ASSER INSTITUTE
Centre for International & European Law

ELSA LAB DEFENCE

RESEARCH REPORT

From Development to Deployment: A Framework for Evaluating the Legal Acceptability of Military AI Systems

Jonathan Kwik and Marta Bo

Executive Summary and Key Applications

Military artificial intelligence (AI) systems are increasingly being developed, acquired, and deployed, yet determining whether they can be used lawfully under international humanitarian law (IHL) remains a significant challenge. This report introduces a **framework** and accompanying **toolkit** for evaluating the legal acceptability of military AI systems—from development to deployment—with a particular focus on autonomous weapon systems (AWS) and AI-enabled decision-support systems (DSS). It addresses two central questions: *when can a military AI system be considered legally acceptable under IHL and related norms, and how should that assessment be conducted in practice?*

Current practice reveals two structural obstacles. First, many legal and policy norms are expressed at a **high level of abstraction**. While concepts such as "indiscriminateness", "context-appropriate human judgment and control", and "human agency" provide important legal and policy standards, they offer limited practical guidance for those designing, procuring, reviewing, or deploying military AI systems. As a result, these requirements are often difficult to translate into concrete technical and operational criteria, creating uncertainty for both Defence and industry.

Second, the legal assessment of military AI systems is **inherently context-dependent**. The lawfulness of an AI system cannot be determined by reference to a single technical feature or performance metric. Rather, it depends on the complex interaction between the system's characteristics (e.g. performance, training data, human-machine interface), operational conditions (e.g. type of operation, rules of engagement, temporal, and spatial constraints), and environmental factors (e.g. civilian presence, adversary behaviour, and third-party activities). Existing research has begun to address these questions but often focuses on a narrow subset of variables or on AWS alone, leaving the broader spectrum of military AI—and the interaction between variables—only partially understood. Moreover, because military AI systems are continually updated, retrained, and re-configured, legal assessments must also be **continuous and iterative** throughout the system lifecycle.

This report responds to these challenges by developing a **structured methodology and accompanying toolkit** that translate abstract IHL principles into practical, context-sensitive criteria for assessing military AI systems. Drawing on existing literature, a dedicated expert workshop held in The Hague on 28–29 January 2026, and engagement with ELSA Lab Defence partners, the framework identifies a set of **variables** relevant to the legal assessment of military AI systems. These variables encompass technical, operational, environmental, human, and institutional factors, providing a common basis for consistent and transparent legal assessment. Rather than asking evaluators to assess abstract legal concepts directly, the toolkit focuses on observable and measurable factors (e.g. false positive rates, data representativeness, the ability to intervene or recall systems, or density of civilians), allowing legal judgments to be made on the basis of shared factual assessments.

The framework links each variable to one or more of three **core legal rationales derived from IHL**: ***D*** *= distinction related concerns* (e.g. the risk of misidentifying and targeting civilians or civilian objects), ***C*** *= civilian harm related concerns* (including proportionality and precautions in attack), and ***P*** *= precautions related concerns* (such as predictability, understandability, human control, and the ability to conduct a meaningful legal assessment in the first place). These rationales explain why particular technical, operational, or environmental factors are legally relevant and how they contribute to legal risk.

The accompanying toolkit provides a colour coded qualitative evaluation scale for assessing the risk of each variable and identifies the relationships between variables, highlighting dependencies, high-risk or disqualifying combinations, potential trade-offs, and opportunities for risk mitigation. Together, these features support a more structured, transparent and context-sensitive legal assessments of military AI.

Key Applications

This framework and toolkit are intended to support a range of stakeholders involved in the development, acquisition, legal review, and operational use of military AI systems. Their principal applications include:

- **Legal review and assessment officers:** Supports formal Article 36 reviews and other legal assessments by identifying legally relevant variables, highlighting high-risk combinations, and documenting legal reasoning and residual risks.
- **Acquisition officers:** Helps translate legal requirements into clear technical and operational requirements, supporting procurement, co-development, and early risk identification.
- **Commanders and operators:** Supports context-sensitive deployment decisions by identifying operational risks, considering mitigation measures, and reassessing systems as operational conditions evolve.
- **Industry and developers:** Helps translate legal requirements from Defence into practical design and development considerations. It supports due diligence, internal training, and customer guidance for highly configurable systems, while helping organisations anticipate legal and operational expectations. In an increasingly competitive market, integrating these considerations early can reduce legal and reputational risk and provide a competitive advantage by demonstrating that products have been designed with legal compliance in mind.

The resulting toolkit is intended to support structured, transparent, and context-sensitive legal assessments. It does not replace legal judgment or decision-makers' responsibility. It helps *legal review officers* identify and reason systematically about compliance risks; *acquisition officers* translate legal requirements into clear technical and operational expectations; *commanders and operators* structure deployment decisions and identify appropriate mitigation measures; and *industry actors* incorporate IHL considerations into design, development, and customer guidance. By providing a **shared vocabulary and common methodology,** the toolkit supports more transparent, consistent, context-sensitive assessments of the legal acceptability of military AI systems throughout their lifecycle.

Table of Contents

Executive Summary and Key Applications **2**

Table of Contents **4**

1. Introduction **5**

2. Project Background **7**

2.1. A Framework Guiding the Continuous and Iterative Evaluations of Military AI Systems 7
2.2. Lawfulness Under IHL: A Key Component of the Legitimacy of Military Operations 9

3. Purpose, Scope, and Intended Use **10**

3.1. For Defence: Methodological Support Instrument 10
3.2. For Industry: Clarity on Legal Requirements 11
3.3. Explaining the Legal Rationales 12

4. Understanding the Toolkit **15**

4.1. Variables: The Units of Evaluation 15
4.2. Linking Variables to Legal Rationales 15
4.3. Assessing Variables Using the Evaluation Scale 17
4.4. Assessing Variables Where Information is Unavailable 18
4.5. Interactions Between Variables 19
4.6. Risk Mitigation 22
4.7. Relationship Chart 23

5. Recommendations and Future Work **24**

The Authors **25**

Acknowledgements **25**

About the ELSA Lab Defence project **25**

1. Introduction

Years of debate on autonomous weapon systems (AWS), followed by more recent discussions on AI-enabled decision-support systems (DSS), have produced growing consensus on the principles that should underpin the responsible use of military artificial intelligence (AI). Across legal, policy, and ethical frameworks, recurring themes include lawfulness, reliability, trustworthiness, bias mitigation, safety, security, explainability, traceability, and human responsibility and accountability.[1]

Yet, translating these high-level principles into practical assessments remains a significant challenge. Among them, the question of lawfulness under international law—and how States and armed forces can ensure the lawful development and use of military AI—has remained central throughout the debate. This challenge permeates across the entire lifecycle of military AI systems, from research and development through testing, acquisition, deployment, and operational use.

Ensuring compliance with international humanitarian law (IHL) is both a normative imperative and a practical necessity. Respect for IHL underpins the legitimacy of military operations and is essential for the responsible adoption of military AI. Conversely, uncertainty about legal requirements can also create inefficiencies throughout the AI system lifecycle, delaying development, complicating acquisition and legal review, and disrupting operational planning.

Against this background, Defence and industry face two recurring challenges when applying legal requirements to the development, review, and use of military AI systems.

Problem 1: Abstract Legal Norms

International treaties, policy documents, and academic literature typically express legal requirements through high-level concepts and principles, such as 'indiscriminateness', '(context-appropriate) human judgment and control', and 'human agency'.[2] While these concepts provide important normative guidance, they are often too abstract to translate directly into system requirements, technical specifications, operational decisions. This creates practical challenges for those responsible for the acquisition, development, legal review, and use of military AI systems:

- ***Defence:*** Defence personnel often struggle to identify the technical, operational, and environmental factors that determine the lawfulness of military AI systems. This can make legal reviews more difficult, complicate acquisition and development, and lead to inefficient expenditure of time and resources.
- ***Industry:*** Industry actors frequently receive legal requirements from Defence that are too abstract to translate into concrete technical specifications. This can create uncertainty during design and

[1] See NATO, 'Summary of NATO's Revised Artificial Intelligence (AI) Strategy' (7 October 2024) <https://www.nato.int/en/about-us/official-texts-and-resources/official-texts/2024/07/10/summary-of-natos-revised-artificial-intelligence-ai-strategy> accessed 23 June 2026; GC REAIM, 'Responsible by Design: Strategic Guidance Report on the Risks, Opportunities, and Governance of Artificial Intelligence in the Military Domain' (Hague Centre for Strategic Studies 2025) <https://hcss.nl/report/gc-reaim-responsible-by-design-strategic-guidance-report/> accessed 19 November 2025 (stating that 'stating that "the development and use of AI in the military domain must comply with international law, guided by widely shared ethical principles, to best ensure the preservation of human life and peace for present and future generations'); US Department of Defense, 'Autonomy in Weapon Systems' (2023) US DoD Directive 300009; Laura Bruun and Marta Bo, 'Bias in Military Artificial Intelligence and Compliance with International Humanitarian Law' (SIPRI 2025) <https://www.sipri.org/publications/2025/other-publications/bias-military-artificial-intelligence-and-compliance-international-humanitarian-law> accessed 29 August 2025; European Parliament, 'Defence and Artificial Intelligence' (202 AD) <https://www.europarl.europa.eu/RegData/etudes/BRIE/2025/769580/EPRS_BRI(2025)769580_EN.pdf> accessed 24 July 2026.

[2] Laura Bruun, 'Towards a Two-Tiered Approach to Regulation of Autonomous Weapon Systems: Identifying Pathways and Possible Elements' (Stockholm International Peace Research Institute 2024) <https://doi.org/10.55163/LPED7967> accessed 25 September 2024; Stop Killer Robots, 'GGE's "Rolling Text": How Will Autonomous Weapons Be Regulated?' (*Medium*, 17 August 2025) <https://stopkillerrobots.medium.com/gges-rolling-text-how-will-autonomous-weapons-be-regulated-2751a0e68487> accessed 24 July 2026.

development, resulting in delays, unnecessary redesign, and avoidable costs where legal expectations are not clearly understood from the outset.[3]

Problem 2: Context-Dependent Legal Assessments

Even where legal norms can be translated into more concrete requirements, assessing compliance remains inherently context-dependent. The lawfulness of a military AI system cannot be determined by reference to a single technical feature or performance metric. Rather, it depends on the interaction between technical (e.g. model training, human-machine interface), operational (e.g. mission objectives, time and space constraints), and environmental factors (e.g. civilian presence, terrain, adversary behaviour).

Several studies have attempted to identify the variables that shape context-specific legal assessments.[4] While these represent an important step forward, most focus on a relatively limited subset of factors—often those most prominent in policy debates—or on AWS alone. As a result, the broader spectrum of military AI applications, such as DSS and AI-enabled cyber-capabilities, and the interactions between the variables that determine legal acceptability, remain only partially understood.

This project responds to these challenges by developing a structured methodology and accompanying toolkit for assessing the legal acceptability of military AI systems. It **translates abstract legal principles into operationally meaningful criteria**, while recognising that legal assessments remain inherently context-dependent. Rather than seeking to eliminate context, this framework is designed to provide a structured way to identify and evaluate the diverse technical, operational, environmental, human, and institutional variables that determine whether an AI system can be used lawfully in a particular situation.

This report focuses on developing a methodology for analysing the *legal* acceptability of AI systems. *Ethical* and *societal* considerations are also important for decision-makers to take into consideration,[5] however, the current framework and toolkit do not (yet) include these normative sources, instead opting to provide a methodological proof-of-concept by relying on legal sources only (see also Section 2.2). The success of this project for the legal domain, subsequently, can provide an indication on how ethical or societal norms can either (a) be analysed in the same way, or (b) be potentially integrated into this existing legal framework. For more on these future directions of research, see Section 5.

This report and framework focus on AWS and AI-enabled DSS, as these are the applications that carry the most critical consequences for civilians. Moreover, these applications are proliferating rapidly because States are increasingly investing in rearmament, especially in AWS and DSS for targeting.[6]

This report and its accompanying toolkit forms part of a main deliverable within the ELSA Lab Defence project by the Asser Institute. Based on an analysis of existing literature and debates, insights from a workshop held in

---

[3] E.g. Netherlands Ministry of Defence, 'Nieuwe challenge voor het uit de lucht halen van snelvliegende drones' (4 March 2026) <https://www.defensie.nl/actueel/nieuws/2026/03/02/nieuwe-challenge-voor-het-uit-de-lucht-halen-van-snelvliegende-drones> accessed 24 July 2026.

[4] E.g. Jonathan Kwik, *Lawfully Using Autonomous Weapon Technologies* (TMC Asser Press 2024) 167–225 <https://doi.org/10.1007/978-94-6265-631-4> accessed 25 September 2024; Bruun (n 2); Herwin Meerveld and others, 'Operationalising Responsible AI in the Military Domain: A Context-Specific Assessment' (2025) 27 Ethics and Information Technology 48 <https://doi.org/10.1007/s10676-025-09865-y> accessed 24 July 2026; Antoine Smallegange, Jurriaan Van Diggelen and Jonathan Kwik, 'Characteristics of the Military Domain and Their Impact on Military AI Applications' *2025 International Conference on Military Communication and Information Systems (ICMCIS)* (2025) <https://doi.org/10.1109/ICMCIS64378.2025.11047950> accessed 1 July 2025.

[5] See Henning Lahmann and others, 'Design Methodologies for Addressing Ethical, Legal and Societal Aspects (ELSA) of Military AI Applications' (Social Science Research Network, 7 November 2024) <https://doi.org/10.2139/ssrn.5116168> accessed 23 June 2026; Sadjad Soltanzadeh, 'Normative Reasoning and Military AI: DILEMA's Dilemmas' in Bérénice Boutin, Taylor Kate Woodcock and Sadjad Soltanzadeh (eds), *Legal, Ethical, and Technical Dilemmas in Military Artificial Intelligence* (TMC Asser Press 2026) <https://doi.org/10.1007/978-94-6265-759-5_2> accessed 23 June 2026.

[6] European Commission, 'EDF Work Programme 2026' (European Commission 2025) <https://defence-industry-space.ec.europa.eu/edf-work-programme-2026_en> accessed 24 July 2026; Tara Copp, Elizabeth Dwoskin and Ian Duncan, 'Anthropic's AI Tool Claude Central to U.S. Campaign in Iran, amid a Bitter Feud' *The Washington Post* (4 March 2026) <https://www.washingtonpost.com/technology/2026/03/04/anthropic-ai-iran-campaign/> accessed 24 July 2026.

The Hague on 28-29 January 2026,[7] responses from a broad stakeholder survey conducted in June 2025, and further feedback from stakeholders and ELSA Lab Defence partners, the goal of this project is to:

- Offer a **versatile theoretical framework** for understanding how different variables interact, to determine the overall 'compliance risk' of a system form a legal perspective;
- Identify **practical mitigation measures** to reduce such risks;
- Produce a **toolkit for practitioners** that translates these insights into a usable cognitive framework; and
- Develop, more broadly, a proof-of-concept of how abstract (legal, ethical, and societal) norms can be **analysed and translated into an accessible methodological framework** for use by practitioners.

Beyond its contribution to the legal discussion on military AI, this project's outputs are particularly relevant to:

Legal Review or Assessment Officers	When assessing an AI system, evaluators must identify the technical, socio-technical, operational, and institutional factors relevant to its lawfulness and understand how those factors interact. This report provides an overview of those factors and a framework for analysing their relationships, which supports structured legal reasoning.
Defence Acquisition Officers	When defining requirements for AI systems during acquisition or (co-) development, acquisition officers must translate legal and operational needs into clear, concrete requirements for industry. This framework and toolkit identify the system and context variables relevant to a particular project and help translate them into requirements that are clear and mutually intelligible to both Defence and industry.
Commanders and Operators	When deciding to deploy an AI system, commanders and operators need to contextually assess this decision's lawfulness based on technical and operational factors. This output offers a cognitive map enabling them to identify legally-relevant factors and how to place them in the bigger operational picture.
Industry and Developers	The framework and toolkit provide a clear and practical overview of factors that should be considered and incorporated into the design of prototypes or products. Considering such factors early improves efficiency and provides a competitive advantage, as it can be demonstrated and explained to Defence customers how the product is already designed with IHL in mind.

This resulting framework and toolkit provide an **accessible common frame of reference** for assessing the legal acceptability of military AI systems. By establishing a shared vocabulary and understanding of complex technical and operational concepts, they are intended to facilitate cooperation between Defence stakeholders at various levels and between Defence and industry, supporting more consistent and transparent legal assessment throughout the AI system lifecycle.

2. Project Background

2.1. A Framework Guiding the Continuous and Iterative Evaluations of Military AI Systems

Military AI systems are evaluated throughout their lifecycle, using both formal and informal processes. Among the most prominent formal mechanisms is the legal review required by Article 36 of Additional Protocol I to the Geneva Conventions (API), under which States must determine whether the use of a new weapon, means, or

[7] Asser Institute, 'Developing Legal Assessment Frameworks for Military AI: ELSA Lab Defence' (*Asser Institute*, 2 June 2026) <https://www.asser.nl/news/workshop-developing-legal-assessment-frameworks-for-military-ai-elsa-lab-defence/> accessed 24 July 2026.

method of warfare would be prohibited by international law.[8] In addition, procurement processes often include preliminary or informal assessments prior to formal review (a 'pre-review').[9] Moreover, systems may require re-evaluation after deployment, particularly where they are updated, modified, retrained, or otherwise altered.[10] Whether formal or informal, these evaluations differ from those applicable to traditional ("legacy") systems in at least two important respects.

First, military AI systems are not static. They are continuously updated, retrained, and potentially co-developed by Defence and suppliers over time. Their functionality may evolve through software modifications, new data inputs, or changes in operational parameters. As a result, any meaningful legal evaluation must be continuous and, where necessary, iterative.[11] This marks a departure from more traditional IHL reviews of new means and methods of warfare, which generally focus on a system at the point of review rather than its continued evolution, anticipated use cases, and changing operational conditions.[12]

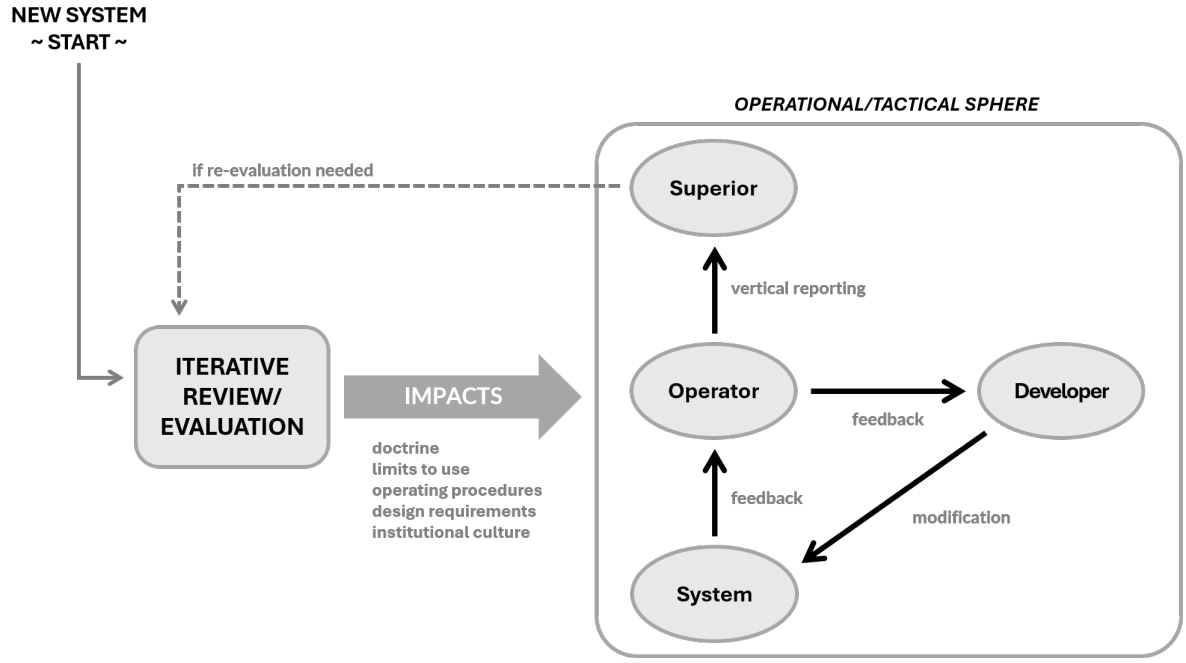


Fig.1: *The iterative development and evaluation process*[13]

Second, evaluating the legal acceptability of military AI systems heavily depends on the operational context in which they are used. Pre-deployment testing can only approximate real-world conditions and cannot anticipate every situation a system may encounter. As a result, testing, validation, and quality assurance alone are not sufficient.[14] Instead, evaluation must be iterative, drawing on operational experience, feedback from users, technological developments, and evidence from the system's own performance over time.

---

[8] ICRC, *A Guide to the Legal Review of New Weapons, Means and Methods of Warfare: Measures to Implement Article 36 of Additional Protocol I of 1977* (ICRC 2006) <https://international-review.icrc.org/sites/default/files/irrc_864_11.pdf> accessed 25 March 2025.

[9] Netta Goussac and Vincent Boulanin, 'Responsible Procurement of Military Artificial Intelligence' (Stockholm International Peace Research Institute 2026) <https://doi.org/10.55163/YOLG1827> accessed 24 July 2026.

[10] Jonathan Kwik, 'Iterative Assessment for Military Artificial Intelligence (AI) Systems' in Bérénice Boutin, Taylor Kate Woodcock and Sadjad Soltanzadeh (eds), *Legal, Ethical, and Technical Dilemmas in Military Artificial Intelligence* (TMC Asser Press 2026) <https://doi.org/10.1007/978-94-6265-759-5_9> accessed 10 April 2026.

[11] Ulysse Richard and others, 'Before and after Use: A Lifecycle View of Military AI' (UNODA 2026) <https://publications.unoda.org/publications/maps-dialogues-second-webinar-2026> accessed 21 June 2026; Ingvild Bode and Anna Nadibaidze, 'Strengthening Human Agency in the Military Domain: Best Practices Toolkit for Policymakers, Developers, and Users of AI-Based Systems' (Zenodo, 13 January 2026) <https://doi.org/10.5281/ZENODO.17671715> accessed 17 February 2026.

[12] Vincent Boulanin, Netta Goussac and Laura Bruun, 'Autonomous Weapon Systems and International Humanitarian Law: Identifying Limits and the Required Type and Degree of Human–Machine Interaction' (SIPRI 2021) 29–30 <https://www.sipri.org/publications/2021/policy-reports/autonomous-weapon-systems-and-international-humanitarian-law-identifying-limits-and-required-type> accessed 24 July 2026.

[13] Adapted from Kwik (n 10) 228.

[14] ibid 206–209.

These considerations reinforce the importance of evaluating military AI systems in their operational context. Article 36 legal reviews of means and methods of warfare, especially of military AI systems, must be made on the concrete operational circumstances in which they are expected to be used, not solely on their technical specifications. While implementation of IHL at design stage and 'lawful by design' initiatives are important,[15] IHL primarily regulates the *use* of means and methods of warfare rather than their design. It is therefore notoriously difficult to conclude that a weapon is unlawful in *all* circumstances, particularly where the behaviour and effects of an AI system largely depend on its interactions with the operational environment.[16]

This raises a central question: *how can military AI systems be evaluated continuously and iteratively—throughout their lifecycle—to determine whether their use is legally acceptable*? This framework and toolkit seek provide structured indications of the **legally relevant factors** that should be considered when (continuously) evaluating military AI systems. It makes explicit a range of factors—such as performance indicators, human control, data dependencies, operational environment, system interaction—and the single or combined effects of these elements on risks to IHL compliance.

2.2. Lawfulness Under IHL: A Key Component of the Legitimacy of Military Operations

This framework supports the assessment of the alignment of military AI systems with applicable legal rules, by identifying the legally salient factors relevant to their use. Its focus is on *ius in bello* considerations, namely compliance with IHL in the context of their deployment. Compliance with IHL is an essential component of the legitimacy of military operations, but it is not the only consideration.[17] Decisions on whether to deploy a military AI system may also be influenced by strategic priorities, operational objectives, military necessity, and broader ethical or societal norms, such as meaningful human control, the dehumanisation of warfare, or the risk that increased speed and scale may encourage less considered use of force. As a result, a system that fully complies with IHL may nevertheless be considered unsuitable for deployment due to ethical, reputational, or strategic considerations.[18]

Adherence to IHL is therefore only one component of the legitimacy of military operations, yet, it is an "indispensable" one.[19] Respect for IHL constitutes an "indispensable foundation upon which any claim to legitimacy must rest".[20] This understanding is reflected in the military doctrines of many States, including the Netherlands, which recognises compliance with IHL during military operations as a central component of operational legitimacy.[21]

While the overall legitimacy of military operations also depends on ethical considerations and societal acceptance, respect for IHL contributes directly to both. Compliance with IHL is closely linked to ethical evaluations of military operations and their acceptance by affected populations, domestic audiences, and the broader public.[22] This framework therefore focuses on legally salient factors relevant to compliance with IHL during deployment. It recognises that compliance with IHL constitutes only one component of the overall legitimacy and acceptability of military operations, albeit an indispensable one.

---

15 See e.g. Damian Copeland, 'Lawful by Design Initiative' (*Article 36 Legal*, 2026) <https://www.article36legal.com/lawful-by-design> accessed 24 July 2026.

16 Laura Bruun, Marta Bo and Netta Goussac, 'Compliance with International Humanitarian Law in the Development and Use of Autonomous Weapon Systems: What Does IHL Permit, Prohibit and Require?' (SIPRI 2023) 13 <https://doi.org/10.55163/DFXR3984> accessed 24 July 2026.

17 Jessica Dorsey, Keeping Our Humanity: Ensuring The Legitimacy Of Military Targeting Operations Through Civilian Harm Mitigation In Increasingly Autonomous Warfare, Phd Thesis, at 41 https://publication-online.com/p/jessica-dorsey/

18 See e.g. Netherlands Ministry of Defence, 'Integraal Risicomanagement (IRM)' (Netherlands Ministry of Defence 2025) 4.

19 Geoff Corn and Robert Ashley, 'LOAC and Legitimacy: When Combat Becomes a Supporting Effort to Information' (*Lieber Institute West Point*, 18 January 2022) <https://lieber.westpoint.edu/loac-legitimacy-combat-supporting-effort/> accessed 24 July 2026.

20 Jessica Lynn Dorsey, 'Keeping Our Humanity: Ensuring the Legitimacy of Military Targeting Operations Through Civilian Harm Mitigation in Increasingly Autonomous Warfare' (Doctoral Thesis, Utrecht University 2026) 43 <https://doi.org/10.33540/3308>.

21 Netherlands Ministry of Defence, 'Nederlandse Defensie Doctrine 2025' (Ministerie van Defensie 2025) 24 <https://www.defensie.nl/documenten/2025/02/28/nederlandse-defensie-doctrine> accessed 24 July 2026.

22 Amichai Cohen and David Zlotogorski, *Proportionality in International Humanitarian Law* (Oxford University Press 2021) 17–18 <https://doi.org/10.1093/oso/9780197556726.001.0001>.

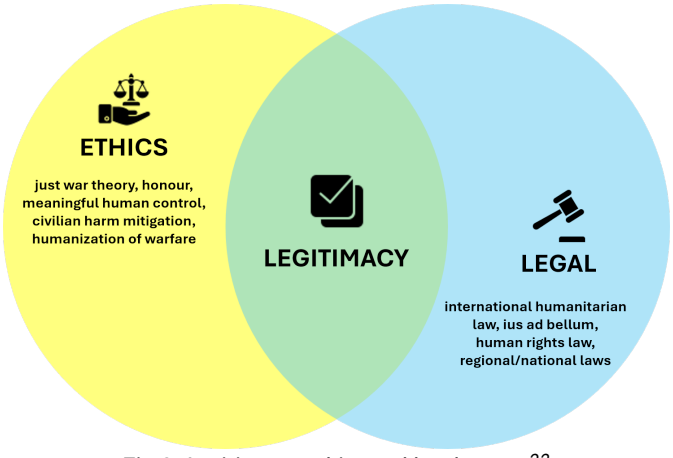


Fig.2: *Legitimacy, ethics and legal norms*[23]

As discussed above, this framework focuses on *legal* compliance (the right-hand side of Fig. 2), and primarily on compliance with IHL. Legal and ethical norms may diverge on certain issues, such as the whether the dehumanisation of warfare associated with AI systems is acceptable.[24] Even within the legal domain, different bodies of law may point in different direction. For example, the collection of personal data to improve targeting may be permissible under IHL, but is restricted under international human rights law (IHRL).[25] Incorporating multiple normative frameworks would therefore require the methodology to accommodate potentially conflicting evaluations of the same facts, significantly increasing its complexity. This framework accordingly adopts a deliberately narrower scope, by focusing on IHL alone.[26] Section 5, nevertheless, considers how the framework could be extended to include other bodies of law, including *ius ad bellum* and IHRL, as well as ethical and societal considerations.

3. Purpose, Scope, and Intended Use

This framework and toolkit are intended for both Defence and industry. The following sections describe how they can support each stakeholder throughout the acquisition, development, legal review, and deployment of military AI systems, while promoting a common understanding of the factors relevant to legal acceptability

3.1. For Defence: Methodological Support Instrument

The framework is intended primarily to support structured legal reasoning by relevant Defence stakeholders and decision-makers, including those involved in acquisition, legal review, and operations. It supports structured legal reasoning by helping users identify the factors relevant to IHL compliance and assess how they interact in the use of military AI systems. It is not intended to replace human judgment in context-specific legal assessments and decisions.

Among these, **legal reviewers and other pre-deployment evaluators** assessing a new AI system should be aware of the key factors relevant to AI evaluation—factors that may differ from those applicable to traditional systems—and how operational variables can play an important role in legal review. They should also understand how these

---

[23] Adapted from Dorsey (n 20) 45.

[24] This question is often answered in the negative from an ethical perspective, but in the positive from a legal perspective, provided that sufficient precautions are taken to avoid mistakes. See e.g. Giovanni Sartor and Andrea Omicini, 'The Autonomy of Technological Systems and Responsibilities for Their Use' in Nehal Bhuta and others (eds), *Autonomous Weapons Systems* (Cambridge University Press 2016) 72 <https://doi.org/10.1017/CBO9781316597873.003>; Masahiro Kurosaki, 'Toward the Special Computer Law of Targeting' in Claus Kreß and Robert Lawless (eds), *Necessity and Proportionality in International Peace and Security Law* (Oxford University Press 2020) 412 <https://doi.org/10.1093/oso/9780197537374.003.0014>.

[25] See e.g. Leah West, 'Face Value: Precaution versus Privacy in Armed Conflict' in Russell Buchan and Asaf Lubin (eds), *The Rights to Privacy and Data Protection in Times of Armed Conflict* (NATO CCDCOE 2022).

[26] Stephen Coleman, 'Ethical Challenges of New Military Technologies' in Hitoshi Nasu and Robert McLaughlin (eds), *New Technologies and the Law of Armed Conflict* (TMC Asser Press 2014) 30–31 <https://doi.org/10.1007/978-90-6704-933-7_3>.

factors may be interrelated or mutually influential, and how evaluations may need to be revisited following updates or modifications during deployment, making legal review an integral part of the iterative evaluation cycle.

For **commanders**, the tool helps make explicit the (legal) bases underpinning operational decisions. Engaging with such a framework may form part of the precautionary obligations under IHL, insofar as it structures *ex ante* consideration of factors relevant to lawful targeting and deployment. It also supports **accountability**, as undertaking structured legal reasoning in advance enables decision-makers to explain and justify their assessments should outcomes later be questioned.

Legal advisers may equally benefit, particularly where they are not technical experts in AI systems. This framework can assist in identifying which system characteristics or contextual elements require closer legal scrutiny. It may also be relevant for weaponeers and those involved in capability integration.

By contrast, **end-users and operators** will often rely on assurances that a system has been cleared for use and these actors may focus primarily on performance and whether the system functions as intended. For that reason, the framework is likely to be more useful 'upstream' (e.g. during procurement, doctrine development, capability development, and legal review processes).

The framework does not replace legal evaluations, for example, within procurement or legal review processes. Rather, it is **designed to support and structure the assessment of compliance with IHL within established processes**. In practice, there is no standardised way of conducting legal reviews of AWS nor evaluations of DSS,[27] and this framework has been designed to be sufficiently versatile in order to support any such reviews or assessments, regardless of the specific format.

In addition to formal Article 36 reviews, there may be "pre-reviews" or preliminary assessments of certain systems within procurement processes, particularly at early stages of development or acquisition. Moreover, procurement is increasingly characterised by rapid acquisition cycles, driven by the perception that existing procedures are excessively bureaucratic and that potential AI-enabled advantages may otherwise be lost. In such contexts, structured guidance on legally relevant factors becomes even more important.[28] The framework may also assist in situations where a system appears manifestly problematic. Having a systematised range of legally relevant factors can help articulate and substantiate a determination that a particular use would be unlawful.

Importantly, the framework **does not lead to definitive legal determinations**. It is not prescriptive in its outcomes and does not produce binary "yes/no" answers. Legal assessments cannot be reduced to compliance-based risk scoring exercises, for they entail contextual and normative judgments inherent in targeting operations, for which human decision-makers remain responsible.[29] In the fog of war, 100% certainty is unattainable and no compliance framework can absorb or eliminate all risks, as legal responsibility is not discharged by merely ticking boxes or risk scoring. This framework supports structured reasoning, but it does not replace the judgment of decision-makers or absolve them of responsibility and accountability.

3.2. For Industry: Clarity on Legal Requirements

This framework provides **structured guidance on the legally relevant factors that industry actors should take into account** when developing military AI systems. It may be useful to industry actors in several ways.

---

[27] Ioana Puscas, *AI and International Security: Understanding the Risks and Paving the Path for Confidence-Building Measures* (UNIDIR 2023) 39 <https://unidir.org/wp-content/uploads/2023/10/UNIDIR_AI-international-security_understanding_risks_paving_the_path_for_confidence_building_measures.pdf> accessed 15 July 2026; Klaudia Klonowska, 'Article 36: Review of AI Decision-Support Systems and Other Emerging Technologies of Warfare' in Terry D Gill and others (eds), *Yearbook of International Humanitarian Law, Volume 23 (2020)* (TMC Asser Press 2022) <https://doi.org/10.1007/978-94-6265-491-4_6> accessed 4 October 2024.

[28] See Goussac and Boulanin (n 9); Samuel Bendett and David Kirichenko, 'Battlefield Drones and the Accelerating Autonomous Arms Race in Ukraine - Modern War Institute' (*Modern War Institute*, 10 January 2025) <https://mwi.westpoint.edu/battlefield-drones-and-the-accelerating-autonomous-arms-race-in-ukraine/> accessed 24 July 2026.

[29] See Taylor Kate Woodcock, 'Human/Machine(-Learning) Interactions, Human Agency and the International Humanitarian Law Proportionality Standard' (2024) 38 Global Society 100 <https://doi.org/10.1080/13600826.2023.2267592> accessed 19 January 2026.

First, not all companies—particularly small or medium-sized enterprises (SME) or non-traditional defence contractors—have in-house legal expertise in IHL. Second, demand from Defence is not uniform as operational contexts, available resources, effects sought, and national doctrines differ significantly. There is no standardised model for acquisitions of AI systems, and in many cases no sufficient specification from Defence actors at early stages of engagement.

However, achieving "clarity and certainty about what military clients expect from the AI capabilities they procure"[30] requires legal requirements to be communicated in a clear, accessible, and operationally meaningful way. Doing so offers several important benefits for both Defence and industry. First, decisions affecting lawful use and respect for IHL during operations are often made during the design stage (see also Toolkit, **[Engagement in Development]**). Second, for companies, anticipating legally relevant constraints may be strategically valuable. For example, doing so could ensure that preliminary evaluations of proposals or prototypes are successfully passed, and doing so may also facilitate development processes and streamline cooperation with Defence.

More generally, anticipating legal requirements enables companies to remain competitive. In an increasingly crowded defence market, demonstrating that legal compliance has been considered throughout the design and development process can provide a competitive advantage by reducing development risk and giving acquisition authorities greater confidence in the product's suitability. Therefore, a structured framework supports this by helping companies identify and address IHL considerations early, before they become costly barriers to procurement or deployment.

This framework may be particularly useful in contexts of rapid procurement, closer defence-private sector collaboration, or co-development between defence and industry. In such settings, **a shared framework and common terminology** regarding factors relevant to IHL compliance can facilitate cooperation and ultimately respect for IHL. Where industry actors are more deeply involved in design, adaptation, and iterative system updating and evaluation of AI systems, questions of IHL risk awareness, due diligence, and accountability become increasingly important. Companies therefore have an interest in ensuring that employees understand the factors that may contribute to unlawful or otherwise harmful outcomes. As such, this framework and toolkit can support the internal training of staff, while also providing a structured means of documenting due diligence and internal risk governance, thereby helping to demonstrate that IHL-related risks have been identified and appropriately managed throughout the development process.

Moreover, this framework can support companies when providing **training or instruction to customers**. For example, where products are highly customisable—such as decision-support platforms that allow clients to configure their own systems—clarifying which parameters and factors may affect IHL compliance can help guide end-users through the customisation process and reduce the legal or reputational risks of the developer. In these respects, the framework supports industry actors seeking to engage responsibly in the development and supply of military AI systems. It also complements broader efforts to strengthen responsible innovation, risk governance, and legal compliance across the defence AI sector.[31]

3.3. Explaining the Legal Rationales

Evaluations of military AI systems are highly context-dependent. Understanding the legal rationales that underpin them helps to ensure that legal principles are applied consistently across different contexts and avoids "reinventing the wheel" in each new case. Our analysis—drawing on academic literature, an expert workshop held at the Asser Institute,[32] and international discussions, including those within the framework of the CCW Group of Governmental Experts (GGE) in Geneva[33]—suggests that evaluations of the lawfulness of use of military AI systems tend to revolve around three main sets of concerns.

Within the framework, the legal rationales are represented by the indicators **D-C-P**, as follows:

---

[30] Goussac and Boulanin (n 9) 26.

[31] E.g. UNIDIR, 'Framework of Responsible Industry Behaviour for AI in the Military Domain' (29 January 2026) <https://unidir.org/framework-of-responsible-industry-behaviour-for-ai-in-the-military-domain/> accessed 24 July 2026.

[32] Asser Institute (n 7).

[33] Stop Killer Robots (n 2).

D DISTINCTION RELATED CONCERNS	The **principle of distinction** requires parties to an armed conflict to distinguish at all times between civilians and combatants, and between civilian objects and military objectives.[34] Delegating distinction-related functions to military AI systems therefore raises concerns about whether those systems can reliably distinguish lawful military targets from protected persons and civilian objects in the envisioned operational context, particularly in complex and dynamic environments.[35] Where an AI system cannot make the requisite distinctions required by IHL in its intended context of use, its use in that context is not legally permissible unless the relevant, operational or system conditions are modified to ensure compliance.[36]
C Civilian PresenceILIAN HARM RELATED CONCERNS	The **principle of proportionality** prohibits attacks in which the expected incidental harm to civilians or civilian objects would be excessive in relation to the concrete and direct anticipated military advantage.[37] For example, DSS raise significant questions about whether they enable users to make the contextual and qualitative judgments required by IHL to conduct genuine proportionality assessments.[38] Both AWS and DSS also raise concerns about failures to detect protected persons or civilian objects, potentially resulting in inaccurate collateral damage estimates and, consequently, flawed proportionality assessments. Moreover, even where an AI system 'correctly' identifies a legitimate target, an attack may still cause excessive incidental civilian harm, including through direct or reverberating effects. All these considerations are central when evaluating whether, and under what conditions, an AI system may be lawfully used.
P PRECAUTION RELATED CONCERNS	Beyond the substantive rules of distinction and proportionality, IHL also requires parties to take all feasible **precautions in attack**.[39] This is a standalone obligation requiring a number of precuationary steps to be taken, as well as maintaining constant care during operations more generally, to spare civilians and civilian objects. These include the obligation to verify targets and refrain, cancel, or suspend attacks.[40] This principle is procedural in nature, in that it comprises measures and processes to be *performed* by decision-makers before and during an attack, rather than just results to be achieved. Many concerns surrounding the **predictability and understandability** of AI systems can be understood through the lens of this obligation.[41] These epistemic challenges include whether the behaviour and effects of an AI system can be predicted with sufficient confidence in the envisaged operational context, whether its outputs can be understood and critically assessed by human decision-makers, and whether appropriate human control can be exercised throughout the targeting process. Where epistemic uncertainty about the system's behaviour and its likely effects are too great, decision-makers may be unable to determine whether use

[34] *Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts* (8 June 1977) art 48,51,52 <https://ihl-databases.icrc.org/en/ihl-treaties/api-1977>.

[35] Kwik (n 4) 236–240; Marta Bo, 'Autonomous Weapons and the Responsibility Gap in Light of the Mens Rea of the War Crime of Attacking Civilians in the ICC Statute' (2021) 19 Journal of International Criminal Justice 275 <https://doi.org/10.1093/jicj/mqab005>; Abhimanyu George Jain, 'Autonomous Cyber Capabilities and Individual Criminal Responsibility for War Crimes' in Rain Liivoja and Ann Väljataga (eds), *Autonomous Cyber Capabilities under International Law* (NATO CCDCOE Publications 2021).

[36] For example, deployment may be delayed until the density of civilian concerns in the operational environment has decreased. See Toolkit, **[Civilian Presence], [Environment Complexity]**.

[37] *Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts* (n 34) art 51(5)(b), 57(2)(a)(iii).

[38] Abhimanyu George Jain, 'Autonomous Weapon Systems, Errors and Breaches of International Humanitarian Law' (2023) 21 Journal of International Criminal Justice 1005 <https://doi.org/10.1093/jicj/mqad043> accessed 24 July 2026; Jessica Dorsey, 'Proportionality under Pressure: AI-Based Decision-Support Systems, the Reasonable Commander Standard and Human(e) Judgment in Targeting' [2025] GC REAIM Expert Policy Note Series 1 <https://hcss.nl/wp-content/uploads/2025/05/Dorsey.pdf> accessed 7 April 2026.

[39] *Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts* (n 34) art 57.

[40] Marco Sassòli and Anne Quintin, 'Active and Passive Precautions in Air and Missile Warfare' in Fania Domb, Yoram Dinstein (ed), *Israel Yearbook on Human Rights, Volume 44 (2014)* (Brill | Nijhoff 2014) 6–7 <https://doi.org/10.1163/9789004423954_003> accessed 24 July 2026.

[41] Arthur Holland Michel, *The Black Box, Unlocked: Predictability and Understandability in Military AI* (UNIDIR 2020) <https://doi.org/10.37559/SecTec/20/AI1> accessed 27 June 2025.

	would comply with IHL.[42] In such circumstances, additional precautions, testing, validation, or operational restrictions may be necessary before the system can be lawfully deployed. It is thus on decision-makers to undertake those actions necessary that would allow them to reasonably predict how the AI system will react in its use-context.

By linking variables to these legal rationales, the framework also promotes **legal transparency** by explaining why particular technical, operational, or environmental factors are legally relevant. Without such a framework, these relationships are not immediately obvious, particularly where variables affect compliance indirectly or only on combination with other factors.

For example, consider the issue of *modifications* (see also Toolkit, **[Modification]**). Concern is frequently expressed regarding the ease with which AI systems can be modified, and regulatory proposals often identify this as a potential risk to legal compliance.[43] But why *specifically*? The framework clarifies that modifications pose legal risk, but for different reasons. For example, under **D** and **C**, the legal issue arises from the *technical* danger (i.e. that said modifications may introduce performance drops). In contrast, for **P**, the main issue is the *unpredictability* of the effect, as it is not always foreseeable *when* a modification may alter the behaviour or performance of the AI system, or *how* (see also Fig. 3).

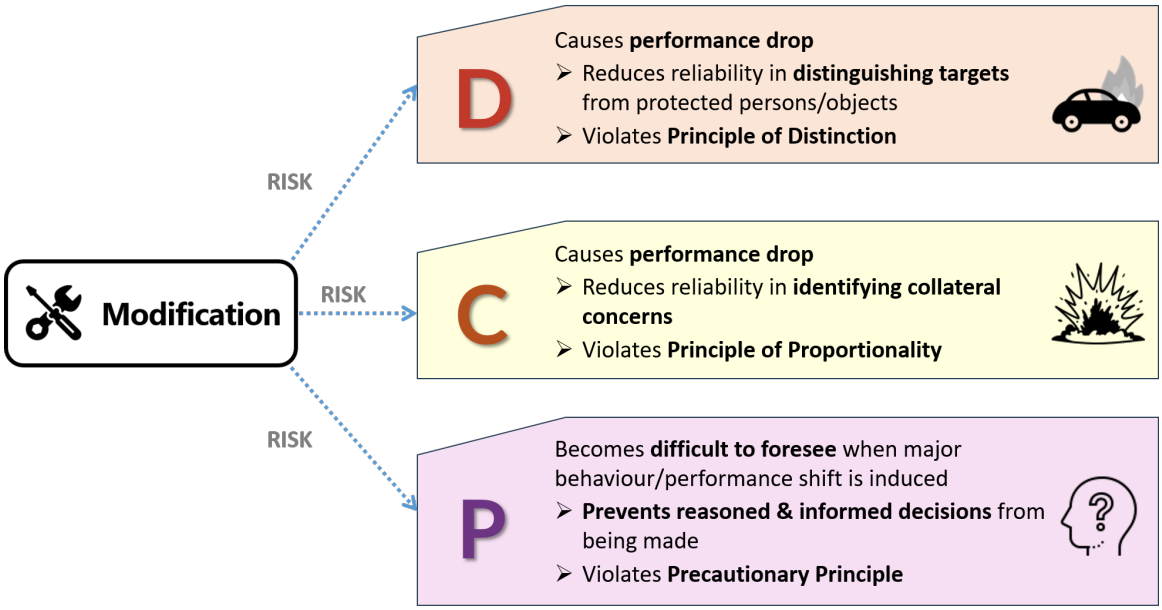


Fig.3: *Example breakdown of IHL-related risks*

Ultimately, the process set out in this framework and toolkit **supports informed decision-making and discussion**. It enables users to explain why a particular circumstance or factor is legally problematic and to explore ways of mitigating the associated risk (see Section 4.5).

[42] Kwik J and Van Engers T, 'Algorithmic Fog of War: When Lack of Transparency Violates the Law of Armed Conflict' (2021) 2 Journal of Future Robot Life 43 <https://doi.org/10.3233/FRL-200019>

[43] Group of Governmental Experts on Lethal Autonomous Weapons Systems (GGE on LAWS), 'Report of the 2023 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems' (2023) CCW/GGE.1/2023/CRP.2 s Annex III, para 34 <https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_(2023)/CCW_GGE1_2023_CRP.2_12_May.pdf> accessed 4 May 2026; US Department of Defense (n 1) 15.

4. Understanding the Toolkit

This section provides a complete explanation of all features of the toolkit. For a shorter, step-by-step practical guide on how to use the toolkit in practice, see Toolkit, p i-ix.

4.1. Variables: The Units of Evaluation

'Variables' are the fundamental units of evaluation in the framework. They represent the technical, operational, environmental, and organisational factors that influence whether an AI system can be used in compliance with IHL. **Each variable is documented in the accompanying toolkit (annexed to this report) as a self-contained entry**. Every entry includes a description of the variable, its relationship to the three legal rationales (D-C-P; see Section 4.2), guidance on assessment (see Section 4.3-4.4), associated risks (alone or in combination with other variables, see Section 4.5 and Section 4.7), and, where appropriate, measures that may help mitigate associated risks (see Section 4.6).

At the time of publication, the toolkit contains 27 variables are in use organised into seven thematic clusters (see Fig.4).

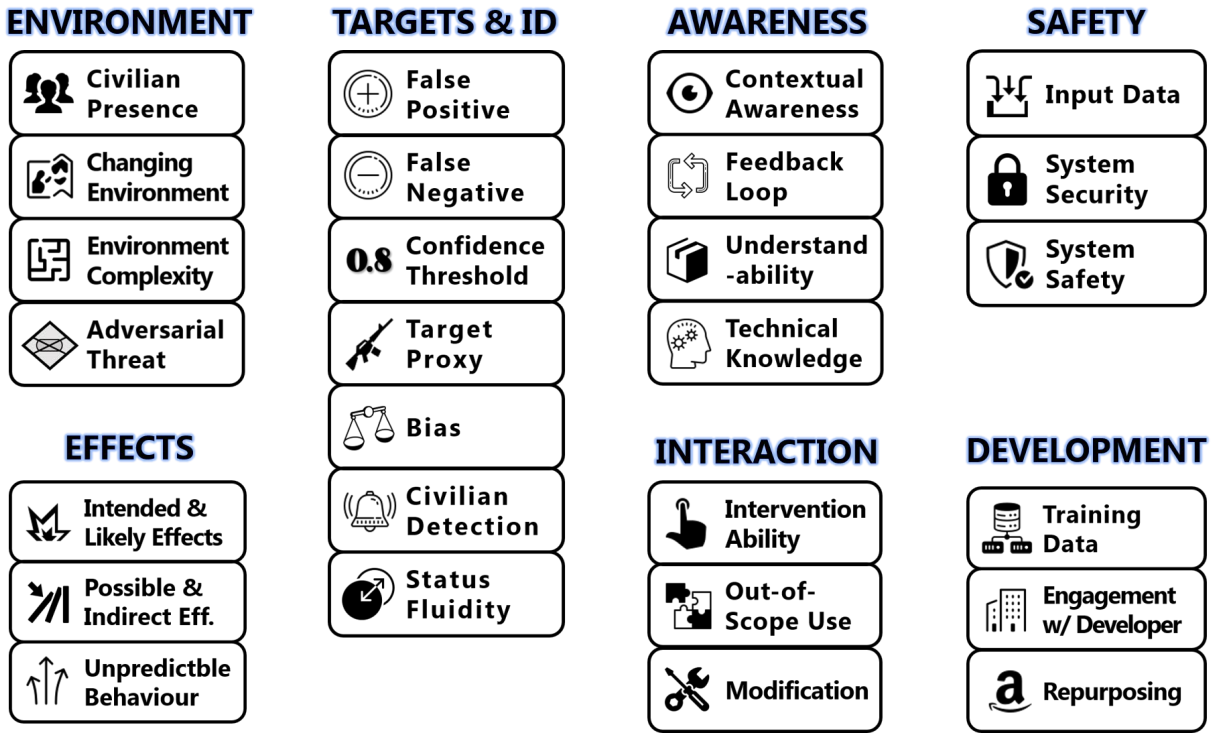


Fig.4: *Variables and clusters*

The variables extend beyond the characteristics of the AI system itself. They also encompass environmental conditions, operational choices, human factors, and organisational factors relating to system development, modification, and governance. This broader approach is necessary because legal acceptability is inherently context-dependent (see Section 1). Assessments must therefore consider not only the technical characteristics of the AI system, but also the wider operational, and socio-technical, and institutional (e.g. internal policy, working relationship with industry) context in which it is developed and used.

4.2. Linking Variables to Legal Rationales

Each variable in the toolkit is linked to one or more legal rationales to explain why it is relevant to assessing the legal acceptability of military AI systems. As explained in Section 3, the framework is intended to support *legal* reasoning while leaving the ultimate assessment to the judgment of the human evaluator. The variables included in this toolkit were therefore selected because they represent **objective factors that are legally relevant**.

Objective Factors

Variables are **not themselves linked to any legal rules or ethical norms**. For example, the framework does not include variables such as the 'ability to distinguish', which would more directly correspond to the IHL rule prohibiting indiscriminate weapons.[44] There are two main reasons for this.

First, the variables are designed to be evaluable. Legal determinations are difficult and require legal reasoning and interpretation, and reasonable assessors may reach different conclusions based on the same facts. The variables instead focus on observable features of AI systems that can be assessed more consistently across different users. This, in turn, facilitates discussion, comparison, and problem-solving. For example, while experts may disagree about whether a particular system is 'indiscriminate', they are more likely to agree on its False Positive Rate (see Toolkit, **[False Positive Rate]**) as measured through testing. That information can then inform the subsequent *legal* assessment of whether the system is indiscriminate.[45]

Second, the context-dependent nature of IHL compliance often requires **multiple variables to be considered together before a legal assessment can be made**. To expand on the previous example, a high False Positive Rate is not, by itself, sufficient to determine (un)lawfulness. Consider the classic counterargument from the literature: what if the system were deployed in a desert devoid of civilians?[46] In that context, a high False Positive Rate may not result in unlawful attacks on civilians. However, when considered alongside another variable, such as the presence of civilians and collateral concerns are present (see Toolkit, **[Civilian Presence]**), it becomes apparent that a high False Positive Rate is likely to result in the misidentification of protected persons, and consequently, a clear indication of an unlawful attack.

Legal Relevance

Although the variables are not *themselves* legal rules or concepts, each is relevant to assessing risks to legal compliance under IHL through one or more of the D-C-P rationales identified in Section 3.3. For clarity, each variable includes a brief explanation of why it is legally relevant by linking it directly to the relevant rationale or rationales.

Many variables are **relevant to multiple rationales**. For example, bias (see Toolkit, **[Bias]**) may lead to systematic misidentification—for instance, failing to recognise children or persons with disabilities because of insufficient or unrepresented training data. This may undermine compliance with the principle of distinction and, by affecting the accuracy of collateral damage estimates, also influence proportionality assessments (Fig.5).[47]

D DISTINCTION	Systemically poorer performance for particular groups/categories may increase the risk that protected persons or objects within those groups are misidentified as lawful targets. Where AI outputs are relied upon for target identification or classification, such disparities may therefore undermine the ability to distinguish reliably between military objectives and protected persons or objects.
C CIVILIAN HARM	Systemically poorer performance for particular groups/categories may also result in civilians or civilian objects being consistently under-detected or omitted from assessments of expected civilian harm. Where such outputs inform proportionality assessments, this may lead to civilian harm being underestimated, with particular groups—such as children or persons with disabilities—disproportionately affected.

Fig.5: *Example legal relevance explanation (from the variable **[Bias]**)*

[44] *Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts* (n 34) art 51(4)(a).

[45] There is no universal standard, and thresholds of acceptability may vary between States. Bruun, Bo and Goussac (n 16).

[46] Michael N Schmitt and Jeffrey S Thurnher, '"Out of the Loop": Autonomous Weapon Systems and the Law of Armed Conflict' (2013) 4 Harvard Law School National Security Journal 231, 246 <https://centaur.reading.ac.uk/89863/> accessed 10 July 2026.

[47] Bruun and Bo (n 1).

4.3. Assessing Variables Using the Evaluation Scale

In addition to explaining why a variable is legally important under IHL for risk assessment, each variable also includes guidance on how evaluators can assess the variables' level of risk. This guidance is supported by an evaluation scale (a five-point Likert Scale with descriptive indicators) (see Fig.6). The scale is intended as a qualitative aid to assessment and discussion, rather than as a numerical scoring system.



Very Low Risk	Little or no identifiable risk to IHL compliance in the anticipated use context.
Low Risk	Limited risk. Some attention may be required, but unlikely to raise significant legal concerns on its own.
Moderate Risk	Some risk to IHL compliance, depending on the operational context and its interaction with other variables. Caution and further assessment recommended.
High Risk	High risk to IHL compliance. Mitigation measures or additional safeguards should normally be considered before deployment.
Very High Risk	Significant legal risk. Unless adequately mitigated or justified by the broader assessment, deployment may be inappropriate.

Fig.6: *Evaluation scale and meanings*

Some variables may feature multiple sub-indicators, which may be cumulative (indicated by 'and') or alternative (indicated by 'or)'. Where sub-indicators are cumulative, evaluators should consider them together and make an overall **common-sense assessment of the level of risk**.

Figure 7 illustrates this approach for the variable **[Intervention Ability]**, which evaluates the extent to which operators can intervene when they notice an AI system exhibits problematic behaviour. Imagine that the system provides several redundant options for correcting or recalling it after activation (top row scoring green), but that all of these options take an unacceptably long time to take effect (bottom row scoring red). Under such circumstances, a reasonable evaluator may conclude that there is little or no effective means of correcting or recalling the system after activation, and therefore assign **[Intervention Ability]** an *overall high risk*.

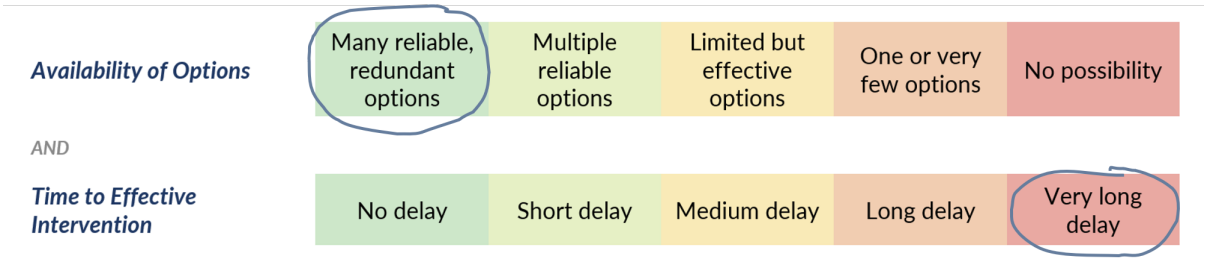


Fig.7: *Example of completed evaluation scale (from the variable **[Intervention Ability]**)*

4.4. Assessing Variables Where Information is Unavailable

In some cases, a variable **cannot be evaluated** because the information necessary to make such an assessment is unavailable. This may arise for several reasons:

1 | Information is Not Available, but Can Be Obtained

The information is not yet available to the evaluator, but can be retrieved. In this case, **additional steps should be taken to obtain it**. For example, assessing performance-related variables (such as **[False Positive Rate]** or **[Bias]**) may require testing data, which may need to be obtained independently or requested from the developer. Similarly, assessing operational variables (such as **[Civilian Presence]**) may require intelligence gathering to establish the current operational situation.

In such cases, where reasonably practical, the evaluation process should ideally be suspended until the requisite information has been obtained. This is recommended over proceeding with incomplete or missing information.

2 | Information is Only Available at a Later Stage

Some variables **can only be assessed at a *later* stage in the lifecycle**. For example, a procurement officer may not be able to assess the extent to which the adversary's tactics pose a threat to the AI system (see **[Adversarial Threat]**), since this depends on the characteristics of the specific adversary—or even the particular unit—encountered during operations. In practice, this assessment can only be made by an operational-level evaluator once further information becomes available. In such situations, two actions are recommended:

> (1) Where possible, the assessment should be based on **reasoned assumptions** derived from the system's intended use-context. For example, an AI-enabled cybersecurity system designed to defend against advanced persistent threats (APTs) can expect to face capable, well-resourced, and crafty adversaries;
>
> *and*
>
> (2) The evaluation should clearly record that the assessment is provisional and identify **the variable for re-evaluation once more information is available**. For example, a note could be addressed to the operational commander indicating that **[Adversarial Threat]** should be reassessed in light of the characteristics of the adversary they may ultimately encounter.

3 | Required Information is Unavailable

It is possible that relevant information should have been obtained at an **earlier stage in the AI system's lifecycle**. For example, an operational-level assessor is unlikely to be able to determine the quality and representativeness of the training data used to develop the system (see **[Training Data]**). In practice, this information should have been obtained during the design, development, or acquisition stages, with said information being communicated further to subsequent evaluators.

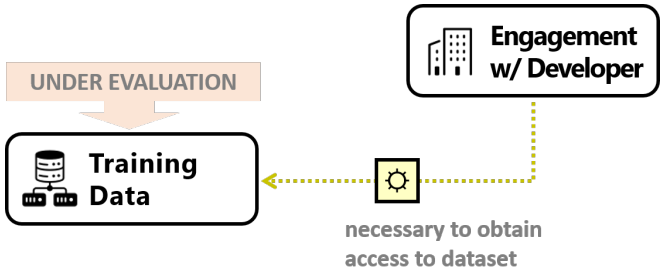
Where reasonably practicable, the evaluation process **should be suspended until the missing information can be obtained**, particularly where the unavailable information is important to the overall evaluation. For example, if an evaluator concludes that the threat posed by the adversary **[Adversarial Threat]** is high but cannot assess the system's resilience to enemy exploitation (see **[System Security]**), it may be appropriate to delay deployment until that information has been obtained. Where this is not possible, the evaluator may **proceed on the basis of clearly documented assumptions.** For example, an evaluator might record that they are assuming that the training data was sufficiently vetted during the acquisition and legal review process, because there is no reason to doubt that this occurred. Whether such an assumption is reasonable will depend on the circumstances. It may be justified within a well-established military with a strong acquisition and review process, but less so where the system has been recently acquired from a new supplier with an unknown track record.

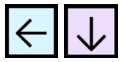
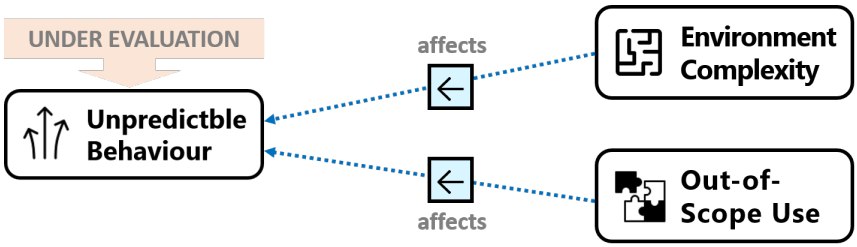
4.5. Interactions Between Variables

Variables should not be assessed in isolation. As explained earlier (see Section 1), compliance with IHL is inherently context-dependent. The legal acceptability of an AI system often depends on the interaction between multiple variables, rather than any single factor. Additionally, some variables are correlated, influence one another, or provide indications about the likely state of other variables.

During the development of the framework and toolkit, several types of relationships between variables were identified. The first two concern how a variable should be assessed (*when evaluating*), whereas the remaining relationships help evaluators interpret their significance within the wider context of the system (*after evaluating*).

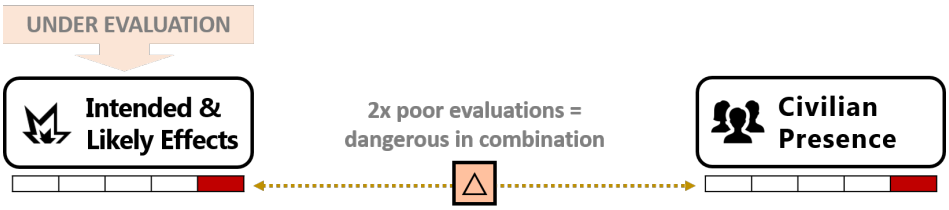
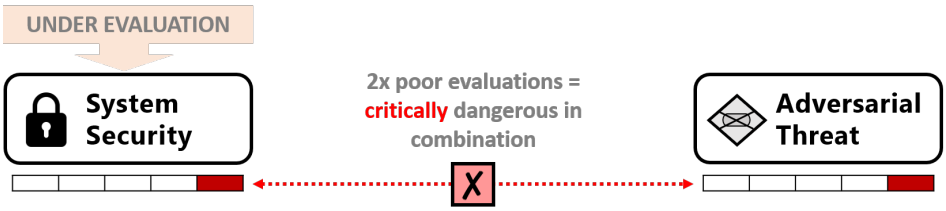
<table>
<tr><th colspan="2">BEFORE EVALUATING</th></tr>
<tr><td>(1)

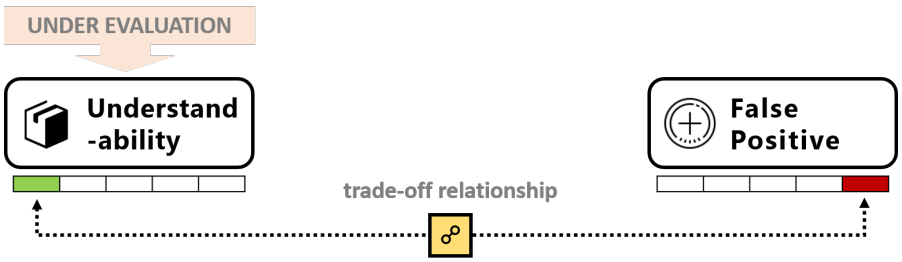
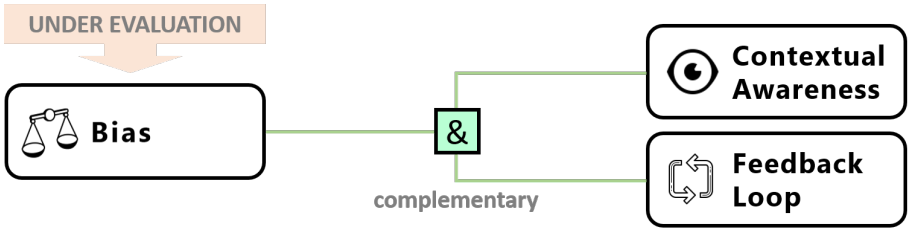
Enables evaluation</td><td>Certain variables may be prerequisites to the ability to evaluate other variables (see also Section 4.4). Identifying these information bottlenecks allows decision-makers to prioritise what relationships, organisational structures, or technologies to invest in beforehand, knowing that they will potentially enable important evaluations to be made ‘downstream’.


For example, obtaining access to training datasets to evaluate the quality and safety of [Training Data] may require a closer working relationship with the developer or a data transparency arrangement (a trait captured in [Engagement in Development]). In another example, predicting how a system may behave in response to less-common inputs (captured in the variable [Possible & Indirect Effects]) may require installing tools to render the AI model less opaque (captured in [Understandability]).</td></tr>
<tr><td>(2)

Correlation and Causation</td><td>Variables may be affected or influenced by other variables. By mapping out these correlations, this toolkit makes clear to decision-makers which other factors to cross-reference in the process of assessing particular variables.


For example, [Unpredictable Behaviour] may be more likely to occur if the environment is very complicated and features multiple interacting agents (a trait captured in [Environment Complexity]) or if the system is deployed in a use-context different than it was originally designed for (a trait captured in [Out-of-Scope Use]). When evaluating the risk of the former, therefore, it is useful to also consider the status of the latter variables and their potential impact.</td></tr>
</table>

AFTER EVALUATING

(3) △ **Dangerous combinations**	Certain **combinations of variables introduce aggravated risk if both evaluate poorly on the evaluation scale**. Identifying these relationships helps decision-makers remain aware of other variables they may need to urgently cross-check, to ensure that overall risk remains acceptable.  As a classic example, systems that cause harmful direct and secondary effects, such as drones carrying high-explosive munitions, are challenging to use lawfully in environments with a high density of civilians. A person evaluating **[Intended & Likely Effects]** who notices a high-risk score on the scale, therefore, may want to also reconsider **[Civilian Presence]**. If both evaluate poorly (as in the example above), action may be needed to mitigate this risk (see Section 4.6).
(4) ✗ **Critically dangerous combinations**	In an amplification of the risks stemming from the relationship above, certain combinations of variables may **disqualify the use of the system altogether if both score poorly** on the scale.  For example, a system that is poorly secured against spoofing, hacking, or manipulation (= *poor evaluation in **[System Security]***) is not, in itself, ideal. However, knowingly deploying such a system would be unacceptable in combination with the knowledge that the enemy is prone to using anti-AI spoofing and manipulation tactics (= *poor evaluation in **[Adversarial Threat]***). Thus, in this example, if it becomes clear that **[System Security]** and **[Adversarial Threat]** both evaluate poorly, this may be grounds for instantly rejecting the system unless one (or preferably both) of these risks can be addressed.

(5) **Trade-offs**	Certain variables are inversely correlated with each other. In such a case, **reducing the risk profile of one variable may increase risk in another**. This presents interesting dilemmas in terms of which variable to prioritise, as optimising for one may involve introducing additional risks elsewhere. The toolkit makes these inverse-relationship pairings more explicit, thereby allowing evaluators to weigh and decide upon which opportunity risk to take.  For example, it is generally considered as legally desirable to improve a system's [Understandability], as this allows more informed decisions to be made when employing systems. Often, however, this may worsen [False Positive Rate], which is undesirable as it increases the probability of civilian objects/persons being misclassified as targets.
(6) **Complementary variables**	Some variables are complementary, meaning that **mitigating risk in both works synergistically and add an additional layer of safety and redundancy**. This framework highlights those relationships, thereby enabling decision-makers to further mitigate overall risk by looking beyond the variable under evaluation.  For example, an evaluator may make a good-faith assessment that a system exhibits low risk of [Bias] based on extensive test and validation. However, it is still recommended to invest into mechanisms to improve [Contextual Awareness] and [Feedback Loop] to cater for the possibility of certain biases being overlooked and only occurring after deployment.

Variables rarely operate independently. To help evaluators understand how different variables interact, the toolkit identifies a number of relationships relevant to each variable. These relationships are presented in the order in which they are most likely to be consulted during an assessment:

- Variables that might be necessary to evaluate the current variable (*Relationship 1*) (see Section 4.4).
- Variables that may influence the current variable under evaluation (*Relationship 2*).
- Particularly dangerous or disqualifying combinations of variables (*Relationship 4*).
- Undesirable combinations that require urgent need action (*Relationship 3*).
- Potential trade-offs with other variables (*Relationship 5*).
- Complementary variables that can help reduce risk (*Relationship 6*).
- Variables that may themselves be affected by the current variable, and should therefore be considered next (*Relationship 2*).

To assist users, the Toolkit shows a relationship matrix that provides a quick reference to all identified relationship variables (see Fig.8). The symbols in the matrix are arranged to reflect the recommended sequence in which relationships should be considered.

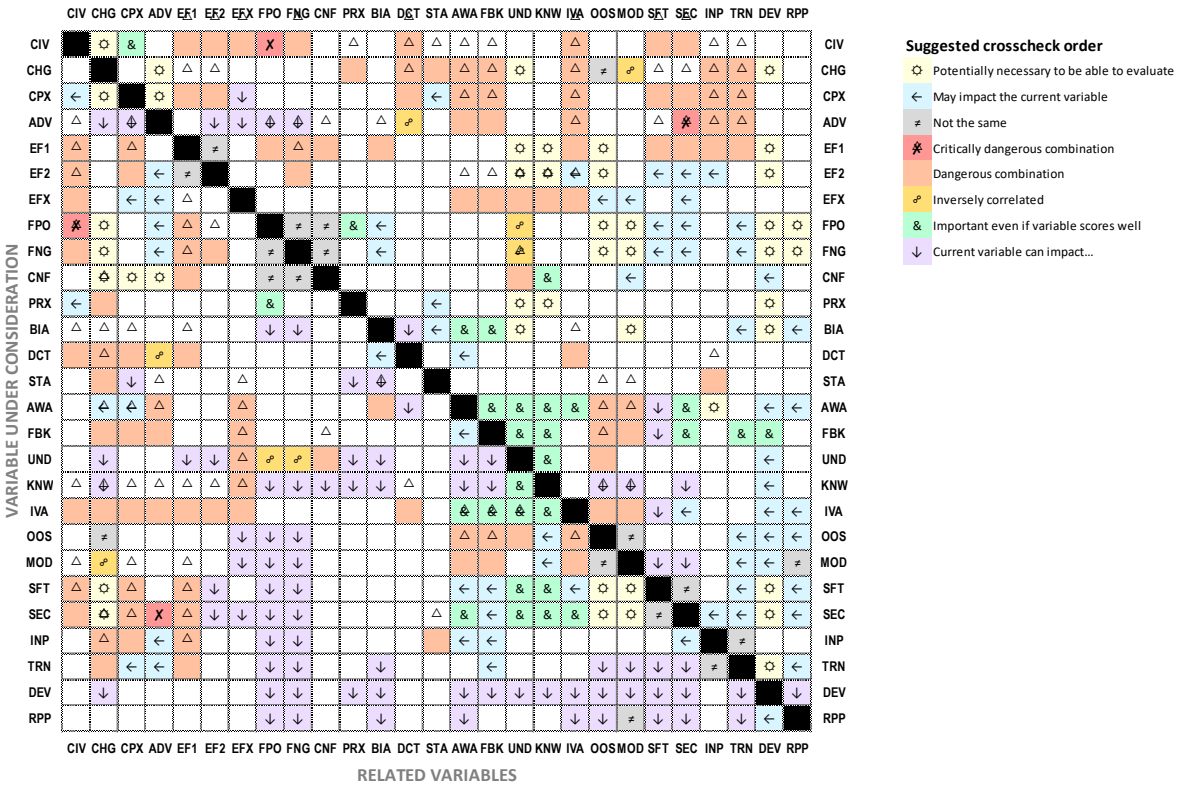


Fig.8: *Crosscheck matrix*

4.6. Risk Mitigation

After assessing a variable and considering its relationships with other variables (Section 4.3 and 4.5) an evaluator may conclude that the overall level of **risk is too high** for the AI system to be used under the envisaged conditions. Where this is the case, three broad choices are available.

First, the evaluator may attempt to **reduce the level of risk associated with the variable itself**. For example, if **[Understandability]** is assessed as poor, the developer could be asked to integrate explainable (XAI) AI tools into the system's interface to improve understandability. Similarly, if **[Intended & Likely Effects]** raises concerns, the explosive munitions carried by the system could be replaced with a lower-yield alternative. Where appropriate, the framework provides suggestions for measures that may help mitigate identified risks.

It is **not always possible to reduce the level of risk associated with a variable directly**. This may be the case where:

- The variable cannot realistically be controlled. For example, many environmental variables—such the rate at which the operational environment changes, its level of complexity, or the adversary's tactics (see **[Changing Environment]**, **[Environment Complexity]** and **[Adversarial Threat]**)—cannot be influenced by the evaluator; or
- The variable could be mitigated in principle, but *not at the stage at which the evaluation is taking place*. For example, an operational commander may identify that the AI system lacks the ability to detect civilians around targets (see **[Civilian Detection]**), a functionality that could support compliance with the obligation to take precautions in attack. However, introducing this functionality would require changes at the design or development stage: an option likely not available to the operational commander.

Where the risk-score of a variable on the evaluation scale cannot be reduced directly, the second option is to **mitigate this by strengthening a *related* variable** that can compensate for the identified deficiencies. The framework identifies such opportunities through its cross-reference guidance and relationship matrix (see Section 4.5). Continuing the previous example, where a commander is operating an AI system without **[Civilian Detection]** functionality available, they may instead seek to strengthen the ability of operators to intervene whenever civilians are unexpectedly identified near a target. A strong score in **[Intervention Ability]** may compensate for the poor score in **[Civilian Detection]**, potentially making the system legally acceptable for use in that operational context.

Finally, the evaluator may conclude **the AI system should not be used** in the envisaged circumstances and instead select an alternative system, which may or may not involve AI. This course of action is likely the appropriate one where a critically dangerous combination of variables (✗) is observed, or where multiple variables present unacceptable levels of risk that cannot be adequately mitigated through changes to the system, its mode of operation, or other complementary safeguards.

4.7. Relationship Chart

To improve usability, each variable is accompanied by an interactive relationship chart that brings together the key information discussed in the preceding sections. For each variable, the chart identifies related variables (Section 4.5) and highlights opportunities for risk mitigation (Section 4.6), allowing evaluators to understand at a glance how the variable fits within the wider framework. The recommended order of considering these relationships (corresponding to the optimal crosscheck order identified in Section 4.5) is indicated through numbered icons. An example is shown below in Fig.9.

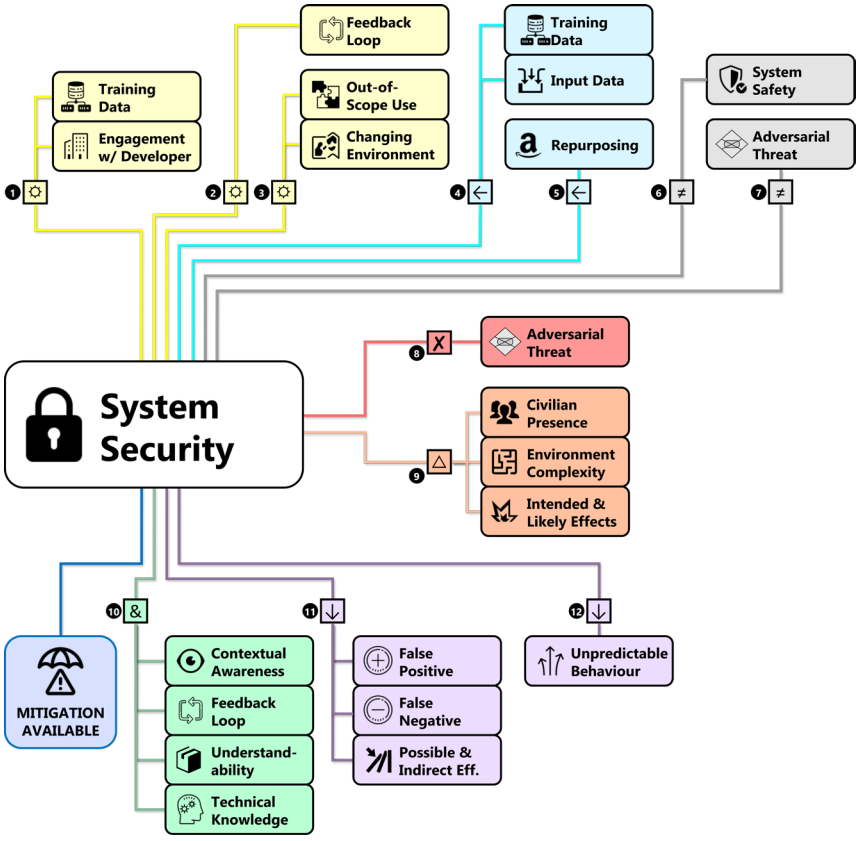


Fig.9: *Relationship Chart (from the variable **[System Security]**)*

Finally, a Chart Legend explains the meaning of each of the relationships identified in the chart, clarifying how the variables interact or affect each other (see Fig.10).

RISKY PAIRINGS

8 ✗ It is **likely prohibited to have simultaneous poor evaluations in [System Security] and [Adversarial Threat]**. Poor security in combination with an adversary who is adaptive and prone to employing counter-AI tactics is likely to **render the AI system unreliable to the point of becoming indiscriminate**.

9 △ A **poor evaluation in [System Security] is dangerous if** many civilian objects/persons are likely to be affected by the system (= *poor evaluation in [Civilian Presence]*), the environment is complex, multi-agent or cluttered (= *poor evaluation in [Environment Complexity]*), or if errors due to exploitation by adversaries can potentially lead to harmful direct and secondary effects (= *poor evaluation in [Intended & Likely Effects]*).

Fig.10: *Example clarifications of the Relationship Chart (from the variable [System Security])*

5. Recommendations and Future Work

This report has presented a framework and accompanying toolkit for assessing the legal acceptability of military AI systems under IHL. By identifying the technical, operational, environmental, and organisational factors relevant to legal compliance, together with the relationships between them and the ways in which identified risks may be mitigated, the framework provides a structured methodology for supporting legal assessment throughout the AI system lifecycle. It is intended to support structured legal reasoning, not to replace legal judgment or the responsibility and accountability of decision-makers.

The framework is designed to support a wide range of stakeholders involved in the development, acquisition, review, and deployment of military AI systems. By providing a common vocabulary and a shared understanding of the factors relevant to legal acceptability, it can facilitate communication between legal advisers, acquisition authorities, commanders, operators, developers, and industry partners. Used in this way, it can contribute to more consistent legal assessments, better-informed procurement and development decisions, and earlier identification and management of legal risks.

We recommend that Defence organisations and industry partners adopt and further refine the framework as a shared reference point throughout the acquisition, development, legal review, testing, training, and deployment of military AI systems. In particular, it should be used to support continuous, context-sensitive evaluation throughout the system lifecycle rather than static, one-off reviews. Integrating legal considerations earlier in the design and acquisition process can improve communication between Defence and industry, reduce uncertainty during development, and help ensure that legal requirements are translated into clear and operationally meaningful technical expectations.

Although this report focuses on compliance with IHL, the methodology has broader potential. Future work should validate and refine the variables and relationships presented in the toolkit, expand the evidence base underpinning them, and improve the guidance available to users. The framework could also be extended to incorporate other legal regimes, ethical principles, and societal considerations, providing a more comprehensive methodology for assessing the broader acceptability of military AI systems.

Ultimately, the contribution of this project lies not only in the variables it identifies, but in demonstrating that the legal assessment of military AI systems can be supported by a structured, transparent, and context-sensitive methodology. By providing a common framework for identifying relevant factors, documenting legal reasoning, and considering appropriate risk mitigation measures, the framework offers a practical foundation for supporting the responsible development and use of military AI systems in accordance with international law.

The Authors



Dr Jonathan Kwik is a Researcher in International Law at the T.M.C. Asser Institute, specialised in the law of armed conflict and disruptive technologies. He is also a member of the Drafting Team for the *AI in Warfare Manual*, a member of the Board of Experts of the *Asia-Pacific Journal of International Humanitarian Law*, and an Advisory Board member for the *STEADFAST Project*.
Webpage: https://jonathankwik.com/



Dr Marta Bo is a Senior Researcher in International Law at the T.M.C. Asser Institute, specialising in the legal implications of artificial intelligence and disruptive military technologies. She is also an Associate Senior Researcher in SIPRI's *Governance of AI* team and Co-President of the *Antonio Cassese Initiative*.
Webpage: https://www.asser.nl/who-we-are/dr-marta-bo-llm/

Acknowledgements

The authors extend special thanks to:

- **Ms Cheya Took**, for extensive support in the form of both substantive review and the copyediting of the report and toolkit.
- **Dr Clara Maathuis** and **Dr Jurriaan van Diggelen**, for their peer-review of, and substantive input to, the report and toolkit.
- **Ms Célestine de Zeeuw, Ms Colette Wibaut,** and other **colleagues at TNO** who provided valuable substantive suggestions that helped shape many elements of the toolkit.
- **All participants of the expert workshop** "*Compliance Risk Framework for Assessing the Legal, Ethical and Societal Acceptability of Military AI Systems*" of 29-30 January 2026 for their insights, which substantially shaped the final design of this toolkit.
- **All respondents of the survey** "*Building Bridges: Insights for Cross-Sector Dialogue on Military Artificial Intelligence*" of June 2025, which formed the initial impetus of this project and the ensuing toolkit.
- **Ms Eylül Kabakcı**, for her support with the layouting of the report and toolkit.
- **Mr Felix Hariyanto Sugianto**, for his support in the design of the Relationship Charts.

About the ELSA Lab Defence project

This project forms part of the Asser Institute's deliverable within the wider ELSA Lab Defence project, an interdisciplinary consortium bringing together academia, government, industry, and civil society to advance the responsible development and use of military AI. ELSA Lab Defence is premised on the view that responsible military AI requires the continuous integration of legal, ethical, and societal considerations throughout the system lifecycle. To answer these questions, ELSA Lab Defence is developing an independent, consultative ecosystem that supports the responsible development, assessment, and use of AI in the defence domain. Website: https://elsalabdefence.nl/