MCGSMUN 2025

OCULUS PROCALLAE

Eye of The Hurricane

The United Nations Advisory Body on Artificial Intelligence

Establishing Global Regulatory Framework for the Ethical Use of AI in Autonomous Weapons Systems

LETTER FROM THE EXECUTIVE BOARD

Distinguished delegates,

It is an honor to welcome you all to the United Nations Advisory Body on Artificial Intelligence (UNAB-AI). As AI continues to reshape the global landscape, its integration into military applications, particularly autonomous weapons systems (AWS), has sparked ethical, legal, and security concerns. This committee aims to formulate a globally accepted regulatory framework that ensures AI-driven weapons adhere to humanitarian principles while preventing escalation and misuse.

The Background Guide provides you with an extracted collection of essential basic information relevant to the agenda, including past actions as well as issues that need to be addressed. We expect that every bit of research that has been put into the making of this background guide is utilised and is reflected in your prep. In addition to the information given, it is encouraged for members to research extensively on the agenda and have a reflection of their understanding on this topic. The final recommended reading lists outlines some of the critical documents in the field, and broad questions on looking at the agenda.

We urge all representatives to engage in constructive dialogue, leveraging their national/ individual perspectives to achieve a consensus on governance structures, ethical parameters, and accountability measures. Your insights will contribute to a historic framework that balances innovation with responsibility.

We look forward to your deliberations.

Sincerely,
Muskan Ajitsaria,
Chairperson

Meet Goel
Vice-Chairperson

List of Abbreviations

Abbreviatin	Full Form
AI	Artificial Intelligence
AWS	Autonomous Weapons Systems
UN	United Nations
UNAB-AI	United Nations Advisory Body on Artificial Intelligence
LAWS	Lethal Autonomous Weapons Systems
GGE	Group of Governmental Experts
IHL	International Humanitarian Law
CCW	Convention on Certain Conventional Weapons
MHC	Meaningful Human Control
ML	Machine Learning
ICT4Peace	Information and Communication Technology for Peace Foundation
ZHET	Zurich Hub for Ethics and Technology
AP I	Additional Protocol I (to the Geneva Conventions)
ICC	International Criminal Court

Glossary

1. **Lethal Autonomous Weapons Systems (LAWS)** – A subset of AWS designed to select and attack targets, raising ethical and humanitarian concerns.
2. **Convention on Certain Conventional Weapons (CCW)** – A UN treaty aimed at restricting the use of weapons that cause unnecessary suffering.
3. **Group of Governmental Experts (GGE)** – A UN-established group that discusses and provides recommendations on emerging security technologies, including LAWS.
4. **Meaningful Human Control (MHC)** – The principle that humans should retain control over AI-based weapons to ensure accountability and ethical use.
5. **International Humanitarian Law (IHL)** – A set of rules that regulate armed conflict and aim to protect civilians and combatants.
6. **Machine Learning (ML)** – A branch of AI that enables systems to learn from data and improve decision-making without explicit programming.
1. **International Criminal Court (ICC)** – A global tribunal that prosecutes individuals for war crimes, genocide, and crimes against humanity.

INDEX

i. INTRODUCTION

In October 2023, United Nations (UN) Secretary-General António Guterres stated during the formation of the High-Level Advisory Body on Artificial Intelligence, "It is already clear that the malicious use of Artificial Intelligence could undermine trust in institutions, weaken social cohesion, and threaten democracy itself." Artificial intelligence (AI) is becoming highly prominent in multiple sectors of the international community - from healthcare to research and the military- and is now widely accessible to the general public.

Within two months of its release in 2022, ChatGPT, an artificial intelligence chatbox, received an unprecedented 100 million users.[1] According to experts, AI has the potential to add up to $15.7 trillion USD to the global economy across various sectors by 2030.[2] Subsequently, an analysis by the International Monetary Fund estimates that AI is expected to impact 60% of jobs in advanced economies.[3]

Most artificial intelligence technologies are dual-use. They are incorporated into both peaceful civilian applications and military weapons systems. Most of the existing codes of conduct and ethical principles on artificial intelligence address the former while largely ignoring the latter. But when these technologies are used to power systems specifically designed to cause harm, the question must be asked as to whether the ethics applied to military autonomous systems should also be taken into account for all artificial intelligence technologies susceptible of being used for those purposes.

While the positive effects of AI in research, healthcare, and global economies cannot be understated, its negative externalities pose an opportunity to outweigh it. In June 2023, for the first time, the UN Security Council discussed AI, its associated risks, and its potential to threaten international peace and security, showcasing AI's newly realized potential harm among international organizations. [4]Autonomous weapons systems (AWS) are "weapon systems that select targets and apply force without

---

[1] Milmo, Dan. "ChatGPT reaches 100 million users two months after launch." The Guardian. www.theguardian.com/technology/2023/feb/02/chatgpt-100-million-users-open-ai-fastest-growing-app.

[2] "Sizing the prize - PwCs Global Artifical Intelligence Study: Exploiting the AI Revolution." PwC. https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html.

[3] Mauro Cazzaniga, Florence Jaumotte, Longji Li, et al. "Gen-AI: Artificial Intelligence and the Future of Work." International Monetary Fund. https://www.imf.org/en/Publications/Staff-Discussion-Notes/Issues/2024/01/14/Gen-AI-Artificial-Intelligence-and-the-Future-of-Work-542379?cid=bl-com-SDNEA2024001

[4] Tokariuk, Olga. "Ukraine's Secret Weapon - Artificial Intelligence." Center for European Policy Analysis.https://cepa.org/article/ukraines-secret-weapon-artificial-intelligence/.

human intervention – pose serious humanitarian, legal, ethical and security concerns. The International Committee of the Red Cross describes AWS as an "immediate cause of humanitarian concern and demand[s] an urgent, international political response." Additionally, AI has shown its potential to spread global disinformation by creating hyper realistic deep fakes of world leaders.[5]

However, while a freeze in investigations is neither possible nor desirable, neither is the maintenance of the current status quo. Comparison between general-purpose ethical codes and military ones concludes that most ethical principles apply to human use of artificial intelligence systems as long as two characteristics are met: that the way algorithms work is understood and that humans retain enough control. In this way, human agency is fully preserved and moral responsibility is retained independently of the potential dual-use of artificial intelligence technology.

ii. COMMITTEE HISTORY AND MANDATE

Establishment of UNAB-AI: The United Nations Advisory Body on Artificial Intelligence (UNAB-AI) was established in 2023 to address global challenges associated with AI development and deployment. Functioning under the guidance of the UN Secretary-General, this body serves as an advisory platform for member states, industry leaders, and civil society on AI governance, ethics, and international security concerns.

The body operates with a multi-stakeholder approach, integrating perspectives from UN agencies, governments, academia, and technology firms. The increasing role of AI in military applications led to a dedicated subcommittee on autonomous weapons systems (AWS), tasked with developing ethical, legal, and policy frameworks.

Role and Objectives: The United Nations Advisory Body on Artificial Intelligence (UNAB-AI) was established in 2023 to address global challenges related to the development, deployment, and regulation of artificial intelligence (AI). It functions as a high-level advisory entity under the guidance of the **United Nations Secretary-General**, focusing on AI governance, ethics, and security concerns.

The primary objectives of UNAB-AI include:

a) **Developing AI Governance Frameworks**

i. Establishing **global regulatory standards** for AI applications, particularly in military contexts like **Autonomous Weapons Systems (AWS)**.

[5]Quinlan, Matthew. "AI is finding its voice and that's bad for democracy." World Economic Forum. https://www.weforum.org/agenda/2023/11/ai-is-finding-its-voice-and-that-s-bad-for-democracy/.

ii. Providing recommendations to **UN member states, policymakers, and international organizations** on responsible AI use.

b) **Ensuring Ethical AI Development and Use**

i. Promoting AI systems that comply with **international human rights standards and humanitarian principles**.

ii. Encouraging transparency and **human oversight** in AI-based decision-making.

c) **Addressing Security and Legal Implications**

i. Assessing the risks posed by **Lethal Autonomous Weapons Systems (LAWS)** and their compliance with **International Humanitarian Law (IHL)**.

ii. Preventing AI misuse in warfare, cyber threats, and autonomous military operations.

d) **Facilitating International Cooperation**

i. Acting as a **neutral platform** for multilateral discussions on AI governance.

ii. Bringing together **governments, tech companies, researchers, and civil society organizations** to create a consensus-based regulatory framework.

e) **Monitoring AI's Impact on Global Stability**

i. Evaluating AI's influence on economic, social, and security landscapes.

ii. Recommending **preventative measures** to mitigate AI-related risks, such as algorithmic biases, misinformation, and unintended escalations in conflicts.

Key Stakeholders: UNAB-AI operates using a **multi-stakeholder approach**, ensuring a balanced representation of interests across governments, academia, industries, and civil society. The **key stakeholders** involved in UNAB-AI's mission include:

1) **United Nations Bodies and Agencies**
    a) **UN Secretary-General's Office**: Oversees and provides strategic direction for UNAB-AI's initiatives.
    b) **United Nations Office for Disarmament Affairs (UNODA)**: Focuses on AI's role in arms control and military applications.
    c) **International Telecommunication Union (ITU)**: Establishes AI-related technical and ethical guidelines.
    d) **Office of the UN High Commissioner for Human Rights (OHCHR)**: Ensures AI policies uphold fundamental human rights.
2) **National Governments and Defense Agencies**
    a) Governments that are investing in AI-driven military technologies, such as the **United States, China, Russia, and the European Union**, are key participants in discussions about AI regulation.
    b) Defense organizations and ministries involved in **AI research, cybersecurity, and autonomous weapons policies**.
3) **Academia and Research Institutions**
    a) Leading universities and think tanks specializing in AI ethics, international security, and governance, such as:
        i) **MIT Media Lab (USA)**
        ii) **Oxford Internet Institute (UK)**
        iii) **Swiss AI Lab IDSIA (Switzerland)**
        iv) **Zurich Hub for Ethics and Technology (ZHET)**
4) **Technology Companies and Industry Leaders**
    a) Major AI developers, including:
        i) **Google DeepMind, OpenAI, Microsoft, IBM, and NVIDIA** – Working on AI research and machine learning advancements.

ii) **Defense contractors like Lockheed Martin, BAE Systems, and DARPA** – Developing AI-powered military applications.

b) Private-sector representatives contribute insights on **AI innovation, risks, and regulatory compliance**.

5) **Civil Society and Human Rights Organizations**

a) **International Committee of the Red Cross (ICRC)**: Advocates for humanitarian safeguards in AI-based warfare.

b) **Human Rights Watch (HRW)**: Calls for bans or restrictions on **fully autonomous lethal weapons**.

c) **Future of Life Institute (FLI)**: Focuses on AI risk mitigation, comparing AI governance to nuclear arms control.

d) **Campaign to Stop Killer Robots**: A coalition urging global bans on AI-driven weapons without human oversight.

6) **Intergovernmental and Regional Organizations**

a) **European Union (EU)**: Has proposed regulations for AI in military and civilian applications.

b) **NATO**: Engages in AI-related military research and cybersecurity measures.

c) **G20 and G7 Nations**: Discuss AI ethics and regulatory policies at international summits.

iii. TIMELINE OF KEY DEVELOPMENTS

1. **2013** – UN discussions on Lethal Autonomous Weapons Systems (LAWS) initiated under the Convention on Certain Conventional Weapons (CCW).
2. **2017** – The Group of Governmental Experts (GGE) on LAWS was established under the UN.
3. **2019** – EU Parliament calls for a global ban on AI-powered lethal autonomous weapons.
4. **2021** – UN Secretary-General urges states to regulate AI in weapons systems.
5. **2023** – Creation of UNAB-AI as a dedicated advisory body.
6. **2025** – Proposed framework discussion on ethical AI governance in AWS.

iv. THE AGENDA:

Establishing a Global Regulatory Framework for the Ethical Use of AI in Autonomous Weapons Systems

We can break down our agenda into three segments:

1. Artificial Intelligence (AI)
2. Autonomous Weapons System
3. Global Regulatory Framework

A. ARTIFICIAL INTELLIGENCE

AI are two letters that represent the financially most lucrative scientific field that currently exists. Moreover, they represent something that is often regarded as the fuel of the fourth industrial revolution, which is taking place at an unprecedented pace compared to any other in human history.[6] However, the question of what AI really is most often receives a rather vague and elusive answer.

The reason for this lack of clarity may be two-fold. First, the term 'Artificial Intelligence' includes the term 'intelligence.' 'Intelligence' originally has been used as a characteristic of humans. However, there neither exists a general understanding of this natural trait, nor a standard definition, despite a long history of research and debate. Precisely due to the growing research on AI, there exist strong incentives to define what the term 'intelligence' shall mean. This need is especially acute when artificial systems are considered that are significantly different to humans.

This is the reason why researchers at the Swiss AI Lab IDSIA (Instituto Dalle Molle di Studi sull'Intelligenza Artificiale) created a single definition based on a collection of 70 definitions of 'Intelligence' by dictionaries, psychologists and AI researchers. They state that 'intelligence measures an agent's ability to achieve goals in a wide range of environments.' This general ability includes the ability to understand, to learn and to adapt, since those are the features that enable an agent to solve a problem in a wide range of environments.

[6] Kelnar, David, 2016, The fourth industrial revolution: a primer on Artificial Intelligence (AI), Medium.com

It must be highlighted that the driving force behind the above-mentioned definition was to create a definitional reference point useful for both human as well as technological artefacts. This ignores the fact that the term 'intelligence' was originally used to refer to a natural human capacity; and without a clear understanding of this human trait, we could possibly risk an overvaluation of technology and a devaluation of human beings.

And second, a reason for confusion about the meaning of AI may lie in the fact that the term AI is used to refer to two distinct but interrelated understandings. The distinction of these two possible understandings of AI will be highlighted here by two definitions of AI. However, we do not claim for these definitions to gain universal validity, as they would merely increase the existing pool of possible choices of such definitions. Yet, they should provide the reader with a first sense of caution when dealing with the application of originally 'human terms' such as 'intelligence' or 'autonomy' to technological artefacts. At first glance, it might seem accurate and comprehensive to apply originally human terms to technological artefacts, since the latter are increasingly capable to perform 'actions' that resemble those of humans.

Bearing in mind the above-mentioned risk of devaluating humans in creating a definition of (artificial) intelligence without a human reference, AI shall here be understood as:
"a scientific undertaking that is aiming to create software or machines that exhibit traits that resemble human reasoning, problem-solving, perception, learning, planning, and/ or knowledge."

Core parts of research on AI include: 'knowledge engineering,' which aims at creating software and machines that have abundant information relating to the world; 'machine learning', which is the modern probabilistic approach to AI and that studies algorithms that 'learn' to predict from data; 'reinforcement learning', a sub-discipline of machine learning and currently the most promising approach for general intelligence that studies algorithms that learn to act in an unknown environment through trial and error; 'deep learning', a very fast-moving and successful approach to machine learning based on neural networks, which has enabled recent breakthroughs in computer vision and speech recognition; 'machine perception', which deals with the capability of using sensory inputs to deduce aspects of the world, 'computer vision', the capability of analysing visual inputs; and 'robotics', which deals with robots and the computer systems for their control.[7]

B. AUTONOMOUS TECHNOLOGY (AT)

[7] Techopedia.com, Artificial Intelligence, at: https://www.techopedia.com/definition/190/artificialintelligence-ai

AT is a result of research in the fields of AI and robotics, but also draws on other disciplines such as mathematics, psychology and biology. Currently, there exists no clear understanding and no universally valid definition of the term 'autonomous' or AT in the context of AI and robotics. However, there do exist different attempts.

Sometimes a purely operational understanding of 'autonomy' is used. In this sense, the term 'autonomous' may refer to any outcome by a machine or software that is created without human intervention. This could include, e.g., a toaster's ejection of a bread slice when it is warm. In this form, autonomy would be equivalent to automation22 and would not be limited to digital technology but could be used in analog technology or mechanics as well. Hence, this understanding does not locate AT exclusively within the research field of modern AI.[8]

The theoretical AI approach that is at the core of AT in its narrow understanding, and that enables technological systems to perform the above-mentioned actions without a human operator, is deep learning. Deep learning software tries to imitate the activity of layers of neurons in the human brain. Through improvements in mathematical formulas and the continuously increasing computing power of computers, it is possible to model a huge number of layers of virtual neurons. Through an inflow of a vast amount of data, the software can recognize patterns in this data and 'learn' from it.[9]

This is key for 'autonomous' systems' reaction to unanticipated changes: due to new data inflow, the software can recognize new patterns and adapt to a changing 'environment'. Thereby, an autonomous system can modify its actions in order to follow its goal or agenda.

C. LETHAL AUTONOMOUS WEAPONS SYSTEMS (LAWS)

In addition to promising applications of AT, autonomous software can be (and arguably already is) integrated into robots that can select and engage a (military) target (e.g. infrastructure and potentially also combatants) without a human override.[10] Often called Lethal Autonomous Weapons Systems (LAWS), as yet, there exists no agreed definition of LAWS. One reason for this lack of definition is that there exists, as highlighted above, no general understanding of the term 'autonomy' in AI and robotics.

[8] Atkinson, David J., 2015, Emerging Cyber-Security Issues of Autonomy and the Psychopathology of Intelligent Machines, Foundation of Autonomy and Its (Cyber) Threats: From Individuals to Interdependence.
[9] Watson, David P., and Scheidt, David H., 2005, Autonomous Systems, Johns Hopkins APL Technical Digest 26(4), 368-376
[10] Roff, Heather, and Moyes, Richard, 2016, Dataset: Survey of Autonomous Weapons Systems, Global Security Initiative: Autonomy, Robotics & Collective Systems, Arizona State University

The general idea is that a LAWS, once activated, would, with the help of sensors and computationally intense algorithms, identify, search, select, and attack targets without further human intervention. Whether the human being can still overpower or veto an autonomous weapon's 'decision' in order for it to be called a LAWS, is also debated.[11] However, military operational necessity precisely seems to require weapons systems that can function once human communication links break down.
Furthermore, state-of-the-art research on AI is currently creating software which can 'learn' entirely on its own and even 'learn' to 'learn' on its own. Hence, (precursor) technologies for creating fully 'human-out-of-the-loop' weapons systems already exist.[12]

From a military perspective, LAWS have many advantages over classical automated or remotely controlled systems: LAWS would not depend on communication links; they could operate at increased range for extended periods; fewer humans would be needed to support military operations; their higher processing speeds would suit the increasing pace of combat; [13] by
replacing human soldiers, they will spare lives; and with the absence of emotions such as self-interest, fear or vengeance, their 'objective' 'decision-making' could lead to overall outcomes
that are less harmful.[14]

However, the use of LAWS may also generate substantial threats. Generally, LAWS may change how humans exercise control over the use of force and its consequences. Further, humans may no longer be able to predict who or what is made the target of an attack, or even explain why a particular target was chosen by a LAWS. This fact raises serious legal, ethical, humanitarian and security concerns.[15]

From a humanitarian and ethical point of view, LAWS could be regarded as diminishing the value of human life as a machine and not a human being 'decides' to kill. Also, the physical and emotional distance between the programmer or engineer of a LAWS and the targeted person may generate an

[11] ' Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, Christoph Heyns, UN doc. A/HRC/23/47, § 38; Human Rights Watch (HRW)
[12] Adams, T., 2002, Future Warfare and the Decline of Human Decision making, Parameters, U.S. Army War College Quarterly, Winter 2001-02, 57-71.
[13] Thurnher, J., 2014, Examining Autonomous Weapons Systems from a Law of Armed Conflict Perspective, in: Nasu, H., and McLaughlin, R. (eds.), New Technologies and the Law of Armed Conflict, TMS Asser Press, 213-218.
[14] ICRC, 2011, International Humanitarian Law and the Challenges of Contemporary Armed Conflicts, Official Working Document of the 31st International Conference of the Red Cross and the Red Crescent, November 28 – December 1, 2011.
[15] Geneva Academy, 2017, Autonomous Weapons Systems: Legality under International Humanitarian Law and Human Rights,
https://www.geneva-academy.ch/news/detail/48-autonomous-weapon-systems-legality-under-internationalhumanitarian-law-and-human-rights

indifference or even a 'Gameboy Mentality' on the side of the former.[16] From a security perspective, LAWS could be dangerous because they may also be imperfect and malfunction. 51 Moreover, the greater the technology advances, the more the level of autonomy of a LAWS increases. This, further, leads to an increased unpredictability of outcomes of LAWS and may enable the interaction of multiple LAWS as e.g. self-organizing swarms.[17]

D. GLOBAL REGULATORY FRAMEWORK

The rapid advancement of artificial intelligence (AI) in autonomous weapons systems (AWS) has raised significant ethical, legal, and security concerns. While AI-driven military technologies offer operational advantages, their deployment without robust regulatory oversight poses risks to human rights, international stability, and compliance with humanitarian law. AI-driven autonomous weapons are increasingly becoming a reality, with nations investing heavily in research and deployment. However, the absence of a universally accepted regulatory framework has led to concerns over accountability, compliance with international humanitarian law (IHL), and the potential for unintended escalations in armed conflicts.

The UN has been debating the legal, social, ethical and security implications of AWS for over a decade. The publication of the Report of the UN Special Rapporteur on Extrajudicial, Summary or Arbitrary Execution by Christof Heyns in 2013 drew attention to these issues. Writing in the context of the widespread use of drones for targeted killings, Heyns warned that the use of algorithms to make targeting decisions could have far-reaching consequences. His concerns were primarily legal – such as the issue of compliance with international humanitarian law (IHL) – but not exclusively so. Drawing on work by philosophers and civil society groups, Heyns highlighted that the delegation of life and death decisions to algorithms touched on fundamental moral issues about the act of killing and our relationship to technology. This included the prospect that, even if AWS could be used in a way that met all relevant legal requirements, nevertheless "as a matter of principle [they] should not be granted the power to decide who should live and die."[18]

---

[16] Sassòli, Marco, 2014, Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified, International Law Studies Vol. 90, 308-340, 317.

[17] ICRC, 2014, Expert Meeting on 'Autonomous weapons systems: technical, military, legal and humanitarian aspects', March 26 – 28, 2014, Report of November 1, 2014, available at: https://www.icrc.org/en/document/report-icrc-meetingautonomous-weapon-systems-26-28-march-2014#

[18] Christof Heyns, *Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions*, UN Doc. A/HRC/23/47 (2013). https://documents-dds-ny.un.org/doc/UNDOC/GEN/G13/118/43/PDF/G1311843.pdf?OpenElement

Heyns's report galvanized the international community, and within six months of its publication governments approved the first multilateral meeting on AWS, taking place in 2014 at the UN in Geneva under the framework of the Convention on Certain Conventional Weapons (CCW). Since then, ethics has been a prominent part in the debate on AWS, and many states, international organizations, civil society groups, and prominent figures have raised ethical concerns about AWS. This includes the humanitarian impact of AWS on civilians; the dehumanization and violation of the dignity of those targeted; the impact on operators' capacity for exercising moral judgement; and the risks of algorithmic bias.

Ethical and Legal Considerations

Despite the prominence of ethics in multilateral debate, there has been notably little progress on understanding and addressing the ethical challenges of AWS, especially when compared to progress made on legal aspects. Indeed, ethics in the regulatory debate is currently in a state of limbo: there is widespread recognition that ethics has a role to play in the debate, but deep uncertainty about what that role is. This limbo is well illustrated by the guiding principles adopted by the CCW Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapon Systems (GGE) in 2019: the relevance of ethics is affirmed in the preamble, but absent from the principles themselves. There are three main reasons for this limbo.

The first reason is the novelty of ethics factoring in arms control discussion. To be sure, ethical considerations have often preceded and motivated the development of new international legal constraints on means and methods of warfare. But after progressing to formal multilateral fora, these types of concerns are typically subsumed either by core IHL concerns – such as whether a weapon is inherently indiscriminate or causes unnecessary suffering – or security concerns, such as proliferation risk. Specific to the international debate on AWS is that ethics has remained a distinct and salient feature of the debate after its progression to multilateral fora. It appears there is little historical precedence for this.

This points to the second reason: despite the recognized importance of ethics, the regulatory debate is dominated by legal-based arguments. Since 2017, the centre of gravity for debate on AWS has been the CCW GGE at the UN in Geneva. Because the CCW is an instrument of IHL, the GGE has not been a particularly sympathetic forum for deepening ethics-based argument about AWS. Indeed, the mandate of the GGE directs it to draw on legal, military, and technological expertise, but not on ethics. To find purchase in debate under the CCW framework, ethics was often hitched to concepts with currency in

legal argument, meaning in turn that the health of debate about ethics became dependent on the continued popularity of such concepts.

One example of this is 'meaningful human control' (MHC). At the 2014 CCW informal meeting of experts, MHC emerged as a point of common ground for addressing ethical and legal issues associated with AWS, and with the 2016 establishment of the GGE, MHC came to act as a proxy for discussions on ethics. However, once a 'lightning conductor' for debate on AWS, MHC has lost much of its salience. If discussion on ethics at the GGE is to progress, it may need a new concept to carry it. But with debate at the GGE progressing towards substantive discussion on risk mitigation measures, safeguards, and the elements of a two-tier regulatory approach, finding a new concept for ethics may become more challenging.

Third is the way ethics has been conceived in the debate. Broadly, ethics is the study of moral phenomena: it investigates what people ought or ought not to do, and what justifications can be given for such claims. Ethical reflection is particularly useful when novel technologies create new conditions for action, when the morally right thing to do may not be immediately clear. On this basis, ethics can help formulate morally good solutions, particularly as a source of reasoning about law: it can motivate us to create new regulations, correct or remove deficit regulations, or provide principles for action over and above existing regulations.[19]

However, in the regulatory debate on AWS, this view of ethics as an open-ended enquiry is sometimes eclipsed by a view of ethics as a set of fixed rules or positions. This reflects the character of multilateral debate: time for deliberation is limited. But this view creates the risk that ethics appears as a parallel body of directives or rules in competition with law, rather than a source of moral reflection on law and the content of regulation. This can frequently leave parties to debate treating ethics as an opaque source of additional requirements or obligations to be wary of, rather than as a useful aid to identifying whether legal concepts require elaboration and concretization.

Challenges in Establishing a Global Framework

Divergent national interests present a significant challenge, as technologically advanced states may resist stringent regulations to maintain military superiority, while others advocate for strict

[19] Blanchard Alexander, "The road less travelled: ethics in the international regulatory debate on autonomous weapon systems," International Committee of the Red Cross, April 25, 2024, https://blogs.icrc.org/law-and-policy/2024/04/25/the-road-less-travelled-ethics-in-the-international-regulatory-debate-on-autonomous-weapon-systems/.

prohibitions. Verification and compliance remain complex, requiring effective monitoring mechanisms and international cooperation to ensure AWS development and use adhere to agreed standards. Additionally, technological ambiguities pose difficulties in regulation, as rapid advancements in AI may outpace legal and ethical frameworks, necessitating adaptable policies that can evolve with technological progress.

Three lessons for the regulation of AWS by Marta Bo and Vincent Boulanin[20]

Lesson 1. Discerning IHL violations in the development and use of AWS will remain challenging without further clarification on what IHL permits, requires, and prohibits

The first lesson is that legal clarification will be needed to ensure that the legal framework governing accountability can be effectively triggered.

The rules governing State responsibility for internationally wrongful acts and individual criminal responsibility for war crimes are linked to IHL. Both the establishment of State responsibility for IHL violations and individual criminal responsibility for war crimes depend on normative standards established by IHL rules. The fact that the debate on IHL compliance in the development and use of AWS is still unsettled presents, in that context, a fundamental challenge. Many questions remain about what IHL requires, permits, and prohibits, for instance in terms of human-machine interaction. This means that the basis for establishing that a State or an individual violates IHL is still, in some cases, unclear, or at least subject to different interpretations.

AWS bring also into new light old and unresolved legal disputes around the standards of conduct that would trigger State responsibility or individual criminal responsibility for war crimes (or both). For instance, it has been debated to what extent a violation of the principle of distinction has to be 'deliberate' for State responsibility to arise. And it is an open question whether recklessness or omission satisfy the mental and material elements of perpetrating or participating in the commission of a war crime. The fact that AWS are pre-programmed weapons, which are ultimately triggered by the interaction with the environment rather than direct user input, gives these debates new resonance but also new scenarios to deal with. For instance, would a failure to suspend an attack involving an AWS that is expected to harm civilians be considered a deliberate attack on civilians and amount to a war crime?

These questions and controversies underline the need for the policy process on AWS to achieve more precision and a common understanding of IHL compliance. In particular, they invite the GGE to

[20] Vincent Boulanin and Marta Bo, "Three lessons on autonomous weapon systems and IHL," *International Committee of the Red Cross*, March 2, 2023, https://blogs.icrc.org/law-and-policy/2023/03/02/three-lessons-autonomous-weapons-systems-ihl/.

elaborate on standards of intent, knowledge and behaviour that are demanded on the part of the user(s) of AWS. Clarifying what the user(s) of an AWS should be able to reasonably foresee and do to ensure that the AWS attack is directed at a specific military objective and the effects of the weapon are limited as required by IHL would make it easier to determine whether a violation has been committed intentionally or that the user engaged in risk-taking behaviour that could give rise to State responsibility, and individual criminal responsibility or both.

Lesson 2. Elaboration on what constitutes a 'responsible human chain of command' could help with the attribution of responsibility

The second lesson is that the policy process needs to unpack the notion of 'responsible human chain of command'. Elaboration on how such a chain may look could dramatically facilitate the attribution of responsibility, be it to the State or individual.

Some States and experts have expressed the concern that, in the case of a harmful incident involving an AWS for instance, it could be difficult to identify whose conduct is blameworthy given that the operation, performance and effect of an AWS were determined in part by decisions and actions of multiple individuals involved in the development and use of the systems; as well as the interaction of the system with the environment.

We argue in this context that it would be extremely useful if States could elaborate on what a scheme of responsibility for the development and use of AWS could look like. Such a scheme would provide more clarity on how the roles and responsibilities for IHL compliance may or may not be distributed in practice: who should do what, when and where the roles and responsibilities of the different individuals start and end and how might these interact with one another. Such an effort would be doubly beneficial. On the one hand it would strengthen IHL compliance by providing clearer expectations for the users of AWS. On the other, it could make it easier to detect who engaged in unlawful conduct that could give rise to State responsibility, individual criminal responsibility (or both).

Lesson 3. Traceability is a critical component for the regulation of AWS

The third lesson is that traceability – understood here as the ability to trace the operation, performance and effect of an AWS back to people involved in its development and use – should be regarded as a critical component of further regulation of AWS. It should inform the identification of new limits and requirements on the design and use of AWS – for two reasons.

First and foremost, traceability is a practical requirement for complying with States' obligations under international law. Under AP I, States are obliged to repress war crimes, including searching for individuals responsible, and suppressing any other violations of IHL. To be able to perform these obligations, States need to be able to determine whether illegal conduct took place and, if so, identify blameworthy individuals. Second, it is also a practical requirement to assess and impose State responsibility, individual criminal responsibility or both.

If an attack involving an AWS results in the deaths of civilians – both the States with jurisdiction over the incident and other States and institutions that are entitled to investigate the incident, such as the ICC or fact-finding commissions, would need to determine whether the deaths were caused by a technical failure or unlawful conduct on the part of the user(s) and/or developers of the AWS. This demands a practical ability to scrutinize the operation, performance and effect of AWS and trace back whether and how these result from decisions and actions made by people involved in the development and use of AWS.

Certain emerging technologies in the area of AWS, such as certain approaches to artificial intelligence and machine learning (ML), could make the task of investigating the cause of an incident difficult. Machine learning methods, such as deep learning, could offer military benefits but they are also opaque in their functioning. As they stand, current ML techniques used in target recognition software are not explainable, which means that a programmer or a user cannot fully understand how they learn to recognize a target type. This opacity could make it difficult to determine after the fact what caused a system to strike civilians or civilian objects. Even in situations where a technical problem can be excluded, attribution problems could also emerge as the operation, performance and effect of the AWS is determined by decisions made by multiple people at different points in time and, in part, depends on the interaction of the AWS with the environment. Tracing back whose conduct is blameworthy could be difficult.

The takeaway here for the regulation of AWS is two-fold. Should States decide to explicitly prohibit AWS that are incompatible with IHL or otherwise posing unacceptable risks to civilians and other protected persons, such a prohibition should make explicit that technical characteristics and forms of human-machine interaction that preclude the ability to trace back the cause of a harmful incident are off-limits. That could include unexplainable machine learning algorithms. Efforts to codify lawful uses of AWS could, on the other hand, make traceability a critical requirement for the design and use of an AWS. On the technical side, that could entail that algorithms on which the targeting functions are based should be transparent, explainable, and interrogable enough to identify legal/illegal conduct and blameworthy individuals. On the organisational side, that could entail, as suggested, developing and

using, a scheme of responsibility, but also mechanisms to record and trace back decisions in the development and use of AWS.

Exploring how accountability for IHL violations involving AWS would be ensured may seem to some actors premature if not irrelevant, as the use of AWS is – depending on one's understanding of an AWS – not yet an operational reality. With this post we hope to have demonstrated that it is a useful and much-needed exercise for the policy process on AWS, as it provides a lens to explore what is, or should be, demanded, permitted, and prohibited in the development and use of AWS.

v. CONCLUSION

The establishment of a global regulatory framework for AI in AWS is imperative to balance military advancements with ethical and legal obligations. While challenges exist in harmonizing international perspectives, multilateral cooperation, technological safeguards, and enforceable legal standards can mitigate risks. Moving forward, an inclusive dialogue involving governments, technology experts, and human rights organizations is crucial for shaping a sustainable and ethical future for AI in warfare.
Consequently, the endeavour to minimize risks of AI and AT must not focus on definitional questions regarding LAWS but concentrate on binding principles for responsible AI research.
This alternative track would take into account the fact that 'autonomy' for technological artefacts, e.g. LAWS, can and should be regarded as a proxy term for the loss of human control and responsibility for outcomes of technological processes. Principles guiding AI research could require programmers and engineers only to develop technological artefacts whose outcomes will stay controllable for humans, and for which the latter would, hence, always bear responsibility. Initiatives of professional organizations as well as representatives of the private sector have led to several lists of principles for responsible/ ethical research on AI and autonomy (ANNEX). It would be advisable to bundle those principles and create an international body that would supervise compliance.
Consequently, an open discussion on whether or not humanity accepts the fact that technology is already crossing a threshold after which its creations might not be controllable for humans anymore, must be encouraged. Luckily, the UN CCW's debate on LAWS has brought this crucial moment into the public spotlight. Yet, for a purposeful discussion of this broader question, both the architecture of the CCW forum as well as its limited mandate of LAWS are unsuitable.

Humanity is striding into a future where machines and software will have an unprecedented role in almost all aspects of our lives. Moreover, future technology may have immense potential for humans to define what they want to become. If we want to navigate wisely through a future that we might share

with artefacts with cognitive abilities, we need to discuss some serious questions on 'autonomy', 'responsibility,' 'privacy' and 'identity'– and we have to do it now.

vi. KEY QUESTIONS TO CONSIDER:

1. How can we establish a universally accepted definition of "autonomous weapons systems" to ensure clarity in regulation?

- Examine existing definitions and their limitations.
- Analyze the role of "human control" in these systems.
- Consider the implications of varying levels of autonomy.

2. What are the most pressing ethical dilemmas posed by the development and deployment of AI in weapons systems?

- Investigate the potential for algorithmic bias and discrimination.
- Assess the psychological impact on soldiers and civilians.
- Debate the moral implications of machines making life-or-death decisions.

3. How can we ensure transparency and accountability in the development and use of AI for military purposes?

- Explore mechanisms for monitoring and oversight of AI development.
- Consider the role of international cooperation in enforcing regulations.
- Discuss the challenges of attributing responsibility for AI-driven actions.

4. What existing international legal frameworks can be applied to AI in weapons systems, and where are the gaps?

- Analyze the applicability of International Humanitarian Law (IHL) and other treaties.
- Identify areas where new legal norms may be needed.
- Consider the challenges of enforcing legal standards in a rapidly evolving technological landscape.

5. How can we prevent an AI arms race and ensure that AI is used for peaceful purposes?

- Examine the role of export controls and technology transfer restrictions.
- Consider the potential for arms control agreements specific to AI.
- Discuss the importance of promoting international dialogue and cooperation.

6. What are the potential risks and benefits of using AI in autonomous weapons systems for national security?

- Assess the potential for increased efficiency and precision in military operations.
- Consider the risks of accidental escalation and unintended consequences.
- Debate the long-term strategic implications of AI-powered warfare.

7. How can we foster collaboration between governments, industry, and civil society to develop ethical guidelines and best practices for AI in weapons systems?

- Explore the role of multi-stakeholder initiatives and public-private partnerships.
- Consider the importance of public awareness and education.
- Discuss the need for ongoing dialogue and adaptation as technology evolves.

8. What international legal instruments should govern the use of AI in autonomous weapons?

- Analyze the adequacy of existing treaties like the Convention on Certain Conventional Weapons (CCW).
- Consider the need for new protocols or agreements to specifically address AI in warfare.
- Examine how international law can adapt to rapid technological advancements in AI.

9. How can AI weapons comply with International Humanitarian Law (IHL)?

- Discuss the challenges of ensuring AI systems adhere to principles of distinction, proportionality, and necessity.
- Explore the concept of "meaningful human control" and its implications for IHL compliance.
- Consider mechanisms for testing and verifying the compliance of AI weapons with IHL.

10. What measures should be taken to ensure human oversight of AWS?

- Define clear criteria for human oversight and control in the use of AI weapons.
- Explore technological solutions for maintaining human-in-the-loop or human-on-the-loop control.
- Consider the role of training and protocols in ensuring responsible human oversight.

11. Should there be a global ban on certain forms of AI-driven weaponry?

- Debate the merits of a preemptive ban versus a regulatory approach.
- Identify specific types of AI weapons that may warrant a ban due to their inherent risks.
- Consider the challenges of verifying and enforcing a global ban.

12. How can AI-related military developments be monitored effectively?

- Explore the role of transparency and information sharing in monitoring AI development.
- Consider the use of verification technologies and mechanisms for arms control.
- Discuss the challenges of monitoring dual-use AI technologies.

13. What role should private-sector AI developers play in regulation?

- Examine the responsibilities of AI developers in ensuring their technologies are not used for harmful purposes.
- Discuss mechanisms for industry self-regulation and ethical guidelines.
- Consider the role of government oversight and potential collaboration with the private sector.

REFERENCES AND FURTHER READINGS (Consider The Footnotes as Well)

- **United Nations: The role of AI in international security** Retrieved from https://www.un.org/en/role-ai-international-security
- **United Nations Office for Disarmament Affairs: Lethal autonomous weapons systems** Retrieved from https://disarmament.unoda.org/the-convention-on-certain-conventional-weapons/background-on-laws-in-the-ccw/
- **Human Rights Watch (2020, August 10): Stopping killer robots: Country positions on banning fully autonomous weapons and retaining human control.** Retrieved from https://www.hrw.org/report/2020/08/10/stopping-killer-robots/country-positions-banning-fully-autonomous-weapons-and
- **International Committee of the Red Cross: Artificial intelligence and warfare** Retrieved from https://www.icrc.org/en/document/artificial-intelligence-and-warfare
- **Future of Life Institute_ Governing AI: Lessons from nuclear arms control** Retrieved from https://futureoflife.org/governing-ai-lessons-nuclear-arms-control
- Artificial Intelligence: Autonomous Technology (AT), Lethal Autonomous Weapons Systems (LAWS) and Peace Time Threats By Regina Surber, Scientific Advisor, ICT4Peace Foundation and the Zurich Hub for Ethics and Technology (ZHET)

- Air and Space Power Review Vol 22 No 3, Viewpoint, An Ethics Framework for Autonomous Weapon Systems, Professor Peter Lee

---