sipri

RESPONSIBLE MILITARY USE OF ARTIFICIAL INTELLIGENCE

Can the European Union Lead the Way in Developing Best Practice?

VINCENT BOULANIN, NETTA GOUSSAC, LAURA BRUUN AND LUKE RICHARDS

STOCKHOLM INTERNATIONAL PEACE RESEARCH INSTITUTE

SIPRI is an independent international institute dedicated to research into conflict, armaments, arms control and disarmament. Established in 1966, SIPRI provides data, analysis and recommendations, based on open sources, to policymakers, researchers, media and the interested public.

The Governing Board is not responsible for the views expressed in the publications of the Institute.

STOCKHOLM INTERNATIONAL PEACE RESEARCH INSTITUTE

Signalistgatan 9
SE-169 70 Solna, Sweden
Telephone: +46 8 655 97 00
Email: sipri@sipri.org
Internet: www.sipri.org

About the authors

Dr Vincent Boulanin (France) is a Senior Researcher leading SIPRI's research on emerging military and security technologies. His focus is on issues related to development, use and control of autonomy in weapon systems and the military applications of artificial intelligence (AI). He regularly presents his and SIPRI's work and engage with governments, UN bodies and international organizations, research institutes and the media. Before joining SIPRI in 2014, he completed a doctorate in political science at the École des Hautes Études en Sciences Sociales, Paris. His recent publications include *Artificial Intelligence, Strategic Stability and Nuclear Risk* (SIPRI, 2020 co-author), *Limit on Autonomy in Weapon Systems: Identifying Practical Elements of Human Control* (SIPRI/ICRC, 2020 co-author).

Netta Goussac (Australia) is an Associate Senior Researcher within the Armament and Disarmament research area with particular expertise in legal frameworks related to the development, acquisition and transfer of weapons. Before joining SIPRI in 2020, she worked as an international lawyer for over a decade, including for the International Committee of the Red Cross (2014–20) and the Australian Government's Office of International Law (2007–14), and as a lecturer at the Australian National University. She has provided legal and policy advice related to new technologies of warfare, including autonomous weapons, military applications of AI, and cyber and space security. Since 2017, Netta has participated in the United Nations' Group of Governmental Experts on lethal autonomous weapon systems. Her recent publications include, *Limit on Autonomy in Weapon Systems: Identifying Practical Elements of Human Control* (SIPRI/ICRC, 2020 co-author); and 'Safety net and tangle web: legal reviews of AI in weapons and warfighting', Humanitarian Law and Policy Blog, ICRC, 18 Apr. 2019.

Laura Bruun (Denmark) is a Research Assistant working on emerging military and security technologies. Her focus is on how emerging military technology, notably autonomous weapon systems, affects compliance with—and interpretation of—International Humanitarian Law. She has a background in both Middle Eastern Studies and International Security and Law. Before joining SIPRI, Laura worked with Airwars in London, where she monitored and assessed civilian casualty reports from US and Russian airstrikes in Syria and Iraq.

Luke Richards (United Kingdom) is a Research Assistant working on emerging military and security technologies. His current focus is on the responsible innovation and ethics of AI alongside broader technology governance issues. Prior to joining SIPRI, he worked at the International Institute for Strategic Studies (IISS) while finishing an MSc in Science and Technology Policy, writing his master's dissertation on the 'The Civil-Military Entanglement of Global Innovation'. He has previously worked on diverse projects such as the development of a methodology to rank the cyber power of states through to the security implications of human enhancement.

RESPONSIBLE MILITARY USE OF ARTIFICIAL INTELLIGENCE

Can the European Union Lead the Way in Developing Best Practice?

VINCENT BOULANIN, NETTA GOUSSAC, LAURA BRUUN AND LUKE RICHARDS

STOCKHOLM INTERNATIONAL
PEACE RESEARCH INSTITUTE

November 2020

Contents

Acknowledgements

This report was produced with the generous support of the German Federal Foreign Office. It is part of a research project on 'Governing the Opportunities and Risks of AI for International Peace and Security', which seeks to provide input on the topic of governance of military artificial intelligence in the context of Germany's reflection as part of the German presidency of the Council of the European Union on 'Rethinking Arms Control'.*

The authors are indebted to all the participants who shared their knowledge and experience under Chatham House Rule at the SIPRI online workshop held on 8–9 September 2020 on 'Governing the risks and opportunities of AI for international peace and security: what role for the EU?'.

The authors wish to thank the anonymous peer reviewer and SIPRI colleagues Dr Sibylle Bauer, Kolja Brockmann, Dr Lucie Béraud-Sudreau and Giovanna Maletta for their comprehensive and constructive feedback. Finally, we would like to acknowledge the invaluable work of the SIPRI Editorial Department.

The views and opinions in this report are solely those of the authors and do not represent the official views of SIPRI or the funder. Responsibility for the information set out in this report lies entirely with the authors.

Vincent Boulanin, Netta Goussac, Laura Bruun and Luke Richards

* For information on the rethinking of arms control initiative, see German Federal Foreign Office, '2020. Capturing Technology. Rethinking Arms Control', 2020. For information on Germany's presidency of the European Union, see 'Germany's presidency of the Council of the European Union'.

Executive summary

During the Finnish Presidency of the Council of the European Union (EU) in 2019, several EU member states went on record calling for greater national deliberation and collaboration between member states on the topic of artificial intelligence (AI) in defence. This report explores in that regard whether it is politically desirable and possible for the EU and its member states to propose a more concerted and identifiable perspective on the responsible military use of AI, from a legal, ethical and technical safety standpoint. The three main findings can be summarized as follows.

First, the EU and its member states would benefit politically, strategically and economically from developing principles and standards for the responsible military use of AI—although it will be a difficult process involving juggling competencies, political sensibilities and wills of the various EU actors (institutions and member states). From a political standpoint it would allow them to gain traction in the global discussion on the military use of AI and its governance and effectively defend a vision that is aligned with their shared values and interests. From a strategic standpoint it would help them maintain some autonomy regarding the military technology they want to use and how they want to use it. From an economic standpoint it would benefit the development of the European military-industrial base and operational capabilities in the field of AI.

Second, the groundwork for the development of a European view on responsible military use of AI has already been laid. The work already conducted by the European Defence Agency (EDA) and the European Parliament, as well as several EU member states, demonstrates a growing political interest for the topic. The work initiated by the European Commission on trustworthy AI provides a useful, substantial basis for engaging in conversations on the responsible military use of AI. In that regard, the report provides a series of concrete proposals for EU institutions and EU member states to consider openly, methodically and collaboratively developing a joint perspective on the responsible military use of AI. The report notably suggests that the EU and its member states should hold three tracks of expert discussions on (*a*) legal compliance focusing on legal reviews and challenges posed by AI in the interpretation and application of international law; (*b*) ethics, which focus on the development of shared principles for human-centric military use of AI and (*c*) AI safety, which focus on standards of transparency, explainability and reliability of AI systems.

Third, there are multiple ways for the EU and its member states to carry out these expert discussions and discuss responsible military use of AI in general. The EDA already provides a forum for both high-level political discussions between defence ministers of the member states and technical deliberations between national experts on military AI. The EDA's work on the development of AI capability could be expanded to explicitly include issues related to legal compliance and ethics. The preparatory bodies of the Council of the EU, such as the EU Council's Working Party on Public International Law (COJUR), the Working Party on Non-Proliferation (CONOP), the EU Council's Working Party on Global Disarmament and Arms Control (CODUN) and the EU Council's Military Committee Working Group (EUMCWG), also provide forums for states to share their perspectives, coordinate their views and potentially reach consensus on principles and best practices for compliance with international law and ethics. Further, the European Parliament provides a forum for public democratic discussion on fundamental ethical issues raised by using AI in the military. Finally, the European Commission, could use the European Defence Fund (EDF) to fund projects that explore ways of meeting safety standards for AI at a technical level.

In light of the above, this report's main recommendations for consideration of EU member states and EU institutions are the following:

- Engage more openly and transparently in national and intra-European deliberations related to the opportunities and risks of AI for the military. For instance, make national reports and strategic documents or summaries of their key content publicly available.
- Create an ad hoc expert group, made up of national experts, to discuss how the High-Level Expert Group on AI (AI HLEG) recommendations may apply in the military context, with the aim of proposing European principles and guidelines for the responsible development and use of AI in the military.
- Engage with academia and industry and use the European AI Alliance to engage broader society, the business sector and civil society on issues related to the development, use and control of military AI.
- Fund research projects on end-to-end 'ethical design' or methodologies that would allow for ethical issues to be considered throughout the cycle of technology or capability; the transparency, explainability and traceability of military AI systems; the development of a European framework for testing and evaluating military AI systems; the design of methodologies for pooling data at the EU level according to key ethical principles.

The report also recommends to use the preparatory bodies of the Council of the EU to foster deeper intra-European discussion by organizing special meetings on:

- Questions of interpretation and applications of international humanitarian law (IHL) with respect to military uses of AI, which could include conducting legal reviews of weapons, means and methods of warfare within COJUR.
- Possible ethical principles and safety guidelines for military use of AI by European armed forces via the EUMCWG.
- The EU's position on human control in the debate on lethal autonomous weapon systems (LAWS) in Geneva and other arms control forums while military use of AI could be discussed within the framework of CODUN/CONOP.

Abbreviations

AI	Artificial intelligence
AI HLEG	High-Level Expert Group on AI
CCW	Convention on Certain Conventional Weapons
CODUN	EU Council's Working Party on Global Disarmament and Arms Control
COJUR	EU Council's Working Party on Public International Law
CONOP	Working Party on Non-Proliferation
DG DEFIS	European Commission's Directorate General for Defence Industry and Space
EDA	European Defence Agency
EDF	European Defence Fund
EU	European Union
EUGS	EU Global Strategy
EUMCWG	EU Council's Military Committee Working Group
GGE on LAWS	Group of Governmental Experts on emerging technologies in the area of lethal autonomous weapon systems
IHL	International humanitarian law
JCR	Joint Research Centre
LAWS	Lethal autonomous weapon systems
PESCO	Permanent structured cooperation

1. Introduction

AI has become the focus of competition between the world's great powers, notably between the United States and China.[1] During the Finnish Presidency of the Council of the EU in 2019, several EU member states went on record calling for greater national deliberation and collaboration between member states on the topic of AI in defence.[2] In 2020 the European Parliament has also been deliberating over the role that the EU should play in the governance of military AI, and the idea that the EU should aim to become a role model in that area, as it already is a proponent of responsible use of AI in the civilian sphere.[3] The question of whether it is in the interest of the EU and its member states to work more intensely, collaboratively and openly towards the development of common principles on the responsible military use of AI is at the core of this report.

This report is intended to serve as a resource for the German Presidency of the Council of the EU and the ongoing initiative of the German Federal Foreign Office on 'Rethinking Arms Control'. It explores why the EU and its member states would benefit politically, strategically and economically from developing principles and standards for the responsible military use of AI. The report maps out what has already been done on the topic and how further expert discussion within the EU on legal compliance, ethics and technical safety could be conducted.

The report is based on the authors' desk research, a series of background interviews and the authors' takeaways from an online workshop organized by SIPRI, which included scholars and practitioners from the EU and other international organizations, including the North Atlantic Treaty Organization and the United Nations Office for Disarmament Affairs.

The report aims to inform and provide practical and actionable recommendations to practitioners and decision makers from the EU and its member states who work on issues of military capability development, arms control and technology governance. In order to do so, the report first discusses the rationale for the EU and its member states to engage more concertedly in the global conversation on responsible military use of AI (chapter 2). Then it outlines how the EU and its member states could lead the way in terms of best practice, focusing on three areas: legal compliance, ethics and technical safety (chapter 3). It concludes by summarizing the key findings and recommendations (chapter 4).

[1] Roff, H. M., 'The frame problem: the AI "arms race" isn't one', *Bulletin of the Atomic Scientists* vol. 75, no. 3 (May 2019), pp. 95–98; and Atherton, K. D. 'We are on the verge of a no-win AI arms race, warns NGO', C4ISRNET, 9 May 2019.

[2] See the food-for-thought paper by Finland, Estonia, France, Germany and the Netherlands, 'Digitalization and artificial intelligence in defence', 17 May 2019.

[3] European Parliament Committee on Legal Affairs, 'Opinion of the Committee on Foreign Affairs for the Committee on Legal Affairs on artificial intelligence: questions of interpretation and application of international law insofar as the EU is affected in the areas of civil and military uses and of state authority outside the scope of criminal justice', 2020/2013(INI), 9 July 2020; and High-level Expert Group on Artificial Intelligence (AI HLEG), 'Ethics guidelines for trustworthy AI', 8 Apr. 2019.

2. The European Union and the responsible military use of artificial intelligence

Why should the EU and its member states work towards a more concerted position on what constitutes responsible military use of AI? Given the EU's limited competencies in the fields of armament and arms control, it is legitimate to question the relevance of bringing the issue of responsible military use to the EU level. However, this chapter finds that it would be in the interest of the EU and its member states for three reasons. First, it would allow EU member states to weigh in more heavily in the global debate on governance and AI, while ensuring that the military use of AI develops in a way that is aligned with their shared values. Second, it would support the EU's ambition to foster greater interoperability between EU armed forces and improve European collaboration in military research and development.[4] Third, EU member states have an economic need to collaborate in the development of next-generation military technologies, which will likely have to rely on AI.

I. The alignment and promotion of values of the European Union and member states at the global level

The USA and China are currently shaping the way militaries across the globe perceive the future military use of AI. Their leadership is the result of high levels of investment in AI specifically, and in military technology in general.[5] Over the last decade or more, they have also dedicated a significant amount of resources to conceptualizing how advances in AI could and should be leveraged to their military advantage, at the strategic, operational and tactical levels.[6] Such thinking includes considerations of the impact of AI on their relative technological superiority, their ability to lower the risks to their personnel and military assets, and the possibility of processing more information more quickly than their adversaries.[7]

For the EU and its member states, the fact that the USA and China are racing ahead could be problematic in at least two regards. First, it undermines the ability of EU member states to develop a genuine view on why and how their militaries should use advances of AI. There is a risk that the USA or China could pre-shape the views of the EU member states, and consequently those of the EU.[8] Second, and more importantly, there is the risk that the USA, China and other major actors, such as Russia, could adopt military AI policies, doctrines or uses that challenge the EU's values or interests. One concrete scenario is that the USA and China, as a result of their reciprocal anxiety over each other's capability in the field of AI, could engage in a negative spiral that challenges the EU and its member states' views of ethically acceptable military use of

[4] European Commission, *European Defence Action Plan*, COM(2016) 950 Final, (European Commission: Brussels, 2016).

[5] In 2018 the top 2 500 software and computer service firms in China and the USA spent $11.8 billion and $77.4 billion respectively on R&D in general compared to $10.1 billion in the EU. Both China and the USA dwarf the EU in raising money from venture capital and private equity. For an insight into levels of investment in AI and related areas, see Castro, D., McLaughlin, M. and Chivot, E., 'Who is winning the AI race: China, the EU or the United States?', Center for Data Innovation, 19 Aug. 2019; in 2019 the USA spent $732 billion and China $261 billion on military expenditure. For further details see SIPRI Military Expenditure Database, <https://www.sipri.org/databases/milex>.

[6] For an overview of China's thinking, see Kania, E. B., 'Battlefield singularity: artificial intelligence, military revolution, and China's future military power', Center for a New American Security, 28 Nov. 2017; for the USA, see Sayler, K. M., 'Artificial intelligence and national security', Congressional Research Service (CRS) Report for Congress R45178 (US Congress, CRS: Washington, DC, 26 Aug. 2020); and Boulanin, V. et al., *Artificial Intelligence, Strategic Stability and Nuclear Risk* (SIPRI: Stockholm, 2020).

[7] Work, B., 'Remarks by Deputy Secretary Work on Third Offset Strategy', US Department of Defense, 28 Apr. 2016; Allen, G. C., 'Understanding China's AI strategy', Center for a New American Security, 6 Feb. 2019.

[8] Franke, U. E., 'Not smart enough: the poverty of European military thinking on artificial intelligence', European Council on Foreign Relations, 18 Dec. 2019.

AI.[9] This could cause the EU to revisit its demand for human control of AI. A position the EU has stated on several occasions at the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapon Systems (GGE on LAWS) and is currently a major point of consensus within the EU.[10]

In other words, it is in the interest of EU member states to take a proactive approach rather than a reactive one to what they deem appropriate standards for responsible military use of AI. Some EU member states have already considered this topic in detail.[11] Their contribution to the UN debate on LAWS reflects their level of engagement and the maturity of their views (see box 2.1). The deliberations in Geneva, however, have revealed that a few EU member states championing proposals might not be sufficient to influence the position or actions of the USA or China regarding the military use of AI. Moreover, despite the apparent consensus in the EU bloc, some differences remain in the way EU member states perceive the opportunities and risks associated with the military use of AI.[12] Arguably, further intra-European discussion on the topic of responsible AI, or related topics, such as the regulation of LAWS, could allow the EU to present a more coherent—hence robust and influential—front in international discussions related to the military use of AI (see chapter 3).

II. Strategic autonomy, interoperability and more effective collaboration at the European Union level

There are also very practical reasons for a more substantial intra-European conversation about responsible military use of AI, mainly internal to the economic and military future of the EU. These relate to the EU's ambition regarding its strategic autonomy, the interoperability of EU armed forces and the progress of European research and industrial collaboration.

The notion of strategic autonomy in Europe has recently gained greater salience within European political discourse, due to increased global instability and geopolitical tensions, and the idea that Europe can no longer rely on global institutions and partners as it has done in the past.[13] Yet the concept remains vague and contested within Europe.[14] It rests on the perceived need for the EU to protect and develop its technological, industrial, defence and economic interests.[15] Access to technology is one central aspect of the debate around Europe's desire to gain strategic autonomy, and the development of AI can fall within this.[16] A coherent EU position around the responsible use and development of AI for use within the military could feed into this wider debate.

Historically, the competence of the EU institutions on defence has been limited, but now the EU and its member states are committed to fostering greater cooperation

[9] On how the risk of strategic competition may affect states' behaviour with regard to development and use of AI, see Boulanin et al. (note 6).

[10] European Union (EU), EU statement at the Group of Governmental Experts (GGE) on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems (LAWS) within the Convention on Certain Conventional Weapons (CCW), 25–29 Mar. 2019; and EU, EU statement at the Group of Governmental Experts (GGE) on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems (LAWS) within the Convention on Certain Conventional Weapons (CCW), 27–31 Aug. 2018.

[11] Boulanin, V., *Mapping the Debate on LAWS at the CCW: Taking Stock and Moving Forward*, EU Non-proliferation Paper no. 49 (SIPRI: Stockholm, Mar. 2016); PAX, *Convergence? European positions on lethal autonomous weapons* (PAX: Ultrecht, Nov. 2019); and PAX, *Crunch Time—European positions on lethal autonomous weapons* (PAX: Ultrecht, Nov. 2018).

[12] For a discussion on these differences, see Franke (note 8).

[13] Breton T., Speech at the Committee on Security and Defence, European Parliament, 25 June 2020.

[14] For a discussion of the contested meanings, see Franke, U. and Varma, T., 'Independence play: Europe's pursuit of strategic autonomy', European Council on Foreign Relations, July 2019.

[15] Borell, J. and Breton, T., 'For a united, resilient and sovereign Europe', European Commission, 10 June 2020.

[16] Brattberg, E., Csernatoni, R. and Rugova, V., 'Europe and AI: leading, lagging behind, or carving its own way?', Carnegie Endowment for International Peace, Working paper, 9 July 2020.

Box 2.1. The Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapon Systems (GGE on LAWS)

Since 2014, the legal, ethical and security challenges posed by lethal autonomous weapon systems (LAWS) have been subject to intergovernmental discussions within the framework of the 1981 Convention on Certain Conventional Weapons (CCW) under the United Nations. Since 2017, the GGE on LAWS has led the discussions, with the mandate to 'explore and agree on possible recommendations on options related to emerging technologies in the area of LAWS, in the context of the objectives and purposes of the Convention, taking into account all proposals—past, present and future'. In 2019, the GGE on LAWS adopted 11 guiding principles that aim to guide its work on the governance of emerging technologies in the area of LAWS, including technological developments in the field of military AI. These guiding principles reflect areas of convergence between the high contracting parties. Among other things, the guiding principles establish that international humanitarian law applies fully to LAWS; that humans remain responsible for decisions on the use of weapon systems; that states should conduct the required legal reviews of new weapons, means and methods of warfare; and that humans and human-machine interaction must be considered in the entire life cycle of the weapons, although taking various forms.

Sources: Peldán Carlsson, M. and Boulanin, V., '13. Conventional arms control and new weapon technologies', *SIPRI Yearbook 2020: Armaments, Disarmament and International Security*, pp. 502–503; and Certain Conventional Weapons Convention, Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems, Report of the 2019 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems, annex IV to CCW/GGE.1/2019/3, 25 Sep. 2019.

between its member states in this area.[17] In 2016 the Global Strategy for the Foreign and Security Policy of the EU, known as the EU Global Strategy (EUGS) was released and EU members states' commitment to work together on defence matters has increased alongside an evolving EU common security and defence policy.[18] EU member states have increased cooperation in a range of measures, from the creation of the EDA in 2004 to the activation of the permanent structured cooperation (PESCO) in 2016. Given the EU's competencies much of the work is led by member states rather than the European Commission or other EU institutions.[19] Such cooperation allows member states to coordinate their views on capability development needs, prevent duplication and reduce the cost of European intellectual and industrial investments related to defence innovation and work on projects that will foster greater interoperability between EU armed forces. AI already plays a role—and is bound to play an even larger one in the future, as it is considered a key domain—in projects conducted by the EDA and PESCO.[20] Alongside this, the newly established EDF, under the responsibility of the European Commission, will look to bolster European defence research and industrial projects.[21] EU guidelines about what constitutes responsible military use of AI could aid the AI-related work that is conducted within these mechanisms by helping to avoid contradictions between projects and optimizing the use of financial investments.

III. Scaling up capabilities while ensuring cost-efficiency at the national level

EU members states will also need to work more with one another towards the development of responsible military use of AI for economic reasons. For example, no individual EU member state has the resources to keep up with the USA and China's level of investments in the military domain. In order to acquire next-generation

[17] European Council, Council of the EU, 'European cooperation on security and defence', 20 Oct. 2020; Denmark is the only EU member state to have opted out of EU defence integration projects.

[18] EU, *Shared Vision, Common Action: A Stronger Europe A Global Strategy for the European Union's Foreign And Security Policy* (EU: Luxembourg, 2016); For an overview of what has changed since the EUGS in 2016, see EU, 'The European Union's global strategy three years on, looking forward', 13 June 2019.

[19] PESCO, 'PESCO'; and European Commission, 'Internal market, industry, entrepreneurship and SMEs'.

[20] European Defence Agency (EDA), 'The EU capability development priorities', 3 Dec. 2018, p. 18.

[21] Besch, S, 'The European Commission in EU defense industrial policy', Carnegie Europe, 22 Oct. 2019.

military systems, EU member states will need to conduct joint collaborative projects, such as the Future Combat Air System (FCAS) that is being developed collaboratively between France, Germany and Spain.[22] Given that AI is expected to play a part in virtually all major future military capabilities, EU member states will need to align their views on what should be considered responsible standards for the development and use of AI systems. An agreement between them could allow for the EU's pooled resources to be put to better use.

[22] Airbus, 'Future combat air systems: owning the sky with the next generation weapons system', 17 June 2020.

3. The European Union as a leader in best practices

Having considered the many reasons why the EU should develop its views on what constitutes responsible military use of AI (see chapter 2), the question now is how this can be done given the institutional limitations of the EU on defence matters and the sensitivity of some member states regarding the importance of maintaining defence as a matter of national sovereignty? In this context it is important to acknowledge that various EU frameworks are already in place and EU agencies have engaged in work around the military use of AI, which could be either expanded upon, become more public or made more collaborative. Through the framework of the European Commission, the EU has also orchestrated a lot of the work on the governance of AI in the civilian sector, which could provide a useful basis for future projects by the EU and its member states on responsible military use of AI. Therefore, this chapter explores how the EU and its member states could advance their reflection on responsible military use of AI by focusing on three areas: legal compliance, ethical acceptability and safety.

I. Expanding the European Union's work on the military use of artificial intelligence

The European Defence Agency's work on capability development

According to the EDA, there is already a structured conversation between EU member states on the development and use of AI in defence.[23] The EDA has reportedly been working on developing a joint perspective on AI capability development since 2016.[24] It has coordinated several research and development (R&D) projects that touch on AI. In addition, it has also worked towards the production of: (*a*) a common AI definition, taxonomy and glossary to resolve the possible conceptual differences between states on the concept; (*b*) a shared view of relevant application areas for European Capability Development Plan, within the wide spectrum of capabilities that can be enabled with AI; and (*c*) an action plan for EU collaboration on AI in defence, which is to be presented to the EU defence minister in December 2020.[25]

However, these documents have not yet been made public. It would be valuable if the outcome of the work conducted so far could be made available to the public in order to determine where EU member states stand in terms of responsible military use of AI.

According to an expert who presented at the SIPRI workshop on the topic of this report on 8 September 2020, the EDA's AI-related projects have focused on issues concerning the lawful and ethical use of military AI.[26] They have not discussed responsible use as a standalone topic but rather have addressed it within the framework of specific capability development and R&D projects. The rationale behind this is that the discussion is more concrete and operational. According to the same expert, one thing that the defence ministers of EU member states have been able to agree on is that humans should remain in control of the use of AI-enabled systems.[27] This concurs with the position of the EU presented at the UN debate on LAWS in Geneva.[28]

[23] EDA, 'Joint quest for future defence applications', *European Defence Matters*, no. 19 (2020).
[24] EDA (note 20); and EDA (note 23), p. 36.
[25] EDA (note 23), p. 36.
[26] The presentation was held at a SIPRI workshop organized on 8–9 September 2020 on 'Governing the opportunities and risks AI pose to international peace and security: what role for the EU?'. The workshop was under the Chatham House Rule.
[27] SIPRI workshop (note 26).
[28] EU, EU statement, 25–29 Mar. 2019 (note 10); and EU, EU statement, 27–31 Aug. 2019 (note 10).

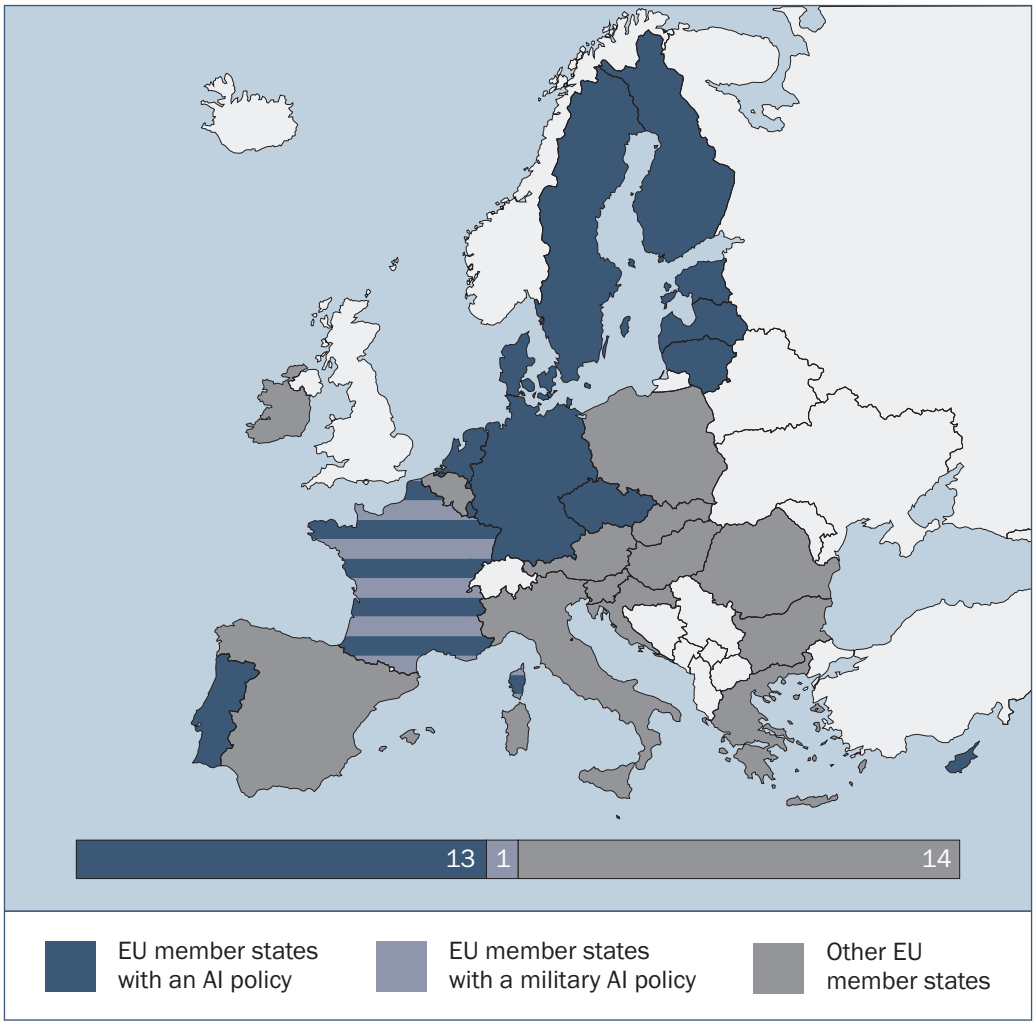

Figure 3.1. AI policy development in the European Union

AI = Artificial intelligence; EU = European Union.

Source: Authors' compilation based on European Union member states' policy documents.

The European Parliament's work on the regulation of artificial intelligence and military use

The European Parliament has increasingly followed the military use of AI. In 2018 it notably adopted a resolution calling for the ban of lethal autonomous weapons, defined as weapons without meaningful human control.[29] This resolution demonstrated that political representatives from EU member states could agree on a red line independent of their nationality and political affiliation. However, the resolution has had little effect on the way EU member states have engaged in the debate of the GGE on LAWS at the CCW. It has only conveyed the opinion of the European Parliament and does not impose any obligation on EU member states.

The interest of the European Parliament in ethical issues related to autonomous weapons specifically and military use of AI more generally has continued to grow. In 2019, as part of the trilateral discussion with the European Commission and the Council of the EU about the creation of the EDF, members of the parliament conditioned their approval of the EDF's budget with an obligation to not allocate funds to R&D projects

[29] European Parliament, Resolution of 12 Sep. 2018 on autonomous weapon systems, 2018/2752(RSP); and European Parliament, Decision of 18 June 2020 on setting up a special committee on artificial intelligence in a digital age, and defining its responsibilities, numerical strength and term of office, 2020/2684(RSO).

on lethal autonomous weapons as defined in the 2018 resolution calling for a ban on them.[30]

In June 2020, the European Parliament agreed to set up a special committee on AI, which has the mandate to investigate issues related to military applications of AI.[31] The European Parliament's Committee on Legal Affairs and the Committee on Foreign Affairs have also have published two draft reports, one on the ethical aspect of AI and the other, which the European Parliament has not yet formally adopted, on AI and IHL.[32] The latter document calls on the EU not only to actively formulate its own technical and ethical standards for the adoption of AI-enabled military systems but also to promote them in the multilateral forums.

European Union member states' national strategies and intra-European cooperation

According to an expert from within the EU who participated in the workshop hosted by SIPRI, the work of EU member states on the military use of AI has been limited; in fact, according to the EDA only three member states have produced national strategies on the military use of AI (see figure 3.1). Of these, France is the only state that has published its strategy publicly.[33] This being said, a small group of states do seek greater coordination and national reflection on the topic, as evidenced by the food-for-thought paper that Finland published in 2019 while presiding over the Council of the EU.[34] The paper calls on EU member states to reflect on the strategic role of AI for the future of EU defence. It points out that ethical considerations need to play a key role in such deliberation, and it outlines a series of questions that EU states should consider as part of their national reflection.

II. Building on the European Commission's work on the governance of civilian uses of artificial intelligence

The European Commission has done a lot to support the development of an approach to governance of AI in the civilian sector. In 2018 it formed AI HLEG, whose mandate was to provide guidelines on trustworthy AI.[35] The group's recommendations that were published in June 2019 then served as the basis for a white paper published by the European Commission in 2020, entitled 'On Artificial Intelligence—A European Approach to Excellence and Trust'.[36] These documents provide a basis for the potential work of the EU on responsible military use of AI in three regards.

First, they share what should be the objectives of the work of the EU and its members on responsible military use of AI: making the EU a role model for the responsible development, adoption and use of AI while ensuring the competitiveness of EU organizations, universities, research laboratories and companies alike in the field.

[30] Brzozowski, A. 'European Defence Fund agreed amid ethical concerns', Euractiv, 22 Feb. 2019; and Campaign to Stop Killer Robots, 'No killer robots for European Defence Fund', 20 Feb. 2019.

[31] European Parliament, 2020/2684(RSO) (note 29).

[32] European Parliament, Committee on Legal Affairs, 'Draft report with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies', 2020/2012(INL), 21 Apr. 2020; and European Parliament, Committee on Legal Affairs, 'Opinion of the Committee on Foreign Affairs for the Committee on Legal Affairs on artificial intelligence: questions of interpretation and application of international law insofar as the EU is affected in the areas of civil and military uses and of state', 2020/2013(INI), 9 July 2020.

[33] French Ministry of the Armed Forces (MAF), *L'intelligence artificielle au service de la défense* [Artificial intelligence at the service of defence], Report of the AI Task Force (MAF: Paris, Sep. 2019).

[34] Food-for-thought paper by Finland, Estonia, France, Germany and the Netherlands (note 2).

[35] AI HLEG (note 3).

[36] European Commission, 'White paper on artificial intelligence—A European approach to excellence and trust', COM(2020) 65, 19 Feb. 2020.

Second, they outline what could be the thematic pillars of the EU's position on responsible military use of AI, namely that AI technology needs to be: (*a*) lawful; that is, it should show respect for all applicable laws and regulations, including the EU charter of fundamental rights; (*b*) ethical and 'human-centred', meaning it should be respective of the EU's principles and values and preserving human agency and oversight in the use of AI systems; and (*c*) robust; that is, it should be technically safe while taking into account the social environment in which AI technology is used.

Third, the inherently dual-use nature of AI technology also makes some of the technical recommendations of the AI HLEG relevant for the military sector. Those related to 'high-risk' AI applications are particularly pertinent—albeit they might need to be adapted to the military context.[37] The EU and its member states could explore the extent to which the principles and recommendations designed for the civilian sector in these three areas—legal, ethical and technical safety—could be applied to the military context and determine how they might be adapted or supplemented accordingly.

III. A road map for the European Union on responsible military use of artificial intelligence

As shown above, the groundwork for the development of a European view on responsible military use of AI has already been laid. The work conducted by the EDA and the European Parliament, as well as several EU member states, demonstrates a growing political interest. The work initiated by the European Commission on 'trustworthy' AI provides a useful, substantial basis for engaging in conversations on the responsible military use of AI.[38] This section provides a series of concrete proposals for EU institutions and EU member states to consider how they might engage in an open, methodical and collaborative way to develop a joint perspective on the responsible military use of AI. This section suggests that the EU and its member states should hold three expert discussions on legal compliance, ethics and AI safety.

Legal compliance: Supporting compliance through legal reviews and focused international humanitarian law discussions

It is beyond dispute that the baseline for what constitutes responsible military use of AI in armed conflict is compliance with international law, including IHL. This is the case for all means and methods of warfare. Participants of the UN's intergovernmental expert discussion on emerging technologies in the areas of LAWS recently recognized and affirmed this fact.[39] AI military applications encompass a broader range of tools than LAWS, but this principled support for the applicability of international law to a subset of AI systems bodes well. Indeed, France has already indicated that compliance with international law is an element of its military AI strategy.[40]

Acknowledging the applicability of international law is an important first step, but it must be accompanied by concrete and practical measures to ensure such compliance in practice. This can involve a range of measures that include conducting robust legal reviews of new weapons, means or methods of warfare, developing appropriate

[37] European Commission (note 36).

[38] AI HLEG (note 3).

[39] United Nations, General Assembly, Group of Governmental Experts of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects, 'Draft report of the 2019 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapon Systems', CCW/GGE.1/2019/CRP.1/Rev.2, 21 Aug. 2019.

[40] Parly, F., Minister of the Armed Forces, 'Intelligence artificielle et défense' [Artificial intelligence and defence], French Ministry of the Armed Forces, Speech at the Saclay Campus, 5 Apr. 2019.

doctrine and rules for the use of AI systems, training personnel on the proper use of such systems and ensuring that processes exist for investigating and enforcing possible violations of the law that involve the use of AI.[41]

However, the challenge is addressing how IHL should be interpreted and how such compliance measures should be applied with respect to AI military applications, given the novel characteristics of those systems.

While this is a matter for each EU member state, it is also a collective concern among states who share the same legal obligations (since all EU member states are party to the Geneva Conventions and the First and Second Additional Protocols). A coherent and consistent approach to legal compliance would be advantageous, given the EU's military cooperation, arms and dual-use export policies and collaborative defence research activities.[42] This approach would amplify the EU's voice and minimize interoperability challenges.

Two possible ways for EU member states to work more closely on issues of legal compliance are (*a*) through increased cooperation on conducting legal reviews of new weapons, means and methods of warfare, and (*b*) through focused discussions among EU member states on interpretation and application of IHL concerning AI military applications.

Legal reviews

The obligation to carry out legal reviews of new weapons, means and methods of warfare is set out in Article 36 of the 1977 Additional Protocol to the 1949 Geneva Conventions. EU member states widely recognize the importance of conducting legal reviews according to Article 36, and it is increasingly emphasized given ongoing developments in technology. It is an essential tool for ensuring that states conduct hostilities in accordance with their international obligations.[43]

However, legal reviews are national procedures and not subject to international oversight. There are no established standards about how EU member states should conduct legal reviews. In fact, the majority of states are not known to have any standing legal review mechanism.

In the GGE on LAWS, a number of participants (including the EU and many member states) have aired their views on the critical role that legal reviews play in preventing the development and adoption of inherently unlawful technologies.[44] As reflected in an EU statement in 2018, member states share the conviction that greater cooperation and information sharing on legal reviews could contribute to wider compliance with Article 36.[45] They also believe that it could facilitate identifying best practices that could help states to navigate the legal complexities of reviewing weapons, means and methods of warfare based on new technologies such as AI.[46]

[41] Boulanin, V. et al., *Limits on Autonomy in Weapon Systems: Identifying Practical Elements of Human Control* (SIPRI: Stockholm, June 2020).

[42] European Council, Council of the EU, 'European cooperation on security and defence', 20 Oct. 2020.

[43] International Committee of the Red Cross (ICRC), 'A Guide to the legal review of new weapons, means and methods of warfare: measures to implement Article 36 of Additional Protocol I of 1977', Jan. 2006; Boulanin, V., 'Implementing Article 36 weapon reviews in the light of increasing autonomy in weapon systems', SIPRI Insights on Peace and Security no. 2015/1, Nov. 2015; and Boulanin, V. and Verbruggen, M., 'SIPRI compendium on Article 36 reviews', SIPRI Background paper, Dec. 2017.

[44] United Nations, General Assembly, Group of Governmental Experts of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects, 'Draft report of the 2019 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapon Systems', CCW/GGE.1/2019/CRP.1/Rev.2, 21 Aug. 2019.

[45] EU, EU Statement at the Group of Governmental Experts (GGE) on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems (LAWS) within the Convention on Certain Conventional Weapons (CCW), Geneva, 9–13 Apr. 2018.

[46] EU (note 45).

This discussion provides a firm foundation for further work to ensure that legal reviews—not only of LAWS but of all new technologies of warfare including military AI applications—are conducted effectively.[47] This technical, complex and sensitive issue lends itself to expert discussions among selected participants (government, industry and academic) with experience in or knowledge of assessing the functioning and effects of AI technologies. EU member states could hold a separate thread of discussions that address core questions of how to establish robust legal review mechanisms, especially among states that do not currently conduct such reviews. Information sharing and cooperation among EU member states about legal reviews—be it in general or in relation to AI-enabled military capability—could be beneficial in three regards.

First, it would signal to the world that EU member states are committed to legal compliance and determined to ensure that their national review processes are an adequate, hence legitimate, mechanism for ensuring the responsible development and military use of AI—at least from a legal standpoint.

Second, it could help states to learn from each other and assist EU member states that wish to establish legal review mechanisms or strengthen them. Currently, according to a SIPRI survey, only 11 EU member states appear to have formalized a legal review process, and only 5 have published detailed public information about it (see figure 3.2).

Third, it provides an opportunity for EU member states to share potentially sensitive information about past or current review processes and to have transparent discussions on the interpretation and application of the rule of international law when assessing new technologies like AI. Over time, this could assist in building expertise among those involved in legal reviews. It could also help states develop a common approach to the questions raised by assessing the legality of AI systems, such as what types of AI technologies should be reviewed, which legal standards should apply to this technology, how they should be interpreted, and when and how to review and test systems like machine learning that undergo constant modification (see box 3.1).[48]

Focused discussions on international humanitarian law

The UN process on LAWS has revealed that states, including members of the EU, continue to develop their views on how international law should be interpreted and applied given the unique characteristics of LAWS. Diverging views may characterize any multilateral discussion of IHL compliance with respect to AI military applications. Translating IHL rules for the use of new technologies can raise unique challenges and reopen old debates, particularly with respect to the rules on the conduct of hostilities. In the absence of a dedicated international forum for discussing IHL compliance, such differences can complicate multilateral policymaking and distract from efforts to address challenges and harness opportunities brought by new technologies. While divergent positions among states are not of themselves problematic, they can signal more significant shifts in how states view IHL's inherent balance between humanity and military necessity. This is of critical concern, given the important role IHL plays in setting the boundaries of responsible innovation and use of AI in the military realm.

[47] Belgian Commission for the Legal Review of New Weapons, Report presented at the CCW Informal Meeting of Experts on LAWS, 13 Apr. 2016; the Netherlands and Switzerland, 'Weapons Review Mechanisms', Paper submitted at the 2017 Meeting of the Group of Governmental Experts of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects, Geneva, 13–17 Nov. 2017, CCW/GGE.1/2017/WP.5; and German Federal Foreign Office, Commentary on operationalizing all eleven guiding principles at a national level as requested by the chair of the 2020 GGE on Emerging Technologies in the Area of LAWS within the CCW, 24 June 2020.

[48] Lewis, D. A., 'Legal reviews of new weapons, means and methods of warfare involving AI: 16 elements to consider', Humanitarian Law and Policy Blog, ICRC, 21 Mar. 2019.

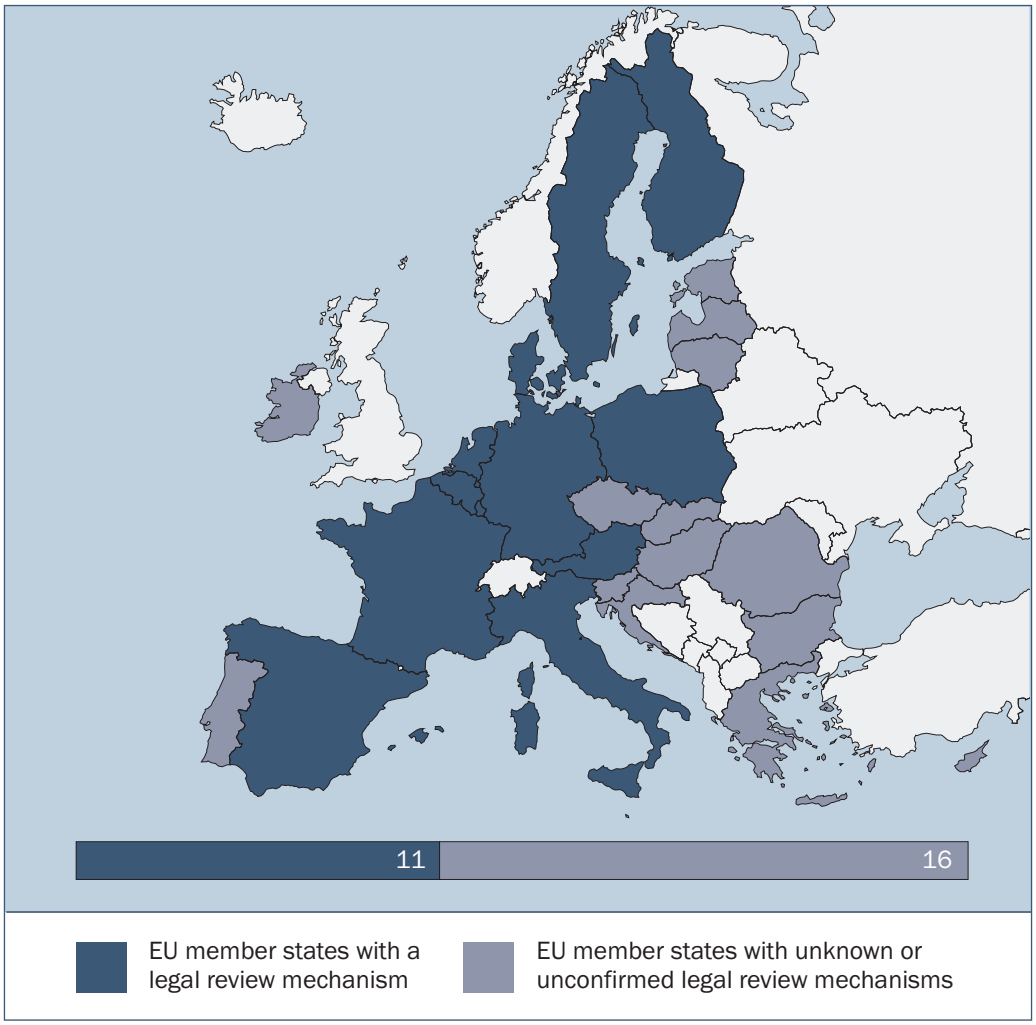

Figure 3.2. Sample of statements by European Union member states on human control

EU = European Union.

Source: Compilation based on survey conducted by the authors.

More intense and complex expert discussions among EU member states on the application of IHL in the context of military usage of AI is necessary to clarify the legalities underpinning responsible military use of AI. The UN expert process on LAWS provides an avenue for such deliberation, albeit within limits (i.e. with respect to a subset of AI military applications only, with relatively few legal experts, and under the strain of competing agendas and the frozen state of disarmament diplomacy).[49] EU member states should, nonetheless, engage in such a discussion for two reasons. First, from a practical standpoint, the EU provides a forum for frank and focused discussions, while reducing the risk of politicization. This is not to say that aligning the views of 27 EU member states will be easy, but it will surely be more manageable than those of 125 High Contracting Parties and four states signatory to the UN CCW. Second, from a political standpoint, focused discussions could allow EU member states to present a more consolidated view at the UN, including in the GGE on LAWS. The question then raised is how and where that intra-European discussion on the interpretation and the application of IHL, including legal reviews, could take place.

The EU framework offers several opportunities for collaborating on these issues. For example, COJUR provides a forum to share information and exchange views about

[49] Regehr, E. 'Disarmament diplomacy in the age of Trump and Putin', Center for International Policy Studies, 11 Feb. 2020.

Box 3.1. Challenges associated with assessing the lawful use of artificial intelligence in military applications

The unique characteristics of artificial intelligence (AI) in military applications pose new challenges to conducting a legal review. Some of the key challenges include:

What to review. Besides the obligation to review newly acquired weapons, means and methods of warfare (whether hardware or software, self-contained or in systems), states are also obliged to review weapons, means or methods modified after acquisition.[a] Because the software that underpins AI military applications can be easily modified, states may need to conduct legal reviews more often, perhaps even leading to a constant cycle of assessment and reassessment. However, it remains unclear what kind of modifications should trigger the need for a new legal review. For AI systems that change their functioning after activation, the challenge is ensuring that the legal review does not become invalid immediately upon the use of the system.[b]

Predictability and certainty. For a weapon, means or method of warfare to pass a legal review, the reviewer needs to be satisfied that it will be used in compliance with international law in some or all circumstances. Achieving this certainty, mainly through testing, involves understanding the capabilities of the system and reliably predicting its effects in the likely or expected circumstances of its use. However, foreseeing the effects of complex AI applications can pose major challenges. Not only is it difficult to fully anticipate all elements of a battlefield in a testing environment, but some algorithms may be so complex or opaque that they are inscrutable. The high expense involved in conducting tests may also pose difficulties.[c]

The applicable criteria. When conducting a legal review, the reviewer will need to be satisfied that the reliance on AI by the weapon, means or method does not interfere with the eventual user's ability to comply with international humanitarian law (IHL). This includes the ability of a commander employing an AI application to exercise the context-specific evaluative legal judgements required when conducting hostilities. Exactly how commanders can ensure their ability to make these judgements when using AI is an issue that may require further clarification.[d]

[a] International Committee of the Red Cross (ICRC), 'A Guide to the legal review of new weapons, means and methods of warfare: measures to implement Article 36 of Additional Protocol I of 1977', (June 2020).

[b] Goussac, N., 'Safety net or tangled web: legal reviews of AI in weapons and war-fighting', Humanitarian Law and Policy Blog, ICRC, 18 Apr. 2019.

[c] Boulanin, V., and Verbruggen, M., 'Article 36 reviews: dealing with the challenges posed by emerging technologies', SIPRI Report, Dec. 2017, pp. 24–5.

[d] Lewis, D. A., 'Legal reviews of new weapons, means and methods of warfare involving AI: 16 elements to consider', Humanitarian Law and Policy Blog, ICRC, 21 Mar. 2019.

what may be deemed best practices as well as for producing guiding documents for member states. The European Parliament, through the committees for foreign affairs and legal affairs, could also establish an ad hoc parliamentarian expert group, which could meet regularly for a definite period (e.g. 12 months) to discuss issues related to information sharing in the area of legal reviews. Other opportunities may exist through specific projects under the responsibility of EDA, EDF and PESCO, given the links to IHL, governance of defence innovation and operational cooperation.

Ethical acceptability: Defining shared principles for human-centric artificial intelligence in the military

Human agency as a central issue in responsible military use of artificial intelligence

AI raises many different ethical issues that are relevant in both the civilian and military sectors. These issues range from how data serving the development and training of AI systems is sourced, used and protected, to the biases and assumptions that appear in AI programming, to the potentially dehumanizing effect of reducing the real world to data points, to the consequences of non-transparent or inexplicable decision making for accountability and responsibility.[50]

Many of these issues feed into a higher-level discussion on how AI is challenging human agency. In this regard, the EU member states seem to already agree on the fundamental premise that human agency—be it in the civilian or military sector—is necessary for responsible use of AI. Finland is one EU member state that has given

[50] ICRC, 'Ethics and autonomous weapon systems: an ethical basis for human control?', Report, 3 Apr. 2018.

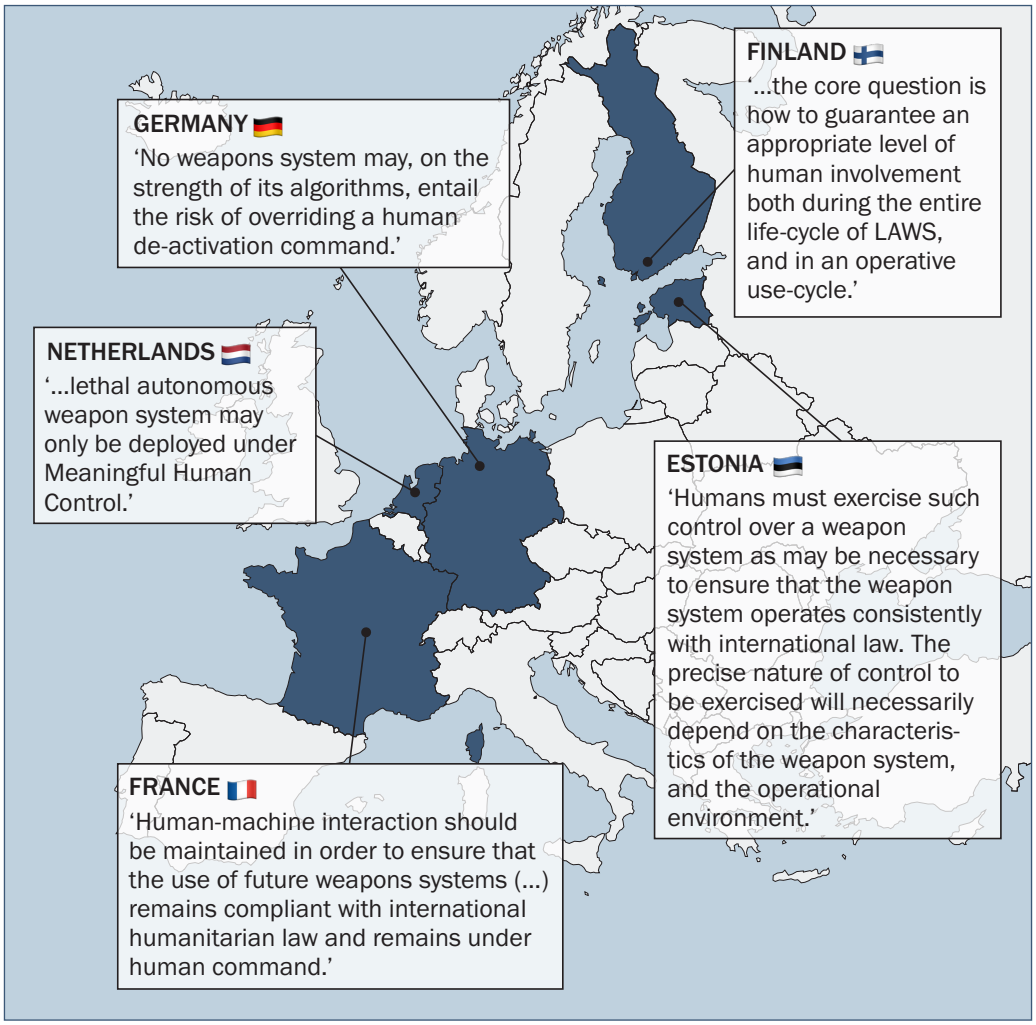

Figure 3.3. Sample of European Union member states on human control

LAWS = Lethal autonomous weapon systems.

Sources: Authors' compilation based statements by European Union member states.

this issue a lot of thought. In its submission to the GGE on LAWS 2020 it stated, for example, that 'the easier an advanced technology is to apply, the easier it will be to use it for harmful purposes. In the future, we might well see armed autonomous civilian capabilities used for military purposes', arguing for the necessity of establishing a framework for human involvement.[51]

AI should remain a tool for humans to make decisions, and it should not strip humans of their obligation and ability to do so. That principle is a cornerstone of the strategy of the European Commission on 'human-centric AI'.[52] Members of the European Parliament also agreed in 2018 that the litmus test for determining if an autonomous weapon would be acceptable is whether it remains under 'meaningful human control'.[53] In the context of the GGE's debate on LAWS, EU members states also supported the adoption of the guiding principle which states that 'human responsibility for the decision on the use of force must be retained' (see box 2.1).[54] Many of the EU states have made the concept of human control a pivotal point of their national position in

[51] Finland, 'Considerations on the appropriate level of human involvement in LAWS', Food-for-thought paper submitted to the LAWS Group of Governmental Experts of the High Contracting Parties to the Convention on Certain Conventional Weapons, June 2020.

[52] AI HLEG (note 3).

[53] European Parliament, Resolution of 12 Sep. 2018 on autonomous weapon systems, 2018/2752(RSP).

[54] EU, EU statement, 25–29 Mar. 2019 (note 10).

the debate on LAWS, such as Germany that has stated for instance that 'no weapons system may, on the strength of its algorithms, entail the risk of overriding a human de-activation command' (see figure 3.3).[55]

While EU members agree that meaningful human control should be maintained throughout the entire life cycle of the weapon, the question of how human agency needs to be exercised in practice, however, remains unresolved.[56] While some proposals have been made in the context of the discussion of GGE on LAWS, they are limited in scope as they do not have the intent to guide other possible military uses of AI, from logistic and maintenance to mission planning, training and recruitment of personnel.[57] The general principles and recommendations laid out by the AI HLEG could, in this regard, provide a useful baseline to think about how human-centred AI could look like in the military sector, beyond the sole case of LAWS.

EU member states would have a lot to gain from engaging in an intra-European expert discussion on human control. First, on the issue of legal compliance for example, it would allow them to discuss, deepen, align or, at least, remove the potential for conflict in their reflections on the case of LAWS. Second, it would allow them to find a clear direction and to determine their limits in developing collaborative defence acquisition projects.

Fostering intra-European discussion on human-centred artificial intelligence in the military

Deliberations on human-centred AI in the military could take place in different forums at the same time. The choice of forum should be guided by concrete political aims.

The Council of the EU already has a number of preparatory bodies that would allow delegates from member states to coordinate their views on human control in relation to discussion on LAWS at the UN in Geneva, namely, CONOP, CODUN and EUMCWG. The regular meeting of these bodies provides an opportunity for the EU External Action Service to identify the current common position of EU member states on this issue. The meetings, however, are meant to cover a wide range of topics every time, which limit the possibility for states to discuss topics at a very granular level. It could be meaningful in that regard to organize a series of extra meetings specifically dedicated to the issue of human control.

When it comes to developing more general political guidelines or technical standards for the development of joint defence acquisition or capability development projects involving AI, then the EDA and the European Commission could both play a role. EU member states could task the EDA to coordinate an expert reflection on human-centred AI for defence, which would require technical, military and arms control experts. Such an initiative would need to connect with both the work on human-centred AI in the civilian sector and with the discussion on human control within arms control forums. The European Commission could utilize the EDF to support multidisciplinary research projects that would explore issues around human control at a more technical level and feed into the development of AI systems that can be deployed responsibly.

[55] German Federal Foreign Office (note 47).

[56] EU, EU statement, 25–29 Mar. 2019 (note 10).

[57] As per the commentaries submitted to the GGE on LAWS in 2020, France, for example, argues that if feasible, the commander should be able to give warnings, communicate and deactivate the system. Also, France, Germany and Finland argue that limits in time and space should be imposed, and in some cases, the types of targets should be limited. See France, 'Operationalization of the 11 guiding principles at national level—Comments by France', Paper submitted at the Group of Governmental Experts (GGE) on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems (LAWS) within the Convention on Certain Conventional Weapons (CCW), 2020; German Federal Foreign Office (note 47); and Finland (note 51).

It would also be valuable if EU member states could engage in the issue at a higher political level and grapple with more fundamental ethical and political issues. Such a discussion could take place in the context of an ad hoc expert group of the European Parliament and directly at the level of the ministers of foreign affairs and defence and the EU High Representative of the Union for Foreign Affairs and Security Policy at the European Council.

Artificial intelligence safety: Mainstreaming artificial intelligence safety as part of joint European collaborative defence efforts

AI safety as the foundation for responsible military use of AI

AI safety is an issue that the EU already has given a great deal of consideration. It is the pivotal point of the European Commission's 2020 white paper on AI. According to the white paper, AI technology needs to rely on technical standards that make it safe for the user while allowing the user to adhere to the legal and ethical rules and norms. The white paper has already delineated a set of fundamental principles that should be the core of AI design: human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, societal and environmental wellbeing, and accountability.[58] The Joint Research Centre (JCR) of the European Commission has also worked on the topic of AI safety at a more technical level. In 2020 it published a technical report on the robustness and explainability of AI, which provides some technical and policy recommendations for AI safety.[59]

AI safety is equally, if not more, important for the military than the civilian sector. The military context is adversarial by definition: an intelligent enemy will purposefully try to defeat the system by attacking it or exploiting its weaknesses. Therefore, military systems generally need to abide by high safety and security standards, even with applications that are not directly intended for combat use (e.g. logistics); AI technology is no exception. The latest developments in AI, which are based on machine learning for the most part, raise safety challenges that are particularly problematic in 'life-critical applications' (applications whose use could endanger human life or put safety at risk) and are the case for many military applications of AI—if not the majority of them. These could lead to unintended or wrongful uses of force and thereby undermine the military user's ability to behave lawfully and ethically. For example, algorithmic bias in AI-powered intelligence, surveillance and reconnaissance (ISR) systems could lead a military commander to select the wrong type of targets or disproportionately focus on certain types of targets.[60] The black-box problem that is characteristic of deep learning systems makes the behaviour of the systems hard for the user to understand and predict, and difficult to investigate when things go wrong.[61] In sum, the military community has a vested interest in the progress of AI safety, whether from an operational, legal or ethical standpoint.

In February 2020, the US Department of Defense adopted ethical principles and guidelines for the acquisition of AI systems for military use.[62] France also outlined

[58] European Commission (note 36).

[59] Hamon, R., Junklewitz, H. and Sanchez Martin, J., *Robustness and Explainability of Artificial Intelligence—From Technical to Policy Solutions*, EUR 30040 (Publications Office of the EU, Luxembourg, 2020).

[60] Boulanin et al. (note 41).

[61] The 'black-box problem' refers to the opacity of computer systems programmed using machine learning. It is difficult to know why they do what they do or how they actually work; the inputs and outputs are known and can be understood, while the process that turns the former into the latter is unknown. For further detail see Zednik, C., 'Solving the black box problem: a normative framework for explainable artificial intelligence', *Philosophy and. Technology* (Dec. 2019).

[62] US Department of Defense, 'DOD adopts ethical principles for artificial intelligence', 24 Feb. 2020.

some general principles in its national strategy on military AI.[63] Other European countries such as the Netherlands are also known to have set up specific commissions that explore this topic. Given that AI technology will necessarily play a role in future EU military-industrial and capability development plans, it would make sense for member states to at least coordinate their reflections, in order to eventually reach a common set of standards for joint projects.

Searching for common safety standards in the military use of AI

There are two possible avenues for an intra-European discussion on AI safety applied to the military context. The first and main avenue would be through the EDA, which can take action at multiple levels. As the host of regular meetings of the EU ministries of defence, it provides an appropriate forum for high-level political discussion among member states on this topic. At the more technical level, the EDA can support cooperative defence projects through (*a*) the Industry, Synergies & Enablers (ISE) Directorate; (*b*) the Capability, Armament and Planning (CAP) Directorate; and (*c*) the Research, Technology and Innovation (RTI) Directorate. The EDA has already engaged in reflecting on the use of AI in defence.[64] At the request of member states, the EDA could start an ad hoc working group of national experts to explore whether or not AI safety guidelines, which are already part of the report of the AI HLEG and publication of the JCR, are applied to the military sector.[65] The conclusion of such projects would then serve as a basis for political discussion among member states and could possibly define the European safety standards for use of AI in the defence sector.

The second avenue is the European Commission's Directorate General for Defence Industry and Space (DG DEFIS), which will likely manage the EDF. DG DEFIS could use the EDF to fund R&D projects that explore ways of meeting safety standards for AI at a technical level. In concrete terms, this could include projects that would create or improve testing and evaluation methods of systems based on machine learning, or projects that would develop technical methods to improve the explainability and interpretation of AI among its military users.

[63] French Ministry of the Armed Forces (MAF), *L'intelligence artificielle au service de la défense* [Artificial intelligence in the service of defence], Report of the AI task force (MAF: Paris, Sep. 2019).

[64] EDA (note 23), p. 36.

[65] AI HLEG (note 3); and Hamon, Junklewitz and Sanchez Martin (note 59).

4. Key findings and recommendations

This report has dealt with the question of whether it is politically desirable and possible for the EU and its member states to propose a more concerted and identifiable perspective on responsible military use of AI. The main takeaway is that the EU and its member states would benefit politically, strategically and economically from developing principles and standards for the responsible military use of AI. It will be far from a simple and straightforward process, as such a proposition involves juggling competencies, political sensibilities and wills of the various EU actors (institutions and member states). However, there is a political momentum and substantial work on the legal, ethical and safety requirements for responsible use of AI that provide a basis for the emergence of an EU perspective on responsible military use of AI. The key findings and recommendations of the report can be summarized as follow.

I. Key findings

There are at least three reasons why the EU and its member states would benefit from developing standards for the responsible use of AI. First, from a political standpoint, it would allow them to gain traction in the global discussion on the military use of AI and its governance and effectively defend a vision that is aligned with their shared values and interests. Second, from a strategic standpoint, it would help them maintain some autonomy regarding the military technology they want to use and how they want to use it. Third, from an economic standpoint, it would benefit the development of the European military-industrial base and operational capabilities in the field of AI.

The groundwork for the development of a European view on responsible military use of AI has already been laid. The work conducted by the EDA and the European Parliament, as well as several EU member states, demonstrates a growing political interest. The work initiated by the European Commission on 'trustworthy' AI provides a useful, substantial basis for engaging in conversations on the responsible military use of AI. In this regard the report provides a series of concrete proposals for EU institutions and EU member states to consider how they might engage in an open, methodical and collaborative way of developing a joint perspective on the responsible military use of AI. Notably it suggests that the EU and its member states should hold three tracks of expert discussions on (*a*) legal compliance focusing on legal reviews and challenges posed by AI in the interpretation and application of international law; (*b*) ethics, which focus on the development of shared principles for human-centric military use of AI; and (*c*) AI safety, which focus on standards of transparency, explainability and reliability of AI systems.

There are multiple ways for the EU and its member states to have these expert conversations and work towards developing European standards for the responsible military use of AI. In fact, multiple processes need to be conducted in parallel. The EDA already provides a forum for both high-level political discussion between defence ministers of the member states and technical deliberation between national experts on military AI. Its work on the development of AI capability could be expanded to explicitly include issues related to legal compliance and ethics. The preparatory bodies of the Council of the EU, such as COJUR, CONOP, CODUN and the EUMCWG, also provide forums for states to share their perspectives and potentially agree on joint high-level standards regarding compliance with international law and ethics. Further, the European Parliament provides a forum for public democratic discussion on fundamental ethical issues raised by using AI in the military. Finally, the Euro-

pean Commission, with the could use the EDF to fund projects that explore ways of meeting safety standards for AI at a technical level.

II. Recommendations

In light of the above, this report makes the following recommendations for consideration of EU member states and EU institutions:

European Union member states

Contribute to open reflection on the responsible military use of AI

EU member states could follow the example of the USA and be more open and transparent about their national and intra-European deliberations related to the opportunities and risks of AI for the military. This could entail:

1. Making national reports and strategic documents or summaries of their key content publicly available.
2. Responding to the invitation made to member states during the Finnish Presidency of the Council of the EU in 2019 to draft a concept paper on digitalization and AI in defence.[66] The paper could address not only questions outlined in the food-for-thought paper but also the perspectives of member states on the application of IHL, ethics and safety standards in the use of AI-enabled military capabilities.
3. Allowing the EDA, through the steering board of the defence ministers, to make internally agreed documents and resolutions on AI publicly available as a way of increasing awareness of the EDA's work on the responsible use of AI.
4. Tasking the EDA to create an ad hoc expert group, made up of national experts, to discuss how AI HLEG recommendations may apply in the military context, with the aim of proposing European principles and guidelines for the responsible development and use of AI in the military.

European Union institutions

Council of the European Union: Foster deeper intra-EU discussion on the legal, ethical and technical bases for responsible military use of AI

Use the preparatory bodies of the Council of the EU to foster deeper intra-European discussion by organizing special meetings on:

1. Questions of interpretation and applications of IHL with respect to military uses of AI, which could include conducting legal reviews of weapons, means and methods of warfare via COJUR.
2. Possible ethical principles and safety guidelines for military use of AI by European armed forces via EUMCWG.
3. The EU's position on human control in the debate on LAWS in Geneva and other arms control forums where military use of AI may be discussed via CODUN/CONOP.

[66] Food-for-thought paper by Finland, Estonia, France, Germany and the Netherlands (note 2).

European Defence Agency: Engage with academia and industry

1. Continue and expand the ongoing industry engagement activities to include participation from academia, think tanks and civil society.
2. Use the European AI Alliance to engage broader society, the business sector and civil society on issues related to the development, use and control of military AI.

European Commission: Support research on ethical and safety challenges

The EU Commission could use the EDF to fund research projects on:

1. End-to-end 'ethical design' or methodologies that would allow for ethical issues to be considered throughout the cycle of technology or capability.
2. The transparency, explainability and traceability of military AI systems.
3. Developing a European framework for testing and evaluating military AI systems.
4. Designing methodologies for pooling data at the EU level according to key ethical principles.

The EU Commission could also support the AI HLEG's recommendation to 'develop a cross-cutting network focused on Trustworthy AI across European universities and research institutions' and ensure that the fields of international relations, international law and military ethics are part of such a network.[67]

European Parliament: Provide an open forum for the democratic exchange of perspectives on legal and ethical issues related to the military use of AI

The European Parliament's newly established group on AI should discuss the opportunities and risks generated by using AI in defence and the ways that members of parliament should deal with them in multilateral forums. It should also publicly discuss the legal and ethical challenges raised by the use of AI in military applications.

[67] AI HLEG, 'Policy and investment recommendations for trustworthy AI', 26 June 2019, p. 24.

STOCKHOLM INTERNATIONAL PEACE RESEARCH INSTITUTE
Signalistgatan 9
SE-169 72 Solna, Sweden
Telephone: +46 8 655 97 00
Email: sipri@sipri.org
Internet: www.sipri.org