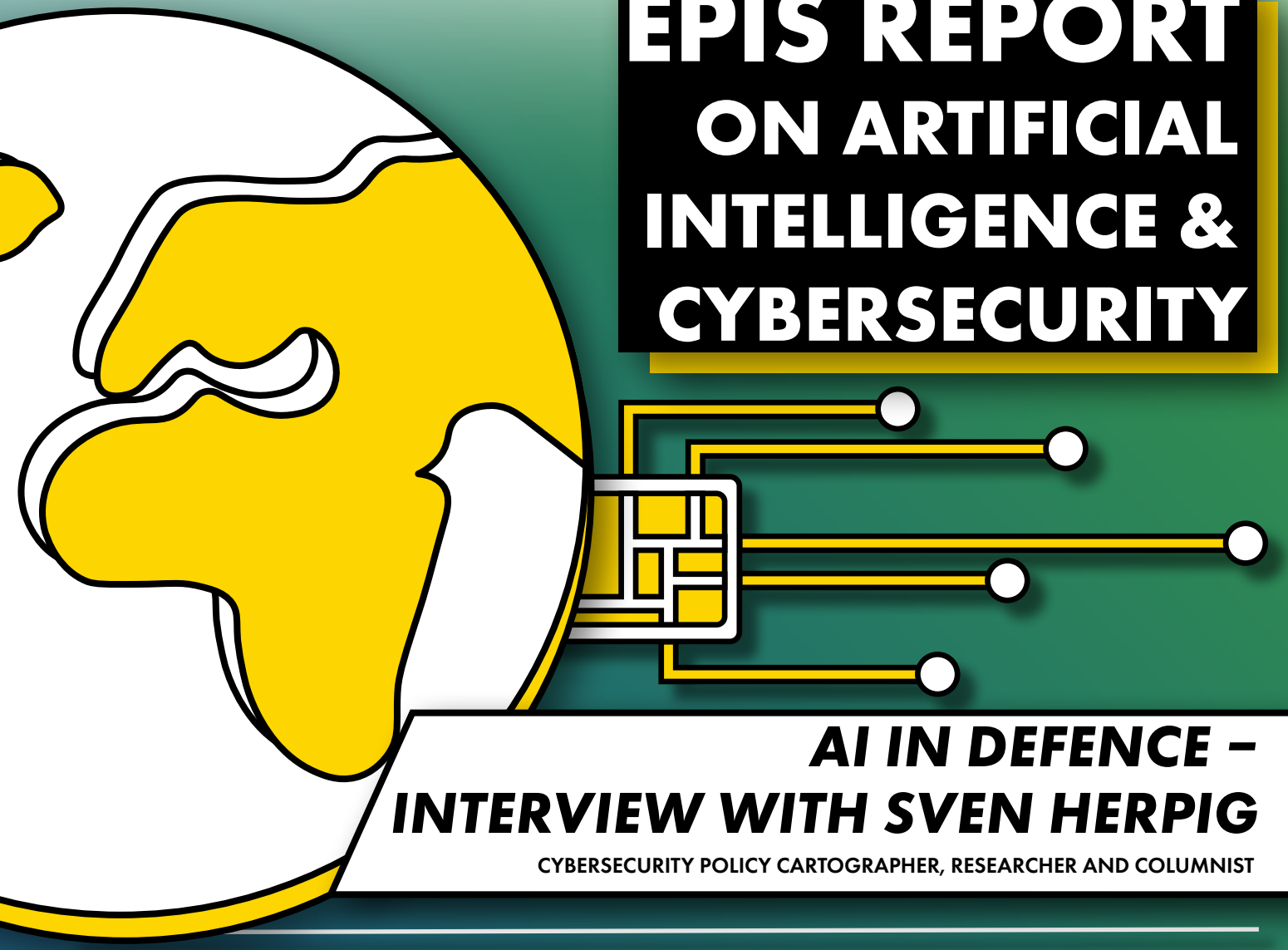
EPIS Thinktank
foreign affairs & security policy

Issue I
October 2025
epis-thinktank.com

EPIS REPORT ON ARTIFICIAL INTELLIGENCE & CYBERSECURITY



AI IN DEFENCE – INTERVIEW WITH SVEN HERPIG

CYBERSECURITY POLICY CARTOGRAPHER, RESEARCHER AND COLUMNIST

AI and the Future of Warfare

Main question: How does AI impact modern warfare and IHL compliance? Argument: AI weapons and cyber tools boost military power but risk civilian harm and legal accountability. Conclusion: Human oversight and regulation are crucial to ensure AI respects IHL principles.

AWS Regulation: Accountability & IHL

Main question: Can existing laws adequately govern Autonomous Weapons Systems (AWS)? Argument: A regulatory framework, not a ban, ensures accountability and civilian protection. Conclusion: Current international law can regulate AWS responsibly while allowing strategic use.

Technological Aspect of an Emerging Market

Main question: How is AI changing modern warfare? Argument: AI boosts targeting, intelligence, and autonomy in combat. Conclusion: Militaries must pair AI with oversight and regulation.

Table of Contents

Editorial Team

Editor in Chief / Group Leader

Belen Bringas is pursuing a Master of Science in Global Studies with a major in Political Science at Lund University. She holds a B.A. in International Relations from Malmö University and has experience in policy research, communications, and marketing. Her work focuses on AI governance, gender equality, and peace and security in the EU and Latin America. She plans to pursue a PhD in the future.

Managing Editor / Resort Leader

Valentin Grangier is a French student of International Relations at Leiden University, specialising in East Asia and the Indo-Pacific. Focused on geopolitics, security, and regional dynamics, he analyses power interplay, ASEAN's role, and shifting alliances. Passionate about Japan, with intermediate Japanese, he blends constructivist, realist, liberal, and democratic perspectives to explore the region's political, economic, and security challenges.

Branch Leader

Alek Spieczny is pursuing a B.Sc. in International Organisations and Relations at Leiden University. As the Branch Leader for EPIS Think Tank, he oversees all organisational activities, ensures coordination among all departments, and provides guidance to the Board of Directors. Based in The Hague, he is actively committed to embracing international perspectives and helping foster the growth of EPIS Think Tank.

Layout / EPIS Board – Media Design

Cira Scherenberger is pursuing a B.A. in Information Design at Stuttgart Media University (DE). The EPIS Media Design is responsible for the design and formatting of EPIS' publications, including the EPIS Magazine and EPIS Reports. She ensures visual consistency and readability across all content. Moreover, she works closely with the authors to create professional layouts that enhance the impact of EPIS Thinktank's written work.

& Contributors

Research Fellow
Annabel Iyengar

Research Fellow
Giulia Convertini

Research Fellow
Lennard Raak

Research Fellow
Sofia Zanin

Research Fellow
Liam von der Wiede

Research Fellow
João Pedro Souza Gohla

External Author
Mr. Agamemnon Sotirios Logothetis

External Author
Elias Ricken

External Author
Dr. Sven Herpig

External Author
Brice Lefebvre

Editorial

Editor in Chief / Group Leader

Belen Bringas is pursuing a Master of Science in Global Studies with a major in Political Science at Lund University. She holds a B.A. in International Relations from Malmö University and has experience in policy research, communications, and marketing. Her work focuses on AI governance, gender equality, and peace and security in the EU and Latin America. She plans to pursue a PhD in the future.

AI and Cybersecurity at the Frontline of Global Security

Dear reader,

New technologies create social change; we have seen this countless times throughout history. From the agricultural revolution to what scholars call the second machine age. We are living through a time of rapid and profound change where artificial intelligence (AI) and cybersecurity have moved beyond science fiction plots to change the way we look at international security. These technological advancements are actively shaping the defense sphere in ways we didn't think were possible before.

This first edition of our report covers a range of topics, from peacekeeping, policy recommendations, to asking legal and ethical questions about the use of these technologies in global security. It is crucial to understand their full implications so we are better prepared for the future.

In collaboration with the European Defence Network (EDN), we present a diverse team of students and young professionals from Europe and beyond who have been instrumental in the creation of this report. Through determination and collaboration, we hope to provide thought-provoking articles that showcase both the opportunities and challenges when it comes to implementing AI and emerging technologies in defense and security.

The overarching theme of the report is the following: the urgent need to balance technological innovation with transparency, accountability, and ethical considerations. The responsible use of these technologies is particularly important; while they can provide strategic advantages, we must also consider the negatives and the necessity for regulation and human control.

This inaugural report is the start of an ongoing conversation. It is our sincere hope that it raises awareness and contributes to responsible policymaking in AI and cybersecurity.

Belen Alondra Bringas Machicado
Editor of the EPIS Artificial Intelligence & Cybersecurity Report

About the Article

Main question: How does AI impact modern warfare and IHL compliance? Argument: AI weapons and cyber tools boost military power but risk civilian harm and legal accountability. Conclusion: Human oversight and regulation are crucial to ensure AI respects IHL principles

About the Author

Giulia Convertini is pursuing a M.A. in International Relations with a focus on International Politics and Regional Dynamics at the Università degli Studi di Milano (IT). Her research focuses on EU affairs and digital/tech policies, as well as regional dynamics in Asia and the Middle East, with a particular focus on the US's role in global affairs.

1. Introduction – The Role of AI and Cyber Technologies in Warfare

The integration of artificial intelligence (AI) and cyber technologies into modern warfare represents one of the most consequential shifts in military strategy since the advent of nuclear weapons. AI is now actively shaping battlefield decisions, powering autonomous systems and enabling new forms of digital warfare that transcend physical borders. At the same time, cyberspace has emerged as a contested domain in its own right—where states and non-state actors conduct operations ranging from espionage to infrastructure sabotage, often below the threshold of conventional war and sliding more towards the realm of hybrid warfare. This convergence of AI and cyber capabilities has already begun to transform the character of armed conflict. Russia's war in Ukraine has illustrated the real-time use of AI-enhanced targeting, autonomous drone swarms and coordinated cyberattacks on critical infrastructure, such as energy grids and communications networks. These developments raise urgent questions about the adequacy of existing legal frameworks and the international community's ability to prevent destabilising consequences. The current applications of military AI are very diverse. They range from support in the targeting cycle and the conduct of hostilities in general, use for hostile activities in cyberspace and for intelligence, surveillance, and reconnaissance purposes. AI is also deployed within the context of so-called 'information warfare', launching cyberattacks on communication systems or civilian infrastructures. Another concrete example of 'information warfare' is the use of deep fake videos in the Russia-Ukraine conflict, with the aim of influencing or spreading disinformation to the general public. Research and development into AI for defence is now broad, well-funded and moving fast. Work ranges from blue-sky algorithms to near-term operational integration — with emphasis on trustworthiness, human-machine teaming, and robustness against adversarial attacks. Major defence research agencies and alliances are accelerating programmes while policymakers race to put governance and legal guardrails. An example of that is NATO's revised Artificial Intelligence Strategy, released in July 2024, which stresses both opportunities and risks, reaffirming principles of responsible use such as lawfulness, accountability, explainability, reliability, human oversight, and bias mitigation. Importantly, the strategy acknowledges challenges such as adversarial use and misuse of AI, disinformation, and unintended consequences, positioning NATO to balance innovation with safeguards while shaping global norms for responsible defence applications. The pace of technological advancement far exceeds that of legal reform. International humanitarian law (IHL), while rooted in principles of humanity and military necessity, was not designed to account for self-learning algorithms or invisible cyber operations. Key challenges such as attributing attacks, assigning accountability for autonomous decisions and regulating dual-use technologies, underscore the need for legal reviews. This article explores the legal implications of AI-driven and cyber-enabled conflict. It examines how current International Humanitarian Law engages with these technologies, where significant gaps remain and what emerging risks may demand urgent attention.

LAWS:
AI weapons that autonomously select and engage targets

2. Legal Frameworks: Current landscape and gaps

Autonomous weapons are set to revolutionise warfare. They have the potential to scale up armed conflict to such a fast pace that humans might lose control over it. According to the United Nations Secretary General, Antonio Guterres (2018): "Our challenge is to maximize the benefits of the technological revolution while mitigating and preventing the dangers. The impact of new technologies on warfare is a direct threat to our common responsibility to guarantee peace and security." The development of AI weapon systems would lead to a global arms race, which,

without oversight, could increase risks to global stability. Given the unpredictable behaviour of machine-learning AI systems, which are controlled by algorithms that dictate weapon engagement systems, humans may lose their ability to intervene promptly in case of faulty behaviour, as Jurgen and Altman (2017) argued. Warfare has already integrated AI, mostly to assist physical military hardware with specific functions and tasks such as flight, surveillance and navigation. Autonomous systems are increasingly being deployed both in wars and in law enforcement situations like police operations and cyberspace. Cyberattacks do not directly cause killings but can significantly harm critical infrastructures like electricity grids and hospitals. Autonomous systems can also be hacked. Stuart Russell (2019) warned about how autonomous weapons threaten human security on the national, international, local and personal levels. International Humanitarian Law (IHL), which consists of the four Geneva Conventions of 1949, their additional protocols and customary law, was established to explicitly recognise the need for striking a balance between military necessity and humanity in a situation of armed conflict. This means that there are certain limits as to which actions can be taken, even during wars. A fundamental principle of International Humanitarian Law is that States are constrained in their selection of weapons and methods of warfare by established norms of international law. Specifically, Article 36 of Additional Protocol I to the Geneva Conventions (AP I) requires states parties to assess, during the development, acquisition, or adoption of any new weapon or method of warfare, whether its use would be prohibited under international law. This obligation becomes even more significant and challenging when applied to emerging technologies whose effects on civilians and civilian infrastructure remain uncertain. Article 36 calls on states to evaluate new weapons and methods of warfare through the lens of IHL and all other relevant international legal obligations applicable to them. With the growing recognition of the concurrent application of IHL and International Human Rights Law (IHRL) in armed conflicts, such legal reviews, should ideally assess compliance with both legal frameworks. While many IHL rules apply only in times of armed conflict, Article 36 reviews often occur during peacetime. For states that are party to AP I, this constitutes a procedural obligation. However, it can be argued that even non-party states—if bound by substantive legal limits on the use of certain weapons or methods—should undertake similar pre-emptive legal reviews to ensure they do not violate those substantive rules. For example, regarding the co-application of IHL and international human rights law, the Human Rights Committee's General Comment 36 interprets the obligation to protect the right to life under Article 6 of the International Covenant on Civil and Political Rights (ICCPR) as including preventive measures, such as legal reviews of new weapons. Cyberspace has emerged as a critical domain in modern military operations, with cyberattacks increasingly forming a regular component of armed conflict. The development of new cyber capabilities and tools, whether they represent novel means of warfare or introduce new methods, undoubtedly requires legal scrutiny under Article 36. When cyber operations directly support conventional attacks – for instance, by disabling air-defence systems to enable airstrikes – they function as means of warfare that complement kinetic operations and as such, must be subject to an Article 36 legal review.

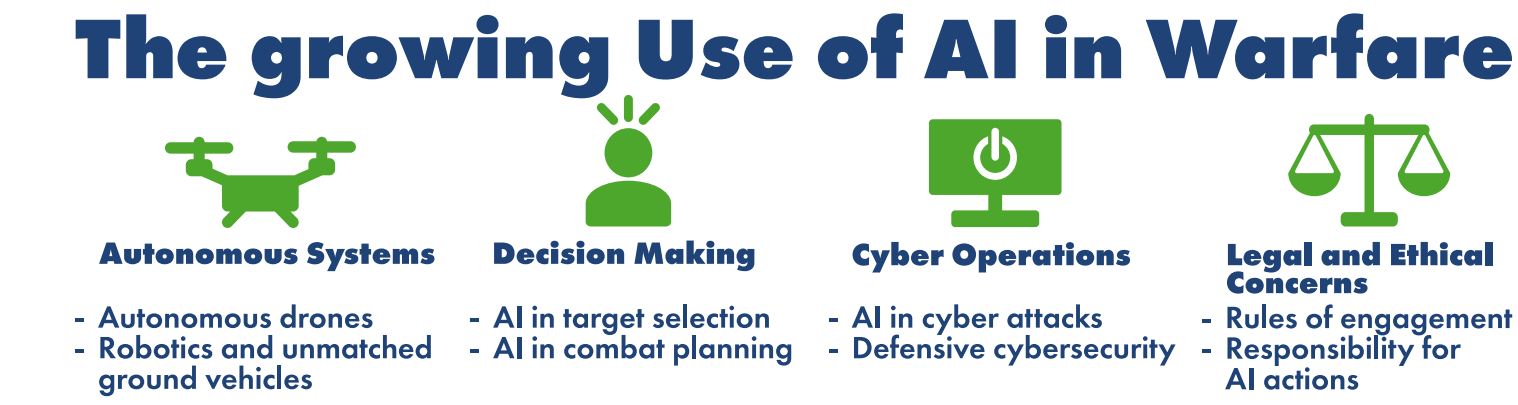


Figure 1: Summary of the growing Use of AI in Warfare

3. The Intersection of Lethal Autonomous Weapons (LAWS), AI and machine learning in conflict within the International Humanitarian Law framework

The most prominent and imaginative use case of AI for military purposes involves the deployment of AI in unmanned physical robotic systems, including lethal autonomous weapon systems (LAWS). The International Committee of the Red Cross (ICRC) defines LAWS as weapons that select targets and attack them without human intervention. According to the ICRC, this means that a human merely activates an autonomous weapon, but at that point does not know specifically who or what it will target, nor where or when it will do so. LAWS will make this decision autonomously based on the observations from sensors and software in the deployment environment, which link this input to a specific 'target profile'. Not all LAWS have machine learning (ML) features, as some of these weapons are rule-based and operate within human-designed scenarios, making their functionalities limited to humans' commands. ML allows LAWS to have a much higher level of autonomy in decision-making, in ways that range from the ability to move through enemy territory to identifying, selecting, locating and attacking particular targets. Within the framework of International Humanitarian Law, the integration of AI in warfare would bring in the opportunity to enhance the respect of IHL, as machine learning processes complex information in a faster way and can take informed decisions while taking into account IHL principles. Ideally, AI-driven LAWS have the potential to take a clear picture of complex scenarios and play an effective part in conflicts. In practice, as the ICRC argues in its publication on Artificial intelligence and machine learning in armed conflict: A human-centred approach (2019), autonomous weapons have no human perception of emotions like fear or anger and preserving this more emotional side of armed conflicts also would allow for the preservation of humanity in this realm. If we look at the case of IHL's proportionality principle, which calls for a balance between the potential civilian harm and military necessity, it indeed requires human, subjective participation in defining military advantage and the level of harm caused to civilians. The core of International Humanitarian Law is to protect people who are not involved in conflicts, so humans cannot feed AI-driven weapons with a way to precisely and objectively evaluate the level of civilian casualties compared to military gains. This puts AI-driven LAWS against IHL, unless human involvement remains a key part of armed conflicts. The application of AI systems in warfare also raises concerns regarding their unpredictability and the issue of explainability. Machine learning can get to a point where humans can't trace how and why an AI-driven system has made a specific decision and acted in a specific way. Not knowing in advance how autonomous weapons might act also raises questions regarding their ability to respect the IHL principle of distinction, which requires parties to the conflict to distinguish at all times between combatants and civilians. Another cause for concern is the fact that AI systems are subject to biases, putting once again at risk the IHL principle of protecting people and places that should not be targeted in armed conflict.

Human oversight: Essential to ensure AI weapons follow IHL

4. Conclusion

Prioritizing the human aspect of military operations calls for a reimagining of the human role within an evolving human–machine cognitive system. Militaries should be equipped to lead diverse, integrated teams across the military infrastructure, encompassing military personnel, government actors and civilian contributors. To do so effectively, they will need a sufficient understanding of their AI-driven tools, enabling meaningful collaboration as well as critical oversight. AI is already reshaping the character of warfare and disrupting long-established human practices. By fostering a human-centric approach to AI

and cyberwarfare, militaries would more effectively prepare for the inevitable transformations ahead without making the world more unsafe. Netta Goussac summarises here the need for legal reviews regarding the integration of AI in warfare: "Today's technological advances in how conflicts are fought mean that robust legal reviews are as critical now as they were when Article 36 was conceived, during the Cold War arms race. While Article 36 does not specify the process by which legality should be determined, in the view of the ICRC, the obligation clearly implies a mandatory standing procedure that assesses all weapons and their normal or expected method of use, against a State's international obligations, including IHL. According to the ICRC's Guide to legal reviews, this entails a multi-disciplinary examination of the technical description and actual performance of a weapon, at the earliest possible stages of its research, development or acquisition. Legal reviews can be a potent safeguard against the development and use of AI weapons that are incapable of being used in compliance with IHL rules regulating the conduct of hostilities, notably the rules of distinction, proportionality and precautions in attack. These rules are addressed to those who plan, decide upon and carry out attacks in armed conflict. It is humans that apply this law and are obliged to respect it. An AI weapon system that is beyond human control would be unlawful by its very nature—a conclusion that would become evident during a legal review." (Goussac, 2019) It is clear that there is an urgent need for global action not only to safely embark on the ongoing AI-driven technological revolution of warfare, but also to make sure that is effectively regulated. This research highlighted the current gaps between the IHL framework and the application of AI in warfare, shedding a spotlight on the need to harness AI responsibility in military contexts, especially with regards to the principles of proportionality and distinction, which are at the core of International Humanitarian Law. This paper ultimately calls for concerted international cooperation to adapt international law to the new challenges and opportunities arising from the development of autonomous weapons and cyber warfare.

Figure 2: Summary of Modern Warfare in the Digital Age: AI, Cyber Capabilities & Legal Gaps

References

Goussac, N. (2019, April 18). Safety net or tangled web: Legal reviews of AI in weapons and war-fighting. ICRC Humanitarian Law and Policy Blog. https://blogs.icrc.org/law-and-policy/2019/04/18/safety-net-tangled-web-legal-reviews-ai-weapons-war-fighting

International Committee of the Red Cross. (2019). Artificial intelligence and machine learning in armed conflict: A human-centred approach. https://www.icrc.org/en/document/artificial-intelligence-and-machine-learning-armed-conflict-human-centred-approach

Altmann, J., & Sauer, F. (2017). Autonomous weapon systems and strategic stability. Survival, 59(5), 117–142. https://doi.org/10.1080/00396338.2017.1375263

EURODEV. (2025, May 19). The future of defense: How AI is transforming the industry. https://www.eurodev.com/blog/defense-industry-ai-transformation

Leuven Centre for Public Law, International humanitarian law and new technologies. KU Leuven. https://www.law.kuleuven.be/ai-summer-school/blogpost/Blogposts/international-humanitarian-law

Lieber Institute. (2023, May 3). IDF introduces AI to the battlefield: A new frontier. West Point. https://lieber.westpoint.edu/idf-introduces-ai-battlefield-new-frontier/

Royal United Services Institute. (2023). Trust in AI: Rethinking future command. https://www.rusi.org/explore-our-research/publications/occasional-papers/trust-ai-rethinking-future-command

United Nations Human Rights Committee. (2018). General comment No. 36: Article 6 of the International Covenant on Civil and Political Rights, on the right to life (CCPR/C/GC/36). https://www.ohchr.org/en/documents/general-comments-and-recommendations/general-comment-no-36-article-6-right-life

UN Secretary-General Discusses Maximizing Security Benefits of Technology, Mitigating Dangers, (3 April 2018)

Russell, S. (2019). Human compatible: Artificial intelligence and the problem of control. Penguin Books.

Summary of NATO's revised Artificial Intelligence (AI) strategy, 10 July 2024. https://www.nato.int/cps/en/natohq/official_texts_227237.htm?utm

About the Article

Main question: Can existing laws adequately govern Autonomous Weapons Systems (AWS)? Argument: A regulatory framework, not a ban, ensures accountability and civilian protection. Conclusion: Current international law can regulate AWS responsibly while allowing strategic use.

About the Author

Sofia Zanin recently graduated a Master's in International Security at the Università degli Studi Internazionali di Roma, and is currently an intern Security Analyst at RileyRisk focusing on Europe and Eurasia. She holds a bachelor's in Legal Studies from the University of Torino.

1. Introduction

Technological innovation is transforming the conduct of war, sparking debate over whether existing accountability frameworks can adequately govern its use. Among the new projects, Autonomous Weapons System (AWS) have been at the centre of an increasingly polarized debate. Humanitarian organizations emphasize the ethical and legal risks, advocating for strict regulation or even a comprehensive ban. In contrast, states highlight the strategic benefits of these technologies and the practical reality that technological innovation is unlikely to be halted. The coexistence of these positions creates a persistent tension: humanitarian concerns retain undeniable validity, yet the continues development and partial deployment of AWS suggest that a complete ban at this stage may be premature. In this context, further research remains essential, both to inform potential regulatory frameworks and to endure that humanitarian considerations are integrated into future decision-making. However, in the absence of a universally agreed definition of what these systems entail, states and stakeholders have been developing multiple definitions arguing on whether these new weapons comply with international standards or are inherently dangerous and should therefore be banned altogether. The United States Department of Defense Directive No 3000.09 on Autonomy in Weapon Systems defined AWS as:

"[a] weapon system that, once activated, can select and engage targets without further intervention by a human operator. This includes human-supervised autonomous weapon systems that are designed to allow human operators to override operation of the weapon system, but can select and engage targets without further human input after activation"
(McDougall, 2019)

Autonomous Weapons System (AWS):
Weapons that can select and engage targets without human input

The European Union, while not having a universally agreed definition, adopts the International Committee of the Red Cross definition as:

„Any weapon system with autonomy in its critical functions. That is, a weapon system that can select (i.e. search for or detect, identify, track, select) and attack (i.e. use force against, neutralize, damage or destroy) targets without human intervention". (Cottier, 2023)

From these two definitions it is evident that Meaningful Human Control (MHC) is a crucial aspect in the determination of whether Autonomous Weapon Systems (AWS), as an umbrella term including both Lethal Autonomous Weapons System (LAWS) and Full Autonomous Weapons (FAW), conform with international law and can, therefore, be further advanced or are inherently too dangerous and their development should be banned. The lack of a universally accepted definition renders the debate on the legality of AWS particularly complex as multiple terms like FAW and LAWS are often used interchangeably. While these two terms do share similarities they have a slightly different scope. Lethal Autonomous Weapons System (LAWS) is a term that emphasizes the use of lethal force without meaningful human control, and it has been popular especially among promoters of a total ban. On the other hand, Fully Autonomous Weapons (FAW) is a term encompassing the most extreme examples of autonomous weapons requiring no human input whatsoever (Lewis, 2015). This paper focuses on autonomous weapons system as a generic term. This article will present the polarized debate between supporters of a pre-emptive ban, who argue that AWS are incompatible with the requirements of international humanitarian law (in particular, the principles of accountability,

distinction and proportionality), and those who contend that such a ban is premature. The latter emphasize that these weapons are not unlawful per se and may, in certain contexts, offer operational advantages, such as enhanced precision and reduced risk to civilians and combatants, when developed and deployed responsibly. To this end, Section 2 outlines the international law framework on accountability; Section 3 explores the arguments supporting a ban, focusing on the accountability gap; Section 4 examines the counterarguments that question the existence of such a gap and oppose a prohibition; Section 5 considers regulatory precedents, including the landmines protocol, as possible blueprints; and Section 6 concludes with reflections on the viability of a regulatory framework for AWS.

Categories	Target Selection & Engagement	Human Input	Lethality	Example
Automated Weapons	No Target selection (just automatic reaction)	High (humans set triggers)	Can be lethal	Landmines
Autonomous Weapon Systems (AWS)	Yes, system selects & engages once activated	Human initiates, but no further control	Mixed	Armed drones with AI target recognition
Fully Autonomous Weapons (FAW)	Yes, without human intervention	None (no meaningful human control)	Mixed	Hypothetical hunter drones
Lethal Fully Autonomous Weapons (LAWs)	Yes, system selects & engages	None	Explicitly lethal	Hypothetical armed robots making kill decisions

Figure 1: Visual comparison of autonomy in weapon systems

2. Accountability in Armed Conflict: the international law framework

The main challenge around autonomous weapons regards what has been called the accountability gap. Due to their nature, AWS cannot be held accountable for criminal conduct as they are not responsible moral agents. Moreover, the increasing separation between the weapon system actions and any proximate human makes it extremely complex to identify the proper agent to be held accountable. In the context of armed conflict and autonomous weapons systems, the debate on accountability revolves around individual criminal responsibility as defined in Artt. 25-28 of the International Criminal Court (ICC) Rome Statute. Art. 30 of the Rome Statute additionally poses a high threshold for the mental element, highlighting that the material elements of each crime must be committed with intent and knowledge (International Criminal Court, 2011). Differently, the Additional Protocol I to the 1949 Geneva Convention provides under art. 85(3) that a behaviour that constitutes a war crime must be

conducted wilfully. Legally speaking, "wilfully" is often interpreted including also the concept of recklessness (Bo et al., 2022). The lack of proximity between a human and the AWS actions makes it particularly hard to establish this requisite mental element on the part of any human (McDougall, 2019). The centrality of the mental element in prosecuting criminal responsibility increases the complexity of ensuring accountability for crimes involving AWS. Due to the difficulty in linking a human directly to wrongful actions committed by an autonomous weapon, proposals arose for a new requirement: Meaningful Human Control. While there is no fully universal definition, this new standard was introduced as to somewhat ensure human oversight and, if necessary, human responsibility.

3. One side of the debate: Accountability Gap and The Case for a Ban

The debate over autonomous weapons gained momentum around 2012-2013 when various organizations and scholars started debating over ethical and legal implications of these systems. Since 2013 multiple countries called for a preventive ban hoping to negotiate a treaty within the Convention on Certain Conventional Weapons (CCW) framework (Sauer, 2016). The petition against autonomous system was presented by the Campaign to Stop Killer Robot, a coalition of civil society and stakeholders led by Human Rights Watch (HRW). The latter published an insightful paper in 2012 leading the anti-AWS debate. Mirroring most of the concerns shared by the Campaign, HRW highlighted how human decision-making in armed conflict entails complex assessments to ensure a discriminate and proportionate application of force in compliance with international humanitarian law. According to the group, such elaborate assessments are unlikely to be replicated in software code, thus raising doubts on whether autonomous system can be designed to operate respecting international standards. From a legal perspective, the accountability gap has been reiterated consistently as a crucial issue as machines cannot be court-martialled leaving no redress to injured or killed civilians. Lastly, from an ethical point of view, HRW argues that allowing autonomous weapons to use lethal force, thus relinquishing life-and-death decision-making to an algorithm that cannot be held accountable for its actions, violates basic human dignity (Human Rights Watch, 2012). In its 2012 consideration, HRW focuses on civilian protection. According to the group, regardless of possible future technological advancement, fully autonomous weapons inherently lack the human qualities necessary to meet the rules of international humanitarian law. Furthermore, by eliminating human involvement fully autonomous weapons undermine other important principles. Following the reasoning provided, fully autonomous weapons are not restrained by human emotions and compassion, by reducing military casualties they lower the threshold for political leaders to engage in armed conflict, and lastly, there is the question about accountability as a form deterrence and remedy for victims. These concerns undermine civilian protection leading HRW to recommend a full prohibition of development, production and use of fully autonomous weapons through an international legally binding instrument (Human Rights Watch, 2012). In light of these considerations, in 2012, as only precursors to Fully Autonomous Weapons existed, advocates called for a preventive arms control for autonomous weapons systems on an international humanitarian law basis. In 2015, the HRW group published another report calling once again for a ban but basing its reasoning on the issue of accountability. HRW argued that no existing body of law „provided adequate accountability of individuals directly or indirectly involved in the use of fully autonomous weapons" (Human Rights Watch, 2015). As the debate evolved, in 2025 HRW published another review on autonomous weapons and implications for human dignity and human rights. In this most recent paper, the group reiterated how the inherent characteristic of AWS of selecting and engaging targets based on sensor processing rather than human inputs infringes multiple fundamental obligations. However, in the

A regulatory framework can ensure AWS use complies with international law and accountability

recommendation section, while a prohibition treaty is still mentioned, the main request seemed to have pivoted towards stronger regulation (Human Rights Watch, 2025).

4. The other side of the debate: Questioning the Accountability Gap

On the other side of the debate on Autonomous Weapons System compliance with international humanitarian law, scholars have maintained that not only it is factually possible to place human responsibility when AWS are unlawfully or negligently deployed, questioning the existence of an accountability issue altogether, but they also openly oppose an outright ban on two different grounds. On one hand a pre-emptive ban risks stopping the development of a weapons that can protect combatants and reduce the risk for civilians, directly complying with two foundational objectives of international humanitarian law; on the other hand, enough consensus among different geopolitical actors is unlikely to be achieved at this stage of AWS development, which makes the case for regulation significantly stronger. Responding directly to the HRW case of 2012, Professor Michael Schmitt presents strong counterarguments to the supra mentioned narrative calling for a ban (Schmitt, 2015). Firstly, Schmitt argues that the foundational assumption that Autonomous Weapons System (understood as "human out of the loop" systems) necessarily violate international humanitarian law because they are inherently indiscriminate and cause unnecessary suffering to combatants and civilians is extremely flawed. Instead, he argues that an assessment on compliance with distinction and proportionality principles must be based on a case-by-case analysis, considering the type of weapon, the environment in which it is deployed and the scope of its deployment. In other words, HRW claims blur the line between IHL prohibition on weapons per se and unlawful use of otherwise useful weapons (Dunlap, 2016). Secondly, he addresses the accountability gap debate. Schmitt clearly demonstrates that across multiple scenarios it is already possible, under international humanitarian law and criminal law, to identify and held humans responsible. For instance, were the AWS to be deployed in an unlawful manner, the commander authorizing its use would be responsible under the relevant law. Similarly, were the AWS to be designed for conduct not compliant with IHL, the developers could be held accountable. Furthermore, should the armed forces of a state use AWS in an unlawful manner, state responsibility could be effectively invoked (Schmitt, 2015). Building on Professor Schmitt's response, Charles Dunlap, Jr went further and questions the HRW assimilation of personal accountability with the legality of the weapon itself (Dunlap, 2016). Citing Art. 36 of Protocol I of the Geneva Conventions, Dunlap asserts that the legality of a weapons should be measured on how the weapon is used and not depending on whether responsibility can be determined. Expanding on previous possible attributions suggested by Schmitt, Dunlap explains further their reasoning. AWS, like any weapon, must be developed and tested so that their intended acts against life and property can be reasonably anticipated in order for designers, commanders, operators, and others associated with autonomous weapons to escape culpability (Dunlap, 2016). The argument furthered by supporters of this side of the debate is that in the remote circumstances that a machine goes rogue, no one can be punished provided that reasonable steps have been taken to avoid such an unexpected result. It follows, then, that if at any point in the development, design, and deployment chain reasonable assessments have not been made to guarantee respect for the IHL principles of distinction and proportionality, those who failed to adequately foresee unlawful conduct may be held personally responsible under international criminal law. Considering the above reflections and being aware that AWS development is to be closely monitored because there are challenges to IHL rules, especially concerning weapons using machine-learning systems, Schmitt and Dunlap conclude that an outright ban based on the inherent nature of AWS as an IHL violating system is unfounded. Additionally, due to the strategic advantage these weapons may offer to states' national security arsenals, they both recommend working on sensible regulation which are likelier to be favourably welcomed by multiple countries.

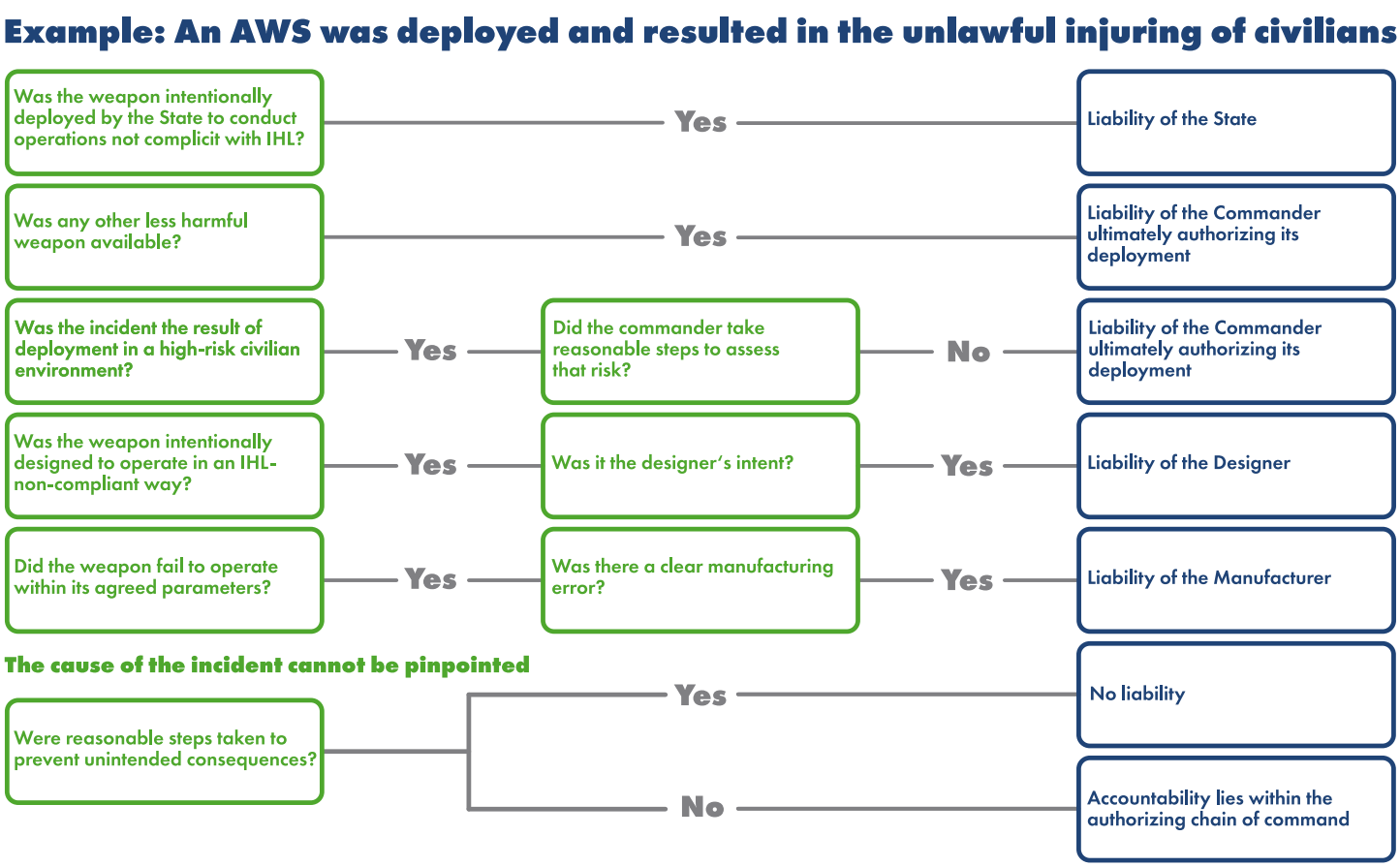


Figure 2: Example of possible accountability attributions in the event of an AWS deployment and subsequent unlawful conduct of the machine.

5. Landmines regulatory scheme as a possible blueprint

Between 2014 and 2015, John Lewis presented a convincing case on using the Amended Protocol regulating landmines as a blueprint for an AWS regulation. The proposal highlights how once activated both weapon systems have the capacity to target and kill without further human input and both weapons are deployed with specific defined parameters thus raising similar questions about distinction. Drawing from the 1996 Amended Protocol on landmines, a model framework was developed for FAWs. The Protocol focuses on a clear definition of lawful environments in which to deploy landmines to allow military advantage while guaranteeing civilian safety and a focus on commander decision-making given the nature of the weapon. Similarly, a framework for FAWs would entail a clear definition of:

- characteristics that AWS must have to ensure distinction and proportionality
- characteristics of the environment distinguishing from remote battlefields to areas in which civilian concentration is higher.
- characteristics of the opposing force especially concerning the enemy evasion techniques which might affect AWSs ability to comply with the distinction principle
- the level of residual human control

Moreover, regarding the issue of accountability, the proposed framework poses the focus of responsibility on the commander's decision to deploy the weapon (Lewis, 2015). Creating a regulatory framework from existing protocols reinforces the argument that existing international law is sufficient to limit the use of AWS in compliance with the law and therefore, a pre-emptive ban is unjustified.

6. Conclusion

Ultimately, while concerns about the level of autonomy in AWS remain valid, existing international laws and principles provide a sufficient framework to regulate the development and deployment of these emerging strategic systems. As technology evolves, it cannot be excluded that a ban or a more structured treaty might eventually become the most appropriate option to ensure compliance with international humanitarian law and international criminal law. However, given the current state of technological development and the legal avenues for attributing accountability, the call for a pre-emptive ban should give way to a regulatory framework that restricts the use of these weapons to contexts where they can achieve military advantage while ensuring civilian protection.

References

Bo, M., Bruun, L., & Boulanin, V. (2022). Retaining human responsibility in the development and use of autonomous weapon systems: On accountability for violations of international humanitarian law involving AWS (Policy Report). Stockholm International Peace Research Institute. https://www.sipri.org/publications/2022/policy-reports/retaining-human-responsibility-development-and-use-autonomous-weapon-systems-accountability

Cottier, D. (2023, January 9). Emergence of lethal autonomous weapons systems (LAWS) and their necessary apprehension through European human rights law (Report No. 15683). Parliamentary Assembly of the Council of Europe. https://pace.coe.int/en/files/31433

Dunlap, C. J., Jr. (2016). Accountability and autonomous weapons: Much ado about nothing? Temple International & Comparative Law Journal, 30(1), 63–76. https://scholarship.law.duke.edu/faculty_scholarship/3592

Human Rights Watch. (2012). Losing humanity: The case against killer robots. Human Rights Watch. https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots

Human Rights Watch. (2015). Mind the gap: The lack of accountability for killer robots. Human Rights Watch. https://www.hrw.org/report/2015/04/09/mind-gap/lack-accountability-killer-robots

Human Rights Watch. (2025, April 28). A hazard to human rights: Autonomous Weapons Systems and Digital Decision-Making. Human Rights Watch. https://www.hrw.org/report/2025/04/28/hazard-human-rights/autonomous-weapons-systems-and-digital-decision-making

International Criminal Court. (2011). Rome Statute of the International Criminal Court. https://www.icc-cpi.int/sites/default/files/RS-Eng.pdf

Lewis, J. (2015). The case for regulating fully autonomous weapons. Yale Law Journal, 124(4), 1309–1325. https://www.yalelawjournal.org/comment/the-case-for-regulating-fully-autonomous-weapons

McDougall, C. (2019). Autonomous weapon systems and accountability: Putting the cart before the horse. Melbourne Journal of International Law, 20(1), 58–87. https://classic.austlii.edu.au/au/journals/MelbJIL/2019/4.html

Sauer, F. (2016). Stopping 'killer robots': Why now is the time to ban autonomous weapons systems. Arms Control Today, 46(8), 8–13. https://www.armscontrol.org/act/2016-09/features/stopping-killer-robots-why-now-time-ban-autonomous-weapons-systems

Schmitt, M. (2015, August 10). Regulating autonomous weapons might be smarter than banning them. Just Security. https://www.justsecurity.org/25333/regulating-autonomous-weapons-smarter-banning/

EUROPEUM

INSTITUT PRO EVROPSKOU POLITIKU
INSTITUTE FOR EUROPEAN POLICY

WHO ARE WE?

EUROPEUM is a Prague and Brussels-based think-tank dedicated to **advancing European integration** and shaping Czech and EU policymaking.

OUR PROGRAMMES

- **Just Europe** *"Integration must be socially just and lead to the convergence of living standards"*
- **Green Europe** *"Our goal is an ambitious climate policy that considers both the planet and its citizens"*
- **Global Europe** *"EU's strong position*
- *in its neighborhoods and partnerships with global actors are key to maintaining position in a changing world"*

WHAT DO WE DO?

Research
Our research and outputs include over **100** policy papers, analyses, reports and other publications yearly

Projects
We partake in projects focused on topics ranging from green and just transformation, digitalisation, migration or EU enlargement up to security or media freedom

Café Evropa

JUSTGREEN

Events and education
We yearly bring important topics into over **80** public debates, workshops, routables and international conferences.

Think Visegrad
Representing Think Visegrad Platform in Brussels

Establishing **network** of partners to maximize the influence of independent research based advocacy

EUROPEUM Brussels Office

EUROPEUM was the first think tank from Central Europe to expand into the heart of the European Union. Our motivation was to follow the debates on the EU agenda closely and to contribute to strengthening the voice of the Czech Republic and other Central and Eastern European countries.

Scan the QR code for more info!

Co-funded by the European Union

About the Articles

Main question: How is AI changing modern warfare? Argument: AI boosts targeting, intelligence, and autonomy in combat. Conclusion: Militaries must pair AI with oversight and regulation

About the Author

Lennard Raak is currently pursuing his BA International Relations & International Relations at the University of Groningen (NL). His research focuses on technological development of information-technology, the increasing influence of Big Tech and the development of new digital landscapes. He is currently a board member of the political youth organisation 'Jong Sociaal Contract' (JSC) in the Netherlands.

About the Author

Mr. Agamemnon Sotirios Logothetis is a Strategic Consultant with over seven years' experience in the Defence Industry and the founder of Lamda Analytics SL, a strategic consulting firm based in Barcelona, Spain that helps clients build pipelines & win contracts in Defence Contracting in the US & EU. He possesses an MSc in International Transportation Management & Supply Chain Management from SUNY Maritime, and a BA in International Studies from Towson University.

Perspective on Writing
From a technological perspective, the increasing use of AI in a military context aligns with the growing importance of data processing. At their heart, defence tech companies that leverage AI are, in essence, data processing companies, and the heart of their product is the data that drives their algorithms.

1. Introduction

Warfare has undergone a profound transformation over the past decade, seeing an influx of technology-driven approaches coming into use. Both state and non-state actors now leverage advanced technologies, particularly artificial intelligence (AI), to gain strategic advantages on and off the battlefield. This shift has far-reaching implications for how wars are fought, how technology is developed, and how both military personnel and civilians are affected. At the same time, the traditional military-industrial complex is evolving. A new wave of players, including tech companies and startups, is entering the defense space, challenging legacy systems and introducing innovative solutions. In this article, we examine the rise of AI, its practical applications in a military context, the emerging products and technological developments because of this, and the companies driving this transformation. Ultimately, we seek to explore a central question: does the proliferation of AI in the military context, and the actors driving it, resemble the traditional defense prime model, or is it instead fundamentally rooted in data, making these emerging defense tech companies essentially data companies at their core, merely weaponizing that data?

2. AI Context Within Warfare

Artificial Intelligence (AI) has seen increased use in recent years in warfare, with a more active role in combat operations being assumed as the years progress. Some of the first known cases of AI use for military application came in the 2010s, with Israel's Iron Dome reportedly incorporating some advanced algorithms of early AI to assist the system in rocket trajectory prediction, providing recommendations for intercept decisions based upon which projectile had a higher probability of impacting populated areas (NPR, 2021). This is followed in 2017 with the launch of Project Maven by the US military, that seeks to train AI algorithms to analyze surveillance video and pictures and automatically identify what is in the frame on the battlefield (Bloomberg, 2024). From 2020 and onwards, AI started to see uses in active warfare. This was seen during the 2020 war between Armenia and Azerbaijan over the region of Nagorno-Karabakh. Azerbaijan, armed with Bayraktar TB-2 drones and Israeli loitering munitions, leveraged AI targeting software in these platforms to assist in target identification kill chain decision. The use of this technology, particularly the drone weapon platforms, were pivotal in the Azerbaijani success in the conflict. The first acknowledged use of an autonomous AI battlefield kill is attributed to have taken place in 2020 during the Libyan Civil War, with a Turkish-made STM Kargu 2 loitering munition drone attributed as having hit retreating forces of the General Haftar's Libyan National Army (LNA) while being flown on automated targeting modality, meaning that this would be the first acknowledged case of AI identifying a target, and autonomously making the "kill" decision (NPR, 2021). In the years from 2021 to present, the use of AI in warfare has proliferated, with it used widely in the Ukraine conflict for targeting assistance by loitering munitions and drones by both sides. The IDF, in the conflicts against Hamas in Gaza has heavily utilized AI for intelligence purposes, to generate target lists and help with identification of them on the battlefield. "Operation Spiderweb", conducted in June of 2025 by Ukrainian forces to strike deep within Russia at the Strategic Bomber bases, was one of the best examples of AI application in warfare to date. Kamikaze drones that were remotely deployed from trucks were reportedly able to fly along "pre-planned routes" with the assistance of AI, and once they reached their target zone the drones identified and engaged targets. All conducted within a location that was subject to target jamming and limited connectivity (Financial Times, 2025).

Military Uses of AI Milestones

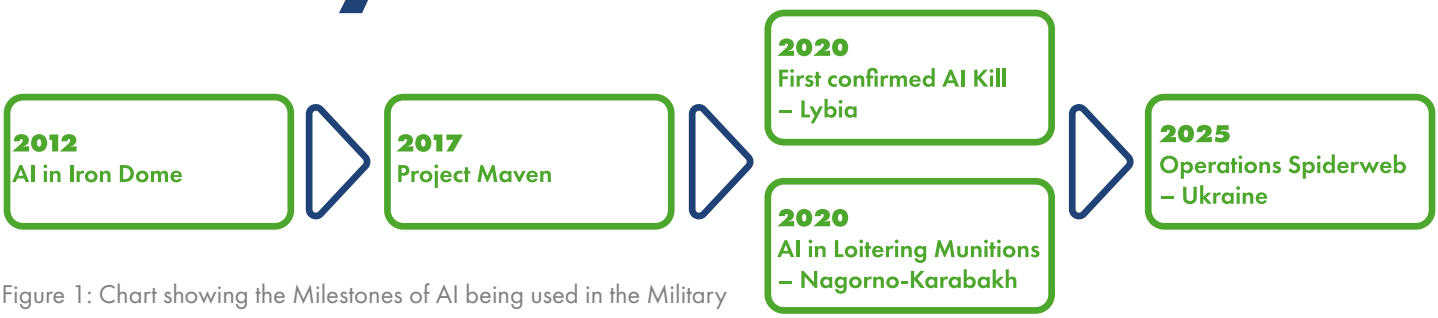


Figure 1: Chart showing the Milestones of AI being used in the Military

The applications of AI within warfare, as seen from the above examples of use cases over the past years, are varied and ever-expanding. Some of the most common applications of AI are for intelligence uses, including Signal Intelligence (SIGINT), Human Intelligence (HUMINT), and Open-Source Intelligence (OSINT). AI has seen heavy use in identification and recognition use cases, assisting the end users identify the targets on the battlefield. This cuts down on the amount of human intervention, making intelligence sorting and analysis quicker and more efficient, leading to a quicker cycle time for the warfighter. Past the battlefield, AI has also advanced in general surveillance, seeing many use cases for facial recognition and tracking as well. Another major application of AI is for target identification and targeting assistance by weapons systems on the battlefield. This has been seen in a multitude of platforms, ranging from UAVs (drones) to more traditional weapons platforms like automatic rifles with optics that may possess target identification assistance. This use of AI is one that has advanced greatly in recent years, and what remains to be seen from this use-case is how much AI progresses from simply assisting in targeting identification and tracking to autonomously making target engagement and kill authority decisions. Another major use of AI that is still largely in an experimentation phase but will inevitably see expanded use is that of autonomous piloting. There was a recent successful test undertaken by Saab and Helsing, where the AI suite of Helsing was utilized to autonomously pilot a Saab Gripen fighter aircraft, with it successfully executing some maneuvers in response to combat-inspired inputs (Saab, 2025). Sixth-generation fighter aircraft that are currently in development such as the Future Combat Air System (FCAS) and the Global Combat Air Program (GCAP) are experimenting with the concept of "loyal wingmen" drone aircraft, that will augment the manned aircraft as part of a broader flying network.

Figure 2: Next Generation Fighter (NGF) Concept, with Manned Aircraft in Control of Unmanned Wingmen

3. Emerging Products / Technological Developments

A multitude of products and technological developments have emerged thanks to AI and autonomous systems. One of the arguably biggest benefactors of AI systems have been Unmanned Aerial Vehicles (UAVs), more commonly known as drones. Although inherently a different system and ecosystem entirely, autonomous systems have begun to be combined with AI software that scale their capabilities. These include enhancements to core functions that are performed by drones such as surveillance mission sets, with AI algorithms being leveraged to assist with automatic target identification. Increasingly, there has been increased testing with, and in some cases even the use of, AI in drones to take their automated identification and targeting software and allow it to make kill-chain decisions as well. This was seen in our example stated above during the Libyan Civil War in 2020. UAVs have not been the only unmanned systems to benefit from the incorporation of AI into their capabilities, however. Unmanned Ground Vehicles (UGVs), Unmanned Surface Vehicles (USVs), and Unmanned Undersea Vehicles (UUVs) have all benefited from the incorporation of AI into their systems. Companies such as the German Defense Tech startup ARX Robotics have been incorporating AI into ground vehicles to create autonomous solutions that can help to make battlefield operations more efficient. Meanwhile, others such as the Delian Alliance from Greece are seeking to create Area Denial solutions using kamikaze USVs combined with kamikaze UAVs, that would lie dormant in pre-positioned locations, being activated when a threat is detected and then autonomously engaging. Although many of these are still in various stages ranging from research to development, to already deployed, it shows how quickly the proliferation of AI has progressed in military applications, and the amount of use cases that it has made its way into in a relatively short amount of time. Another emerging product that has begun to appear increasingly is AI-enabled loitering munitions. These are munitions (be they missiles or drones) that are designed to loiter over a target location, and then able to autonomously identify their targets and engage them. These types of munitions are highly effective in combating ground targets, to include engaging enemy air defense units, as was shown in the engagement in the Nagorno-Karabakh conflict in 2020 between Armenia and Azerbaijan. Frontrunners in the development of AI-enabled loitering munitions include Anduril in the US, who manufacture the Barracuda line of AI-enabled, cruise missile-style loitering munitions (Trevithick, 2024). These munitions are claimed to have the capability to identify sources of increased radar activity, allowing them to coordinate autonomously between themselves to eliminate enemy threats more effectively. More recently in May of 2025, Anduril also unveiled Fury, their unmanned fighter jet that will leverage AI to perform autonomous flight and mission sets (Anduril, 2025). Helsing, a German Defense Tech startup, also manufactures the HX-2, an AI-enabled drone that is designed to be manufactured at much more cost-effective rates than conventional munitions. The design of the weapon enables it to loiter in contested environments, and to autonomously identify and engage enemy targets utilizing the AI algorithms it is programmed with to enable identification and engagement. A technological development that has also come about as part of the development of sixth-generation fighter aircraft has been the testing that is being done on "drone wingmen." The concept behind these is that the manned fighter jet would operate as a sort of central node, and can be supported by a number of autonomous aerial vehicles (AAVs) that would perform as "wingmen." The pilot will be able to give commands to the AAVs on how to support (e.g., stand off support, ground attack, etc.), but they will be programmed to conduct flight and operate in their mission sets autonomously with the support of AI programming.

Autonomous drones:
Unmanned aerial vehicles that operate and attack without human control

4. Companies in Industry Developing AI & New Defense Tech

There have been several companies who have been pioneering AI development in recent years. The main among them have been in the commercial sector, focused around every day, civilian use-cases, with the largest among them being OpenAI & Anthropic, having obtained between them over $81bn in venture funding. (Forbes, 2025) Furthermore, these companies have also been dabbling in the defense space, with OpenAI, Anthropic, Google, and xAI being chosen in July 2025 by the US Department of Defense to help them increase their adoption and use of AI in intelligence analysis, logistics, and data gathering functions (Albon, 2025). There have been additional entrants into the field of AI from China as well, with the launch of DeepSeek in 2024 making a large impact, as the model appeared to mimic the capabilities of western language models at a fraction of the cost. Elon Musk's xAI & recent European startups like Mistral AI have also entered the space, leading to a large confluence of AI capabilities. (Forbes, 2025) The growth of these mainstream AI companies has, in recent years however, overshadowed the growth of a multitude of defense technology companies who have been pioneering AI use specifically in the defense space as well. Some of the most well-known of these companies, that have already been mentioned in this paper include Anduril, Palantir, and Helsing. These companies have also managed to harness the power of AI whilst adapting it to a military use, creating potent products that have significantly changed the nature of modern warfare. There have been different uses of AI by each company, in the case of Palantir harnessing big data obtained through various sources and through monitoring to create predictive insights that can be utilized by government and intelligence agencies. The technology has been touted by the company as having stopped multiple terrorist attacks, and as having helped soldiers in Iraq & Afghanistan avoid ambushes. One of the first investors into Palantir was In-Q-Tel, the CIA's investment arm, underlining its use in the intelligence world. (Rumage & Rodriguez, 2025) Anduril, founded by the founder of Oculus VR, has created an AI platform that combines the inputs from various sensors, radars, etc., and conducts a rapid analysis to provide identification and threat analysis, feeding the data back to the operator, enabling more rapid decisions. The company has branched out, creating cost-effective cruise missiles and drones that also utilize the AI system it has created and can operate in contested environments. The integration of AI to these platforms allows the systems to interpret data from sensors and make decisions regarding engagement. The company has rapidly grown, landing massive contracts from the US DOD, and recently branching into unmanned undersea vehicles (UUVs) as well (Tashjj, 2025). Helsing, like Anduril, is a German startup founded in 2021 that has the Altra Platform, that takes data from multiple drones and sensors, analyzes it, and provides recommendations for battlefield enhancements to operators. The company has recently started building drones, integrating their AI platform into them, allowing the drones to operate in contested environments where there may not be any signal to operations, maintaining the ability to locate, identify, and engage targets (Contrary Research, 2025). The company has also recently These companies, all innovative and pioneers just like their commercial counterparts, have fundamentally altered the nature of military operations and warfare with their application of AI to a military context. The use of AI has enabled governments and militaries to significantly increase the efficiency of data analysis in applications such as intelligence and combat situations. This increase in data analysis facilitates decisions to be made quicker and more efficiently, while also enabling enhanced combat operations and capabilities.

Human oversight: Essential to ensure AI weapons follow legal and ethical rules in combat

5. Conclusion

As we have seen, AI in the context of warfare and military applications has increased at a very rapid rate, particularly within the last years. The technology, having largely started out being applied to conduct large-scale data analysis and recognize trends, assisting human operators in making decisions, has been incorporated into more platforms and systems as time has progressed. AI is now incorporated into everything from surveillance cameras to missiles, unmanned systems, and even is beginning to be incorporated into aircraft. The companies behind the rise in the use of AI have played a large part in this and will continue to play a large part in the further development and incorporation of AI into military uses. As we have seen through the previous study, however, the core of most AI applications lies in data and data analysis. What makes the use of AI in a military context effective is that it enables the rapid analysis and interpretation of the data that is being ingested, and enables decisions to be made off this analysis, whether that means making tactical decisions on the battlefield or a decision on the kill chain. Considering this, to come back to our initial question that was posed at the beginning of this section, are the companies and technologies in this space data companies at their core? The answer would be a clear yes, as the core of AI is data, and the companies and actors who have managed to harness the use of AI most effectively in the defense and military space have been the ones who have the capability to best harness the data and make analysis of it, then weaponizing the output of this analysis. This has led to the automation of data processing, leading to smoother and more streamlined military operations. This has led and will continue to lead to a new emerging group of defense contractors, who at their core are data companies.

References

Government of Luxembourg. (2025, June 17). Mistral AI and Luxembourg enter into a strategic partnership (Press release). The Luxembourg Government. https://gouvernement.lu/en/actualites/agenda.gouvernement2024+en+actualites+toutes_actualites+communiques+2025+06-juin+17-frieden-mistral-ai.html.

NPR. (2021, June 1). A U.N. report suggests Libya saw the first battlefield killing by an autonomous drone. NPR. https://www.npr.org/2021/06/01/1002196245/a-u-n-report-suggests-libya-saw-the-first-battlefield-killing-by-an-autonomous-d?utm_source=chatgpt.com

Saab. (2025). Saab achieves AI milestone with Gripen E (Press release). Saab. https://www.saab.com/newsroom/press-releases/2025/saab-achieves-ai-milestone-with-gripen-e.

King, A. (2024). Digital targeting: artificial intelligence, data, and military intelligence. Journal of Global Security Studies, 9(2), Article ogae009. https://doi.org/10.1093/jogss/ogae009.

Gutman, B. (2021). Palantir's Surveillance Empire: A Story of American Policing, Patriotism, and Profit. [Report / manuscript]. George Washington University / ResearchGate. https://www.researchgate.net/publication/353352542_Palantir%27s_Surveillance_Empire_A_Story_of_American_Policing_Patriotism_and_Profit.

Borat, E. (2025). The Illusion of Neutrality: Algorithmic Vulnerabilities and Corporate-State Collusion in the Age of Surveillance Capitalism. (2025). [Manuscript / working paper — publisher not supplied].

DARPA. (n.d.). AI Next Campaign. Defense Advanced Research Projects Agency. https://www.darpa.mil/research/programs/ai-next-campaign.

DARPA. (n.d.). AI Forward. Defense Advanced Research Projects Agency. https://www.darpa.mil/research/programs/ai-forward.

Bloomberg. (2024, February 29). Inside Project Maven — the US military's AI project (newsletter). Bloomberg. https://www.bloomberg.com/news/newsletters/2024-02-29/inside-project-maven-the-us-military-s-ai-project?sref=W7EBPDtW

Financial Times. (2025). How AI guided Ukraine's drones to hit Russian airfields. Financial Times. https://www.ft.com/content/ccd83e2a-521f-4e35-a5f0-2ec1ef63749e.

Calibre Defence. (n.d.). HX-2: Helsing releases details of AI-enabled loitering munition. Calibre Defence. https://www.calibredefence.co.uk/hx-2-helsing-releases-details-of-ai-enabled-loitering-munition/.

Anduril / Breaking Defense coverage:

Trevithick, J. (2024, September). Anduril unveils new cruise-missile-like weapon, plus voice-controlled drones. Breaking Defense. https://breakingdefense.com/2024/09/anduril-unveils-new-cruise-missile-like-weapon-plus-voice-controlled-drones/.

Anduril Industries. (n.d.). Fury. https://www.anduril.com/fury/.

Forbes. (n.d.). AI 50. Forbes. Retrieved September 16th, 2025, from https://www.forbes.com/lists/ai50/

Albon, C. (2025, July 15). Pentagon taps four commercial tech firms to expand military use of AI. Defense News. https://www.defensenews.com/pentagon/2025/07/15/pentagon taps-four-commercial-tech-firms-to-expand-military-use-of-ai/

Rumage, J., & Rodriguez, A. (2025, August 7). Inside Palantir: The tech giant powering government intelligence. Built In. https://www.builtin.com/articles/what-is-palantir

Tashji, D. (2025, July 8). Anduril's Edge: A New Era of Defense Innovation. PWK International. https://www.pwkinternational.com/2025/07/08/andurils-edge-a-new-era-of-defense-innovation/

Contrary Research. (2025, May 22). Helsing: Business breakdown & founding story. Retrieved from https://research.contrary.com/company/helsing

epis ThinkTank

International Politics Shaped By **You**

EPIS Thinktank

Why Join Us?

- Make Your Voice Heard Through Our Various Formats and Participate in International Politics
- Publish Articles from Early on in Your Academic Career
- Receive Valuable Guidance throughout the whole Writing Process
- Become a Part of Our Network of Likeminded Students and Young Professionals in International Affairs

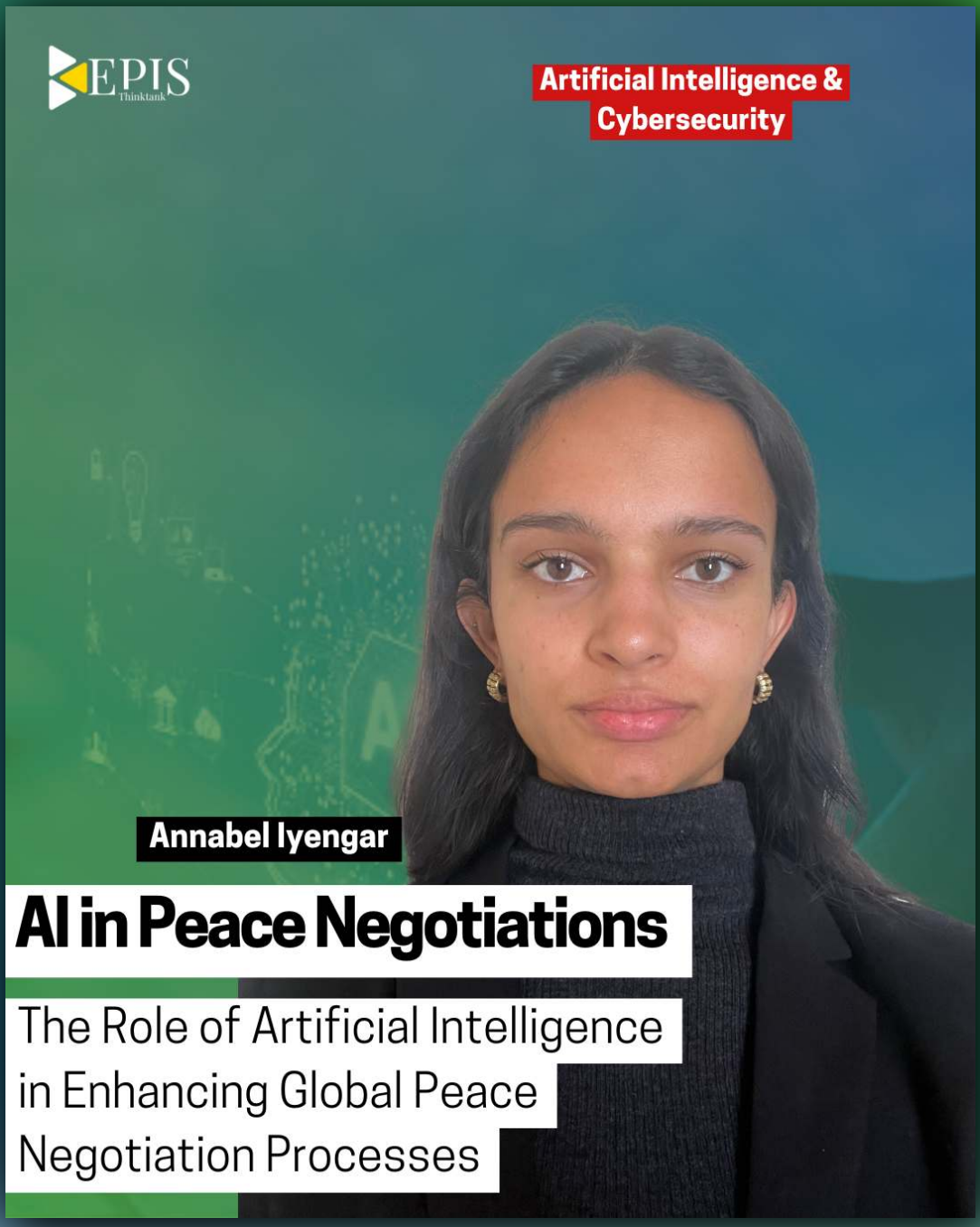
Interested? *Reach Out!*

Contact us on Instagram or LinkedIn or learn more about our work on our website!

@episthinktank

/epis-thinktank

epis-thinktank.de



About the Article

Main question: How can AI support peace negotiations? Argument: AI can enhance conflict analysis, translation, and decision-making, addressing key challenges in co plex peace processes. Conclusion: AI offers significant benefits but must remain a supportive tool under strict regulatory frameworks.

About the Author

Annabel Iyengar is a current MSc Philosophy and Public Policy student at The London School of Economics and a Philosophy graduate from Durham University. Her current research focus is on the value of causal explanations in big data predictive models across healthcare, GDP, and weather forecasting. Within EPIS, she is researching the use of AI models as aids in global peace negotiations. Her goals are to contribute to understanding of the emerging benefits and risks of AI, particularly in cyber warfare.

1. Introduction

Several global entities have begun the process of incorporating AI into their operations. Examples of this include the United Nations Educational, Scientific and Cultural Organisation (UNESCO) recommendations on AI, the Organisation for Economic Co-operation and Development (OECD) AI principles, the G7 statement on the Hiroshima AI process, and the European Union (EU) AI Act (Giovanardi, 2024). As part of this global adoption of AI, it has begun to be established as a potential tool for peacebuilding, standardly divided into three primary domains of opportunity:

- As an assist to conflict analysis
- As an early warning system predicting tensions before they erupt
- To support human communication (Mäki, 2020).

Retrieval-Augmented Generation (RAG):
AI method combining LLMs with external data for context.

Present studies find that around 60% of all wars conclude through some form of compromise (Reynolds & Jensen, 2025). However, peace negotiations traditionally suffer from a host of challenges which obstruct their success in both conflict resolution and establishing policies for long-term stability. Challenges for peace negotiations typically include: dilemmas of resolution enforcement, identity differences, ideology incompatibility of warring parties, vulnerabilities of parties and a general lack of credible guarantees (Walter, 1997). It is in recognition of such challenges to the success of peace negotiations that AI has begun to be introduced as a diplomatic aid, such as the 2020 and 2021 deployment of the Ramesh AI platform used as a dialogue tool in Yemen and Libya by the UN Innovation Cell in DPPA (Alavi, et al., 2022). This article will examine to what extent AI can be of aid to obstacles in peace negotiation processes and identify some issues that may arise from such implementation. From this analysis, recommendations will then be constructed aiming to maximise the benefits of AI for peace-negotiation processes whilst minimising the risks this may incur. Negotiations among warring parties are some of the most critical and sensitive of all bargaining procedures (Wanis-St.John, 2008, p. 1). Even when such negotiations result in agreement, this alone does not guarantee resolution of the underlying conflict and despite on-going efforts towards the structural enhancement of peace negotiations, they continue to evade predictability (Wanis-St.John, 2008, p. 1). Given their importance and notoriously challenging nature, it is worth outlining exactly how 'peace negotiations' should be understood and traditionally what structural elements compromise processes generally constitutive of 'peace negotiations.' Standardly, "an agreement or accord is a formal commitment between hostile parties to end a war" (Anderlini, 2012, p. 1). Furthermore, traditional attempts at achieving this goal through peace negotiations involve three key phases: pre-negotiations, negotiations and post-negotiations implementation (Anderlini, 2012, p. 2). Within said stages, important issues to establish include: logistics, location, security for each party involved, participants, time frame, meditators and their responsibilities, setting achievable goals, building trust and agreement on agenda topics (Anderlini, 2012, p. 2). Despite this generally established structure peace negotiation processes remain non-linear and messy, with talks regularly commencing just to break down and be restarted (Farquhar, 2024). Peace negotiation processes habitually involve several rounds of talks, ceasefires and agreement revisions, for example there were 39 ceasefires in the Bosnian conflict from 1992-1995 (Farquhar, 2024).

Major Historical Peace Negotiations (1815-2016)

Figure 1: Timeline of Major Historical Peace Negotiations (1815-2016)

Furthermore, during the extensive process of these discussions, negotiators can often lack means of effectively gauging the responses and opinions of the superiors for whom they act (Economist, 2025). A further challenge is therefore the maintenance of steady and effective communication channels between the represented and their representors as discussions progress or alter rapidly. At present, pauses often need to be made to regularly inform the represented of developments, which can break momentum and gives other parties time to regroup (Economist, 2025). AI could prove a tool for overcoming this inconvenience. Lastly, translation is one of the fundamental aspects of peace negotiations as "conflict zones are characterised by linguistic diversity" (GSI, 2023). This can be a source of opportunity for strength but also poses an on-going trigger for conflict as misunderstandings could aggravate and even escalate existing violence (GSI, 2023). Therefore, as linguistic diversity often characterises conflict zones for which peace negotiations are necessary, translation comprises a key factor of success and possible further conflict. Some of the key limits and challenges of peace talks, as with those in the above diagram, are thus: informational requirements, cultural and linguistic assimilation and communication within parties. The following section will outline how the emergence of AI as a peace negotiation aid can be of benefit to challenges in standard peace negotiation processes.

2. How AI could be of aid to challenges in traditional peace negotiations

As highlighted above, one of the key challenges and limits in existing peace negotiation processes are the intricate and extensive informational requirements needed to appropriately navigate generating solutions which are realistic, sustainable and mutually beneficial to involved parties. It is this first challenge which has been identified as an area for which AI can be an aid, since AI can hold vast quantities of data and use said data to provide AI-assisted data-analytics (Giovanardi, 2024). Specifically, AI can process data-driven decision-making through data analysis of across social media, vast datasets, diplomatic texts and speeches as well as fact-checking discussion material (Giovanardi, 2024, pp. 41-48). Diplomatic strategy is enhanced by machine-learning algorithms which can analyse geopolitical data, historical treaty negotiations and the content of real-time diplomatic engagements (Pasupuleti, 2025, p. 4). In doing so, optimal negotiation tactics can be identified, and strategies can be established that best align with national interests whilst mitigating the interests of other interested parties (Pasupuleti, 2025,4). Negotiators therefore have a tool for monitoring and staying up to date with the ever-shifting dynamics of geopolitics and can incorporate this information into peace negotiation processes. Such methods were employed in the analysis of negotiations prior to the establishment

of a peace deal in early 2022 for Yemen (Arana-Catania, et al., 2022, p. 4). Yemen's regionalised war involved increasingly fluid and fractured coalitions, national and international actors, divergent goals and continuously shifting party positions (Arana-Catania, et al., 2022, p. 4). To aid future consensus building, 177,789 words from dialogue sessions between Yemen's key stakeholders were systematically analysed and used to generate significant contextual information which could then guide conflict analysis and mediation strategy (Arana-Catania, et al., 2022). This data collection and analysis can not only keep involved parties updated on global political dynamics in the present, but through extensive data analysis AI can also provide predictive scenario modelling through risk forecasting and analysis of potential outcomes concerning a range of actions. In peace processes characterised by uncertainty and instability, scenario planning becomes a powerful tool for navigating uncertainties through providing holistic projections of future outcomes generated by present-day decisions (Hao, et al., 2024, p. 1). In existing research on peace negotiations for Ukraine and Russia retrieval-augmented generation (RAG) techniques of large-language AI models were prompted to generate various versions of peace agreements with distinct parameters as an assist to for analysts in projecting the impact of various deals and the comprehensiveness of each potential option (Reynolds & Jensen, 2025). In another Ukraine-Russia Peace Agreement simulator, outcome preferences could be entered under groups including territory and sovereignty, economic conditions and justice and accountability (Economist, 2025). A draft agreement is then output according to the input parameters along with scores 1-10 for how acceptable such a deal would be to Russia, Ukraine, America and Europe (Economist, 2025). AI is therefore emerging as a tool not only for informing present negotiations but also for solutions formation and analysis. AI is also proving to be of considerable aid as a dialogue assist to the translation aspect of peace discussions. The UN and its partners have begun to use natural language processing and machine learning techniques for dialogue across thousands of individuals in local dialects and as a means of identifying points of agreement in conflict settings such as Libya and Yemen (Brown, 2021). In 2020 and 2021, the UN innovation cell in DPPA deployed the Ramesh AI-platform as a dialogue tool through which up to 1000 participants could anonymously engage in 'large-scale digital-dialogues' (Alavi, et al., 2022). Dialogue can take place in local dialects to enable greater inclusivity in peace negotiation processes as moderators can gauge opinions on actions and outcomes across demographics (Alavi, et al., 2022). AI assist in translation can also reduce translation times, handle greater documentation volume and diminish the traditional costs of manual translation (Farquhar, 2024, p. 1). Additionally, AI's role in translation can provide an intuitively unbiased analysis whilst facilitating communication between parties, as found in efforts by the 'Carnegie Endowment for International Peace" to study how digital technologies such as AI can benefit complex conflict mediation processes.

AI-assisted data analytics:
AI can process vast datasets to guide negotiators in strategy

3. Risks of AI implementation in peace negotiations

However, the benefits that AI could offer existing peace negotiation processes bring with them risks and concerns for exactly how AI aids will be implemented and how this could undermine or complicate peace negotiation efforts. Three main areas of AI-risk to peace and security are identified as: miscalculation, escalation, and proliferation (Giovanardi, 2024). Debate continues over whether the use of AI in negotiations is even a desirable step towards increased rationality or a disputable move away from unique human wisdom (Zia & Waks, 2025). Relying solely on information from past peace deals may restrict creative human input in complex problem-solving, guiding negotiators towards solutions which merely appear successful but fall short operationally or politically. It is further possible that, despite AI generally being heralded as 'objective and impartial,' it presents a biased or

flawed operational picture which could catalyse the deterioration of international relations during peace negotiation processes (Giovanardi, 2024). Akord.ai, a model developed by Conflict Dynamics International, has been trained on 1500 documents about Sudan with a focus on past peace agreements as a tool for greater inclusivity in peacebuilding concerning the ongoing Sudanese civil war (Wilmot, 2024). However, the use of such technology for predicting and generating solutions has been warned against in the acknowledgement that any outputs will inevitably reflect biases both in training data and algorithms (Wilmot, 2024). Even when data is heralded as operationally 'impartial,' the historical content of negotiations and outcomes can be heavily contested as narratives conflict and facts constitutive of 'truth' become filtered (Zia & Waks, 2025). The quality of solutions, and AI outputs generally, is therefore limited by the quantity and quality of training data. AI's role, despite being labelled as objective and impartial, could perpetuate and aggravate existing power imbalances through data biases and subtly partial algorithms. Issues of biases and partiality in AI models stem into wider concerns of accountability, transparency and possible manipulation in models used for operations as consequential and delicate as peace negotiations. AI models are often considered 'black box' models due to the internalisation of data by algorithms in ways currently 'inauditable' by, and inaccessible to, human understanding (Bathaee, 2018, p. 901). A significant aspect of peace negotiation is the ability of negotiators to justify their decisions and positions to both their counterparts and their domestic constituencies, so such inaccessibility obstructs requisite justification. Public decision-making is often required to be both morally and legally transparent, so if AI cannot prove sufficiently transparent and explainable, then it either cannot or should not have a role in decision-making (Maclure, 2021). Under the transparency obligations of Article 13 – 14 of the EU AI Act, systems are obligated to be "developed and used in a way that allows appropriate traceability and explainability" and 'high-risk' AI systems must be interpretable (EU ACT, 2024). Thus, whilst there is regulatory momentum for transparency, explainability and interpretability in models, it is not yet the case that this is standard or even possible in current AI models. Finally, AI models pose increasing cybersecurity threats to models used in peace negotiation discussions and for geopolitical conditions underpinning said negotiations. Using LLMs, hackers can devise "social-engineering assaults" to manipulate human behaviour (Economist, 2025). AI is also being used to make existing malware more aggressive and a greater threat to international security systems, as employed in recent cyber-attacks on Ukraine's security and defence systems in July 2025 (Economist, 2025). Overall, some pressing issues in AI models comprise transparency, interpretability, bias, and attack or manipulation. The use of AI models in peace negotiations therefore requires a united regulatory framework which can mitigate these risks and challenges whilst reaping the above-mentioned benefits.

Figure 2: Visualisation of risk categorisation for AI models passed in the EU AI Act 2021 CITATION Peñ24 \l 2057 (Peñalver, 2024) https://www.nemko.com/blog/a-quick-dive-into-the-eu-ai-act

4. Policy recommendations

At present, options for the incorporation of AI into peace negotiations comprise either full use, partial use as an aid, and no use of AI in favour of solely human effort. Acknowledging that nations increasingly rely on digital infrastructure, and that AI has become a "critical tool" for the maintenance of international peace and stability, the inclusion of AI into peace negotiation processes appears both desirable and inevitable (Pasupuleti, 2025, p. 5). However, it is generally agreed that AI should need remain a 'tool' which can support negotiations whilst not replacing the centrality of human strategic judgement in political decision-making (Reynolds & Jensen, 2025). The following recommendations aim to retain AI's role as a tool for negotiation enhancement whilst implementing regulatory measures to mitigate risk.

- Adhering to the OECD AI Principles of 2024, AI actors should commit to transparency and responsible disclosure of AI systems including fostering an understanding of the capabilities and limitations of AI and providing enough information on relevant AI operations that those adversely affected by its involvement may challenge its output (OECD, 2024).

- Abide by and enforce Article 10 of the EU AI Act 2024 which states that data providers are obligated to evaluate whether their training, validation and testing datasets meet quality criteria including the examination of biases in data and correction measures (van Bekkum, 2025).

- International audits and sanctions: aligning with the 2020 UN Digital Coordination Roadmap, AI models should be auditable by international, third-party regulatory bodies (Rafi, 2025).

- Models used in peace negotiations should be classified as high-risk under the EU AI 2024 act due to the likely impact on human life. Deviation from this categorisation should be subject to individual evaluation and justification on a case-by-case basis. (EU ACT, 2024)

- Development of a set of metrics for measuring international digital inclusion based on the fundamental premise that everyone should have equal access to empowerment through ICT: a measure for acting against the use of AI models in peace negotiations aggravating existing global power imbalances (UN, 2020).

5. Conclusion

The emerging inclusion of AI into peace negotiations provides clear benefits both to negotiators themselves and to wider negotiating parties. However, uncertainties around the exact capabilities of AI, coupled with the risks and challenges of its application demand tight regulatory conditions. This is particularly important given the gravity and volatility of peace negotiations processes, wherein misuse of AI models could instead catalyse the deterioration of international relations. Furthermore, the use of AI in geopolitical settings remains consciously limited to that of a tool for augmenting, rather than replacing, human decision-making and rationale (Goldfarb & Lindsay, 2022, p. 28). To maximise benefits and manage risk, AI models should be subject to a global unified regulatory framework which emphasises transparency, auditability, accountability and inclusion.

References

Alavi, D. M., Wählisch, M., Konya, A. & Irwin, C., 2022. Using Artificial Intelligence for Peacebuilding. Journal of Peacebuilding and Development, 17(2), pp. 239-243.

Anderlini, S. N., 2012. Peace Negotiations and Agreements, London: International Alert: Women Waging Peace.

Arana-Catania, M., van Lier, F.-A. & Procter, R., 2022. Supporting peace negotiations in the Yemen war through machine learning. Data & Policy, 4(e28).

Bathaee, Y., 2018. The Artificial Intelligence Black Box and the Failure of Intent and Causation. Harvard Journal of Law and Technology, 31(2), pp. 890-938.

Brown, D., 2021. The United Nations is turning to artificial intelligence in search for peace in war zones. [Online] Available at: https://www.washingtonpost.com/technology/2021/04/23/ai-un-peacekeeping/ [Accessed 27 July 2025].

Economist, 2025. AI models could help negotiators secure peace deals. [Online] Available at: https://www.economist.com/science-and-technology/2025/04/16/ai-models-could-help-negotiators-secure-peace-deals?utm_medium=cpc.adword.pd&utm_source=google&ppccampaignID=18156330227&ppcadID=&utm_campaign=a.22brand_pmax&utm_content=conversion.direct-respons [Accessed 2 August 2025].

Economist, 2025. How AI-powered hackers are stealing billions. [Online] Available at: https://www.economist.com/business/2025/08/19/how-ai-powered-hackers-are-stealing-billions [Accessed 30 July 2025].

EU ACT, A., 2024. Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence and amending prior regulations, s.l.: The Official Journal of the European Union.

Farquhar, A., 2024. Utilisation of Artificial Intelligence in Accurate Translation of Peace Agreements: A Practical Assessment, Edinburgh: PeaceRep: Peace and Conflict Resolution Evidence Platform.

Giovanardi, M., 2024. AI for peace: mitigating the risks and enhancing the opportunities. Data & Policy, 6(e41).

Goldfarb, A. & Lindsay, J. R., 2022. Prediction and Judgement: Why Artificial Intelligence Increases the Importance of Humans in War. International Security , 46(3), pp. 7-50.

GSI, 2023. Language as a Bridge to Peace. [Online] Available at: https://gsiassociates.com/language-as-a-bridge-to-peace/ [Accessed 10 August 2025].

Hao, H., Wang, Y. & Chen, J., 2024. Empowering Scenario Planning with Artificial Intelligence: A Perspective on Building Smart and Resilient Cities. Engineering, Volume 43, pp. 272-283.

Mäki, N., 2020. Between Peace and Technology- A Case Study on Opportunities and Responsible Design of Artificial Intelligence in Peace Technology. s.l.:s.n.

Maclure, J., 2021. AI, Explainabilty and Public Reason: The Argument from the Limitations of the Human Mind. Minds and Machines, 31(3), pp. 421-438.

OECD, 2024. AI principles. [Online] Available at: https://www.oecd.org/en/topics/sub-issues/ai-principles.html [Accessed 10 August 2025].

Pasupuleti, M. K., 2025. Research Report: „AI in Global Strategy: Harnessing Game Theory and Reinforcement Learning for Diplomatic Innovation". International Journal of Academic and Industrial Research Innovation, 5(3).

Peñalver, M. F., 2024. Unpacking the EU AI Act: A Milestone in AI Regulation. [Online] Available at: https://www.nemko.com/blog/a-quick-dive-into-the-eu-ai-act [Accessed 20 August 2025].

Rafi, M. U., 2025. Ideal Regulations of AI: Safeguarding Peace. [Online] Available at: http://dx.doi.org/10.2139/ssrn.5309675 [Accessed 3 August 2025].

Reynolds, I. & Jensen, B., 2025. Machine Learning Meets War Termination, s.l.: Center for Strategic International Studies .

UN, 2020. Roadmap for Digital Cooperation. [Online] Available at: https://www.un.org/digital-emerging-technologies/content/roadmap-digital-cooperation#:~:text=The%20High%2Dlevel%20Panel%20on,PROTECT [Accessed 15 August 2025].

van Bekkum, M., 2025. Using sensitive data to de-bias AI systems: Article 10(5) of the EU AI act. Computer Law & Security Review , Volume 56.

Walter, B. F., 1997. The Critical Barrier to Civil War Settlement. International Organization, 51(3), pp. 335-364.

Wanis-St.John, 2008. Peace Processes, Secret Negotiations and Civil Society: Dynamics of Inclusion and Exclusion. International Negotiation, Volume 13, pp. 1-9.

Wilmot, C., 2024. Can AI Bring Peace to the Middle East. [Online] Available at: https://www.thebureauinvestigates.com/stories/2024-12-19/can-ai-bring-peace-to-the-middle-east [Accessed 19 July 2025].

Zia, L. & Waks, L., 2025. Rethinking Diplomatic Negotiations in the Age of AI, California: USC Center of Diplomacy.

About the Article

Question: How has Turkey's politics shaped its information ecosystem? Argument: AKP/Erdoğan control media and manipulate narratives. Conclusion: Media is regulated, polarized, and limits press freedom

About the Author

As a dedicated Radboud University Premaster Business student, **Liam von der Wiede** am passionate about expanding his academic knowledge and skills in International Politics, and Business Administration. Moreover, he is proud to be a member of the 22nd United Netherlands Delegation, where he engages and improves his skills in diplomacy, international relations, and teamwork. His experiences have deepened his interest in learning languages and getting to know people with diverse cultural and ethnic backgrounds. Focused on developing soft skills, particularly in teamwork, leadership and negotiation.

1. Introduction

In March 2025, mass protests erupted in Istanbul. Two million participants went to the streets and called for the release of the imprisoned mayor of Istanbul, and opposition party candidate, Ekrem Imamoglu (Michaelson 2025). While the immediate reason for this protest was the imprisonment, it must be seen as a broader historical reflection of what many critics see as democratic backsliding within Turkey (Toksabay and Erkoyun 2025). Over the past two decades, the systematic erosion of institutional checks and balances has undermined the judiciary, executive, and legislative branches of governance. Yet it is perhaps the press that has experienced the most profound transformation, especially after the Turkish 2022 Disinformation Law. Since then, the Turkish media landscape has been increasingly reshaped under government-led constraints on freedom of the press, and the growing influence of both domestic propaganda and foreign disinformation campaigns. As such, Turkey offers an important case for examining how disinformation, democratic resilience and media freedoms interact with democratic governance. Therefore, this article will grapple with the question: In what ways has Turkey's political evolution in the last 20 years shaped its information ecosystem and the control of public narratives? The paper argues that the AKP, the ruling party, and the president effectively reshaped the Turkish media landscape, undermined critical political discourse, oppressed the opposition, and consolidated political power. Foreign actors, while gaining prominence, are only of secondary importance. First, the paper will introduce the political background of Turkey over the past 20 years. Then it will elaborate on the rise of the information age, with a particular focus on social media and the COVID-19 infodemic. Next, it discusses how the AKP and President Erdoğan increase their state control by controlling the information landscape. Lastly, the paper will shed light on the importance of foreign actors and their impact on the Turkish information ecosystem.

2. Political Background

Before 2001, Turkey witnessed a series of di culties, including continuous political unrest, financial crises in 2001, and economic turbulence (Kubilay 2022, 2-3). The Justice and Development Party (AKP) was founded in 2001 and successfully took advantage of this widespread public frustration and presented itself as a party of political stability and economic reform (Kubilay 2022, 2-3). By appealing to conservative and centrist voters, the AKP gained broad popular support among Turks and won all national elections since 2002, which made Erdogan prime minister in 2002 (Yeşilada 2016, 21; Esen 2024, 7). When Erdoğan became prime minister, the presidency was largely symbolic while both the judiciary and the military were still powerful political actors. In the next decade, Erdoğan reduced the influence of the judiciary over politics through institutional reforms (Esen 2024, 12). Even though these changes were framed as efforts to strengthen civilian governance, they were highly contested and critics argued that it weakened judicial independence and limited media pluralism. In 2007, a constitutional amendment was introduced that allowed direct presidential elections, which set the stage for Erdoğan's 2014 election victory as Turkey's first popularly elected president (Yeşilada 2016, 23; Esen 2024, 12). Later, in 2016, a failed military coup served as a turning point and enabled Erdoğan and the AKP to justify a constitutional referendum on transitioning Turkey to an executive presidential system. Moreover, it gave the government the possibility to detain more than 77,000 people and suspend about 150,000 civil servants and military personnel (BBC 2019). With the support of the majority of the public, the new system was enacted in 2018 and created a presidency with unprecedented political power over the executive and judiciary branches (Kersting and Grömping 2021, 224; Esen 2024, 14). Together with the AKP's parliamentary majority, the new system effectively centralized power around Erdogan. Turkish politics was therefore transformed from a parliamentary system limited by judicial and military oversight into a personalized presidential system supported by party control (Esen 2024, 15).

3. The Information Age

In the middle of the political transformation in Turkish politics, social media became increasingly more popular in the mid-2000s. The rise of social media marked a dramatic shift in how people across the globe, and also in Turkey, receive and share information. While the internet had already existed for some time, the widespread adoption of platforms such as WhatsApp, YouTube, Facebook, and Twitter (now X) in the 2010s fundamentally changed the political and social landscape globally. Initially, it was seen as a democratizing force that would allow collective organization and resistance against authoritarianism. A prominent example is the Arab Spring in 2011, in which protestors organized themselves through social media to call for democratic reforms across the MENA region. However, over time, social media evolved into something that could also endanger democracies, and again, Turkey was no exception. In Turkey, platforms became saturated with partisan content, bots, and trolls (Kirdemir 2020, 6), while emotional appeals rooted in anger and fear contributed to escalating polarization. The digital sphere not only increased the competitive and polarized tendencies of traditional media (Kirdemir 2020, 4) but also further undermined public trust, already weakened by longstanding structural problems within Turkey's media system (Yurdakul 2020, 4). Social media helped to spread rumors, xenophobic narratives, and disinformation campaigns. Even more, it left Turkish society fragmented and vulnerable to manipulation in an increasingly volatile information ecosystem. This newly emerging breeding ground of social change was followed by the global COVID-19 health crisis that hit Turkey hard and exacerbated the ongoing political turmoil. Not only did the entire world and Turkey fall into a pandemic, but also an infodemic (Kirdemir 2020, 12). The pandemic negatively influenced the already fragile Turkish information environment to a significant extent in that it accelerated mis-, dis-, and malinformation flows of any kind (Yurdakul 2020, 3). Before and during the spread of COVID-19 in Turkey, many cases of false information seen in other countries also appeared on Turkish-language platforms. Prevention and cures, the nature of the virus, conspiracy theories relating to origins and nature of the pandemic, false claims about 5G communication technology, biological weapons, and grand conspiracies to control or curb the world population were the most frequent narrative types in this category (Kirdemir 2020, 15). Examples of these sorts of false information included that COVID-19 could be cured with garlic, vinegar, herbal cures, and saltwater, that only Asians could get coronavirus, or that it was caused by Chinese culture or race (Kirdemir 2020, 16). This multi-dimensional misuse of information from various actors harmed the Turkish information environment, since it undermined trust in traditional media outlets, which had to grapple immensely with the flood of misinformation. An additional problem is that fact-checking is still rather new, or not as popular (Yurdakul 2020, 11; Kirdemir 2020, 18; Bek 2025, 222-223), leading to a situation where misinformation almost spreads completely unregulated. Often enough, Turkish media outlets were taking over false narratives and also published them online; all in all, at much higher rates than media outlets in most other countries (Kirdemir 2020, 14). Besides, this rise of misinformation also intensified the political polarization in Turkish politics along party lines. Fake accounts, trolls, and bots took over online discourses on social networks and shaped political narratives. The outcome was a self-reinforcing feedback loop between three factors: polarization, toxicity, and false information. Each factor fueled the other two, thereby creating a vicious circle: greater polarization encouraged the spread of false information; false information deepened polarization; and both drove increasingly toxic discourse (Kirdemir 2020, 6-7). In sum, the COVID-19 pandemic showed how an already fragile information environment, combined with political polarization, was highly prone to any form of mal-, mis-, and disinformation and fundamentally hurt Turkish public discourse.

Disinformation:
False information spread to manipulate public opinion

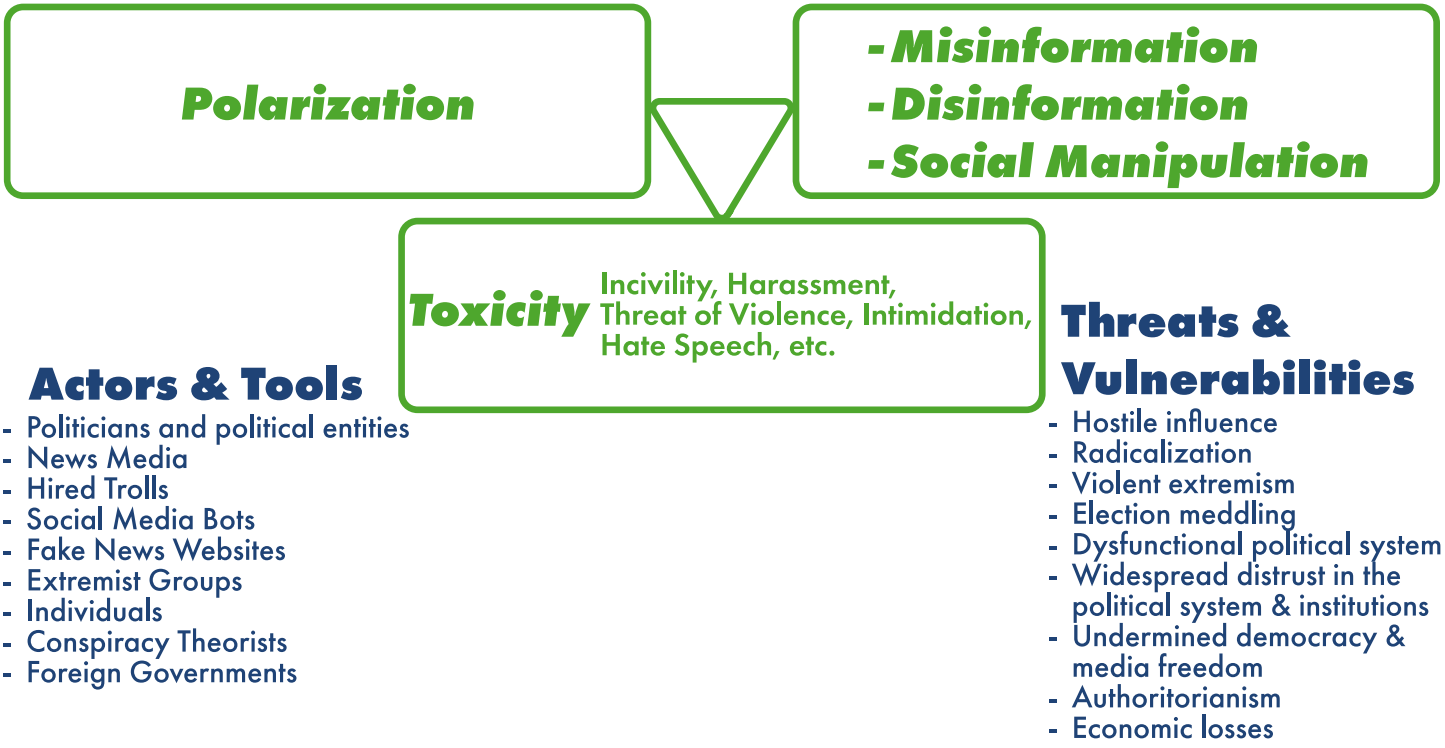


Figure 1: The self-reinforcing system of false information, polarization, and toxic discourse in Turkey (Kirdemir 2020, 7)

4. Autocratization of the Information Landscape

Since the early 2010s, the AKP government has increased its control over the press and social media. While in the early stages, social media was declared as a „source of evil" by Erdogan and other AKP politicians, later it was used to shape public narratives with the help of trolls (Bek 2025, 220). In 2013, during the military coup, where Fethullah Gülen, a clergyman and previously a friend of Erdogan and the AKP, was arrested, the government revealed that FETÖ, the terrorist organization that tried to overthrow the ruling party, promoted disinformation and propaganda on social media to create chaos within Turkey (Bek 2025, 222). Critics, therefore, argue that disinformation is used to portray enemies of the party clearly, at the same time, when the AKP is also using social media as its domain of influence (Bek 2025, 222). With the rise of COVID-19, the Turkish government blocked, in 2020 alone, 467011 websites, domain names were blocked, 22554 news articles were blocked, and 15,832 news articles were deleted (Bek 2025, 225). Moreover, the Free Web Turkey project reports that in 2021, around 11,050 URLs, domain names, and social media posts were blocked, while 49 news websites were banned (Free Web Turkey 2022). In 2022, five media representatives were detained, 20 media representatives were attacked, and 126 media representatives went to court between July and August (Bek 2025, 225). With the enactment of the Amendment to the Press Law (number 7418), published on the offcial government website on October 13, 2022, a major change in press freedom appeared.Article 29 stipulates that anyone who publicly publishes false information concerning the internaland external security, public order, and public health of the country with the aim of creating anxiety, fear, or panic among the public shall be punished with imprisonment of 1 to 3 years (Turkey 2022, art. 29). Additionally, Articles 3 and 4 state that news websites and social media platforms need to store news content for a certain period of time and make it available upon request (Turkey 2022, art. 3, 4). While the government claims that this law is in place to protect the public from disinformation, critics argue that the law must be considered as criminalising journalism and limiting freedom of information, since it is unclear what mis-, or disinformation, or even danger means, and is only up to the courts to decide — the same courts that are overly represented by the AKP and its allies (Bek 2025, 225). On the other hand, Erdogan and his allies continue to spread

misinformation themselves. During the 2023 national election, which Erdogan won once again, a widely circulated video depicted Kılıçdaroğlu encouraging people to vote, followed by an endorsement from Karayılan, a co-founder of the PKK, which helped the AKP to spread their narrative that the opposition party is working together with the Kurdish terrorist organization (Andı et al. 2025, 7; Bek 2025, 225). The same narrative was used to crack down on the opposition candidate for the upcoming election in
2028, Imamglou, who was arrested in March 2025 due to alleged support for the PKK, and „several financial crimes" (Fraser 2025). The arrest enflamed nationwide protests, which led to an even more crackdown on opposition voices (Aslan 2025). Today, Turkey has experienced around two decades of democratic backsliding, alongside polarization. About 90% of all news outlets are controlled directly or indirectly by the government, while social media is now regulated by the new Disinformation Law (Andı et al. 2025, 7). When it comes to press freedom, Turkey ranks globally in 159th place, on the same level as Sudan and Venezuela (RSF 2025). Moreover, it is one of the six most rapidly autocratising states worldwide, aside from Brazil, Hungary, India, Poland, and Serbia (Boese et al. 2022, 990).

5. Geopolitics and Foreign Influence

Turkey is at the center of many regions, with various state and non-state actors that have various interests. It borders not only countries in the Caucasus, such as Armenia and Georgia, but also in the Middle East, such as Syria, Iran, and Iraq, and Europe, including Greece, Bulgaria, and Cyprus. Moreover, it borders two seas: the Mediterranean and the Black Sea. On top of that, Turkey upholds a complex web of relationships with various parties, which especially concerns security and economic partnerships. On the one hand, Turkey is a key NATO partner with one of the biggest armies in the alliance (Turak 2024), has close ties to Washington, and provides military equipment to Ukraine (Notte and Kane 2022, 5). On the other hand, Turkey does not shy away from following its own security interests even against allied Greece in the Mediterranean or the US when it comes to the Kurds in Syria (Notte and Kane 2022, 5). Moreover, it also has strong economic partnerships with Russia and refrains from joining Western sanctions against Russia (Notte and Kane 2022, 4). With Turkey also being on the rim of many contemporary wars, such as in Ukraine, Syria, or Israel, it seems trivial that Turkey is also impacted by foreign disinformation campaigns that try to steer political narratives inside the

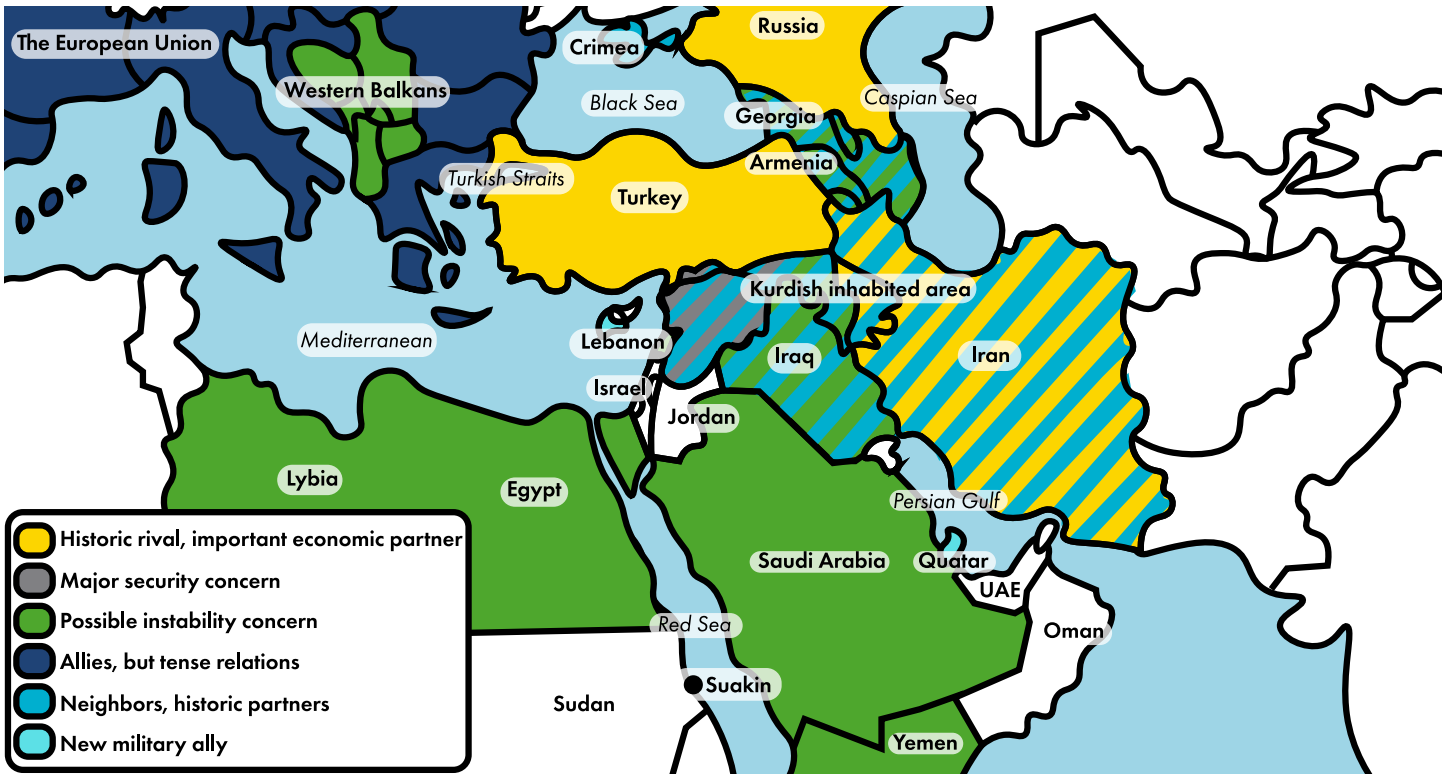


Figure 2: Strategic Geopolitical Map of Turkey and its neighboring countries (Uzunköprü 2019) Source: Geopolitical Intelligence Services

country. „The most frequent sources of such campaigns and narratives, as discovered by social media monitoring outlets so far, originate from the Middle Eastern or Russian entities (Kirdemir 2020, 9)." The best case for such a disinformation intervention occurred in 2015, when a „Russian aircraft violated the Turkish airspace by about 2.19 kilometers" and got shot down by Turkish authorities, killing one of the two pilots (Ünver 2018, 14). This event led to a high polarization within the internet, in which two narratives appeared: Narrative A focused on blaming Russia for violating Turkish airspace, and Narrative B focused on blaming Turkey for shooting down the jet outside of Turkish airspace. After a week, however, a new narrative was coming up that presented Turkish government offcials, including Erdoğan's family, as being involved in smuggling oil from ISIS during the peak of the Syrian war (Ünver 2018, 17; Costello 2018, 4-5). This narrative, although proven to be incorrect, was highly effective in its reach and impact. Even Western media outlets took up the narrative, portraying Turkey as an untrustworthy NATO partner (Ünver 2018, 17; Kirdemir 2020, 9-10). Besides, other Turkish military operations in the region (Ünver 2018, 22-23; Kirdemir 2020, 10), as well as other events such as the military coup in 2016, or the assassination of the Russian ambassador in Turkey, became subjects to manipulative campaigns by various actors, most importantly Russia (Castello 2018, 8-9, 11). Russian trolls by the Internet Research Agency were also involved in promoting „opposition to President Erdogan and Turkey in general" and made jokes about „Erdogan planning the refugee crises" (Harry Collins et al. 2022, 152). All in all, however, research remains unclear on whether the Turkish information ecosystem, since it is so infected with originated mis-, dis-, and malinformation, is more or less prone to Russian disinformation (Ünver 2018, 44; Kalathil 2020, 37)

Turkey's government controls media to shape public opinion and suppress dissent

6. Conclusion

To summarize, Turkey's political evolution over the past two decades has strongly reshaped its information ecosystem. The AKP and President Erdoğan have been able to centralize control over traditional and digital media, crack down on press freedom, and criminalize dissent through the 2022 Disinformation Law. This created an environment where opposition voices are suppressed and public narratives are controlled. The rise of social media and the COVID-19 infodemic supported the spread of misinformation, strengthened political polarization, and undermined public trust in traditional media outlets. While the 2022 Disinformation Law was framed as a measure to combat harmful misinformation, the law effectively criminalized critical journalism and constrained press freedom. By controlling both traditional and digital media channels, the AKP has created a highly regulated information environment in which political power and public perception are increasingly inseparable. It also enabled the government to suppress critical opinions, as illustrated by the arrests of prominent opposition figures and the nationwide crackdown on media outlets. With Turkey's unique strategic position in foreign and security affairs, it appears to be, from time to time, a target of information attacks from foreign actors, such as Russia and Middle Eastern entities. However, with Turkey having a largely infected information environment itself, it remains unclear what the concrete impact of those attacks is. Future scholarship is needed on the effects of disinformation on the 2023 national election.

References

Andı, Simge, Ali Çarkoğlu, Lemi Baruh, and Zsofia Bocskay. 2025. "Authoritarians Do It Better? Belief in Misinformation in Turkey." The International Journal of Press/Politics. https://doi.org/10.1177/19401612241307812.

Aslan, Kemal. 2025. „In Turkey, ‚We Are No Longer Afraid to Speak out about the Things That Have Become Unbearable for Us.'" LeMonde, 31 March 2025. https://www.lemonde.fr/en/international/article/2025/03/31/in-turkey-we-are-no-longer-afraid-to-speak-out-about-the-things-that-have-become-unbearable-for-us_6739708_4.html.

BBC. 2019. „Turkey Election: Erdogan Disputes Results in Major Cities." 2 April 2019. BBC. Accessed 30 September 2025. https://www.bbc.com/news/world-europe-47785095.

Boese, Vanessa A., Martin Lundstedt, Kelly Morrison, Yuko Sato, and Staffan I. Lindberg. 2022. "State of the World 2021: Autocratization Changing Its Nature?" Democratization 29(6): 983–1013. https://doi.org/10.1080/13510347.2022.2069751.

Collins, Harry, Robert Evans, Martin Innes, Eric B Kennedy, Will Mason-Wilkes, and John McLevey. 2022. "Disinformation and Misinformation." In The Face-to-Face Principle. Science, Trust, Democracy, and the Internet. Cardiff University Press. https://www.jstor.org/stable/jj.14308229.14.
Costello, Katherine. 2018. „Russia's Use of Media and Information Operations in Turkey: Implications for the United States." RAND Corporation. https://www.jstor.org/stable/resrep19906.

Echeverría, Martin, Sara García Santamaría, and Daniel C. Hallin. 2024. „State-Sponsored Disinformation Around the Globe: How Politicians Deceive Their Citizens." Routledge. https://doi.org/10.4324/9781032632940.

Esen, Berk. 2025. "Judicial Transformation in a Competitive Authoritarian Regime: Evidence from the Turkish Case." Law & Policy 47 (1): e12250. https://doi.org/10.1111/lapo.12250.

Fraser, Suzan. 2025. "What We Know about Istanbul's Mayor and Why He Was Arrested." AP News. 19 March 2025. https://apnews.com/article/turkey-crackdown-istanbul-mayor-imamoglu-arrest-erdogan-a4b5e63c626df17b7075ccf6b6d9cdc8.

Free Web Turkey. 2022. Free Web Turkey Website. "Free Web Turkey Report: Another year of increasing censorship and surveillance." Accessed 30 September 2025 https://www.freewebturkey.com/free-web-turkey-report-another-year-of-increasing-censorship-and-surveillance/.

Kalathil, Shanthi. 2020. "The Evolution of Authoritarian Digital Influence: Grappling with the New Normal." PRISM 9 (1): 32–51.

Kersting, Norbert, and Max Grömping. 2022. "Direct Democracy Integrity and the 2017 Constitutional Referendum in Turkey: A New Research Instrument." European Political Science 21 (2): 216–36. https://doi.org/10.1057/s41304-020-00309-3.

Kirdemir, Baris. 2020. „Exploring Turkey's Disinformation Ecosystem: An Overview." Centre for Economics and Foreign Policy Studies. https://www.jstor.org/stable/resrep26087.

Kubilay, Murat. 2022. "The Turkish Economy Under The Presidential System". Middle East Institute (MEI). https://www.mei.edu/publications/turkish-economy-under-presidential-system.

Michaelson, Ruth. 2025. "Turkish Opposition Rallies in Defence of Jailed Istanbul Mayor in Mass Protest." The Guardian, March 29. https://www.theguardian.com/world/2025/mar/29/turkish-opposition-calls-mass-rally-in-defence-of-jailed-istanbul-mayor.

Notte, Hanna, and Chen Kane. 2022. „Russian-Turkish Relations and Implications for U.S. Strategy and Operations." James Martin Center for Nonproliferation Studies (CNS). https://www.jstor.org/stable/resrep47183.4.

RSF. 2025. Reporters Without Borders Website. Accessed 25 September 2025. https://rsf.org/en/index.

Toksabay, Ece, and Ezgi Erkoyun. 2025. "Turkey Detains Istanbul Mayor in What Opposition Calls 'Coup.'" Reuters, March 19. https://www.reuters.com/world/middle-east/turkish-authorities-order-detention-istanbul-mayor-some-100-others-2025-03-19/.

Turak, Natasha. 2024. "Turkey Is Back in from the Cold with NATO and F-16 Moves, but Thorny ssues Remain." CNBC, February 21. https://www.cnbc.com/2024/02/21/turkey-is-back-in-from-the-cold-with-nato-and-f-16-moves.html.

Turkey. 2022. Law on Press and Amendments to Some Laws (Law No. 7418, adopted October 13, 2022). Offcial Gazette, October 18, 2022. https://www.resmigazete.gov.tr/eskiler/2022/10/20221018.pdf.

Uzunköprü, Süleyman. 2019. "Figure 2. Turkey's Geopolitical Position." Research Gate. Accessed September 30, 2025. https://www.researchgate.net/figure/Turkeys-Geopolitical-Position_fig2_336679879.

Ünver, H. Akin. 2019. „Russian Digital Media and Information Ecosystem in Turkey." Centre for Economics and Foreign Policy Studies. https://www.jstor.org/stable/resrep21042.

Yeşilada, Birol A. 2016. "The Future of Erdoğan and the AKP." Turkish Studies 17 (1): 19–30. https://doi.org/10.1080/14683849.2015.1136089.

Yurdakul, Afsin. 2020. „The Impact of Polarization on Turkey's Information Environment." Centre for Economics and Foreign Policy Studies. https://www.jstor.org/stable/resrep26090.

About the Article

Main Question: How can AI be used safely in NATO and EU cyber operations? Argument: AI boosts cyber defence but needs oversight, security, and coordination. Conclusion: Combining AI with regulation and collaboration strengthens defence and accountability

About the Author

João Pedro Souza Gohla is a Master's student in Law and Security at NOVA School of Law in Lisbon. He holds a Bachelor's degree in Political Science with a focus on International Relations and History from Goethe University Frankfurt. His current research examines security policy, migration, and state fragility within contemporary global governance frameworks.

1. Introduction

Artificial intelligence (AI) has transitioned from an emerging technology in the cyber realm to an omnipresent aspect of operational reality. Its ability to analyse large datasets, identify anomalies instantly, and even automate specific decision-making tasks presents unparalleled chances for enhancing cyber defence. However, the same abilities that render AI beneficial also introduce new dangers, especially when used in delicate military situations. These dangers are heightened in cyberspace, where activities can occur rapidly, transcending borders, and often lacking immediate identification. For NATO and the European Union (EU), cyber capabilities enhanced by AI present a twofold challenge. On one side, they can greatly improve group resilience against advanced cyber threats. While, simultaneously posing intricate legal, ethical, and political dilemmas concerning autonomy, accountability, and control. In contrast to traditional weapon systems, AI employed in cyber operations is typically hidden from public view and can be used secretly, complicating oversight efforts. At present, NATO does not carry out offensive cyber operations these are solely the responsibility of individual member nations. These abilities are regarded as „national resources" that can be utilized for alliance operations on a voluntary basis (Shape n.d.). Although this setup honours national sovereignty, it leads to coordination issues. Sensitive data regarding the extent, techniques, and preparedness of national assets is frequently kept under strict control, hindering NATO's capacity to organize and execute genuinely coordinated cyber operations. This issue of „secrecy" can obstruct trust and delay decision-making in times of crisis. The EU encounters a distinct structural situation. It lacks a military command structure similar to NATO but wields significant influence via its regulatory capabilities, research financing, and coordination tools such as the EU Agency for Cybersecurity (ENISA). EU cybersecurity efforts typically emphasize standardization, strengthening resilience, and enhancing capabilities, which can support NATO's operational functions (Trimintzios et al., 2017). Nevertheless, the absence of a direct operational mandate implies that EU-level actions must depend on the implementation by member states. Considering these variations, the subsequent suggestions aim to enhance AI-driven cyber defence at the national, NATO, and EU tiers. Every suggestion tackles both operational and governance aspects, guaranteeing that technological progress is accompanied by strong policy structures.

2. Clarify NATO's Role and Capabilities in Cyber Operations

Why it matters:

NATO's existing cyber defence strategy is based on the idea that member nations maintain authority over offensive abilities (Shea, 2025). Though politically essential, this setup may hinder prompt collective reactions to rapidly evolving cyber threats especially those driven by AI. In situations where an opponent launches AI-based assaults on several NATO nations at the same time, lags in the coordination of national resources may enable the attacks to intensify without restraint. The categorization of national capabilities introduces an additional layer of complexity alliance planners might remain unaware of available tools until a crisis arises.

Recommendation:

NATO ought to expand the mandate of its Cyber Operations Centre to strengthen the operational integration of AI-driven tools, while still maintaining national authority over their usage. This may require. The creation of a protected, classified database of AI-driven cyber capabilities possessed by member countries, available solely to approved NATO strategists. Establishing interoperability standards for AI technologies, guaranteeing that national systems operate cohesively during collaborative missions. Developing pre-approved operational playbooks for specific types of cyber defence measures, minimizing the necessity for prolonged political discussions in critical situations.

Supporting platform:

The NATO Cyber Operations Centre provides a central hub for alliance cyber activities, while the CCDCOE supports joint exercises and training (NATO CCDCOE, 2023). These institutions could serve as the operational and conceptual anchors for AI integration.

3. Develop International Rules for AI in Cyber Warfare

Why it matters:

Artificial intelligence in cyber warfare presents distinct regulatory issues. In contrast to kinetic weapons, AI cyber tools can be created and utilized with minimal physical infrastructure, making them more challenging to oversee within current arms control systems (Dykstra, Inglis, & Walcott, 2020, p. 116-118). The lack of global regulations leads to a strategic void where nations might feel compelled to create and implement offensive AI technologies proactively.

Recommendation:

Engage with multilateral platforms such as the United Nations Group of Governmental Experts (UN GGE), NATO, the EU, and the G7 to develop practical, enforceable standards for AI applications in military cyber activities. These must, outline banned applications of AI, especially fully autonomous offensive cyber tools that can operate independently without human supervision. Set up essential transparency standards, including mechanisms for reporting before and after operations. Steer clear of impractical universal bans that might unfairly impact liberal democracies, prioritizing practical protections instead.

Supporting example:

The Tallinn Manual 2.0 (Schmitt, 2017) offers a general starting point but is a non binding resource for legal advisers and policy experts dealing with cyber issues. Updating this framework to include AI-specific scenarios would bridge a critical gap (NATO Cooperative Cyber Defence Centre of Excellence, n.d.).

4. Require Human Oversight and Explain ability in Military AI

Why it matters:

Lack of transparency in AI decision-making poses a major risk for governance. In military cyber operations, where immediate decisions can have significant strategic impacts, the lack of clarity in an AI system's actions diminishes both operational trust and democratic responsibility. This is especially pronounced in the EU, where the regulatory environment highlights transparency and rights safeguards, and in NATO, where political agreement necessitates that member countries have confidence in each other's systems

Recommendation:

Enforce legal and policy standards that require AI utilized in cyber defence to be understandable to operators and policymakers. Creates a record of choices and measures implemented. Is subject to significant human supervision at critical decision moments. NATO might integrate these demands into its procurement criteria for shared initiatives, and the EU could broaden the reach of the Artificial Intelligence Act to specifically include military AI, ensuring uniform oversight processes among member states (European Commission, 2021).

Artificial Intelligence (AI):
Machines performing tasks that normally require human intelligence

Supporting model:

The European Union's Artificial Intelligence Act (AI Act) explicitly excludes applications used solely for military, defence, or national security purposes, as set out in Article 2(3) and Recital 12. In the civilian sphere, however, the Act establishes stringent requirements for high-risk systems, including those that process biometric data. Extending comparable oversight mechanisms to military AI could help bring defence technologies into line with established civilian standards. Alongside this, the European Parliament has adopted policy resolutions encouraging the development of AI-driven cyber defence capabilities, encompassing both defensive and offensive measures

provided they comply with international law. These initiatives, however, remain political guidance rather than binding provisions of the Act itself (European Parliament, 2022; European Parliament, 2023; Sierra-Tango, 2023).

5. Build AI Tools That Are Secure by Design

Why it matters:

AI systems are naturally susceptible to threats like adversarial inputs, data poisoning, and model inversion (Federal Office for Information Security, 2023, p.5-11). Within a NATO framework, a breached AI defence mechanism might generate weaknesses among various member nations if interoperability functions are misused. In the EU, insecure AI solutions created in the private sector might be incorporated into defence supply chains, bringing systemic risks.

AI enhances cyber defence but requires strict human oversight to prevent misuse

Recommendation:

Implement secure-by-design principles as a mandatory criterion for all AI systems employed in defence (Cybersecurity and Infrastructure Security Agency, n.d.). This involves, strict adversarial testing throughout development. Incorporating cybersecurity protocols into the AI framework from the beginning. Implementing ongoing surveillance to identify and address emerging risks. Although this might raise upfront expenses, savings over time will be achieved by lowering the necessity for expensive retrofits and minimizing the chances of major failures.

Supporting example:

Ukraine's Delta situational awareness system demonstrates the practical importance of secure-by-design architecture in contested settings (Bondar, 2024, p.7-12).

6. Promote International AI-Cyber Threat Intelligence Sharing

Why it matters:

Cyber threats powered by AI function at machine speed, allowing minimal time for human detection and response. In the absence of real-time intelligence sharing, even the most proficient individual state can become overwhelmed. Intelligence-sharing mechanisms are present in NATO but may be hindered by classification obstacles. The EU has put resources into civilian cyber threat-sharing systems, yet these frequently lack connection to military networks.

Recommendation:

Create cohesive, AI-focused threat intelligence networks at NATO and EU levels, connected via secure gateways. They ought to provide not just threat information but also AI-generated analytical results. Utilize standardized taxonomies for AI-driven threats to enhance interoperability. Perform collaborative NATO–EU training drills that replicate AI-assisted assaults on essential infrastructure. As AI-enabled cyber threats move at machine speed, integrated threat-intelligence channels are critical for both NATO and EU cyber resilience. Figure 1 illustrates a conceptual NATO–EU intelligence-sharing network, highlighting how existing military and civilian channels could be linked through AI-driven analytical hubs. This proposed structure aims to reduce classification bottlenecks, standardise threat taxonomies, and enable near real-time data exchange during cyber incidents. This diagram shows military (NATO), civilian (EU), member state, and AI analysis nodes, along with proposed secure gateway links for AI-driven cyber threat intelligence exchange.

Integrated NATO-EU AI Threat Intelligence Sharing Network

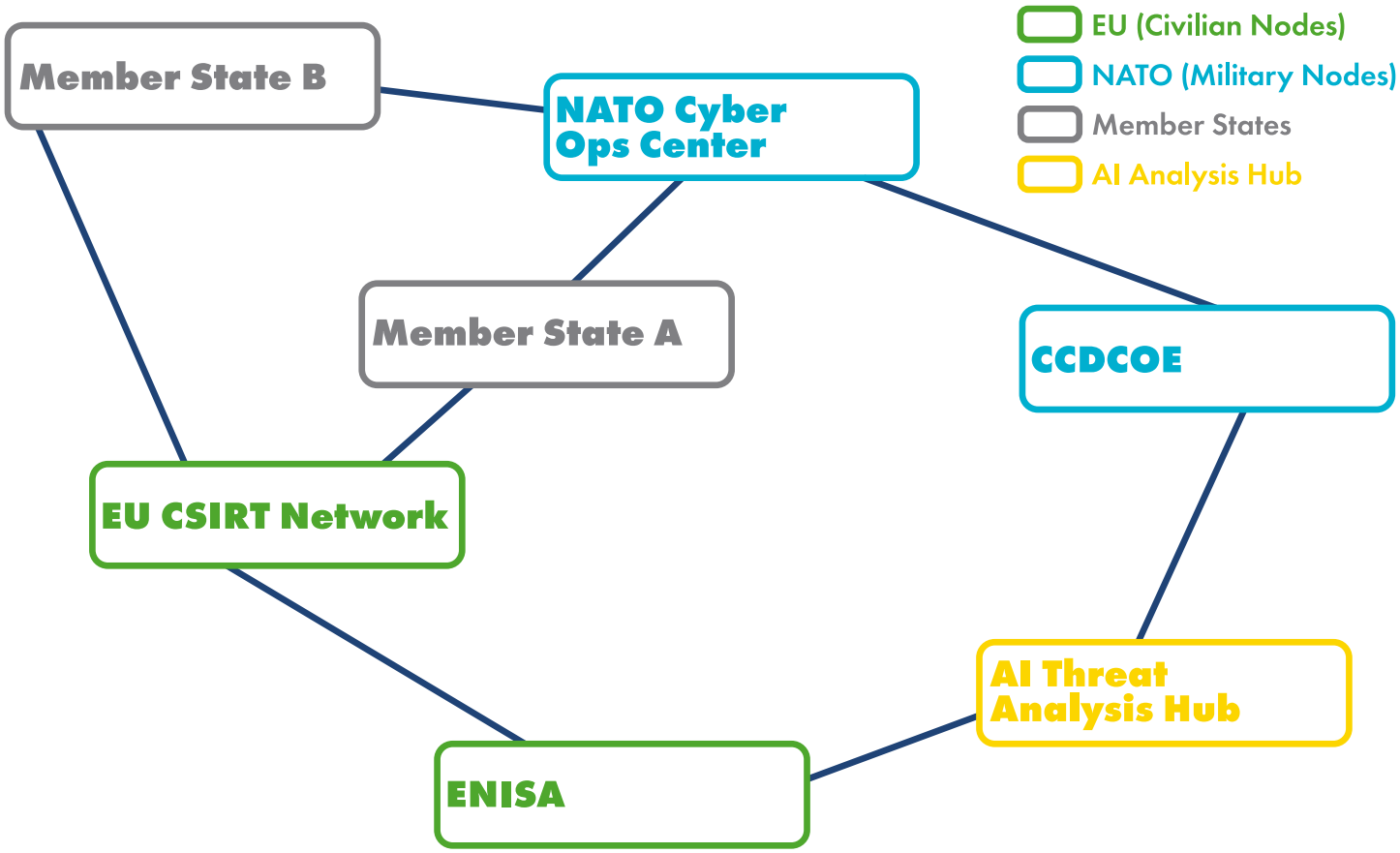


Figure 1. Integrated NATO–EU AI Threat Intelligence Sharing Network.

Supporting platform:

The NATO CCDCOE's Locked Shields exercise is a prime venue for testing these capabilities, and the EU's CSIRT Network could serve as the civilian counterpart.

7. Close the Skills Gap in AI and Cybersecurity

Why it matters:

NATO and the EU both encounter a lack of experts skilled in the convergence of AI, cybersecurity, and defence strategy. Lacking adequate expertise, even the most sophisticated policy frameworks will be ineffective in real-world application. NATO's power is rooted in its operational training framework, whereas the EU possesses more sway in educational policies and research financing.

Recommendation:

Initiate collaborative NATO–EU talent programs that create unified curricula addressing AI, cybersecurity, and military uses. Provide shared scholarships and exchange initiatives for professionals from partner countries. Encourage collaboration between public and private sectors to train and keep specialists in the defence industry. Even the most advanced AI-enabled cyber capabilities will falter without a sufficiently trained workforce. The EU currently employs approximately 9.37 million ICT specialists (Eurostat, 2024), but is on track to reach only 12 million by 2030—far below the EU's target of 20 million. Figure 2 visualises this projected shortfall, illustrating the urgent need for joint NATO–EU talent development programmes. Data source: Eurostat (2024) and European Commission (2023). The EU faces an estimated shortfall of approximately 8 million ICT specialists by 2030, underscoring the urgency of coordinated training and recruitment initiatives.

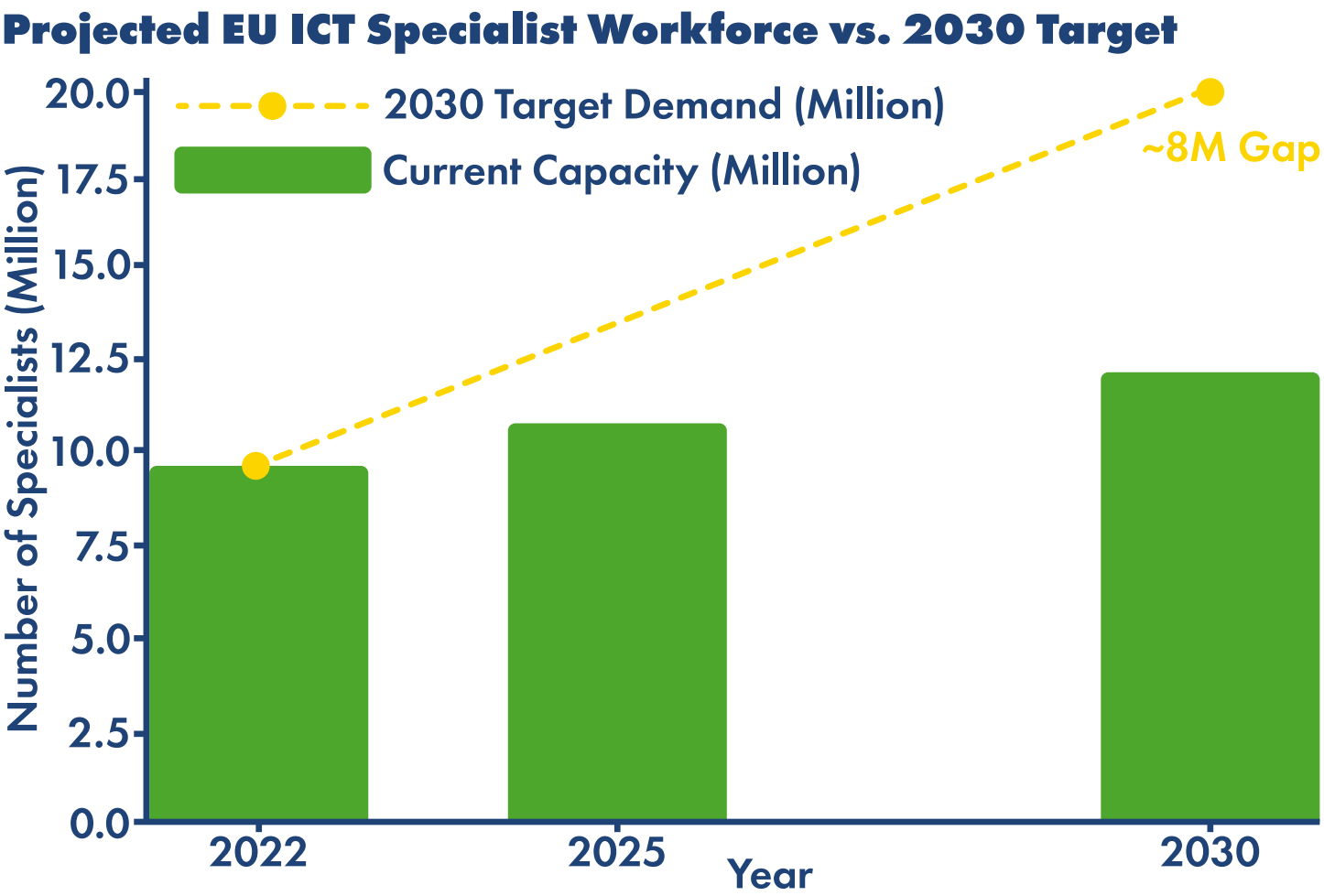


Figure 2. Projected EU ICT Specialist Workforce vs. 2030 Target.

Supporting example:

India's post-crisis investment in AI-cybersecurity training offers a model that could be scaled for multinational application (Bharadwaj, 2025).

8. Regulate Offensive AI Use with Political Oversight

Why it matters:

Offensive AI in cyber warfare represents a highly sensitive political aspect of military technology policy. In NATO, where consensus is essential for decision-making, any ambiguity regarding the use of offensive AI may lead to tension among member states. Within the EU, the lack of a direct military mandate does not eliminate the ability to exert influence especially via export regulations, procurement policies, and industrial strategies.

Recommendation:

At the national level, establish explicit criteria for parliamentary or legislative consent prior to utilizing offensive AI in cyber operations. At the NATO level, establish common definitions and political oversight processes that promote transparency among allies while maintaining operational security. At the EU level, create export regulations for offensive AI systems, in line with wider international standards.

Supporting framework:

NATO's Responsible AI principles already emphasise accountability and chain of command these could be formalised into binding political oversight structures (NATO Innovation Hub, 2021).

9. Conclusion

The incorporation of AI into military cyber operations presents significant strategic benefits as well as serious governance issues. For NATO, the main focus is operational integration making sure that national assets can be aligned swiftly and efficiently without compromising sovereignty. For the EU, the main focus is on regulatory consistency establishing elevated benchmarks for transparency, security, and accountability that member nations and industries are required to uphold. NATO and the EU can enhance their collective defence strategy and maintain the democratic values they aim to safeguard by defining roles, establishing practical global standards, ensuring human oversight, integrating security into AI technologies, exchanging threat intelligence, addressing the skills gap, and regulating offensive AI with political supervision.

References

Bondar, K. (2024). Understanding the military AI ecosystem of Ukraine. https://www.csis.org/analysis/understanding-military-ai-ecosystem-ukraine

Cybersecurity and Infrastructure Security Agency. (n.d.). Secure by design. CISA. Retrieved August 19, 2025, from https://www.cisa.gov/securebydesign

Dykstra, J., Inglis, C., & Walcott, T. S. (2020). Differentiating kinetic and cyber weapons. Joint Force Quarterly, 99, 116–123. National Defense University Press.

European Commission. (2021). Proposal for a regulation laying down harmonized rules on artificial intelligence (Artificial Intelligence Act). https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence

European Parliament. (2022). Artificial intelligence in a digital age [Resolution 2022/2040(INI)]. EUR-Lex. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A52022IP0140

European Parliament. (2023). European Parliament legislative resolution on the proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html

Eurostat. (2024). ICT specialists—statistics on hard-to-fill vacancies in enterprises. Retrieved from https://ec.europa.eu/eurostat/statistics-explained/index.php?oldid=679049

Federal Office for Information Security. (2023). AI security concerns in a nutshell: Practical AI-Security Guide 2023. BSI. https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/KI/Practical_AI-Security_Guide_2023.pdf

NATO Cooperative Cyber Defence Centre of Excellence. (2023). About CCDCOE. https://ccdcoe.org/about-us/

NATO Innovation Hub. (2021). Responsible AI in the military: A framework. https://innovationhub-act.org

Schmitt, M. N. (Ed.). (2017). Tallinn Manual 2.0 on the international law applicable to cyber operations. Cambridge University Press.

Shea, J. (2025). Non-traditional security threats. In J. Sperling & M. Webber (Eds.), The Oxford handbook of NATO (Oxford Handbooks). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198851196.013.29

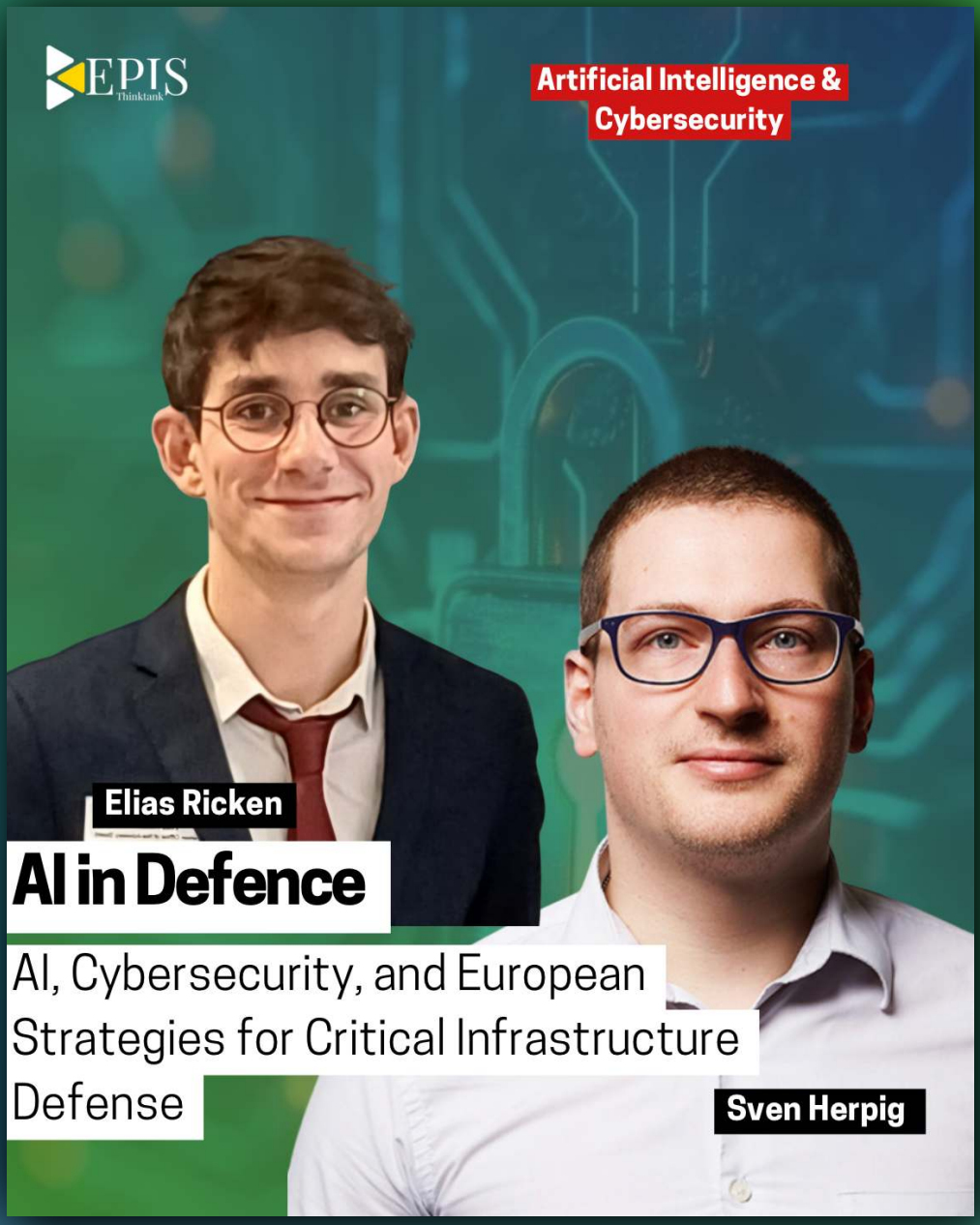
Sierra-Tango. (2023). AI Act: What impact for defence? https://sierra-tango.eu/en/ai-act-what-impact-for-defence/

Supreme Headquarters Allied Powers Europe. (n.d.). Cyber defence. NATO. Retrieved August 15, 2025, from https://shape.nato.int/about/aco-capabilities2/cyber-defence

Bharadwaj, S. (2025, May 25). How pro-India hackers defended country during cross-border cyberattacks amid Op Sindoor. The Times of India. https://timesofindia.indiatimes.com/city/hyderabad/how-pro-india-hackers-defended-country-during-cross-border-cyberattacks-amid-op-sindoor/articleshow/121385229.cms

Trimintzios, P., Chatzichristos, G., Portesi, S., Drogkaris, P., Palkmets, L., Liveri, D., & Dufkova, A. (2017). Cybersecurity in the EU Common Security and Defence Policy (CSDP): Challenges and risks for the EU (STOA Study PE 603.175). European Parliament. https://www.europarl.europa.eu/RegData/etudes/STUD/2017/603175/EPRS_STU(2017)603175_EN.pdf

United Nations Group of Governmental Experts (UN GGE). (2021). Reports on responsible state behavior in cyberspace.
World Economic Forum. (2024). Strategic Cybersecurity Talent Framework. Retrieved from https://www.weforum.org/publications/strategic-cybersecurity-talent-framework



About the Interview

Main Question: How is AI used in cybersecurity and critical infrastructure defense in Europe? Argument: AI amplifies attacks and defenses, widening threats and needing better talent and operational practices. Conclusion: Europe must improve skills, operations, and selective cooperation to strengthen cyber resilience

About the interviewee

Dr. Sven Herpig leads Cybersecurity Policy and Resilience at the expert organization Interface. His work spans projects on local government cybersecurity and national UN cyber norms implementation. His research focuses on active cyber defense, state responses to cyber operations, government hacking, vulnerability management, AI security, open-source resilience, and election protection.

About the Interviewer

Elias Ricken is a Master Student in International Affairs at Hertie School in Berlin. He specializes on Defence and Security Policy in Germany and France and hybrid threats. To this end, he also works as a Research Assistant at the German Council of Foreign Relations (DGAP) and is an active member if the association European Defence Network.

Elias Ricken:

Thank you, Mr. Herpig, for being here today with us for this interview. I would like to discuss three overarching topics with you today. The first one deals with Cyber Security, AI and Defence. The second one is on critical infrastructure, and the third one is about on European perspectives. Then the first question of the general segment would be: How is AI used today in cyberspace and information space?

Sven Herpig:

I am mainly looking at AI from a perspective on the intersection of machine learning and AI with cybersecurity. You have basically three sub-intersections in that area.

The first is using AI or machine learning to overcome security mechanisms. This is, for example, used in automatically trying to scan for vulnerabilities and helping to create more authentic-looking phishing emails. Then the second intersection is the entire opposite. It consists of using machine learning or AI elements to prove and bolster cybersecurity, which includes, for example, automated log scanning, and anomaly detection. So cybersecurity is very much based on large chunks of data, and that's where machine learning shines: It can help to make sense of large amounts of data in very little time. That's why it helps to bolster some existing cybersecurity mechanisms. And then the third of the section, I personally found the most interesting research topic, is how including any machine learning or AI-enabled part in software and infrastructures, increases and changes the threat landscape and the supply chain for any given IT environment. The distinction between the three is helpful, because in practice you have multiple AIs for different things: For example those helping the attackers, those helping the defenders, and those that assist you in running the systems and make your infrastructure more vulnerable.

Elias Ricken:

Speaking of defenders and attackers, as I understand it, the means of attacking, which you just mentioned, phishing, scanning for vulnerabilities and others, remain very similar, no matter if the defender is an entire state, or just an individual with his computer at home. Is that correct?

Sven Herpig:

Well, in general, yes, but the more complex the infrastructure is, the more the attack surface is going to differ. That means that attackers can leverage different segments of your infrastructure or attack surface and exploit parts of it. If you're running your laptop at home, you have a small amount of services and hardware that you're using. So the attacker is basically limited to exploit the few vulnerabilities in those devices. But if you're running a network with 10,000 services, 20,000 computers and servers and whatnot, then of course, the attack surface is much larger. So the attacker can find different entries and vulnerabilities. So you have a larger attack surface to secure.

Elias Ricken:

Okay, thank you. We also want to talk about critical infrastructure in this interview. In general, for someone who is not specialized in cybersecurity, could you explain how these kind of cyberattacks process? How can I visualize them, in order to understand them better?

Sven Herpig:

First you have to understand that the entire ecosystem, especially in the realm of cyber criminals, but also the cooperation between criminals and intelligence and security agencies, for example, is very much a division of tasks and expertise. You have groups of people that program the malware, those that get access to a system and those that conduct the operations once the malware is in the system and so on and so forth. When the attacking actor is a state, there's a lot of resources invested from certain states to carry out these operations and campaigns. When we look at the criminal ecosystem, the criminals make a lot of money by conducting cybercrimes, so they also have a lot of money to spend on improving their tools and to hire talent. So, the first thing that we need to understand is that there can be a lot of resources on the attacking side.

Now, having said that, the attackers don´t necessarily need to use all that sophistication in order to conduct their operation. Instead they will look for the easiest point of entry into a system. This is important, because we need to take notice of the fact that it's of course important to secure machine learning enabled software. At the same time, if the intelligence agencies and the criminals still get into your infrastructure and your systems by exploiting your traditional software, that you leave rather unprotected, then we need to work on that first. Now, how does it continue from there? First you have to find a path in. That can be finding vulnerabilities, especially the edge devices. These are devices that are directly connected to the internet, such as routers. Exploiting vulnerabilities will get you into the system, first into the router, then pivoting into the internal system, and then you basically continue from there. You establish a bridgehead, if you want to use that synonym or comparison to the military world. Then from there on, you basically have lateral movement inside networks, and you go where you want to go inside the network. Another approach can be phishing. Actually, the most used initial attack vector to infiltrate a system is simply using the legitimate credentials. So to figure these out by trying to convince someone to give them your password and your account name and maybe sometimes your two factor authentication. There's different ways of getting to the system, but the initial access is where the operational attack starts. With that, there's two common ways to do that: Number 1 is tools and malware that were either custom made for your operation or can be more like a commercial off the shelf product. Number 2 is what we call "living off the land", the stealthier of the two approaches, that consists of actually using legitimate software that is already installed to go from A to B. Both ways enable to accomplish various goals, such as surveying a system or copying data.

Elias Ricken:

What would be the role of AI in such a development? What is the difference between a cyberattack using AI and a cyberattack not using AI? Or does the latter one even exist today?

Sven Herpig:

Well, here we can also look at the three intersections: First, if I'm the attacker and I'm leveraging AI, what it mainly helps me to do is being much more efficient: By using AI I can create better and more authentic phishing emails. I can scour through larger amounts of data in shorter time, such as vulnerability reports or internet connections of that network where I want to get it. I can customize my malware so that it evades detection much more efficiently than if I could manually do it. Then secondly, on the defender side I can get much more efficiently through log data and it helps me to spot anomalies faster. AI almost always works an efficiency multiplier and will be used to go through vast amounts of data much faster and more reliable than a human could. A basic Tier 1 AI agent can run through your data, flag all behaviour that it thinks is anomalous and hand it over to human analysts. Just as if saying: "Look, this is what I found. What is your action? This is what I would suggest." And then the third element is the machine learning components in your IT infrastructure, which diversify and extend your attack surface as compared to if you wouldn't use it. It basically makes the defenders need defend another spot.

Elias Ricken:

So AI is more of an amplifier and something that renders these attacks more efficient, but that doesn't necessarily change them in the process or in the sequence of attacking?

Sven Herpig:

Yeah, I would say 80% to 90% of the cases, it just makes already existing procedures more efficient. There are about 10% to 20% of the cases, where using AI enables a new operational approach: Let's look at Microsoft recall for example, the program that takes screenshots of everything you're doing on you screen and later on helps you better to search and find different things on your computer through a machine learning-enabled application. These screenshots might entail your type of your password or your bank card information. Without the machine learning component used in Microsoft recall, this data

wouldn't exist. But when an attacker gets into your system and that program, they all of a sudden have direct access to this very sensitive amount of data and therefore it constitutes a new attack vector. There are certain operational aspects of using AI in cybersecurity that are changing in this three intersections that following the interview, but the main game is still the same.

Elias Ricken:

Would you say that, speaking from a perspective of protecting critical infrastructure, that are there certain sectors of critical infrastructure, such as energy infrastructure or hospitals, that are most vulnerable to these attacks?

Sven Herpig:

Well, I think in general, the attackability pre or post-leveraging AI remains the same. If your hospital is insecure before leveraging the tools, it will remain insecure when you're doing so and the other way around. For example, the banking sector is well known for high compliance and better securities than other critical infrastructures. It doesn't really change with AI because it follows the same pattern: If you know how to secure your conventional environment, you will also then know how to secure it against AI-supported operations. That follows the same paradigm. I think one thing that also makes sense to explain is that while these critical infrastructures might be using and will be using machine learning enabled software in the office environment, the vulnerable part where the critical operational things happen is the operational technology. Everything that moves, such as the industry robots and the MRT in the hospital for instance. Normally, these pieces of operational technology are the ones that you need to keep running in order to remain operational, because that's where the essential services are actually happening: It's not where the emails are written. The operational technology parts of most critical infrastructure so far are not yet leveraging machine learning in a critical sense. However, if you conduct an interview with people working in critical infrastructures and ask them "Are you using machine learning?", most would probably they would say "yes". And what they refer to is something like co-pilot or generative AI for their office environment. What they don't mean is that they have machine learning software running into their operational technology. If these are tampered with, their infrastructure will be severely disrupted. So I think that's another important distinction of where to look when we talk about protecting our critical infrastructure. If we just stick to the legal terms, a small hospital is not considered critical infrastructure. But if things don't work, there, then people might still die. A big hospital, is considered critical infrastructure, because it has more capacity and therefore holds more patients. Similarly as solid waste management: Some sites are considered critical infrastructure and it's sure bad if they wouldn't work for one or two days, but it's different than if an entire power grid goes out of operation for the same duration. Therefore critical infrastructure itself is a spectrum.

AI-enabled Cybersecurity:
Using artificial intelligence to detect, or exploit cyber threats

Elias Ricken:

Regarding these attacks, we know that they come from a certain direction of this world. My question would be, how can defenders, which are in our case, the European nations and companies keep pace with these adversaries who sometimes seem to have much larger means and much larger know on how to weaponize these AI tools in cyberdefence than we do. How can we react to that?

Sven Herpig:

I don't think it's going to be an AI versus AI match. I think it's cross-domain responses. First we have to define at what point we want to respond, either as an individual government or as the European Union. What responses are appropriate for what behaviour. When red lines are crossed, how do we react. That is interconnected, of course, with the geopolitical environment and international relations. For example within the current scale of cyber operations, if it's Chinese threat actors, that are after economic secrets or political espionage, there's only so much an individual state can do. Simply because China is a very big strategic competitor. Therefore a response is more likely to be tied

to the European Union. Most probably we would call out this kind of behaviour in a public attribution. That kind of response remains the status quo until today, with the level of Chinese interference. With Russia it's a different story. They're at war with a friendly country and regularly make active threats to wage war against the EU member states as well. In this political situation and the current scale of Russian cyber campaigns on European countries, there's nothing we would do as a response to cyber operations that we are not already doing to help Ukraine. We're going to do it either way because they're killing people in Europe and not because of something that happens in the cyber domain. Now, if Russia would shut down a power grid in Germany, for example. I think that scale of attack would warrant us to respond more severely. However, we haven´t reached that point before. Then we're talking about countries like North Korea and Iran, where we also at the maximum of sanctions that we can put on these countries. Regarding these countries, our additional geopolitical responses are pretty limited. EU-wide, we also need to get better secure our systems and be more resilient, to make sure that if our systems go down, we get them back running as fast as possible and as smoothly as possible. To this day, that likely remains the best strategy we have against strategic competitors. But for example, a couple of years ago, there was an espionage operation from Vietnam against Vietnamese citizens and dissidents in Germany. This is a case in which even a singular country can employ a significant toolset of responses, inside and outside the cyber domain, to make sure that they don't try that again. But to sum up, your toolset of responses largely depends on the international relations and the general geopolitical environment that we are in right now, as well as on how strong the competitor is that you're trying to respond to, and how serious the operation was that was carried out.

Elias Ricken:

You stated that cyberattacks range all across the EU and that in addition to that, the member states are very closely interlinked networking-wise. Wouldn't therefore the EU be the right framework to establish norms and to defend and respond to cyberattacks?

Sven Herpig:

Well, I think When we're talking about the strategic or the normative approach, the European Union is pretty close interlinked and it works well. But again, I don't think that really helps us to operationally defend ourselves. And operational defence is a matter of national security. National security is member state prerogative, especially in cybersecurity. We see that operational cooperation within the European Union is not working very well. You have instances where it works well, but normally these instances are either ad hoc or they are in smaller scale: For example, Germany decides to form a new group with the Netherlands, France and Luxembourg and they exchange a cybersecurity defence concept, with success. Then five years later, the European Union will decide to do that EU-wide and that every member state has to participate. Then the same concept doesn't work anymore because the level of trust to share sensible and security related data is not there in this larger EU-wide scale. On the other hand, there is fairly good reason for some of the mistrust: We have countries in the European Union that are not totally averse towards Russia and therefore I wouldn´t share my information about Russian cyber threats if I have certain European countries with me in the room. This complicates European Union-wide operational cooperation. At the same time, these very same countries speak with one voice at the UN for example, so they are very much political allies, but on the operational level, that is not so easily done. And that's where the difficulty with EU decision making in cybersecurity is: On one hand, operational cooperation is something that we need to get much better within the European Union, because this is also where our main strength, to pull our resources together, lies: On the other hand, we also need to acknowledge the realities. The best solution I believe, would be to have certain operational mechanisms in fixed groups only, so that we can rely on the people we are working with.

AI amplifies cyberattacks, making defenses more complex but not fundamentally changing them

Elias Ricken:

Remaining this European perspective or also the national security perspective, you just mentioned that cyberattacks are a matter of national security. And how do the cyber security strategies on a national level in the EU then address this convergence also of AI and cyberattacks?

Sven Herpig:

We do certainly not have a lack of strategy or legislation, for that matter. On the contrary, among European member states, we have the tendency to overregulate and overstrategize. Instead, we have an implementation problem. Allow me to give you an example: Shortly after the Federal elections in Germany, it was leaked that only 10% of the governments network operators have geographical redundancy of their data centres and know how to use backups and actually do tests with their backups. Still, this is regulated. Every strategy, everyone who has ever worked in IT security knows if we have big data centres running, they need to be redundant. If there is a break down, we need to have another one to take over. Even from private life we all know how important backups are, but still this leak shows that there is a lack of implementation in a major European country on a the federal level. And again, it is not the strategic level that is missing: We have best-practices, we have ISO-standards, we have strategies, we have regulations, we have all of that. Now we need to mass-operationalize a standard of IT security and that's where we need to catch up drastically.

Elias Ricken:

Why are we not doing exactly that? What reforms or political steps would be needed to implement these operational changes?

Sven Herpig:

For one part it is talent shortage. We face a six to seven-digit shortage in IT security personal in the European Union. ISACA, the Information Systems Audit and Control Association, has a good data on that. When we address a talent shortage in the EU, normally we start creating professorship positions and all that stuff. But we don't need people who can write encryption algorithms in 10 years from now. We need people who can configure firewalls and load balancers in six months' time. So addressing the talent shortage is also not something that we've done pretty well in the past, I think. The other reason is mostly in internal prioritization, such as political agenda setting and resources allocation. Security is not a very prestigious topic, because if it works, you won't see anything of it. You only see it when it fails. That is in my experience a lot of the political reason why we don't have enough resources for operational implementation of security standards. That is where our cybersecurity construct eventually fails. I mean, I've rarely talked to someone who responds, "Yes, my organization has enough people and who have the possibility to push through with security implementation of their measures." The responses mostly is somewhere around "We don't have people, we don't have money. It's not as important as other things for the company, institution or organization."

Elias Ricken:

SO what could potentially be the role of EU institutions such as the European Defence Agency or the EU Cyber Defence Competence Center. We already mentioned that it's foremost a matter of national security, but how can these institutions help the EU member states in operationalizing their defence in cyberspace?

Sven Herpig:

I think of two points: First they need to improve the ecosystem for talent. I mean, talent that between 6 and 18 months can actually work in companies and doing IT security. If that is a matter of changing the incentives or setting up more facilities that can train people is up to debate. I think that we have to figure out, but we need those people and we need them fast. How to do that is something that we might want to do actually on the European level as well. The second point is that EU-agencies such as ENISA, the European Cyber Security Agency, should figure out why operational cooperation is not working as well as we need it to work. We need to figure out why

it's not working the way it should be working. And we should address that whatever change will be part of that. And then the third point is that agencies like the European Defence Agency, need to bring the Ministries of Defence together and discuss two things right now: One is if we want to have joint active cyber defence operations? So cyber operations from European member states together to either disrupt the attacking infrastructure in, for example, Russia or whatever, and/or attribute those attacks better. The second question we need answered is: Do we want to have a European cyber command? Do we want to have a command structure on the European level? Right now, it would be NATO level. Do we want to have that command structure on the European level so that Europe is actually able to conduct their own cyber campaigns and cyber operations, especially, of course, with a focus on defence and resilience.

Elias Ricken:

Thank you, Mr Herpig for the interview.

epis ThinkTank

International Politics Shaped By **You**

EPIS Thinktank

Who We Are

EPIS is a young think tank on foreign affairs and security policy. We publish scientific articles, send members to international conferences, and maintain a network of: students & young professionals.

The deal:

- You professionalize yourself in your field
- We help you start your career

What We Do

EPIS Magazine

- In-Depth Analyses of Political Issues of Your Choice
- 80 Pages
- 3x/Year

EPIS Working Groups

- Monthly Briefings on Political Developments in Eight World Regions

EPIS Talks

- Deep Dive into the Articles of our Magazine with the Authors

EPIS Blog

- Short Analyses of Political Issues of Your Choice
- Weekly Release

Conclusion

President of the European Defence Network (EDN)

Brice Lefebvre is a Strategic Consultant for the French Aerospace and Defence industry, manager at STEP Consulting firm. Before that, with a background of aerospace engineering, he gained both technical experience in military programmes at Dassault Aviation and operational as tactical coordinator in the French Naval Aviation. He is also Co-founder and President of the European Defence Network, European-scale association for youth engagement and connection in defence.

EPIS Report on Artificial Intelligence & Cybersecurity

Dear reader,

As President of the European Defence Network (EDN), I am proud about the result of this first cooperation of our members with the EPIS Think Tank.

The EDN is a network of students and young professionals willing to engage actively with like-minded people in the field of defence and security with a European perspective. As youngsters, we witness the evolution of the world and technologies with a more acute sense of foresight and responsibility. The security environment in which we will evolve in our whole carrier and life is taking shape now. Therefore, having the opportunity to work and express our thoughts on such critical topics as AI and cybersecurity, which are only just starting to shape our future, was for us paramount to bring our stone to the debate and build our minds.

In that sense, AI covers it all. While its definition and potential are ever changing, what can only be said for certain is that it will revolution our security environment, and with all revolutions, the comfortable certainty of the past leaves the place to many questions to be answered. In the defence field, these questions are similar than in civilian applications, but are raised more acutely - with the specificity of not triggering another sudden "Oppenheimer moment" as for nuclear weapons but invading all the historical fields of weaponing. This raises new questions, that we have investigated in this report. At the core of AI is technology. Present in a basic and controlled form in guidance systems with image recognition, the first official use in a military operation in the 2010s in Israël for automatic targeting, and further developments until today's conflict in Ukraine, created a dynamic and a sense of urgency from which mainly newcomers benefited. This is an opportunity to move lines in the defence industry sector, and also for countries with weaker defence industrial bases to take their share in the European market. Technology developments also triggered the entry of those systems inside legacy human prerogatives, such as targeting decision. However, the inherent explainability gap of AI is a challenge for keeping those systems within the Rule of Law and humanitarian ethics in place in Europe. Hence our recommendation 3 on Explain ability. The potential accountability gap of Lethal Autonomous Weapon Systems is also conflicting with the principles of the International Humanitarian Law. Work has started to update regulatory and legal frameworks, but the standard pace for diplomacy and consensus is not compatible of the evolution pace of AI. On top, the conflicting geopolitical landscape is not auspicious for common ground. This is why I would like to open, in this conclusion, on the strong need for European leadership on this new field, to set the new standards while allowing innovation and breakthrough to secure our future defence capabilities. This can only be done by recognising the specificities of the defence applications, and potentially preparing exceptions for high-intensity conflicts.

We are confident this report, fruitfully born from this rich collaboration with the EPIS Think Tank, has brought some piece or opened new paths in the debate. This shows our members are skilled and engaged, and is a sign of hope for what the future brings!

Brice Lefebvre,
President of the European Defence Network


episthinktank
EPIS ThinkTank | Foreign- & Security Policy
409 posts
757 followers
65 following
Community
Thinktank & Network for Students
Connect with Experts & Publish your Articles
Based in
Join EPIS Now
See Translation
www.epis-thinktank.com and 4 more
Follow
Message
Email
Delegations
W.Groups
Basics
Blog
Donate
About us
Network
Working Groups
Formats
Report Groups
EPIS Thinktank
The Think Tank for Foreign- and Security Policy
Berlin Security Conference 2024
What is Happening to People in El Salvador?
Announcement
Informal Talks
Dr. Johann Schmid

Imprint

Editor-in-chief: Belen Alondra Bringas Machicado
ViSdP: Theodor Himmel
Publisher: EPIS ThinkTank e.V.
Contact: board.external@epis-thinktank.com
ISSN: 2944-747X

Are you interested in our work?

Epis ThinkTank e.V. welcomes your support. As a member, author or supporter you can get involved. We have been participating in the political debate for several years. As an assosiation, we are young academics committed to factbased and neutral debate. Our members come from all over Germny and the world.

Find out more on: www.epis-thinktank.de

or visit us on: