.org/10.1007/978-94-6209-128-3_2.

About the AutoPractices team

Dr Ingvild Bode is Professor of International Politics and Director of the Center for War Studies at the University of Southern Denmark (SDU). She is the Principal Investigator of the AutoPractices and AutoNorms projects. Professor Bode's research examines how applications of AI in the military domain change international norms, especially on the use-of-force. Her research has been published in journals such as *European Journal of International Relations*, *European Journal of Public Policy*, *Global Studies Quarterly*, and *Review of International Studies*.

Dr Anna Nadibaidze is Postdoctoral Researcher at the Center for War Studies, University of Southern Denmark and a researcher for the AutoPractices and AutoNorms projects. She holds a PhD in Political Science from SDU. Her research examines military applications of AI and the global governance of AI in the military domain.

Dr Alexander Blanchard is a Senior Researcher in the Governance of Artificial Intelligence (AI) Programme at the Stockholm International Peace Research Institute (SIPRI). His work focuses on issues related to the development, use and control of military applications of AI. Alexander was previously the Defence Science Technology Laboratory (Dstl) Digital Ethics Fellow at the Alan Turing Institute, London.

Shimona Mohan is an Associate Researcher with the Gender & Disarmament and Security & Technology programmes at the United Nations Institute for Disarmament Research (UNIDIR). She was listed as one of the 100 Brilliant Women in AI Ethics for 2024, and holds a Master in International Affairs from the Geneva Graduate Institute in Switzerland. She contributes to AutoPractices in her personal capacity.

Ariel Conn is an Instructor in Computing Ethics at the University of Colorado, Boulder, and she is the founder and President of Mag10 Consulting, where she works with researchers and not-for-profit organizations to address a variety of global ethical and policy issues. Her work is primarily focused on the use of AI in the military and other ethically challenging uses and impacts of AI. She is an expert advisory member for the Global Commission on Responsible AI in the Military Domain, and she was one of 2023's 100 Brilliant Women in AI Ethics.

Dr Hendrik Huelss is Assistant Professor of International Relations at the Center for War Studies, University of Southern Denmark and an associated senior researcher in the AutoNorms project. Hendrik's research combines an interest in norms in International Relations with critical perspectives on technologies in politics. He has published in journals such as *European Journal of International Security*, *International Political Sociology*, *International Theory*, and *Review of International Studies*.

Dr Qiaochu Zhang is a Max Weber Postdoctoral Fellow at the Florence School of Transnational Governance, European University Institute. Her research focuses on global AI governance and Chinese foreign policy. Her work has been published in *International Affairs*, *International Peacekeeping*, and *Global Policy: Next Generation*, among others.

Dr Guangyu Qiao-Franco is Assistant Professor of International Relations at Radboud University and Advisory Board Member at the Leiden Asia Centre. She is Principal Investigator of a Dutch MOFA-funded project on China's export control strategy. Her research focuses on AI governance, export controls, and Global South perspectives in international politics.

Dr. Tom F.A. Watts is a Leverhulme Early Career Fellow based at Royal Holloway, University of London. His current research project examines the relationship between great power competition and military applications of AI. Tom's research has been published in leading International Relations journals including *Cooperation and Conflict*, *Geopolitics*, and *International Politics*, among others.

CENTER FOR
WAR
STUDIES

sipri

AutoNorms

SDU