

Task Force 1: Peace, Security and Global Governance



G7-led Principles for Military AI Governance: Imperatives and Pathways

Sharinee L. Jagtiani, Senior Officer for AI and Democracy, GMF Tech Programme, German Marshall Fund of the United States

Ester Sabatino, Research Associate for Defence and Military Analysis, International Institute for Strategic Studies (IISS), United Kingdom

Abstract

Artificial intelligence (AI) is progressively permeating every sector, ushering in a transformative era for technological applications. In the military domain, AI is hailed as the next frontier. From the battlefield to the boardroom, countries are fervently pursuing advanced AI technology, as AI applications seem to provide militaries with enhanced analytic capabilities, bolstering operational efficiency. Despite various recent initiatives to regulate the application of AI technologies, a global framework to govern AI in the military conspicuously absent. The lack of a common regulatory framework raises concerns about a potentially dangerous race to the bottom. This could lead to the deployment of progressively more destructive systems with insuf-

efficient restrictions. This policy brief strongly advocates for an augmented role of the G7 in shaping the foundation for shared principles for AI-enabled military technology. Establishing a unified regulatory framework at the G7 is imperative to tackle the humanitarian and ethical concerns posed by the use of AI-enabled military technologies.

The challenge

Artificial intelligence (AI) with its growing capabilities and applications, is reshaping almost every facet of society. In the military domain, AI is poised to play a pivotal role not just in downstream tasks such as target selection and engagement but in various upstream tasks that include logistics and decision making (Grand-Clément 2023). A global framework to govern AI in the military is, however, conspicuously absent. Ideally, such a framework would oversee every facet of AI in military operations, from its conceptualization and development to deployment and utilisation. The lack of this framework presents a significant challenge to international peace and stability.

Unlike other more conventional technologies, AI in the military is challenging to regulate as it is a general purpose dual-use technology. Unlike discrete weapons like submarines or gas, determining what qualifies as ‘military AI’ is ambiguous. Attempts to ban AI outright would be akin to prohibiting industrialization in the 19th century – impractical and unlikely.

Military applications of AI often intersect with civilian uses, such as predictive analytics or image processing, blurring the line between acceptable and unacceptable developments in AI technology (Sharre and Lamberth 2022). Clarity is crucial for any effective regulatory framework, but delineating these boundaries is complex. Indeed, military deployment of AI encompasses more than just lethal autonomous weapon systems; it includes operational algorithms designed to aid decision-making, relating to everything from human resource management within the military, to collection and fusion of intelligence, surveillance, and reconnaissance data (Vergun 2023).

Entrusting battlefield decisions to AI can alter the character of war in serious and profound ways.

First, AI-driven systems could accelerate the tempo of combat, potentially outpacing human ability to react (Puscas 2023: 44; Forces News 2024).

Second, the advent of military AI could fundamentally alter the psychology of warfare. With military capabilities becoming increasingly reliant on software-driven systems, decision-makers may find it challenging to accurately assess relative military strengths. Traditionally, understanding adversaries’ behaviour has relied on knowledge of their capabilities and intentions. However, as this judgment is increasingly delegated to AI, whose calculations are opaque, the shift towards AI could erode policymakers’ understanding in both areas. Consequently, this could undermine their comprehension of adversaries’ behaviour, thereby heightening the risk of miscalculation.

(Horowitz and Scharre 2021).

Third, military AI could also alter their risk thresholds, as the perceived precision and reduced risk to human lives afforded by AI systems could encourage more aggressive behaviour (Horowitz and Scharre 2021).

Fourth, algorithms inherently encode historical biases and constraints, whether due to biased training data, flawed assumptions, or other factors. Transplanted into a conflict environment, such systems may contribute to dangerous inaccuracies, such as not identifying a person as human due to the biased data it was trained on (Mohan 2023).

At present, military AI is not merely theoretical; it is already in use on contemporary battlefields, such as those in Ukraine and Gaza (Feldstein 2024). Experts have touted Ukraine as a 'laboratory for AI Warfare' where AI systems are constantly being tested and fine-tuned for immediate deployment (Fontes and Kamminga 2023). In Gaza, the Israeli army has deployed AI technology to intercept enemy drones that regard every soldier – even a blind one – as a sniper (France 24 2024). The use of AI-enabled military systems poses questions on the ethical aspects of conflicts. Examples in the current Israel-Hamas conflict has shown that AI might have implications on the definition of acceptable collateral damages and on the tempo for human oversight of the kill-chain (Fraser 2024).

Among the G7 countries, Canada, France, the United Kingdom and the United States have specific national military AI strategies that in some cases are accompanied by ethical defence ministerial committees to verify a proper use of technology. Nonetheless, specific military AI strategies are lacking in the remaining members of the Group, who have developed and updated their general national AI strategies. The lack of a collective *approach* among the G7 countries on military AI poses additional challenges on the feasibility to reach consensus at the wider international level.

Without a common framework and guardrails in the military field, states risk a dangerous race to the bottom, deploying successively more destructive systems with scant restrictions (Feldstein 2024). The current geopolitical landscape is not conducive to regulating military AI. The escalating strategic rivalry between the US and China has spilled over into the realm of AI, as both nations compete for dominance in AI capabilities, extending into military applications (Csernatonni 2024). At the UN level, Russia and China have proposed guidelines on AI that are not supported by the wider international community. One of the problems in finding agreement on the basis of the proposed guidelines is that divergences of the practical meaning of trustworthy AI may differ among actors (Alberque 2024). Further, the hype around AI in the military implies that states may be less willing to meaningfully regulate this technology or to create international standards verification mechanisms.

Some state-led initiatives, however, have attempted to regulate military AI. In February 2023, the Netherlands hosted the Summit on Responsible Artificial Intelligence in the Military Domain (REAIM). At the Summit, the US launched the Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy. While these initiatives are welcome, it is important that a global framework for the regulation of military AI is rooted in a multilateral approach.

The application of AI in military operations must adhere to relevant international legal standards. Specifically, its use in armed conflicts must align with states' commitments under international humanitarian law, encompassing its core principles. Accountability is essential in employing AI capabilities, ensuring oversight through responsible human command and control structures during military activities. A principled approach to the military implementation of AI involves thorough assessment of risks and advantages, while also mitigating unintentional biases and potential accidents (US Department of State 2023).

Since 2016, the G7 has been proactive in setting principles for AI governance (Japan Ministry of Internal Affairs 2016). At present, the 2019 OECD principles serve as the basis for a global standard for AI regulation, which were also endorsed by the G20 in 2020. Now, the G7 has the opportunity to lead in establishing a unified framework of principles for military AI.

The role of the G7

The G7 has been proactive on the question of responsible AI. It is now imperative that the G7 creates the foundations for a consensus on responsible military AI.

In the establishment of a framework for shared principles for military AI, the G7 can build on its strong track record on proposing AI regulation. The need to regulate AI and facilitate the adoption of emerging technology was first mentioned in a 2016 Joint Declaration (G7 2016a). The focus on AI included in the Declaration and the accompanying Opportunities for Collaboration annex failed, however, to specify principles for a correct and sustainable development and use of AI (G7 2016b). Nonetheless, these documents laid the ground for the first Joint Declaration on AI endorsed in 2018, following a first G7 Innovation Ministers' Statement on Artificial Intelligence in the same year (G7 Research Group 2018).

The Charlevoix Common Vision for the Future of Artificial Intelligence defines a first set of commitments for a responsible development, promotion and use of AI in all sectors (G7 2018). Efforts in identifying Common Principles and Code of Conducts were further strengthened during Japan's Presidency in 2023, which established the Hiroshima AI Process Comprehensive Policy Framework.

The framework includes International Guiding Principles for all AI sectors and a voluntary Code of Conduct for organisations developing advanced AI systems. It is a valuable instrument to influence the design, development, deployment and use of advanced AI. Crucially, the principles advance the OECD's 2019 AI principles.

While the Hiroshima Framework is an important starting point for regulating AI in the military domain, it does not contain any specific language on the use of AI in defence. This is also the case with the Guiding Principles and the Code of Conduct. Specifically, the security-oriented language in these documents centre on malicious actors in suggesting that AI systems should not “facilitate terrorism, promote criminal misuse, or pose substantial risks to safety, security and human rights” (G7 2023).

The G7 Industry, Technology and Digital Ministerial Meeting of early March 2024 re-committed G7 countries to achieve “an appropriate balance between fostering innovation and the need for appropriate guardrails” (G7 2024). This signalled an appetite in member states to regulate AI technology. At the same time, the Declaration did not include any reference to regulation modalities of AI in defence. The threat posed by an unregulated use of AI in the military, in terms of diminished accountability, dehumanisation of conflict, but also as a tool to influence and destabilise the international geopolitical balance, should lead the G7 to address the challenge more purposefully and specifically.

The current Italian Presidency of the G7 has the opportunity to advance the status of AI regulations, by updating the Hiroshima Framework, creating new toolkits for its implementation, and promoting interoperability of regulatory systems from different organisations. The G7 holds a distinctive position in the global landscape and its economic, technological, and military power render the Group a valuable framework for consensus-building. Moreover, the group has evolved since its inception, from a forum for economic coordination to a broader platform for addressing global challenges, making the group the best interlocutor to advance and define principles on the development and use of military AI.

Militarily, the G7 countries spent ca. 1.208,6 billion US dollars in 2023 (IISS 2024), which represents almost 55 per cent of the global defence spending (IISS 2023). The defence companies established in their territories are also among the strongest companies in the sector for military export worldwide (Liang et al. 2023; Wezeman et al. 2024), thus putting the Group at a privileged position for the definition of military technology related regulation.

By coordinating their approaches to the use of AI in defence, G7 nations can establish common benchmarks and requirements for the responsible development and deployment of military AI systems, reducing the risk of proliferation of potentially harmful AI technologies. Given the G7's success in helping establish frameworks for international arms control (Higashi 2020) the forum has the potential to achieve positive results in the realm of military AI.

In sum, a G7-sponsored approach on military AI would first involve fostering a consensus between member states on its responsible and ethical use. Additionally, a G7 approach would increase the likelihood of other countries joining the framework. Six out of the seven countries in the Group are also members of NATO, and three of them are among the founding members of the EU. Their membership in these international organisations gives them an advantage of like-mindedness, as they can promote the adoption of similar, if not stronger, measures within their respective circles.

Policy recommendations

The Italian G7 Presidency represents an important opportunity to create a framework for establishing shared principles underpinning military AI. Possible steps include:

Defining shared principles of military AI through a mission-oriented task force: The G7 should establish a task force to reinforce the emerging global consensus on the principles to guide responsible military AI in line with international humanitarian law. As a basis for developing such principles, the task force should aim to first agree on a definition of military AI that should include both upstream and downstream tasks. Second, agree on a framework that demarcates the risks presented by these technologies so that there is a clearer consensus on what countries are regulating against. It can also conduct AI impact assessments to evaluate the ethical and social implications of AI technologies on the battlefield, particularly on aspects related to collateral damage and the role of human control over AI systems, i.e. ‘humans in the loop.’ The task force can operate in a mission-oriented manner under a one-year sunset clause.

Mapping military AI based on multi-stakeholder input: The work of the task force should be supported by a process for mapping current and potential deployment of AI technologies in the military. This could take the form of a “Responsible AI in the Military” multi-stakeholder forum taking place in the context of a G7 task force meeting and involving technology companies, members of the technical community, and civil society to provide input on priorities and challenges. The G7 should also consider inviting government officials from selected partner countries to task force meetings. These should include like-minded or strategic partners with whom there are strong social, economic, technological, and military ties. A collaborative approach along these lines could ensure that the G7 is constantly updated on new and developing applications of military AI and that they do not escape existing governance mechanisms and oversight.

Supporting the development of standards for the control of military AI: Following the establishment of shared standards on military AI, the G7 should undertake further consensus-building activities among like-minded nations with a focus on technology control for non-proliferation. While AI technology is inherently dual-use, its development and use in conflict must be regulated. Firstly, AI systems should be developed in a trustworthy and secure manner,

with access to relevant technology limited for international actors not adhering to future guiding principles on military AI. Secondly, AI-enabled military systems alter the dynamics of military escalation and its perception, necessitating rules on signalling escalation of force by AI systems. The G7 could facilitate consensus on these issues and adopt supporting documents to pave the way for widespread adoption of control and verification mechanisms for military AI. Such measures would help prevent inadvertent escalation during peacetime or crises.

References

Alberque, William. 2024. Artificial intelligence controls with Simona Soare. *Arms Control Poseur* podcast 18 April. <https://www.iiss.org/podcasts/arms-control-poseur/2024/04/artificial-intelligence-controls-with-simona-soare>

Csernaton, Raluca. 2024. Charting the geopolitics and European governance of artificial intelligence. *Carnegie Papers* March. <https://carnegieendowment.org/publications/91876>

Feldstein, Steven. 2024. AI in war: Can advanced military technologies be tamed before it's too late? *Bulletin of the Atomic Scientists* 11 January. <https://thebulletin.org/?p=109741>

Fontes, Robin, and Kamminga, Jorrit. 2023. Ukraine a living lab for AI warfare. *National Defense* 24 March. <https://www.nationaldefensemagazine.org/articles/2023/3/24/ukraine-a-living-lab-for-ai-warfare>

Forces News. 2024. Top British general explains how AI is aiding military intelligence. *YouTube* 5 February. <https://youtu.be/cIaTpwW6t-Q>

France 24. 2024. Israel deploys new military AI in Gaza war. *France 24* 10 February. <https://www.france24.com/en/live-news/20240210-israel-deploys-new-military-ai-in-gaza-war>

Fraser, Callum. 2024. AI's baptism by fire in Ukraine and Gaza offer wider lessons. *Military Balance Blog* 22 April. <https://www.iiss.org/online-analysis/military-balance/2024/04/analysis-ais-baptism-by-fire-in-ukraine-and-gaza-offer-wider-lessons>

G7. 2016a. *Joint declaration by G7 ICT ministers*. <https://www.g8.utoronto.ca/ict/2016-ict-declaration.html>

G7. 2016b. *Annex: G7 opportunities for collaboration*. <https://www.g8.utoronto.ca/ict/2016-ict-annex.html>

G7. 2018. *Charlevoix common vision for the future of artificial intelligence*. <https://www.g7.utoronto.ca/summit/2018charlevoix/ai-commitment.html>

G7. 2023. *Hiroshima process international code of conduct for organizations developing advanced AI systems*. <https://www.g8.utoronto.ca/summit/2023hiroshima/231030-ai-code-of-conduct.html>

G7. 2024. *Ministerial declaration: G7 Industry, Technology and Digital Ministerial Meeting Verona and Trento 14-15 March*. <https://www.g7italy.it/wp-content/uploads/G7-Industry-Tech-and-Digital-Ministerial-Declaration-Annexes-1.pdf>

G7 Research Group. 2018. *Annex B: G7 innovation ministers' statement on artificial intelligence*. <https://www.g7.utoronto.ca/employment/2018-labour-annex-b-en.html>

Grand-Clément, Sarah. 2012. *Artificial intelligence beyond weapons: Application and impact of AI in the military domain*. Geneva: UNIDIR. <https://unidir.org/?p=14212>

Higashi, Hiromitsu. 2020. *G7 performance on governing arms control*. The Global Governance Project. <https://www.globalgovernanceproject.org/?p=5501>

Horowitz, Michael, and Scharre, Paul. 2021. AI and international stability: Risks and confidence-building measures. *CNAS Reports* January. <https://www.cnas.org/publications/reports/ai-and-international-stability-risks-and-confidence-building-measures>

IISS. 2023. *Military Balance Plus* database [accessed on 19 March]. <https://www.iiss.org/the-military-balance-plus>

IISS. 2024. Chapter one: Defence and military analysis. *The Military Balance* 124(1): 8-15. <https://doi.org/10.1080/04597222.2024.2298589>

Japan Ministry of Internal Affairs. 2016. *G7 ICT ministers' meeting*. https://www.soumu.go.jp/joho_kokusai/g7ict/english/about.html

Liang, Xiao, et al. 2023. The Sipri top 100 arms producing and military services companies, 2022. *SIPRI Fact Sheets* December. <https://doi.org/10.55163/UJNP6171>

Mohan, Shimona. 2023. Managing expectations: Explainable A.I. and its military implications. *ORF Issue Briefs* 570. <https://www.orfonline.org/research/managing-expectations-explainable-a-i-and-its-military-implications>

Puscas, Ioana. 2023. *AI and international security. Understanding the risks and paving the path for confidence building measures*. Geneva: UNIDIR. <https://unidir.org/?p=14365>

Sharre, Paul, and Lamberth, Megan. 2022. Artificial intelligence and arms control. *CNAS Reports* October. <https://www.cnas.org/publications/reports/artificial-intelligence-and-arms-control>

US Department of State. 2023. *Political declaration on responsible military use of artificial intelligence and autonomy*. <https://www.state.gov/political-declaration-on-responsible-military-use-of-artificial-intelligence-and-autonomy-2>

Vergun, David. 2023. U.S. endorses responsible AI measures for global militaries. *DOD News* 22 November. <https://www.defense.gov/News/News-Stories/Article/Article/3597093>

Wezeman, Pieter D., et al. 2024. Trends in international arms transfers, 2023. *SIPRI Fact Sheets* March. <https://doi.org/10.55163/PBRP4239>

About Think7

Think7 (T7) is the official think tank engagement group of the Group of 7 (G7). It provides research-based policy recommendations for G7 countries and partners. The Istituto Affari Internazionali (IAI) and Istituto per gli Studi di Politica Internazionale (ISPI) are the co-chairs of T7 under Italy's 2024 G7 presidency.