

Regulating Further Scientific Development of Mass Destruction Weapons with Regards to Artificial Intelligence

Forum: Disarmament Commission

Student Officer: Angela Kim, Deputy President

Introduction

The prospect of developing fully autonomous weapons is no longer a matter of science fiction and is already fueling a new global arms race. From defensive systems like the Israel's Iron Dome that can stop incoming missiles or torpedoes faster than a human could react to "killer robots" that could revolutionize battlefield, ongoing renaissance of artificial intelligence (AI) is reshaping the world.

One of the fundamental moral questions is the effect of autonomous weapons systems on the security of member states and their people. On this matter, the message of the AI community has been clear: Because they do not require an individual human supervision, autonomous weapons are potentially scalable weapons of mass destruction (WMDs); an essentially unlimited number of such weapons can be launched by a small number of people, which is an inescapable logical consequence of autonomy. Numerous international actions have been made regarding this agenda in the past decade, such as the Conventions on Certain



Iron Dome in Israel.

Conventional Weapons (CCW) and Campaign to Stop Killer Robots, and with it the rise of WMDs with regards to AI is widely regarded as a grave threat to humanity that deserves urgent multilateral action.

Background

The use of autonomy in weapon systems has grown swiftly over the past 50 years. The post-cold war global strategic landscape is currently in an extended process of being redrawn as a binary Russian–US nuclear juxtaposition, a legacy of the old Soviet–US confrontation, and is being gradually augmented by regional nuclear competitions. As the arms control framework that the Soviet Union and the USA created at the end of the cold war disintegrates, the commitment of the two states with the largest nuclear arsenals to pursue strategic stability through arms control and disarmament is in doubt to an unprecedented degree. On top of this comes the impact of a fourth industrial revolution, characterized by rapid and converging advances in multiple technologies including artificial intelligence (AI), robotics, quantum technology, nanotechnology, biotechnology, and digital fabrication. AI is not a weapon per se; it is not even a definite unified technology. It is more accurate to compare AI to electricity than to a specific weapon technology such as missiles or nuclear weapons. Yet, it is beyond dispute that nuclear-armed states seek to exploit these technologies for their national security.

	AI as a Deadly Weapon <i>per se</i>	AI as an Agent of Control	AI as an Agent of Design
Actuality	- Fake news - Hacking	- Missiles - Robots - Conventional arms	- Predicting reactions - Mapping biological mechanisms
Potentiality	- Hostile takeover - Financial crisis - Social unrest - Superintelligence	- Drones - UAVs	- Unstructured learning - Superweapons

Classification of AI as WMDs

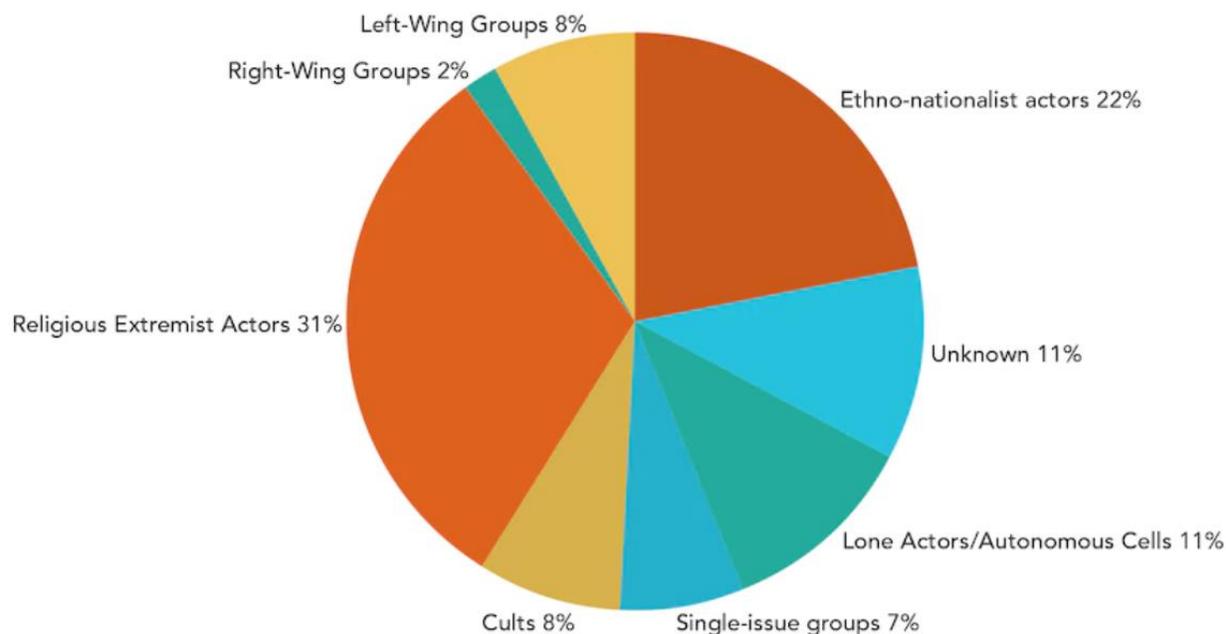
Fears of AI being used to support WMDs can be traced back to at least 1987, where Yazdani and Whitby argued AI at the time was increasing “the possibility of an accidental nuclear war.” These fears carry over to the present and are possibly even more justified as AI technology flourishes.

Concerns over near-peer technological advances have been substantiated by several AI-based developments in Russia. In addition, China’s State Council released a major report on AI, “A Next Generation Artificial Intelligence Development Plan,” in July 2017. This document acknowledges “the world’s major developed countries are taking the development of AI as a major strategy to enhance national competitiveness and protect national security.” Roman Yampolskiy, an associate professor of computer engineering at the University of Louisville, stated in January 2017 that “weaponized AI is a weapon of mass destruction and an AI Arms Race is likely to lead to an existential catastrophe for humanity.” Despite the first discussion of this matter at the United Nations Human Rights Council in 2013, an agreement upon definitions regarding what we are regulating is necessary, and it is determined whether we can achieve humanitarian benefits through the use of this technology.

Problems Raised

Changing Proliferation Threats

WMD development and proliferation occurs in an increasingly complex operating environment. The same venues that facilitate legal global business and spur economic growth can also be utilized by illicit actors. In addition to using circuitous trade routes to ship goods, money, and illegal materials, proliferators use digital financial transactions, cryptocurrencies, and newly established, sometimes one-time-use front companies to circumvent authorities and finance WMD proliferation.



CBRN Incidents by Non-State Actors, 1990-2016

These activities can result in “clean” entities or entities less likely to be monitored through traditional counterproliferation approaches such as sanction lists, or national technical means such as satellite monitoring being involved, in illicit transactions. Such methods of obfuscating identity increase a WMD proliferator’s ability to evade detection, boosting the probability of success. In times of crisis, such as the COVID-19 pandemic, proliferators may seek to take further advantage of shifting national attention, supply chain disruptions, and resource challenges to continue undetected development and proliferation of WMD or related materials.

This is a deeply entrenched phenomenon of contemporary international relations, especially when it comes in forms of Credible Minimum Deterrence. All states have the inherent right to self-defense and consequently the right to acquire arms for their security, including arms from outside sources, as enshrined in the Charter of the United Nations. Furthermore, law of self-defense is justified under the ‘dual conditions’ which is necessity and proportionality as highlighted in the article 49 of the International Law Commission’s Articles on Responsibility of States, where necessity states that the possession of nuclear arms is necessary to prevent further destruction and not simply retaliatory, and proportionality relates to the general requirement that only such force is used as is needed to repel the attack. This idea, if manipulated, could be utilized among nuclear-possessing nations to claim their legal and sovereign rights to proliferate WMDs with the incorporation of AI, which could be a threat to international security.

Potential of Nuclear Escalation

The connection between AI and nuclear weaponry is not new. In fact, AI has been part of the nuclear deterrence architecture for decades. As early as in the 1960s, the United States and the Soviet Union saw that the nascent field of AI could play a role in the development and maintenance of their retaliatory capability—that is, the capability to respond to a nuclear attack, even by surprise. They pursued the development of AI systems that could make their command-and-control process more agile and give decision-makers more time to focus on what really mattered: deciding whether

to launch a nuclear strike or not. Early application of AI included automating threat detection, logistical planning for the transmission of launch orders, and missile targeting and guidance.

However, proliferation of nuclear weapons with regards to AI, in recent decades, acquired a dimension and qualitative characteristics which, together with the increase in illicit arms trafficking, give rise to serious and urgent concerns. Data collection and analysis conducted by AI systems could enable precision strikes to destroy key command, control, and communication assets for nuclear forces. This would be a significant shift from Cold War nuclear strategy, which avoided this type of counterforce targeting. If states can target each other's nuclear weapons and command infrastructure, then second-strike capabilities will be at risk, ultimately jeopardizing mutually assured destruction (MAD).

Accidents are also amplified within the nuclear realm. There are already examples of accidents involving automated conventional weapons systems: in March 2003, U.S. Patriot missile batteries shot down a British fighter plane and a U.S. fighter jet while operating in an “automated mode,” killing the crews of both planes. Accidents are likely to increase as AI systems become more complex and harder for humans to understand or explain. Accidents like these, which carry high costs, decrease overall trust in automated and AI systems, and will increase fears about what will happen if nuclear weapons systems being to rely on AI.



A Trident II D5 Missile Test

Removal of Human Responsibility



Russia's "Dead Hand"

Early on, nuclear-armed states not only identified the appeal of AI for nuclear deterrence, but they also saw its limitations. Given the dramatic consequences that a system failure would have, they were reluctant to hand over higher-order assessments and launch decisions to AI systems: a human had to remain ‘in the loop’. The Soviet Union is the only country that pursued the development of fully automated command and control systems for nuclear weapons. This system, known as the “Dead Hand,” was however meant to be activated only in the exceptional case

of a decapitating attack on the Soviet nuclear command and control.

With the advent of autonomous systems today, however, autonomous unmanned systems such as aerial drones or unmanned underwater vehicles could be seen by nuclear weapon states as an alternative to intercontinental ballistic missiles (ICBMs) as well as manned bomber and submarines for nuclear weapon delivery. Indeed, while distancing humans from the dangers of the battlefield through use of robotic systems may help compensate human inadequacies—especially when it comes to dealing with stress, danger, and lack of stamina—at the same time, new problems are created. The danger is that humans are progressively removed from the chain of decision-making

and responsibility. When experienced in real time, the reasons why an algorithm has taken a decision to act or has preselected and suggested such a course of action, can no longer be comprehended.

International Actions

Biological Weapons Convention (BWC)

Opened for signature on 10 April 1972, the Biological Weapons Convention (BWC), is the first multilateral disarmament treaty banning the development, production and stockpiling of an entire category of weapons of mass destruction. Under the Second Review Conference (1986) and the Third Review Conference (1991), State Parties agreed to implement several confidence-building measures (CBM) in order to prevent the occurrence of ambiguities in the field of biological activities.

Campaign to Stop Killer Robots

As of July 2020, the Campaign to Stop Killer Robots is comprised of 165 nongovernmental organizations in 65 countries. The Campaign is working to ban fully autonomous weapons and retain meaningful human control over the use of force. In the period since Human Rights Watch and other nongovernmental organizations launched the Campaign to Stop Killer Robots in 2013, the question of how to respond to concerns over fully autonomous weapons has steadily climbed the international agenda.

Conventions on Certain Conventional Weapons (CCW)

Concluded at Geneva on October 10, 1980, and entered into force in December 1983, it seeks to prohibit or restrict the use of certain conventional weapons which are considered excessively injurious or whose effects are indiscriminate. The convention currently includes five protocols and bans the use of blinding laser weapons, for example. Demands are now being voiced for agreement on a sixth protocol banning the use of LAWS. While a basic ban on LAWS would be possible in the framework of the CCW, the way it has operated in the past has been by regulating the use of weapons, rather than through banning technologies as such.

Treaty on the Non-Proliferation of Nuclear Weapons (NPT)

The Nuclear Nonproliferation Treaty (NPT) is the only multilateral treaty with the goal of creating a binding commitment of disarmament by nuclear-weapon states. As the article VI of the NPT states, each nation “undertakes to pursue negotiations in good faith on effective measures relating to cessation of the nuclear arms race at an early date and to nuclear disarmament, and on a Treaty on general and complete disarmament under strict and effective international control,” NPT seeks to promote cooperation in the field of peaceful nuclear technology and prevent the spread of nuclear weapons and weapons technology through negotiations. More states have ratified the NPT than any other treaty on arms limitation and disarmament. Every five years, the Treaty’s operations are reviewed.

Possible Solutions

While it is no lie that autonomous weapons of mass destruction could impose a threat to the global society, judicious design and the regulation of LAWS could also lead to the potential saving of

noncombatant lives. If properly developed and deployed, it can and should be used towards achieving that end. Some of the ways in which this could be achieved include:

1. Creating a new binding international treaty banning the development of such weapons with the cooperation of global coalition of nongovernmental organizations alongside consultations among the P5 nations.
2. Regulating autonomous weapons usage instead of prohibiting them entirely; consider restrictions in well-defined circumstances rather than an outright ban and stigmatization of the weapons systems; do not make decisions based on unfounded fears – remove pathos and hype while focusing on the real technical, legal, ethical and moral implications.
3. Deploying autonomous WMDs in concert with soldiers, not as their replacement. Human presence in the battlefield should be maintained.
4. Adopting supply chain risk management practices to enhance CWMD missions. Supply chain network mapping and vendor vetting can ensure the dependability of CWMD countermeasures and equipment, as well as support government's efforts to analyze and mitigate WMD development and proliferation.
5. Understanding the risks and benefits posed by artificial intelligence through confidence-building measures (CBMs) and stakeholder dialogue. The international community could also consider building on existing cooperative efforts concerning cyberspace, such as the U.N.'s work on norms and behavior in cyberspace, the Cybersecurity Tech Accords, and Microsoft, Hewlett, and Mastercard's CyberPeace Institute.

Glossary

Algorithms: An algorithm is a set of instructions that leads to an intended end goal from an established initial situation. Often algorithms have repetitive steps, which is called an iteration. They also generally require decisions, comparisons, or logic to complete the intended task. While algorithms in AI offer great benefits to humanity through self-steering systems, recognizing patterns, speech, image and sound, running robots, etc., it can also pose a threat to potential demise of humanity through autonomous super intelligence.

Artificial Intelligence (AI): Artificial Intelligence, shortly AI, is the simulation of human intelligence in machines that are programmed to think like humans and mimic their action. The concept grew out of work that defeated Germany's "Enigma" code during World War II, primarily pushed by Alan Turing.

Countering Weapons of Mass Destruction (CWMD): The Countering Weapons of Mass Destruction (CWMD) Consortium was formed in response to the Government's expressed interest to establish Other Transaction Agreement (OTA) with an eligible entity or group of entities, to include industry, academic, and not-for-profit partners, for advanced development efforts to support agencies that have a need to prototype new technologies related to Countering Weapons of Mass Destruction.

Lethal Autonomous Weapon Systems (LAWS): Described as the third revolution in warfare after gunpowder and nuclear weapons, LAWS are weapon systems that can identify, select and engage a target without meaningful human control.

MUNiSC 2021 Disarmament Commission

The Fourth Industrial Revolution: Coined in 2016 by Klaus Schwab, the fourth industrial revolution is characterized by a range of new technologies that are fusing the physical, digital and biological worlds, as well as the growing utilization of new technologies such as artificial intelligence, cloud computing, robotics, 3D printing, the Internet of Things, and advanced wireless technologies. Impacting all disciplines, economies, and industries, it opened a new paradigm of global flexibility, but at the same time, raised concerns on that governments could fail to employ and regulate new technologies to capture their benefits, shifting power will create important new security concerns, and inequality may grow. Learning how humankind can benefit from this revolution while addressing its challenges is thus essential.

Weapons of Mass Destruction (WMDs): a nuclear, radiological, chemical, biological, or other device that is intended to harm a large number of people.

Timeline

1942 A.D. – The United States sets up the Manhattan Project to develop the first nuclear weapon.

1945 – The United States conducts a first ever 15-20 kiloton nuclear weapon south of Socorro, New Mexico.

1946 – In its first resolution, the UN General Assembly calls for the complete elimination of nuclear weapons and sets up a commission to address nuclear weapons.

1949 – The Soviet Union explodes a nuclear weapon code-named “First Lightning” in Semipalatinsk, Kazakhstan, becoming the second country to develop and successfully test a nuclear device.

1952 – The United States detonates the first hydrogen bomb at Enewetak Atoll in the Marshall Islands, which is 500 times more powerful than the Nagasaki bomb.

1958 – The Campaign for Nuclear Disarmament in the UK holds its first meeting. Its iconic emblem becomes one of the most widely recognized symbols in the world.

1959 – The Antarctic Treaty opens for signature. It establishes that “any nuclear explosion in Antarctica and the disposal there of radioactive waste material shall be prohibited.”

1963 – A treaty banning nuclear testing in the atmosphere, outer space and underwater is signed in Moscow, following large demonstrations in Europe and America against nuclear testing.

1968 – Under the Nuclear Non-Proliferation Treaty, non-nuclear-weapon states agree never to acquire nuclear weapons, and the nuclear-weapon states make a legal undertaking to disarm.

1996 – The International Court of Justice hands down an advisory opinion in which it found that the threat or use of nuclear weapons would generally be contrary to international law, but the court could not conclude if the use of nuclear weapons would be lawful or unlawful in the case of “extreme circumstance of self-defense, in which the very survival of a state would be at stake.”

1996 – The Comprehensive Test Ban Treaty opens for signature at the United Nations. China, France, the United Kingdom, Russia and the United States all sign the treaty. India says it will not sign the treaty.

Sources

- Carus, W. Seth, and Caves, John P. "The Future of Weapons of Mass Destruction: Their Nature and Role in 2030." *National Defense University Press*. Web. June 2014.
ndupress.ndu.edu/Portals/68/Documents/occasional/cswmd/CSWMD_OccationalPaper-10.pdf.
- Conn, Ariel. "Roman Yampolskiy Interview." *Future of Life*. Web. 18 January 2017.
<https://futureoflife.org/2017/01/18/roman-yampolskiy-interview/>.
- Holland, J.H. "Adaptation in natural and artificial systems." *Ann Arbor, MI: The University of Michigan Press*. Print. 1975.
- Johnson, James, and Krabill, Eleanor. "AI, Cyberspace, and Nuclear Weapons." *War on the Rocks*. Web. 31 January 2020. <https://warontherocks.com/2020/01/ai-cyberspace-and-nuclear-weapons/>.
- Piper, Kelsey. "Death by algorithm: the age of killer robots is closer than you think." *Vox*. Web. 21 June 2019. <https://www.vox.com/2019/6/21/18691459/killer-robots-lethal-autonomous-weapons-ai-war>.
- "Regime Strategy and WMD Timeline Events." *Central Intelligence Agency*. Web. 22 April 2007.
https://www.cia.gov/library/reports/general-reports-1/iraq_wmd_2004/WMD_Timeline_Events.html.
- "Stopping Killer Robots." *Human Rights Watch*. Web. 26 August 2020.
www.hrw.org/report/2020/08/10/stopping-killer-robots/country-positions-banning-fully-autonomous-weapons-and.
- "The Role of Autonomy in DoD Systems." *Department of Defense*. Web. July 2012.
fas.org/irp/agency/dod/dsb/autonomy.pdf.
- "Treaty on the Non-Proliferation of Nuclear Weapons (NPT)." *United Nations*. Web.
<https://www.un.org/disarmament/wmd/nuclear/npt/>.
- Torrens, Linda E. "Conflict in the 21st Century: Counterstrategies for the WMD Terrorist." *Maxwell Air Force Base*. Web. April 1999. <https://apps.dtic.mil/dtic/tr/fulltext/u2/a395718.pdf>.
- Turing, A. M. "Computing Machinery and Intelligence." *Mind*. Print. vol. 59, no. 236, 1950, pp. 433–460. www.jstor.org/stable/2251299.
- "Weapons of Mass Destruction." *International Peace Institute*. Web. no. 3, 2009.
https://www.ipinst.org/wp-content/uploads/publications/wmd_epub_final.pdf.