Lethal Autonomous Weapons Systems and Existing Legal Frameworks

An Analysis of Selected National Regulations and Expert Interviews

Commissioned by: IHEID International Law Department

Abby Naumann
Bram Goede
Shimona Mohan
Shubhangi Priya
Contents

Contents 1
Abbreviations 2
Introduction 3
Background 5
Autonomy and Meaningful Control 7

Part 1: Country Analyses
- France 9
- Israel 10
- Japan 12
- The Netherlands 15
- South Korea 16
- Switzerland 18
- United Kingdom 19
- United States 21

Part 2: Interview Analyses
- Definitional Issues 23
- Sufficiency of Current Regulations on LAWS Development and Attribution of Responsibility 25
- Machine Learning and the Black Box Problem 26
- Potential for Treaties 27

Conclusion 28
Bibliography 30
Appendices 38
Abbreviations

AAR        Association for Aid and Relief in Japan
AI         Artificial Intelligence
AP I       Additional Protocol 1 to the Geneva Conventions
AWS        Autonomous Weapons System
CCW        Convention on Certain Conventional Weapons
C-RAM      Counter rocket, artillery, mortar
DMZ        Demilitarised Zone
DoD        Department of Defence (US)
DoDD       Department of Defence Directive (US)
GGE        Group of Governmental Experts
IHL        International Humanitarian Law
LAWS       Lethal Autonomous Weapons System
LRASM      Long Range Anti Ship Missile
ML         Machine Learning
ROE        Rules of Engagement
SDF        Self Defence Forces
UAV        Unmanned Air Vehicle
WIPO       World Intellectual Property Organization
Introduction

During the late 20th century, artificial intelligence reached a level of sophistication allowing it to be used to confer increasingly autonomous characteristics to weapons systems. During the 21st century this trend continued, and deployment of (partially) autonomous weapons systems in conflicts increased exponentially. These weapons systems range from anti-ship missiles which are independent during the final phases of attack, to new systems such as the AGM-158C long range anti-ship missile (LRASM) which is capable of “conducting autonomous targeting, relying on on-board targeting systems to independently acquire the target without the presence of prior, precision intelligence, to supporting services like Global Positioning Satellite navigation and data-links”[1]

The increasing levels of independence lead to questions regarding the impacts autonomy may have on state accountability and liability for war crimes under international humanitarian law (IHL). To this extent, the LAWS and War Crimes Project seeks to identify gaps and possible solutions in international and national legal frameworks presented by the future development of lethal autonomous weapons systems (LAWS). This project team, an international team of researchers based at the Graduate Institute of International and Development Studies in Geneva and supported by a four-year grant from the Swiss National Science Foundation, is joined by our Capstone Project of graduate researchers.

In legal scholarship there exists a lively debate concerning accountability and its constituting elements regarding war crimes involving LAWS. With regards to International Criminal Law, there is much debate about whether such accountability is currently limited to dolus directus, dolus indirectus, or includes dolus eventualis. Rather than engage in this debate and the hypothetical attribution of responsibility for the war crimes involving LAWS, which are not yet known to exist, our research examines the gaps within the structural attributory mechanisms which underlie the establishment of culpability and intentionality for war crimes involving LAWS in existing national and international legal frameworks.

Through analysing national legislation and exploring the position of scholars and states on accountability for war crimes through semi-structured interviews with professionals in the field of military AI, our team seeks to address and answer the following research questions,

➢ Are current legal frameworks suitable to address issues of attribution for war crimes in the age of artificial intelligence in warfare?
➢ Is there a need to adapt such frameworks and practices in light of the development and deployment of algorithm based new technology in the targeting process and in combat operations based on current and expected use of increasingly autonomous weapons?

We situate this report within the broader scope of the LAWS and War Crimes Project in two parts:

---

In Part One, we present a series of national profiles of states currently developing or employing weapons with various levels of autonomy, or states which are vocal opposers of LAWS development. We examine their national AI use policies, national security policies on LAWS, and the status of AI within the state at this time. These states are France, Israel, Japan, The Netherlands, South Korea, Switzerland, the United Kingdom, and the United States.

In Part Two, we test our findings from the analysis of national standpoints through a series of semi-structured interviews with experts in the areas of military, LAWS, and international law to determine what gaps in the current legal framework may apply to LAWS.
Background

As LAWS do not yet exist, questions remain when defining autonomous weapons and related terms. Different states ascribe different definitions to these concepts, arguably informed by differing views on how these technologies should be regulated. Analyzing state policies towards AI and LAWS specifically allows us to highlight the similarities and differences in such policies.

Definitions: An often found definition of ‘autonomy’ is the ability of a machine to execute a task, or tasks, without human input, using interactions of computer programming with the environment\(^2\). An alternative definition, focussing on technical elements, may refer to the ability of a system to govern its own behaviour as well as actively confront uncertainties in its decision-making process\(^3\). The subtle divergence between these two definitions is typical for the debate surrounding LAWS, where many fundamental definitions have not yet been solidified. During the analysis of country position as well as the later interview, we ourselves also experienced issues regarding unclarity and disagreement of definitions. This report also incorporates the usefulness of these definitions.

Responsibility: The first point to make clear is that LAWS cannot themselves bear legal responsibility. IHL, and rules such as distinction, proportionality and precaution contained within it, address the individuals who use these weapons\(^4\). Consequently, these individuals - namely, operators and commanders - are subjects of possible accountability in the case of a breach of IHL. The increased distance between the person/state accountable for a breach and the act which constitutes the breach raises questions about the effectiveness of the current legal framework. Issues could be envisioned in both legal and practical sense, with actors debating the degree to which current primary norms effectively and comprehensively cover the use of LAWS as well as doubts about our ability to reconstruct and retrace a breach.

The aim of this report, to ascertain whether current legal frameworks and practices are suitable to address issues arising from the development and implementation of LAWS, is a predominantly legal one and we will therefore attempt to appraise the legal relevance of the normative language and rhetoric often found in debates surrounding LAWS.

\(^2\) Boulanin, “Mapping the development of autonomy in weapon systems"
\(^3\) Ibid
\(^4\) Davison, “Autonomous Weapon Systems under International Humanitarian Law”
Autonomy and Meaningful Control

The definition of the terms autonomy and meaningful control create fundamental issues for the international debate surrounding the use of AI in the military. For example, if the threshold for autonomy remains undefined, it becomes unclear which weapons are included in any potential legislation compromising the effectiveness of any agreement reached on the topic. It should also be kept in mind that these are not legal terms but rather tools of rhetoric used to create a norm, which could perhaps be translated into law at a later stage.

Levels of Autonomy: Currently, there are multiple conceptions of ‘autonomy’ in use. An ‘autonomous’ AI system may have a variety of human–machine command-and-control relationships. Systems that require human input at some stage of the task execution can be referred to as ‘semi-autonomous’ or ‘human-in-the-loop’. Systems that can operate independently but are under the oversight of a human who can intervene are called ‘human-supervised autonomous’ or ‘human-on-the-loop’. Machines that could operate completely on their own and where humans are not in a position to intervene are referred to as ‘fully autonomous’ or ‘human-out-of-the-loop’. Although these concepts initially seem clear, one can also find control in the ‘wider loop’, where control is asserted through the setting of boundaries.

The use of ‘loops’ to define human control seems to imply distinct, hierarchical levels of control. However, it is becoming increasingly evident that the level of human control which is required to ensure compliance with IHL differs case by case and should be ascertained on an ad-hoc basis. Statements which call for mandatory ‘human in the loop’, besides the unclarity about the substantive meaning of this term, set an additional requirement in addition to current IHL.

Levels of independence in weapons systems are not a new concept, with torpedoes ‘autonomously’ homing in on the sound of propellers since the second world war. What level of independence (or automation) amounts to ‘autonomous’ thus remains unclear and contested. Autonomy, in the Oxford Dictionary sense of the word means ‘Freedom from external control or influence’. However, control can be exerted in different ways and it is unclear at what level of control the classification of ‘full autonomy’ should apply.

From a legal viewpoint, a commander who employs a weapon should have sufficient control over said deployment to be “reasonable certainty about the effects of an AWS when used in an attack and to be able to limit them as required by IHL”. Although terms such as ‘reasonable’ do not in itself give us much clarity, it is clear that commanders must maintain control, whether direct or indirect, to ensure compliance with IHL. This constitutes the ‘upper limit’ of autonomy. Autonomous weapons of which the behaviour cannot be sufficiently predicted or amended in situ, fail this threshold. The ICRC has confirmed this notion and stated that such systems would be unlawful.

---

5 Ibid.
6 Davidson et al., “Limits of autonomy”
7 International Committee of the Red Cross (ICRC), “Towards limits autonomous weapons”
stated that militaries do not seek autonomous systems beyond control, even if this were allowed under law. However, this later point is again not one of law.

*Article 36*: A large part of ascertaining what level and means of control should be applied can be decided during weapons reviews as prescribed by Article 36 of Additional Protocol I of the Geneva Conventions. These reviews require parties to ascertain whether their weapons will comply with the rules of IHL. This process thus requires, amongst other things, that weapons should not “in some or all circumstances” breach the rules of distinction, proportionality and precaution.\(^8\) This process provides a first hurdle which an AI must pass before being fielded as well as a tool which can be used for establishing state accountability.

*Retrace Efforts*: Scholars suggest guidance control as a strategy to ensure that all actions can be traced back to human considerations as well as designing fail-saves around systems to minimise the chances of errors with grave consequence.\(^9\) The specific features of such systems are however still unclear and methods to monitor and retrace efforts are undeveloped. As AI, and digital technologies in general, have developed rapidly over the last decades, we note that we may not know what LAWS may look like in the future, nor what capacities for monitoring may accompany them.

Several complicating factors at play in a study of the legal framework surrounding LAWS have become clear. Firstly, we are uncertain of what technologies may develop. Secondly, there is a large amount of uncertainty concerning employed terminology. Thirdly, different actors engage with this topic, and they may have different interests. Due to this complex landscape, our team decided to limit our research to an appraisal of national positions. Given the important position of states in this debate, national policy regarding autonomous or semi-autonomous systems can provide valuable insight into how LAWS are controlled and who can be held accountable for their use.

---

\(^8\) International Committee of the Red Cross (ICRC), “Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I)”

\(^9\) Santoni de Sio, “Meaningful human control over autonomous systems: A philosophical account”
Part 1: Country Analyses

France

The French strategy towards the development and use of artificial intelligence considers France as a significant driver of the emerging AI revolution in Europe. The French approach is based on geopolitical and security concerns in the face of competition from the United States and China, and takes into account the increasing need for a national blueprint for action regarding the development and use of AI. In 2018, the French Ministry of Defence announced its plans to devote around 100 million euros to AI research, in an expert paper drafted by Cédric Villani, a member of the French National Assembly. Villani acknowledged the pivotal role of the French State in steering the integration of AI into public policy and ensuring transparency and regulation of AI systems to address ethical concerns emanating from the use of AI.\(^\text{10}\) While there has been agreement regarding increased investment into AI R&D to strengthen the efficacy of security missions, establish strategic dominance over opponents, and maintaining France’s military position compared to its allies; Villani and Florence Parly, the French Minister for Armed Forces, have regularly emphasised the criticality of human responsibility and opposed giving AI the decision-making power to make choices regarding life and death. What has been proposed, instead, is the development of a ‘civil-military complex of technological innovation’, which entails the synthesis of enormous amounts of data to make strategically effective decisions.\(^\text{11}\)

France had introduced the issue of regulation of military AI to the United Nations in 2013 in the framework of the CCW in order to establish a multilateral approach towards the regulation of LAWS. However, French officials frequently stated in 2018 that it may be premature to ban LAWS before they are even in use.\(^\text{12}\) Florence Parly presented France’s military AI strategy through the 2019 Report of the AI Task Force which focuses on the controlled use of AI in the defence sector. The strategy also called for the creation of a ministerial committee to decide matters related to the ethical issues of military AI. Moreover, the report stressed the significance of establishing capability roadmaps to guarantee the controlled integration of AI into the French Armed Forces.\(^\text{13}\) Parly also highlighted in her speech that France would not be involved in developing killer robots or other fully autonomous weapons for military use.

In 2018, the French Director General of Armament (DGA), France’s technology and arms procurement agency, approved the creation of the Defence Innovation Agency (DIA) in order to integrate decision-making in all the actors of the ministry to bolster defence innovation. Not only is the DIA geared towards strengthening cooperation between the various national

\(^{10}\) Villani, Bonnet, & Rondepierre, “Donner un sens à l'intelligence artificielle : pour une stratégie nationale et européenne”

\(^{11}\) ibid.

\(^{12}\) Delcker, “France, Germany under fire for failing to back ‘killer robots’ ban”

bodies involved in defence innovation, it is also responsible for coordinating and implementing ministerial policies concerning defence innovation across public and private actors.\textsuperscript{14} Further, its creation highlights the importance of governmental institutions aimed at regulation of autonomous weapons which arises in the usage of military AI. Other non-governmental stakeholders include the French arms producer Dassault, which has been developing ‘nEUron’, which is designed to function as an unmanned combat air system and has the ability to autonomously fly for more than three hours. Additionally, arms production company Thales is in the process of drafting a charter of ethics on the development and use of military AI. Thales has publicly announced that it would not engage in developing autonomous killer robots.\textsuperscript{15}

Israel

Israel is aware that artificial intelligence is a key technology in both defence and the economic spheres, but did not become attuned to the advantages of AI in the early stages of its development. Israel is now entering the race with redoubled force. According to the recommendations of a committee formed at the request of Prime Minister Benjamin Netanyahu, Israel aspires to become one of the world's five leading countries in AI by harnessing the potential of vast government investments.\textsuperscript{16} In terms of its views on LAWS, Israel is the only other country apart from the United States which has stated that such systems may offer benefits worth exploring\textsuperscript{17}, a belief that is consolidated by its unique and challenging geopolitical positioning and its long history of experimentation and innovation in weapons systems.

Israel is a manufacturer and exporter of advanced weapon systems, the world’s leading exporter of unmanned aerial systems and, the only other country apart from South Korea that makes use of robotic Sentry guns.\textsuperscript{18} Its tryst with autonomous weapons is rooted in the development of its combined drone and anti-radar missile loitering munition called Harpy by Israel Aerospace Industries (IAI) in 1990. While originally built as a UAV which could only target radars, regular upgrades of the Harpy resulted in an autonomous drone that could see and target anything on the battlefield, not just radars\textsuperscript{19}, and function on a fire-and-forget basis.\textsuperscript{20} Harpy’s upgraded version, Harpy 2 or Harop, is an autonomous loitering munition which can dive-bomb radar signals, capable of operating either fully autonomously or under human direction. Harop has reportedly already been used with lethal results on the battlefield. In 2016, it allegedly led to the death of seven people in the disputed Nagorno-Karabakh region, but this has neither been confirmed nor challenged in a court of law.\textsuperscript{21} Apart from UAVs, the Israeli Defence Forces also use target identification systems like Trophy and

\textsuperscript{14} French National Defence Website.
\textsuperscript{15} “State Of AI: Artificial intelligence, the military and increasingly autonomous weapons”
\textsuperscript{16} Berkovitz, “Israel's national AI plan unveiled”
\textsuperscript{17} Sauer, “October 2016 | Arms Control Association”
\textsuperscript{18} Antebi, “The International Process to Limit Autonomous Weapon Systems: Significance for Israel”
\textsuperscript{19} Gao, "The Ultimate Weapon Of War No One Is Talking About"
\textsuperscript{20} "HARPY: Autonomous Weapons for All Weather"
\textsuperscript{21} Perrigo, "A Global Arms Race For Killer Robots Is Transforming Warfare"
holistic defence systems like the Iron Dome, which can identify, prioritise and engage with material (not personnel) targets autonomously. Other autonomous weapons like the Orbiter 1K, Fire Weaver, Rafael Systems’ SPICE bombs, etc are also in use by the IDF.

Israeli policymakers have not yet fully determined how to regulate AI inventions, but efforts are being made to establish a coherent governmental policy on AI. Ethical and institutional approaches will likely determine how AI technology will be governed, and an ethics by design approach is likely to be adopted. In general, the Israel Patent Office (ILPO) considers AI implemented inventions as any software or computer-implemented inventions for purposes of patentable subject matter. In specific terms of AI use in the military, Israel stated to the CCW GGE in 2018 that it supports in-depth discussions in consideration of any possible regulation of LAWS, but such discussions should not hamper progress in civilian research and development of autonomous technologies. It acknowledges that there are differences of opinion regarding definitions or characterisation of LAWS. Furthermore, Israel is of the opinion that the appropriate type and level of human judgment, and the increasing use of human-machine teaming can be discussed prudently.

Israel opposes any discussion on amending the current regulatory framework governing the use of new weapons systems. It also disagreed with the reference to autonomous weapons as systems that ‘make decisions by themselves’. Its position is that all weapons, including autonomous weapons, are operated by human beings. Israel argues that at the research and development stage, human beings have to take account of operational scenarios and obey the laws of warfare. At the planning and operating stage, the commander is responsible for ensuring that their use complies with international law, and if necessary, limiting the use of the systems if they conflict with the currently stipulated law. In view of the security challenges it faces, Israel is likely to maintain its right to develop and use various systems in self defence, including systems based on autonomy and AI, and refuse to participate in the conception of legal instruments which may propose pre-emptive bans or heavy restrictions on LAWS.

Japan

The AI environment in Japan is characterised by an economic focus, proactive policy-making, and an engagement with civil society. Although Japanese civil society voices anti-LAWS sentiments, the official AI policies of the Japanese government focus on economic and labour market AI concerns, reserving few words for security or human rights concerns, and omitting LAWS in national security policy.

22 Boulanin & Verbruggen, “Mapping the Development of Autonomy in Weapons Systems”
23 Frantzman, "Israel Finds An AI System To Help Fight In Cities"
24 Renaud, “Emerging Legal Trends AI: Can Israel Join The U.S. And Europe As A Leader In AI Protections?”
25 “Regulation Of Artificial Intelligence: Middle East and North Africa”
26 ibid.
27 ibid.
Japan’s approach to AI is detailed in a 2017 document issued by the Japanese Cabinet: “Declaration to be the World’s Most Advanced IT Nation and Basic Plan for the Advancement of Public and Private Sector Utilization.” This plan, adopted by the Japanese Diet in 2018, builds upon decades of Japanese IT-focus, dating back to the e-Japan Strategy and e-Japan Strategy II created by the fledgling Strategic Headquarters for the Promotion of an Advanced Information and Telecommunications Network Society in 2001 and 2003, respectively. As early as 2013, the Strategic Headquarters authored its own declaration, indicating that the declaration adopted in 2018 had been in discussion for at least five years.28

The 2017 Declaration sparked a wave of Japanese responsible-use AI policymaking, which was referred to as “human-centric AI.”29 In July 2017, the Ministry of Internal Affairs and Communications facilitated the Conference toward AI Network Society. This conference included professors from esteemed universities such as the University of Tokyo, Keio University, and Kyoto University. Private sector participants included top-level stakeholders from leading tech companies. The conference pointed towards a loosely-governed, economically-focused AI policy without mention of autonomous weapons. Instead, it toted a governance strategy of “a non-regulatory and non-binding approach (soft law).” The conference introduced to Japan “Wisdom Network Society” (WINS), a concept of human-centre, AI-assisted society emphasizing the involvement of stakeholders, collaboration, transparency, and “all people’s enjoyment of benefits.” Yet it made no mention of defence or international governance of AI in the preliminary assessment by usage field. However, under the research and development principles, the conference highlighted that, it is advisable that developers take precautions to ensure that AI systems do not unduly infringe the value of humanity, based on the International Human Rights Law and IHL.” Despite the possibility of this principle to include a ban on AI usage in weapons systems, it is more likely that this inclusion was meant to address issues of data-use privacy and the labour market, both identified in the “human-related issues” under the future challenges chapter.30

However, after the Diet adoption of the Declaration in 2018, Japanese AI policy began to directly address limited security issues. Stressing the importance of private sector involvement and multistakeholder initiatives, contributions from Japanese corporations became significant. The AI Strategy Expert Meeting for Strength and Promotion of the Innovation Decision in February 2019 issued an expansion on the definition of human-centric AI. Again, this conference included academics from esteemed universities but also a higher number of private sector experts such as representatives from Hitachi, Araya, and Sony Computer Laboratories. In addition to highlighting AI education and literacy, data privacy, and fair competition, the AI Strategy experts also included a principle of “ensuring security.” While not naming LAWS specifically, the experts cite the importance of risk assessment, improving safety, and acknowledging the risks of AI to security.31

---

28 “Declaration to be the World’s Most Advanced IT Nation”
29 Hirano, “AI R&D Guideline”
30 “Report 2017 Toward Promotion of International Discussions on AI Networking.”
31 “Social Principles of Human-Centric AI”
Meanwhile, civil society actors gathered in February 2019 to explicitly address LAWS in a conference in Tokyo hosted by the Association for Aid and Relief in Japan (AAR) and the Campaign to Stop Killer Robots, titled “Asia Pacific Regional Conference on Lethal Autonomous Weapons Systems.” A starting point for discussion came from a recording of Undersecretary General and High Representative of Disarmament Affairs of the United Nations Izumi Nakamitsu who had highlighted the potential for LAWS to spur a new arms race. The result of the conference was the issuance of the “Tokyo Statement,” a communique emphasising the regional support in East Asia for an international ban on LAWS, supported by AAR.32 For its part, AAR has been involved in the movement to pre-emptively ban LAWS before the Diet adopted AI strategies, with AAR representative Yukie Osa presenting a statement to the Convention on Conventional Weapons Expert Meeting (CCW GGE) in Geneva in 2016 (Osa, Yukie, 2016).

Perhaps reflecting this civil society inclusion of AI security risks, the G20 Summit in June 2019 held in Tsukuba City, Japan and led by Japanese ministers released a statement expanding the AI conversation in Japanese government. In the AI Principles formulated at the G20, human-centred AI now included the necessity for AI to “respect the rule of law, human rights and democratic values” and to “implement mechanisms and safeguards such as the capacity for human determination,” which seems to express some measure of meaningful human control.33 However, no specific mention was given to LAWS.

At present, the Japanese government does not adequately address AI in security policy outside the issues of data protection and cyberattacks. In the 2020 “Questionnaire on Artificial Intelligence Policies” issued by WIPO, the Japanese delegation lists the “AI Strategy 2019” as its official AI strategy.34 The “AI Strategy 2019: AI for Everyone: People, Regions, Industries, and Governments,” issued in June 2019 does not, however, address autonomous weapons, instead only listing cyberattacks in the brief security section.35 Similarly, the 2013 National Security Strategy also makes mention of cybersecurity and the continued need for Japanese self-defence, however, autonomous or semi-autonomous weapon systems are not discussed.36

The Japanese government may participate in international talks on autonomous weapons, however, they do not support the negotiation of a new treaty addressing the issue, possibly due to pressure from the US.37 In fact, Official Secretary in the Japanese House of Representatives Daisuke Akimoto wrote in The Diplomat in December 2019 that although Japanese ministers have voiced concern over the definition of ‘meaningful human control’ and LAWS, in addition to stating Japan would not develop LAWS at forums such as the CCW, the Japanese government has shied away from discussion of treaties. According to

32 “Towards a World Without Killer Robots (キラーロボットのない世界に)”
33 “G20 Ministerial Statement on Trade and Digital Economy.”
34 “Questionnaire on Artificial Intelligence Policies: Japan.”
36 “National Security Strategy of Japan”
37 HRW, “Japan: Retain Human Control Over the Use of Force”
Akimoto, pressure from the US may stem from AI technology contracts which Japan seeks in research and development with the US.\textsuperscript{38}

Going forward, the question remains as to whether Japan will engage with legal frameworks banning the development and use of LAWS. Japanese Security Strategy struggles with balancing its relationship with the United States and the threat of North Korea with Japan’s image as a ‘peace-loving society,’ alongside the traditional restriction of Article 9 of the Japanese Constitution. Development of LAWS could shatter Japan’s image built on international cooperation, official development assistance, and a limited role for the Self Defence Force, although the SDF already employs and develops artificial intelligence systems for defence in collaboration with the US military.\textsuperscript{39}

At the same time, for legal scholars, Japanese policy may already show signs of applicable anti-LAWS regulation. For example, the 2018 Conference on Principles of Human-centric AI Society published proposed criminal liabilities for traffic accidents caused by automated driving cars. This framework employed five levels of autonomy/responsibility, beginning with levels 0-2 where automated functions exist but a human driver remains primarily in control. At level 3, the system controls most driving functions but requests the human driver to approve actions, requiring the human to remain concentrated on the road. At level 4 and 5, the vehicle performs all operations, and the human driver is not required to concentrate on the road or anticipate accidents. At level 4 and 5, criminal liability is removed from the human driver.\textsuperscript{40} Although weapons are not discussed in this framework, this policy already includes categorizations of control and potential criminal liability, which could be echoed in frameworks applied to liability in the use of LAWS.

Overall, it is likely that Japanese civil society will remain anti-LAWS and continually push for stronger international cooperation in this area, while the willingness of the government to apply unique frameworks to LAWS remains unclear so long as the US-Japan defence apparatus, and therefore AI for defence cooperation, remains in place.

The Netherlands

The Netherlands sees great potential in AI for civilian uses and wishes to invest in building an economy based around this.\textsuperscript{41} However, beyond the entrepreneurial spirit which may drive investment into civilian uses of AI, the Netherlands seems unlikely to take the lead in the development of AI for military applications.

The Dutch navy does not possess weapons with large degrees of autonomy such as the AGM-158C LRASM and instead relies on RIM-162 Evolved Sea Sparrow Missiles (ESSM) and Standard Missile 2 Block IIIA (SM-2), which function through course corrections sent to

\textsuperscript{38} Daisuke, “Clash of Killer Robots? Japan’s Role in Preventing AI Apocalypse”
\textsuperscript{40} Matsuda, “Japan. In AI, Machine Learning, & Big Data (2nd ed.)”
\textsuperscript{41} Adviesraad Internationale Vraagstukken, "Strategisch Actieplan voor Artificiële Intelligentie" (Strategic Action Plan for Artificial intelligence)
missiles and target illumination by an external radar. However, this image is set to change as the Dutch Minister of Defence has started a procurement procedure to replace the aging harpoon missile with a new long range surface-to-surface missile independent from external guidance after firing (so-called fire-and-forget). Although automatic modes have been present in defence systems such as the MIM-104 patriot air defence missile, the new weapon would be the first offensive weapons system which can be regarded as possessing significant levels of autonomy, implying that the Dutch military is not shying away from increasing levels of autonomy in its weapons systems. When such a missile is incorporated into the armoury of the Dutch armed forces, it may raise questions about the readiness of domestic legal frameworks.

The Netherlands’ formal position on LAWS is informed by an advisory paper published in 2015 by the Adviesraad Internationale Vraagstukken, an advisory service of the Dutch government. The main conclusion in this report is that meaningful human control is essential for the use of autonomous weapons, and is informed by Article 36 of Additional Protocol I to the Geneva Conventions, which states that during the “study, development, acquisition or adoption of a new weapon, means or method of warfare” parties have an obligation to ascertain whether the deployment of the weapon could be prohibited by AP I or any other applicable international law.

Due to the independent decision-making processes associated with autonomy, states arguably cannot adequately assess whether weapons will comply with the aforementioned international laws. In order to maintain compliance with international law, the Dutch advocate continued involvement of human operators in target acquisition, selection of weapons and rules as set out in IHL. When control is maintained in these areas, the government considers the autonomous weapon under ‘meaningful control’. This requirement thus does not need ‘in the loop’ or ‘on the loop’ control, but sets a different (and perhaps lower) threshold of control on the ‘wider loop’ where meaningful control is exercised through the setting of boundaries in which a system can function. The Dutch delegation to the 2020 CCW GGE on Lethal Autonomous Weapons restated this standpoint:  

“In the Netherlands’ view, the targeting process enables humans to perform the necessary assessments regarding IHL, in particular the principles of distinction, proportionality and precautions. Deployed under such conditions and with due regard to existing processes, the existing legal regime is adequate to ensure IHL compliance and assign accountability.”

Such a sentiment can also be found in domestic situations in the Netherlands. Fines were upheld when people caught using their phones behind the steering wheel of a car on ‘autopilot’ challenged the notion that they were the ‘driver’ of the vehicle. The underlying

---

42 Ministerie van Defensie, “Behoefstelling van het project” (Requirements for the project)
44 International Committee of the Red Cross (ICRC), “Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I)”
45 Vogelaar, R. “Statement Of The Netherlands At Group Of Governmental Experts On LAWS”
reasoning was that they initiated the autopilot and enabled its continued functioning.\textsuperscript{46-48} In this case, the court thus designated the individual responsible due to prior actions rather than their function at the time of using their phone. Scholars also do not expect the use of AI in financial services such as investment to pose an issue due to the ability to address concerns about accountability through private law.

The Dutch government rejects the development of fully autonomous weapons without meaningful control.\textsuperscript{49} However, they also disagree with the creation of a moratorium on the development of such weapons, based on doubts about the desirability and feasibility of a blanket ban due to the dual-use nature of technology in these systems, and the idea that methods of control will develop in tandem with increasing autonomy. A moratorium could preclude such developments. The advisory paper does set itself a temporal bound, only commenting on developments envisaged between 2015-2025, which should be kept in mind when judging the amount of caution described.

**South Korea**

South Korea has been open about its AI ambitions, owing in part to its cemented reputation as an IT hub, and has been vocal about its interest in the use of AI in the military. It highlighted its support towards the consideration of the advantages and merits of LAWS as opposed to just the risks of their development, claiming that it is ‘premature’ to stipulate new binding regulations that prohibit LAWS, and discussions must continue focusing on promoting a common understanding among states on the weapons systems using autonomous technology at present.\textsuperscript{50} In its National Strategy for Artificial Intelligence released in 2019, South Korea has detailed its aims to create an environment of encouragement for AI development, abiding by the principle of ‘approve first, regulate later’.\textsuperscript{51} To this effect, a large number of private endeavours in the form of AI R&D and other commercialisation activities have been deregulated by the South Korean, and an investment of approximately 1 trillion KRW has been made.\textsuperscript{52}

Apart from its general interest in AI development for civilian purposes, for instance chatbots for the care of its elderly populace like Dasomi (다소미)\textsuperscript{53}, South Korea’s precarious position with respect to its hostile northern counterpart puts pressure on its government to continually upstage the military capabilities of the latter. While it remains wary of fully autonomous weapons systems that remove meaningful human control from the operation loop, due to the

\textsuperscript{46} Gerechtshof Arnhem-Leeuwarden "ECLI:NL:GHARL:2019:6122"
\textsuperscript{47} Gerechtshof Midden-Nederland"ECLI:NL:RBME:2018:5708"
\textsuperscript{48} Gerechtshof Midden-Nederland"ECLI:NL:RBME:2018:5226"
\textsuperscript{49} Ministerie van Defensie “Beantwoording Kamervragen over bericht dat het Europees Parlement killerrobots wil verbieden” (responses to questions raised in the Second Chamber about European Parliament’s intent to ban Killer Robots)
\textsuperscript{50} Park, “Analysis of the Positions Held by Countries on Legal Issues of Lethal Autonomous Weapons Systems and Proper Domestic Policy Direction of South Korea”
\textsuperscript{51} The Government of the Republic of Korea, “National Strategy for Artificial Intelligence”
\textsuperscript{52} Kim, “Korean Government to Repeal Regulations in AI Industry”
\textsuperscript{53} Choi, “Why Covid-19 only accelerates South Korea's AI ambitions"
risk of malfunctioning, potential accountability gap, and ethical concerns\textsuperscript{54}, it has developed weapons with certain degrees of autonomy and either deployed them in the Demilitarised Zone (DMZ), or sold them to interested states.

An example of this is the Super aEgis-II, an automated gun turret that can detect and lock onto human-sized targets from 2.2 kilometres away with the help of a camera and infra-red sensors, developed by Korean military hardware manufacturer DoDAMM in 2010. While it can operate in fully autonomous mode for targeting and engagement purposes, it can also operate in manual mode for more human intervention.\textsuperscript{55} Although it is not clear whether these have been positioned at the DMZ, 30 of these units had been sold to numerous locations in the Middle East as of 2015,\textsuperscript{56} highlighting how the AI weapons race has doubled up as an opportunity for South Korea to profit economically and strategically through diplomatic ties. Another notable example is the SGR-A1 sentry gun developed by Samsung and currently deployed at the DMZ. It has detectors and pattern recognition software that enables it to spot human targets up to two miles, can issue verbal warnings and recognise surrender motions, and may also engage them with a machine gun from up to 800 metres away. Purportedly, the SGR-A1 functions on a highly autonomous human-on-the-loop system, even though the developers insist that a human command is requisite for the machine to fire.\textsuperscript{57}

There are currently debates over the compensation of loss or harm incurred by AI systems in terms of whether existing civil law or criminal law can cover these systems or if they require particular emphasis, since neither AI systems, nor their creators, nor their users can be held liable under the existing liability clauses. Under the conventional criminal liability principle in South Korea, a natural person can be held criminally liable, and curiously, the development of strong or super AI may render it almost equal to a natural person in the eyes of the law. However, AI of that calibre is not a plausible development of the near future, so AI itself cannot be penalised. It is also not possible to penalise the creator, since criminal actions of AI systems may be autonomously performed, either intentionally or unintentionally, and may be out of the creator’s control. Overseers of AI use also cannot be held liable since AI in its present form cannot neatly fit under the definitions of either employee or product. The Korean government seems to be aware of this conundrum and is looking to establish rules to prohibit and ascribe responsibility for criminal use.\textsuperscript{58}

In light of its development of autonomous weapons, South Korea clarifies that it is researching, developing, and investing in military applications of AI and weapons systems with autonomy in their functions, but says it does not possess LAWS and does not intend to develop or acquire them.\textsuperscript{59} This was brought to question in 2018 when Korea Advanced Institute of Science and Technology (KAIST) and defence manufacturer Hanwha Systems were reported to be partnering together, with AI experts boycotting them over claims that this

\textsuperscript{54} "Stopping Killer Robots"
\textsuperscript{55} Blain, "South Korea's Autonomous Robot Gun Turrets: Deadly From Kilometers Away".
\textsuperscript{56} Parkin, "Killer Robots: The Soldiers That Never Sleep"
\textsuperscript{57} Velez-Green, “The Foreign Policy Essay: The South Korean Sentry—A ‘Killer Robot’ to Prevent War”
\textsuperscript{58} Cho & Lee, “AI, Machine Learning & Big Data Laws and Regulations | Korea”
\textsuperscript{59} "Stopping Killer Robots"
partnership would result in the unethical development of killer robots. KAIST professors were quick to deny such claims on the grounds that KAIST would not conduct any research activities that counter human dignity including autonomous weapons lacking meaningful human control. Comments by the Korean government on this issue as well as other developments in the arena of AWS in South Korea have remained conspicuously absent, although their evergreen stance remains one of careful diplomacy.

**Switzerland**

Switzerland’s approach to AI has been detailed in its 2018 digitisation strategy, which focuses on a multi-stakeholder approach towards inclusive and democratic data use and distribution to contribute to a knowledge-based society. The strategy elaborates upon safeguarding critical Swiss systems against risk in cyberspace, and discusses the use of AI only in terms of transparency and verifiability of algorithmic decision-making systems according to prevailing legislations, but does not address the military implications of the use of AI and corresponding regulations.

As an increasing amount of investment is allocated towards military AI R&D, Switzerland has sought to reaffirm the criticality of adhering to IHL and align its military AI development strategy with international norms by utilising a ‘compliance-based’ approach. Switzerland has consistently stressed that exerting human control is essential to embed these systems’ autonomous actions within an acceptable framework of operation to restrict their criteria of engagement. Heavy investment in developing cutting-edge military technology, most notably unmanned aerial vehicles, has also brought to light Switzerland’s prowess in leading innovation in autonomous weapons. The adoption of AI in developing military weapons has witnessed a surge of demand over the past few years and has shaped the business strategy of companies such as Pix4D and SenseFly, which have brought forth the specialisation and expertise essential to design innovative defence and security solutions.

AI is used in Switzerland for military purposes, but without specific national legislations pertaining to the conduct of AI for machine learning or big data. Under the Swiss Criminal Code, there are currently no particular provisions related to felonies committed by AI. It also necessitates the personal culpability of the offender. In the event that an AI robot commits a felony or causes a third party to commit a crime, it will not be held liable under traditional Swiss criminal law. However, if the AI was initially programmed to commit a crime, the creator of the AI will be criminally liable. In addition, even though Swiss Foreign Minister Ignazio Cassis declared that the country would not partake in developing or acquiring LAWS, Switzerland’s approach to LAWS has been less structured than other European countries. In

---

60 Haas 2018.  
61 Swiss Federal Office of Communications, “Digital Switzerland Strategy”  
63 Pax for Peace, “Crunch Time: European positions on lethal autonomous weapon systems”  
64 Fortune Business Insights "Drone Software Market Size To Hit USD 5,961.6 Million By 2027; Increasing Adoption Of AI Technology To Surge Demand, Says Fortune Business Insights™"  
65 Gordon & Gurovitz, “AI, Machine Learning & Big Data”
2017, the Swiss Federal Council decided that it would not call for pre-emptive bans on autonomous weapons, as such technology may prove to be useful in the future, as well as help in avoiding collateral damage resulting out of armed conflict. Even so, Switzerland has called for some form of regulation of autonomy in weapon systems. It presented its working paper to the CCW GGE, wherein it delineated a preliminary working definition of LAWS and stressed on compliance with existing IHL frameworks in the event of armed conflict in order to determine criminal culpability.

United Kingdom

At the CCW GGE meeting in 2019, the United Kingdom (UK) restated its alignment with the European Union and its member states, supporting the idea that accountability cannot be ascribed to a machine. The UK does not currently have any human out-of-the-loop weapons systems and has no intention of attaining them. However, given the current vague nature of terms such as human-in-the-loop, out-of-the-loop or on-the-loop, this statement does not carry too much weight. An example of the extent to which a grey area exists is the use of some Counter Rocket, Artillery and Mortars systems (C-RAM) in automatic mode and the use of weapons systems with the ability to conduct ‘terminal control’ where the weapon is independent in the last stages of attack.

In accordance with the general EU position, it is the UK’s position that current and near-future operationalisation of autonomous weapons systems will not suffer from an accountability gap due to the sufficiency of existing legal frameworks. This position was reiterated at the 2020 CCW GGE on Lethal Autonomous weapons where the UK’s representative stated the following:

“Should a violation of IHL result from the operation of a weapon or weapon system, processes are already in place to conduct appropriate investigations and, if applicable, apportion responsibility. Legal accountability will always devolve to a human being, never a machine – increasing autonomy in weapons or weapons systems does not therefore present the risk of an accountability gap.”

In a working paper submitted by the UK to the 2018 CCW GGE, it further detailed this point stating that due to the “technical standards governing their development, observance of national and international legal obligations and rigorous field testing procedures.”, accountability will always devolve to the operators and is discharged through the chain of command.

66 Zerbe, “Switzerland’s dangerous neutrality on «Killer Robots»”
67 ibid.
68 United Kingdom. “Statement by the United Kingdom at CCW GGE on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems.”
69 Idem
70 United Kingdom. “Human Machine Touchpoints: The United Kingdom’s perspective on human control over weapon development and targeting cycles”
As LAWS and the debate surrounding them are still in their infancy, the UK wishes to withhold from initiating the development of extra legislation.\(^71\) It also sees practical difficulties with preventative legislation because a ban on the development of LAWS is not likely to be accepted and enforced by all countries and actors, making it infeasible. Lastly, it should be noted that the UK seems interested in such developments and has awarded millions of pounds to projects researching use of AI in the armed forces through its Defence and Security Accelerator.\(^72\)

The UK’s understanding of LAWS is subtly different from the others. The definition of an autonomous system involves understanding higher-level intent and direction, leading to independent action to bring about a desired state. These systems may have a human-on-the-loop.\(^73\) This definition, which has a higher threshold for ascribing the term for ‘autonomy’ than the definition employed in this report, may be informed by the UK’s stance on the use of machines in the battlefield. In a working paper published in 2018, the UK outlined the value machines may have, stressing that “within tightly defined circumstances and in response to a specific problem, machines may make better decisions than a human.”\(^74\) The UK thus seems to heighten the threshold for typifying machines as ‘autonomous’ whilst also pushing back against the development of superfluous legislation.

**United States**

The United States insists that LAWS do not exist, according to a 2019 Congressional report, and any policies applicable to LAWS are pre-emptive. The US Congress defines these weapons as “weapons systems that once activated can select and engage targets without further intervention by a human operator.” Congress also insists that the United States is not currently developing LAWS, although the US does not support a ban on LAWS, citing potential humanitarian benefits of robotic weapons as opposed to human-piloted weapons.\(^75\)

In DoDD 3000.09 (2012, 2017) the US Department of Defense (DoD) details the United States’ policy on LAWS, including applicability, restrictions on system design and testing, responsibility of commanders and operators, and the potential exportation of LAWS. This directive requires that systems go through rigorous hardware and software testing to ensure that the systems function in a realistic environment and minimize failures. Such systems must also be safe from potential tampering. Operators of fully and semi-autonomous weapons must be adequately trained in the doctrine, tactics, and techniques of such systems, including applicable IHL treaties and rules of engagement (ROE). Finally, international sales or

---

\(^71\) United Kingdom. “Statement On Lethal Autonomous Weapons Systems To The CCW Meeting Of The High Contracting Parties”

\(^72\) Defence and Security Accelerator, “Innovation for a Safer Future Defence and Security Accelerator (DASA).”

\(^73\) Ministry of Defence. "JDP 0-30.2, Unmanned Aircraft Systems (Change 1)."

\(^74\) *Idem*

transfers of such systems would be approved in accordance with existing requirements and processes applying to other United States weapons systems, explained in DoDD 5111.21.\textsuperscript{76}

The Stockholm International Peace Research Institute (SIPRI) in a 2017 report described the US as the most visible, articulated, and successful research and development state in autonomous weapons technology. Increasingly autonomous technology is a central feature of the US military modernization plans and strategic capabilities.\textsuperscript{77} Current semi-automated defence systems include the counter-rocket, artillery, and mortar (C-RAM) systems, such as the Phalanx Close-in Weapons System (CIWS) used by the US Navy and Land Phalanx Weapons System (LPWS).\textsuperscript{78} In 2016, the US Military released the first prototype of the Sea Hunter, an unmanned vessel that finds and destroys enemy submarines and mines.\textsuperscript{79} In 2013, the U.S. Navy successfully landed an autonomous drone called the X-47B Unmanned Combat Air System (UCAS) on the aircraft carrier USS George H.W. Bush. The X-47B is already equipped with an internal weapons bay, although it does not have fully autonomous offensive capability.\textsuperscript{80}

According to Judge Advocate for the U.S. Air Force and professor Major Jason S. DeSon, the US Air Force envisions certain types of missions as suitable for fully autonomous weapons. Currently, human operators control remotely piloted aircraft (RPA), and these operators are usually pilots who formerly flew manned aircraft such as the F-16 Viper (Fighting Falcon) and F-15 Eagle. After the transition to fully autonomous unmanned aircraft, these systems might accomplish similar tasks as the human pilots. Major DeSon argues that both counter air and counter land missions could benefit from unmanned aircrafts. These might include any currently human-piloted missions, including attack operations, fighter sweep, escort, and suppression of enemy air defences (SEAD). The unmanned aircraft would target and attack enemy aircraft already in the air or surface-based air defences. Counter land missions for unmanned aircraft might include close air support (CAS) where the danger of fratricide is particularly perilous. Such missions would not differ greatly in command from normal human-piloted missions, as in both cases of manned and unmanned aircraft the human or robotic pilot relies on constant feedback from the ground controller.\textsuperscript{81}

The United States has no intention of slowing or limiting American research and development of LAWS, which the US sees as a potentially humanitarian technology. Expert interviewees indicate that any potential learning curves associated with new technology are ultimately worth the ‘growing pains’ which may come with unforeseen consequences. While DOD Directive 3000.09 does not elucidate what a ‘realistic environment’ for testing an autonomous weapon might be, US military experts place their confidence in the review process, stressing the fact that lethal autonomous weapons, per the directive, must undergo

\textsuperscript{76} DoD Directive 3000.09
\textsuperscript{77} Boulainin & Verbrugge, “Mapping the Development of Autonomy in Weapons Systems”
\textsuperscript{78} Atlamazoglou, “Meet C-RAM”
\textsuperscript{79} Stewart, “U.S. military christens self-driving 'Sea Hunter' warship”
\textsuperscript{80} DeSon, “Automating the Right Stuff”
\textsuperscript{81} DoDD 3000.09
additional high-level authorization, and that if at any point that the system no longer appears to operate in its intended use, the system would be pulled from operations and re-evaluated.\textsuperscript{82}

Mitigation measures may include temporal restraints and geographical restraints which cause the system to return to base, shut down, or self-destruct when it exceeds its bounds. A system may also return to base, shut down, or self-destruct when a target cannot be found, the system senses a malfunction, or a communication link is lost with the base. These measures are critical in anti-hacking protection.\textsuperscript{83} In addition, US DOD ethical principles not only prioritize responsible, traceable, reliable, and governable aspects, but also equitable, specifically noting the need to minimize potential bias in AI capabilities.\textsuperscript{84}

Furthermore, responsibility for the appropriate use of systems already lies with the commander within the US chain of command. While a commander may not be held responsible if a system malfunction was determined not to have arisen due to any inappropriate direction from the commander, even in the case where the operator has made an error, the commander may be held accountable within the US military structure. Although this may give rise to the question of how often or how accurately commanders and operators are held accountable for misdeeds, this particular concern is not unique to lethal autonomous weapons and already exists for the use of conventional weapons. For this reason, US military experts are not concerned with attribution of responsibility.\textsuperscript{85}

\textsuperscript{82} Interview with an unnamed military expert
\textsuperscript{83} Interview with an unnamed military expert
\textsuperscript{84} Panda, "US Department Of Defense Adopts Artificial Intelligence Ethical Principles"
\textsuperscript{85} Interview with an unnamed military expert
Part 2: Interviews with Experts and Practitioners

In this section, we compare the various issues brought to light in the country analysis through six interviews which we conducted to gain insight into the practical implications of LAWS. Our interviewees ranged from academic experts to international lawyers and military persons with experience in the field. In our discussions, we aimed to find common themes and perspectives on LAWS as well as military AI, which we then coded, compared, and contrasted to highlight the following persistent issues in the discussion around LAWS.\(^\text{86}\)

Definitional Issues

Pursuant to our own notion of the complexities surrounding key concepts around LAWS, the experts we interviewed highlighted similar convolutions. A military expert interviewed was of the opinion that the definition of autonomy may be “the biggest hang up in this entire discussion” about the potential regulation of LAWS. Most experts interviewed seemed to agree with the definition of autonomy utilised in this report, sourced from the Stockholm International Peace Research Institute (SIPRI), which describes autonomy as the ability of a machine to execute a task or tasks without human input, using interactions of computer programming with the environment\(^\text{87}\). However, it is worth restating here that in the field of LAWS, many actors employ various definitions of autonomy which differ from ours. An expert on LAWS and IHL further opposed the anthropomorphic depiction of AI systems, underlining that human characteristics involving individualistic intentions could not possibly be fully integrated into an autonomous system, which, at the end of the day, would just be a complex piece of machinery.

The question of human control has sparked considerable debate among experts. While our interviewees agree that human control has to be defined in the context of the degree to which AI is autonomous, they emphasise the inconsistency of the meaning of human control across actors. Governments have differed on how they visualise human control in terms of the precise stage in which to exercise it, even if they have agreed that control is a requirement without which military AI may not be deployed. Interviewees support the idea that the sufficiency of human control cannot be a static concept, but is fluid and dependent on circumstances and types of weapons. As such, national definitions of human control are difficult to define. Nonetheless, consensus exists that the use of AI is bound by minimum standards found in international law. When considering regulation, terms such as meaningful human control, autonomy, and ‘in the loop’ may not be helpful for identifying lacunae in the law and consequent definitional unclarity may impede the development of further regulation.

\(^{86}\) Anonymous interviews were conducted with six experts and practitioners in the field of LAWS. The Interview Format may be found in Appendix C. Anonymity of the interviewees was requested and has been ensured.

\(^{87}\) Boulanin & Verbruggen, “Mapping the Development of Autonomy in Weapons Systems”
Currently, what is generally defined as a ‘fully autonomous weapon’ is not in usage. However, uses of military AI can be observed in the defensive strategies employed by militaries. Here, machines have the capability to select targets defensively as well as engage with them. As such, the situations of use of military AI have thus far largely been limited, on a global level, to defensive purposes against inanimate objects. In order for LAWS to be developed and utilised in war in compliance with the laws of armed conflict, their offensive uses must be clearly delineated and tested for compliance with the law and terms of engagement. With respect to terms of use, experts stress the need to align the use of military AI with existing rules of engagement, rather than devising a set of terms uniquely for LAWS. Such a procedure of compliance not only relates to legal considerations, but also contains advocacy on the characteristics of weapons being used, as well as national policies that might impact the outcome of an operation involving LAWS.

**Sufficiency of Current Regulations on LAWS**

**Development and Attribution of Responsibility**

Arguably the most salient question within the discourse around LAWS, however, comes from the dilemma of attribution - who is responsible for the actions undertaken by an autonomous entity? One legal expert interviewed stated that attributing responsibility required the establishing of either negligence, recklessness, or purpose, collectively known as *mens rea*.

The increased independence, and spatial/temporal distance between weapon system and operator may complicate the establishment of these factors. Raising the question whether existing weapons review processes suffice in safeguarding adherence and enabling attribution. In light of such doubt, one could question whether we require a separate legal framework for attribution of responsibility in case of the use of LAWS.

The experts interviewed were of the broad opinion that distinct regulations or regulatory bodies were not necessarily required to govern the development and deployment of LAWS, favouring existing IHL provisions and weapons review processes. Legal experts interviewed recognised that an increase of autonomous capabilities in weapons does not change the inherent constraints that apply to them, which is why national bodies in charge of reviewing weapons for risk assessments and compatibility with the law of armed conflict in general would be sufficient to review LAWS before their deployment as well.

A CCW GGE insider interviewed relayed that there are two relatively distinct camps in terms of the latter question. One expert says current laws, specifically IHL, are sufficient, and that because they implicitly address humans, they already implicitly outlaw any weapon that does not have a responsible human in the operational stage. The other side agrees with this statement, however, they argue that new law is required to make this explicit and clarify that weapons cannot be used without human control or responsibility.

Another legal expert interviewed judged that the normal rules of command responsibility continue to function and apply to the use of LAWS in operations. The commander bears responsibility for ensuring that they have adequate information, that potential targets have
been verified, and to select weapons that are known to operate in compliance with the law in the circumstances of the operation. Despite the precautions taken with any weapons system as part of the military’s standard practice, a breach of international law could occur. However, this would not be distinct from the kind of outcome that may be a result of a technical or human error in a similar operation using ‘conventional’ weapons. Regardless of a human error of judgement or system malfunction, the accountability for a breach of international law would most likely fall upon the field commander in charge of the specific operation, and may be subsequently appraised up the chain of command in accordance with stipulated doctrines of use.

**Machine Learning and the Black Box Problem**

While the experts we interviewed did not find particular limitations within the current legal regime and review processes in terms of LAWS, they were much more apprehensive of the aspect of machine learning (ML) in technologies used by the military. We hypothesize that the primary problem with the deployment of LAWS lies in their ability of self-learning, which is the ability of a programme to alter its behaviour patterns based on data gathered from the environment. This would mean that it may not be possible to sufficiently predict a system’s actions during weapons review procedures, and therefore the existing accountability mechanisms, geared towards systems which retain their characteristics, may be compromised.

According to the interviewed experts, ‘predictability’ and ‘understandability’ are considered to be vital qualities of artificially intelligent systems, put simply: such systems should do what they are expected to do, and they must do so for intelligible reasons.88 When they fail to fulfill these requirements due to the learning process being dictated by algorithms which humans have no way of knowing or controlling, the AI may function as a black box which we cannot look inside or predict. This black box problem thus presents itself in AI used in the military through LAWS, which may become a significant risk to take on. Military operations balance themselves on a delicate line between lawful and unlawful through strict rules of engagement, and any attempt to bend these rules needs to be accompanied by the highest level of certainty of outcome, for which the black box problem may not allow.

Military experts interviewed were wary of the possibility that the machine vetted on day one might not be the machine being operated on day two. One such expert pointed out that if an AI based machine is given certain restrictive commands which comply with the laws of war to engage with an adversary, it may give itself more freedom in its actions when deployed for better results, but these actions may not be compliant with IHL. His solution to this issue is to carefully review the levels of autonomy within particular AI systems based on ML, and then decide which levels to allow and which to ban completely. The motivation behind this ban would not be because of a lack of attribution, but a fear of attribution.

A military expert trained in law made it clear that IHL cares about the result of an operation, and not its causation - if a civilian has been harmed due to the ML aspect of an AI based

---

88 Michel, “The Black Box, Unlocked”
system, someone has to be declared responsible for the actions of the machine regardless of how autonomously the machine functioned. If the machine learns and continues to adapt and operate within the confines of the law, there does not seem to be much of an issue. However, at the point where it cannot be ascertained how the machine will operate, that is the point where our interviewee believes “you stop, pull offline and re-do the process.” If this becomes an eventuality that keeps happening, then the military has to make a decision about whether this is a system that they want to continue using.

Legal experts interviewed state that AI based systems with ML come with a much greater risk of failure. It still seems to be incumbent on manufacturers and weapons reviewers to reach a level of certainty about, and potentially limit, how a weapon system will behave once it's deployed. If an AI based weapon is designed in such a way that it may decide to change the way it behaves, then that propensity to change is also an aspect of its behaviour that needs to be testable and reviewable in advance. From our own perspective, we hypothesise that self learning in important parameters would not pass the Article 36 weapons review processes because the weapon is unpredictable, which may render AI based weapons with ML unlawful.

**Potential for Treaties**

In response to the discussion of the sufficiency of current regulation, concerned parties might consider the potential for treaties and further regulation. However, the unwillingness to agree on a set of definitions is precisely what impedes formal clarification by legal or political means.

As one interviewee shared, current outputs of the GGE, for example, may be commendable for their cooperative quality but carry “watered-down” statements on LAWS, providing little in the way of furthering the discussion. Definitions of full autonomy, human involvement, judgement, supervision, and ‘the kill chain,’ for example, draw endless debate, even among governmental experts. In the view of this interviewee, the only progress in treaty-making would have to come from moving the debate somewhere else in the international arena, which could have the by-product of fracturing what consensus currently exists and pushing the discussion back a decade.

On the other hand, certain like-minded states show promise in their ability to coordinate in small groups. For example, the United States, South Korea, and Japan regularly compare notes in the GGE. Notably, this cooperation comes in the face of concern over the position of China, who these states view as inconsistent in its position on LAWS. While their position exists in opposition to, rather than in cooperation with, China, the capacity for states to team together may signal a potential for coordinated definitions in response to concern about one or a few inconsistent actors. The result could be that although a single definition cannot be reached for a given term, the field of definitions may narrow to a few rather than many.
According to one expert, an additional barrier to cooperation may stem from arms trade. LAWS generate an important question in arms control in the ways states may seek to define the limits of their trade. In a conflict between an actor with LAWS and an actor without, the actor with LAWS in its inventory will almost certainly win. For that reason, states may be unwilling to sign any treaty which limits their ability to develop, employ, or trade LAWS, as a matter of national security.
Conclusion

Naturally, the introduction of a new class of weapons gives rise to pertinent questions concerning legality under international law as well as accountability for the consequences of using these weapons. In pursuit of these questions, our team conducted country reviews and expert interviews, and now draws three primary conclusions:

- The examined states differ in their conceptions of autonomy, and thus differ in the extent to which they consider development and deployment of such systems appropriate.
- Experts in academia, law, and military subject fields converge in agreement that the use of LAWS without machine learning elements do not per se present an issue under international law.
- Auxiliary to these two conclusions is the notion that broader philosophical or non-technical definitional questions distract from the constructive nature of debate, and complicate the formation of meaningful additional norms.

Beyond these core conclusions, it should be noted that our research indicates that states are cognizant of the complications of machine learning and are wary of deploying weapons systems which do not function in a dependable fashion. In the cases of the examined states, states such as South Korea remain boldly ambitious about developing LAWS, and the United States purports that increased efficiency and effectiveness leads to more compliance with IHL and overall ‘humaneness,’ not less. While states such as the Netherlands, the United Kingdom, and France hold reservations about weapons systems where human intervention is not required, none of the examined countries appear in favour of preemptively banning autonomous weapons, citing differences in definitions of autonomy.

Examined states appear to have the position that LAWS can be regulated by existing international legal frameworks and do not complicate the attribution of guilt in the case of violation of the Laws of Armed Conflict. Emergence of new weapons is a continuous process, and national militaries have developed mechanisms, such legal and weapons reviews, to ensure compliance of such systems. An interviewee involved with the CCW GGE was also of the opinion that the UN could potentially perform a secretariat function to convene review conferences or meetings of civil parties, and be a repository for national records and implementation.

Multiple interviewees voiced concern over the state of the academic and policy landscape concerning LAWS. This concern is echoed by our research group. On the one hand, one can find entities advocating for a broad ban on LAWS, a position informed by the humanitarian imperative on which many of these organisations and their supporters stand. On the other hand, one can find states, concerned with national security and wary of undue complications under international law. Noting the significant impact AI can have on warfare, it is understandable that states do not wish to yield a significant advantage to potential adversaries. In the case of a treaty, this means that either every state joins and is subjected to the ban, or no one might be willing to give up a valuable strategic asset, causing negotiations
to fail. Advocating for an overly ambitious ban may frustrate efforts for more moderate regulation by preventing consensus.

The semantic element of this topic further complicates this debate, where there is an absence of similar vocabulary around LAWS. During the course of this project, we encountered numerous unclear and vague definitions, at times employing one contentious definition to define another equally unclear concept. This approach has proven itself un conducive to productive debate, which is why we believe that instead of attempting to define and regulate hypothetical misuses of unknown systems, it may be wiser to outline what ‘stretch’ can be found in existing law. For instance, one area of persistent concern regards Article 36 reviews. Our research concluded that once machine learning is introduced, the lack of predictability potentially compromises the effectiveness of these review processes. Further research is needed to ascertain how legal principles should be applied to the technical specificities of machine learning, and what types of control could be applied to machine learning algorithms beyond philosophical rumination concerning the broad lines of accountability, as well as what effect this would have towards establishing the existence of a war crime.

Supplementary Considerations

Apart from further research directly related to our specific research questions, purposes and output, we believe a number of supplementary considerations on the periphery of our issue-domain may be of interest for researchers and policy-makers who are interested in this topic. In the vein of navigating potential pathways for treaties and political declarations, important consideration must be afforded to these potentially complicating elements. We came across several interesting discussion points and insights both from our own reading and from experts interviewed, some of which we would like to mention briefly before we conclude this report. While not categorically legal, these issues can impede or side-track discussions on definition and regulation.

Psychological dependence, for instance, characterises the issue of lethal autonomous weapons as an human-machine teaming issue. One could also consider the power imbalance whereby increasing technology could exacerbate existing military inequalities. The cost of war and potential for escalation also raise one of the most salient questions of whether the ability to minimise harm to human combatants and the negative domestic political repercussions of loss of life increases the odds of parties engaging in conflict. In our view, existing research does not sufficiently respond to these complicating elements, while outside the scope of this report, such elements remain issues which warrant further investigation.

---

89 A more extensive exploration of these issues can be found in Appendix A of this report.
Bibliography


4. Association for Aid and Relief, Japan. 2019. “キラーロボットのない世界に向けて、東京声明を採択しました.”


19. Frantzman, Seth. 2019. "Israel Finds An AI System To Help Fight In Cities". *C4ISRNET.*
https://www.c4isrnet.com/battlefield-tech/2020/02/05/israel-finds-an-ai-system-to-help-fight-in-cities/

https://www.defense.gouv.fr/aid

https://www.degruyter.com/view/journals/pjbr/9/1/article-p124.xml

https://nationalinterest.org/blog/buzz/ultimate-weapon-war-no-one-talking-about-42497


https://futureoflife.org/2019/05/09/state-of-ai/?cn-reloaded=1

https://www.iai.co.il/p/harpy,


   https://unidir.org/publication/black-box-unlocked


   https://www.rijksoverheid.nl/documenten/beleidsnotas/2019/10/08/strategisch-actieplan-voor-artificiële-intelligentie


   Association for Aid and Relief, Japan.


   https://reliefweb.int/sites/reliefweb.int/files/resources/pax-rapport-crunch-time.pdf


[https://www.unog.ch/80256EDD006B8954/(httpAssets)/6B80F9385F6B505FC12581D4006633F8/$file/2017_GGEonLAWS_WP9_Switzerland.pdf](https://www.unog.ch/80256EDD006B8954/(httpAssets)/6B80F9385F6B505FC12581D4006633F8/$file/2017_GGEonLAWS_WP9_Switzerland.pdf)


[https://lawecommons.luc.edu/cgi/viewcontent.cgi?article=1204&context=lucilr](https://lawecommons.luc.edu/cgi/viewcontent.cgi?article=1204&context=lucilr)


69. United Kingdom. 2015. “STATEMENT on Lethal Autonomous Weapons Systems To The CCW Meeting Of The High Contracting Parties.”
[https://unog.ch/80256EDD006B8954/(httpAssets)/1CBF996AF7AD10E2C1257E260060318A/$file/2015_LAWS_MX_United+Kingdom.pdf](https://unog.ch/80256EDD006B8954/(httpAssets)/1CBF996AF7AD10E2C1257E260060318A/$file/2015_LAWS_MX_United+Kingdom.pdf)


Appendices

Appendix A: Supplementary Considerations

During our research, we came across a number of interesting discussion points and insights from experts regarding LAWS. Although these are not necessarily in line with our research questions, purposes or outputs, we considered that it would be valuable to mention them within this appendix as further food for thought.

In the vein of considering the potential pathway for treaties and political declarations, important consideration must be afforded to potentially complicating elements. Such elements, while not categorically legal, can impede or side-track discussion on definition and regulation. These points include but are not limited to:

- Psychological dependence of operators and commanders on AI
- IHL expertise of civilian-military partnerships
- Imbalance of warfare capabilities of parties in conflict
- Risk of the escalation of conflict absent the loss of human life

Psychological Dependence: One concern comes from the level of trust placed in AI systems, leading to a so-called psychological dependence on the AI system’s decision-making even when a human operator remains involved. As Galliott explains, problems can arise when both too much trust is placed by the operator in an AI system or too little, resulting in putting the operator or others in harm’s way. As a result, the armed forces have currently invested in research on human-machine teaming, although generally focused on the engineering perspective.90 One expert explained that it would be incumbent on armed forces who deploy these systems to put some procedures in place to ensure that only an appropriate level of trust is placed in an autonomous system. As this interviewee also pointed out, “the tendency is there to, maybe, trust the computer more than it should be trusted in some cases, but in other cases, people don’t place sufficient trust in the computer.”91 Thus, the issue of trust is not just one of psychological dependence, but also of trusting the system enough to allow it to function as it should, as unnecessarily overriding the system would negate any of the ‘humanitarian’ benefits of a system with above-human capabilities of targeting and distinction. What constitutes an appropriate level of trust and such levels of trust can be practically monitored and regulated remains a question.

Training may be an appropriate response to questions of AI dependency. One interviewee emphasised the role of training in minimizing the ‘learning curve’ associated with emerging weapons technology, acknowledging that some level of psychological dependence on or misunderstanding of the tech may be unavoidable at first implementation of the technology, but that training of commanders and operators is the greatest tool available to mitigate destructive misuse. Indeed, research in other areas of psychological dependence on

90 Galliott, “The Soldier’s Tolerance for Autonomous Systems.”
91 Interview with unnamed expert, October 2020.
technology reveals the need for proper training. One study tested the psychological dependence of oil and gas engineers on an expert system falsely programmed by the framers of the study to give erroneous reports. Novice operators of the system were more likely to trust the erroneous system reports and report high confidence in the machine than more experienced, expert operators. 92 Other areas of research in psychological dependence on expert technology have included the use of medical robots, management information systems, and online shopping vendors. 93 Effective legal frameworks pertaining to LAWS would incorporate training requirements and identify the comprehensive knowledge necessary for legal implementation by commanders and operators.

Civilian-Military Partnerships: Another point of concern for regulation relates to civilian-military, so called civ-mil partnerships. A criticism of civ-mil partnerships in LAWS regards the immense technical knowledge required to develop and employ them, which may originate from civilian tech developers rather than within militaries themselves. The point of consideration in these set-ups has been assumed to be the increasing dependence on high-tech companies for the acquisition and maintenance of such systems, which may expose the military to new risks.

In response, one expert cited the extensive nature of weapons review with the United States, identifying the costly nature of legal review. According to the interviewee, several factors can increase compliance in civ-mil partnerships. Because research and development take many years and many millions (or billions) of US dollars to fund, both civilian and military partners prioritise compliance from an early stage. Militaries put out bids for system development after identifying a need and establishing that said need could be legally met through tech. Incentivised by such big ticket contracts, civilian companies which regularly engage in civ-mil contracts develop the expertise to comply with IHL requirements. Furthermore, when the weapons systems begin the vetting process within the military approval structure, the R&D project may be terminated at any stage if that system is determined not to operate in compliance with the military’s IHL needs.

A second expert identified the problem of complexity with testing systems for military use today, explaining that there is currently a challenge with acquiring the appropriate expertise. However, this interviewee also argued that this challenge in legal review already exists for current complex software systems vetting, and the challenge will only be expanded by the development of LAWS, not solely created by them.

Regarding civ-mil partnerships, an adequate legal framework may include regulation which sets requirements on the weapons review process, particularly if the civilian partner has little experience in civ-mil contracts, as may be the case in rapidly developing AI technology. As mentioned in the discussion of Additional Protocol II Article 36 on weapons review, a learning machine may overstep traditional concepts of weapons review, directing special attention to how these systems are developed.

92 Will, “True and False Dependence on Technology.”
93 Liu, “Interactivity, Engagement, and Technology Dependence.”
**Power Imbalance:** In regard to the power imbalance between actors who do and do not have access to LAWS, some critics fear that the advantage of LAWS may create a power imbalance similar to the imbalance created by nuclear weapons, and require similar regimes to control and suppress their use. This question is particularly salient for non-state armed groups fighting a war of self-determination, which when following correct channels, could be within their legal right to do. However, a power imbalance between states which have LAWS and those actors which do not is not a new phenomenon. According to one military expert, it is important to remember that whatever technology top militaries develop, non-state armed actors eventually develop on their own. This same expert cited the presence of armed drones operated by the Islamic State during its final moments in Syria.

**Escalation of Conflict:** Lastly, concerning the cost of war, critics of LAWS also fear an ease of escalation into protracted, total wars. As one expert relayed, AI systems reduce collateral damage and the need to put a state’s soldiers in harm’s way. This could be a good change overall, but it also reduces the political cost of a conflict. Another expert dismissed the question, finding it highly unlikely that any military would deploy solely autonomous weapons with no boots on the ground. Even without the heart-breaking image of deceased soldiers returning home, a fully autonomous war would also be a costly operation, which could have its own political consequences. Expert interviewees found this question vague and difficult to adequately respond, as it is an ultimately ethical question and one appropriate for political scientists. For our purposes, both the imbalance of power and the potential for escalation of fully autonomous wars lie outside the scope of our legally-focused discussion. However, we acknowledge that these questions generate insecurity among states and the public, potentially complicating the legal position a state is willing to take in regard to LAWS.
Appendix B: Interview Format

The following list of suggested questions provided a general framework for conducting expert interviews. The questions are not exhaustive. Due to the natural progression of conversation and expertise of each individual, interviews may have included additional information and lines of specific questioning.

1) What uses of AI in the military field is the interviewee aware of?
   a) IF mentioning human machine teaming:
      i) How much human involvement is there?
      ii) Are there special regulations in place for these applications?
      iii) To what extent are these systems autonomous (Human in the loop, on the loop, out of the loop, fire-and-forget)?

   b) IF mentioning AI-based weapons
      i) Is there an a priori defined person responsible for the consequences of using AI?
      ii) How much human involvement is there?
      iii) Are there special regulations in place for these applications?
      iv) To what extent are these systems autonomous (Human in the loop, on the loop, out of the loop, fire-and-forget)?

2) Is the use of AI in the military increasing?
   a) [If increasing]
      i) What does the interviewee see as the driving force behind the increasing use of autonomy?
      ii) Is this trend uniform across uses of AI or skewed towards either human-machine teaming or AI-based weapons
      iii) Would so-called Lethal Autonomous Weapons (LAWS) be a logical progression from this trend?
   b) [If not increasing]
      i) What does the interviewee see as inhibiting factors?
      ii) Do these objections also apply to the development of LAWS?

3) What mechanisms currently exist to attribute responsibilities for crimes committed involving so-called ‘autonomous weapons’ or Human-machine teaming in targeting processes?
   a) Would these mechanisms also work for attributing responsibility in case of future development of LAWS?

4) Does restricting AI-based weapon systems, by for example creating a requirement to have a ‘human in the loop’, detract from possible operational advantages of using autonomous systems?

---

94 The identity of the interview participants has been protected at their request.
a) If yes: which advantages of the integration of AI in the weapon system could be negatively affected?
b) If no: would you see other impediments to the creation of such a requirement?
c) Would regulation have a comparable impact on future development of LAWS?

5) Does the interviewee think the current use of AI in the battlefield to affect the following positively or negatively: what about human machine teaming in target acquisition.
   a) Lethality (capacity to cause death)
   b) Mortality (overall deaths in conflict)
   c) Collateral damage
   d) Conflict duration
   e) Financial cost of the conflict
   f) Political cost of the conflict
   i) [If negative answers given]
      (1) Are the pros of the use of using AI in a military context outweighed by its cons?
      (2) Can this be favourably changed through legislation?
   g) Would there be a difference in these answers when considering future development of LAWS?

6) The CCW GGE on LAWS (2018) has stated that risk assessments and mitigation measures should be part of the design, development, testing and deployment cycle of emerging technologies in any weapons systems. Could you mention any such measures or bodies responsible for taking such measures in states you have researched? (Report of the 2019 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems, Annex IV, G. https://undocs.org/en/CCW/GGE.1/2019/3)

7) To what extent does the interviewee think that the current international legal frameworks suffice in covering questions of responsibility for the use of military AI such as:
   a) AI used in targeting processes
      i) Under IHL
   b) AI based weapons
      i) Under IHL

8) Do you think that existing principles regarding rules of engagement should be interpreted differently when AI is used in targeting processes [or] is integrated in weapon systems?
   a) Would your answer change if LAWS are used?
   b) Do you think LAWS should fall under the same regime or should a separate regime be created to govern their use?
9) Which authoritative bodies (nationally/internationally), if any, does the interviewee consider best-suited to regulate the operations and confines of intelligent algorithms in military use?
   a) If none: does the interviewee believe a different, more specialised body is required for to regulate the military use of AI?
Appendix C: Project Description and GDPR Compliant Consent Form for Interviewees

Information Sheet

The LAWS & War Crimes project analyses the challenges of ascribing criminal responsibility for war crimes raised by the advent of increasingly autonomous weapon systems and human-machine shared decision-making in the targeting process (so-called mixed-initiative systems). In particular, the project focuses on the criminal responsibility of the user of autonomous weapons and of the human-operator in mixed systems.

In order to supplement research of existing literature, data is collected through interviews. The purpose of these interviews is to provide a current image of positions on the use of AI in the military from both professional and academic perspectives.

No data will be collected beyond the Interviewee’s name, publicly available information regarding the interviewee’s background, and data provided by the interviewee during the interview.

The data gathered during the interview is used in research conducted by the LAWS and War Crimes project and may be used in published reports and articles. Data used will be anonymous, but the interviewee has the option to grant permission to use direct quotes with or without name (see consent form).

Data will be stored on the Graduate Institute’s servers and is only directly accessible to members of the LAWS and War Crimes Team. Data will be retained for the duration of the LAWS and War Crimes project and destroyed afterwards.

Contact details for questions about this form:

Bram Goede
MA Candidate
The Graduate Institute, Geneva
bram.goede@graduateinstitute.ch
+316 12756659

Contact details of project lead:

Paola Gaeta
Professor, International Law Department
The Graduate Institute, Geneva
paola.gaeta@graduateinstitute.ch

Form composed 01/09/2020
Consent Form for LAWS and War Crimes Project

Please tick the appropriate boxes

Taking part in the study

I have read and understood the study information dated [01/09/2020], or it has been read to me. □ □

I have been able to ask questions about the study and my questions have been answered to my satisfaction. □ □

I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason. □ □

I understand that taking part in the study involves taking part in an audio-recorded interview which will be transcribed as text. After transcription the recording will be destroyed, and transcript saved on in-house servers. □ □

Use of the information in the study

I understand that information I provide will be used for in research outputs of the wider LAWS and War Crimes Project in the form of reports or publications □ □

I understand that personal information collected about me that can identify me, such as e.g. my name, will not be shared beyond the study team. □ □

I agree that my information can be anonymously quoted in research outputs □ □

(optional) I agree that my real name can be used for quotes □ □

Future use and reuse of the information by others

I give permission for the transcribed responses that I provide to be archived in the in-house database only accessible to members of the LAWS and War Crimes team so it can be used for future research and learning. □ □

Signatures

______________________  __________________  __________
Name of participant [printed]  Signature  Date

I have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands what they are freely consenting to.