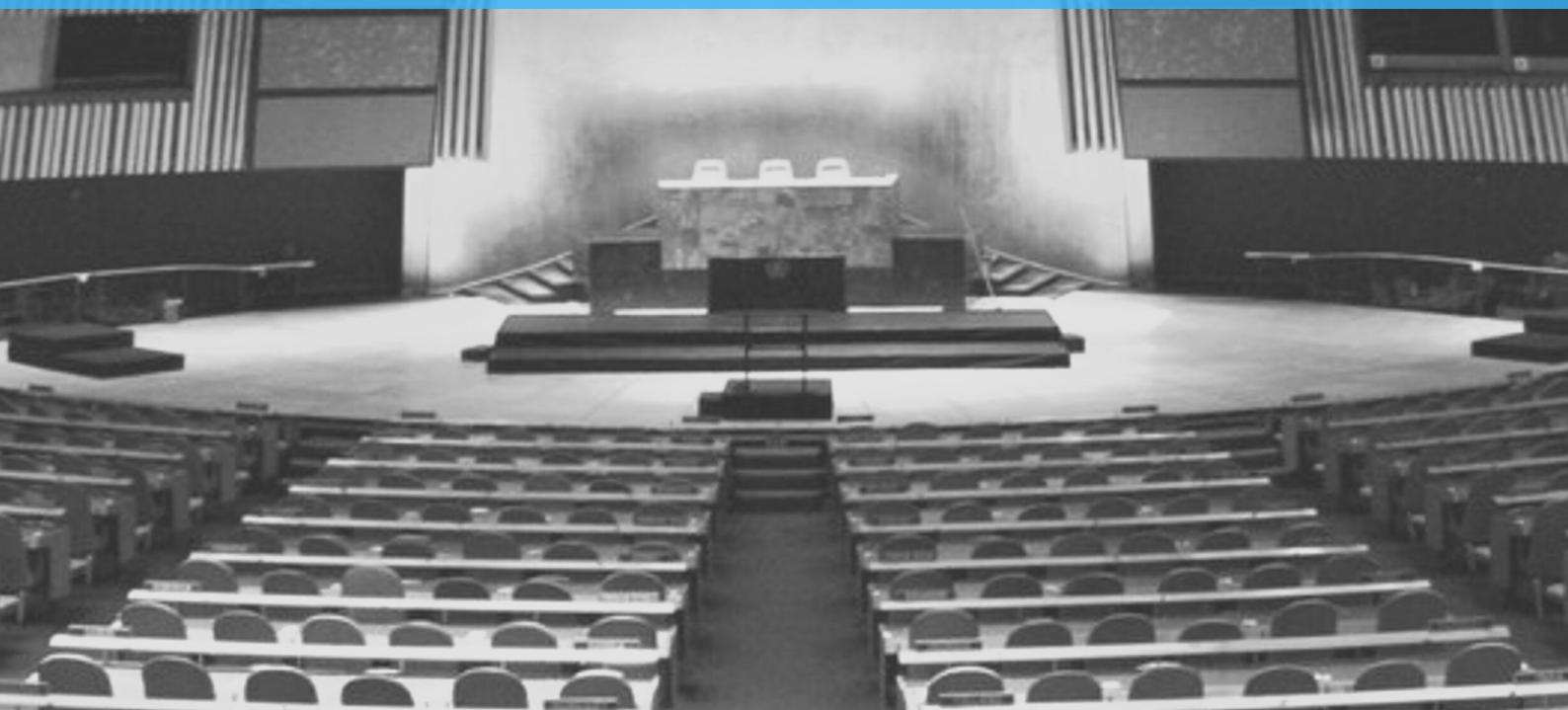


INTERNATIONAL



MUN'19

United Nations General Assembly



*Agenda: Potential Use of Lethal
Autonomous Weapons in War and the
Need for their Regulation*





A WORD FROM THE EXECUTIVE BOARD

Greetings Delegates,

On behalf of the Secretariat of IUMUN'19, I would like to welcome you to the substantive session of United Nations General Assembly. The agendas assigned to the council is in regards with the Potential Use of Lethal Autonomous Weapons in War and the Need for their Regulation. To facilitate a better level of debate the delegates are expected to understand the basic composition/working/procedure of the United Nation and in particular the General Assembly.

We are thrilled to be a part of IUMUN'19 as it will be a platform for discussion, deliberation and consensus over the important and prevalent issues. We promise to deliver an engaging, enticing and exciting council engraved deeply with UN procedures and principles. Diplomacy and adherence to foreign policy will be of the utmost importance. Knowledge of international politics, the state of global affairs is required to make constructive progress to complex and intertwined multi-lateral problems.

The topic we have on the floor are multifaceted and highly relevant in the twenty-first century global scenario. As the Executive Board, we would implore you to explore the issues with multiple lenses and consider diverse viewpoints of various actors, nations and regions impacted directly or indirectly by the issues for a comprehensive approach while researching for the conference.



A WORD FROM THE EXECUTIVE BOARD

The following pages will guide you on the working of the council, and also with relevant information related to the agenda. The Background Guide is intended to give you an insight as to what we as the Executive Board expect from you in terms of what to debate upon, how to frame a solution, etc. Please NOTE once again – this guide may NOT be the only source of your research, and citing arguments, statements, and accusations from the guide will NOT be accepted. Please understand that the UN is not a court of law, discussions and deliberations that happen in the UN chambers although should be done within the ambit of proper legality but the force of law is not the only instrument that is used while negotiating or discussions for the issue at hand.

The entire Executive Board and I are eager to meet you expect the best from all the delegates; I know each one is equally talented to rise to the occasion to solve the agenda at hand. I anticipate you to enjoy the conference make new bonds and friendships in the 3 days of the conference.

Regards,
Pulkit Taneja
Vice-Chairperson

TABLE OF CONTENTS

1. Accepted Sources
2. Mandate
3. United Nations General Assembly
4. Introduction
5. Artificial Intelligence
6. Autonomous Technology
7. Lethal Autonomous Weapons System
8. Debate at the UNCCW
9. Artificial Intelligence
10. Autonomously Generated Profiles
11. Autonomous Technology and Resource Scarcity
12. Moral Argument for Sustainable Environment
13. Conclusion

ACCEPTABLE SOURCES

State – Operated News Agencies

These reports can be used in support of or against the State that owns the news agency. Such credible reports, provided they are substantial, can be used in support of or against any country. However, these can be denied by any other country in the council. Some examples are:

- BBC (United Kingdom)
- Al Jazeera (Qatar)
- RIA Novosti (Russia)
- Xinhua News Agency (PR China)

Government Reports

These reports can be used in a similar way as the State-operated News Agencies reports and can be denied by another country under any circumstances. Note that the Executive Board can still accept a report as credible information despite denial by a certain country. Some examples are:

- Government Websites like the State Department of the United States of America [<http://www.state.gov/index.htm>] or the Ministry of Defense of the Russian Federation [<http://www.mil.ru/en/index.htm>]
- Ministry of Foreign Affairs of various nations like India [<http://www.mea.gov.in/>] or People's Republic of China [<http://www.fmprc.gov.cn/eng/>].
- Permanent Representatives to the United Nations Reports <http://www.un.org/en/members/> (Click on any country to get the website of the Office of its Permanent Representative.)
- Multilateral Organizations like the NATO [<http://www.nato.int/cps/en/natolive/index.htm>], ASEAN [<http://www.aseansec.org/>], OPEC [http://www.opec.org/opec_web/en/], etc.

United Nations Reports

The Executive Board of the IAEA considers all UN Reports as credible sources of information or evidence.

- UN Bodies like the UNSC [<http://www.un.org/Docs/sc/>] or UNGA [<http://www.un.org/en/ga/>].
- UN Affiliated Bodies like the International Atomic Energy Agency [<http://www.iaea.org/>], World Bank [<http://www.worldbank.org/>], International Monetary Fund [<http://www.imf.org/external/index.htm>], International Committee of the Red Cross [<http://www.icrc.org/eng/index.jsp>], etc.
- iii.) Treaty Based Bodies like the Antarctic Treaty System [<http://www.ats.aq/e/ats.htm>], the International Criminal Court [<http://www.iccpi.int/Menu/ICC>]

ACCEPTABLE SOURCES

Wikipedia and Miscellaneous Sources

Sources like Wikipedia [<http://www.wikipedia.org/>], Amnesty International [<http://www.amnesty.org/>], Human Rights Watch [<http://www.hrw.org/>] or newspapers like the Guardian [<http://www.guardian.co.uk/>], Times of India [<http://timesofindia.indiatimes.com/>], etc. are typically not accepted as proof evidence.

However, they may be used for better understanding of any issue and on rare occasions, be brought up in debate if the information given in such sources is in line with the beliefs of a Government.

Furthermore, the information submitted as evidence citing reportage from sources such as specified in this note may be at best, treated as having significance in terms of persuasive value – e.g. to cement one's assertions, but never as binding, indisputable fact.

MANDATE

The General Assembly is one of the six main organs of the United Nations, the only one in which all Member States have equal representation: one nation, one vote. All 193 Member States of the United Nations are represented in this unique forum to discuss and work together on a wide array of international issues covered by the UN Charter, such as development, peace and security, international law, etc. In September, all the Members meet in the General Assembly Hall in New York for the annual General Assembly session.

The General Assembly subsidiary organs are divided into five categories: committees (30 total, six main), commissions (six), boards (seven), councils and panels, working groups, and "other".

The main committees are numbered, 1–6:

- The First Committee: Disarmament and International Security (DISEC)
- The Second Committee: Economic and Financial (ECOFIN)
- The Third Committee: Social, Cultural, and Humanitarian (SOCHUM)
- The Fourth Committee: Special Political and Decolonization (SPECPOL)
- The Fifth Committee: Administrative and Budgetary and General
- The Sixth Committee: Legal

UNITED NATIONS GENERAL ASSEMBLY

Composition

Article 9

- 1) The General Assembly shall consist of all the Members of the United Nations.
- 2) Each Member shall have not more than five representatives in the General Assembly.

Functions and Powers

Article 10

The General Assembly may discuss any questions or any matters within the scope of the present Charter or relating to the powers and functions of any organs provided for in the present Charter, and, except as provided in Article 12, may make recommendations to the Members of the United Nations or to the Security Council or to both on any such questions or matters.

Article 11

- 1) The General Assembly may consider the general principles of co-operation in the maintenance of international peace and security, including the principles governing disarmament and the regulation of armaments, and may make recommendations with regard to such principles to the Members or to the Security Council or to both.
- 2) The General Assembly may discuss any questions relating to the maintenance of international peace and security brought before it by any Member of the United Nations, or by the Security Council, or by a state which is not a Member of the United Nations in accordance with Article 35, paragraph 2, and, except as provided in Article 12, may make recommendations with regard to any such questions to the state or states concerned or to the Security Council or to both. Any such question on which action is necessary shall be referred to the Security Council by the General Assembly either before or after discussion.
- 3) The General Assembly may call the attention of the Security Council to situations which are likely to endanger international peace and security. The powers of the General Assembly set forth in this Article shall not limit the general scope of Article 10.

Article 12

- 1) While the Security Council is exercising in respect of any dispute or situation the functions assigned to it in the present Charter, the General Assembly shall not make any recommendation with regard to that dispute or situation unless the Security Council so requests.

UNITED NATIONS GENERAL ASSEMBLY

2) The Secretary-General, with the consent of the Security Council, shall notify the General Assembly at each session of any matters relative to the maintenance of international peace and security which are being dealt with by the Security Council and shall similarly notify the General Assembly, or the Members of the United Nations if the General Assembly is not in session, immediately the Security Council ceases to deal with such matters.

Article 13

1) The General Assembly shall initiate studies and make recommendations for the purpose of: a. promoting international co-operation in the political field and encouraging the progressive development of international law and its codification; b. promoting international co-operation in the economic, social, cultural, educational, and health fields, and assisting in the realization of human rights and fundamental freedoms for all without distinction as to race, sex, language, or religion.

2) The further responsibilities, functions and powers of the General Assembly with respect to matters mentioned in paragraph 1 (b) above are set forth in Chapters IX and X.

Article 14

Subject to the provisions of Article 12, the General Assembly may recommend measures for the peaceful adjustment of any situation, regardless of origin, which it deems likely to impair the general welfare or friendly relations among nations, including situations resulting from a violation of the provisions of the present Charter setting forth the Purposes and Principles of the United Nations.

Article 15

1) The General Assembly shall receive and consider annual and special reports from the Security Council; these reports shall include an account of the measures that the Security Council has decided upon or taken to maintain international peace and security.

2) The General Assembly shall receive and consider reports from the other organs of the United Nations.

Article 16

The General Assembly shall perform such functions with respect to the international trusteeship system as are assigned to it under Chapters XII and XIII, including the approval of the trusteeship agreements for areas not designated as strategic.

UNITED NATIONS GENERAL ASSEMBLY

Article 17

- 1) The General Assembly shall consider and approve the budget of the Organization.
- 2) The expenses of the Organization shall be borne by the Members as apportioned by the General Assembly.
- 3) The General Assembly shall consider and approve any financial and budgetary arrangements with specialized agencies referred to in Article 57 and shall examine the administrative budgets of such specialized agencies with a view to making recommendations to the agencies concerned.

Voting

Article 18

- 1) Each member of the General Assembly shall have one vote.
- 2) Decisions of the General Assembly on important questions shall be made by a two-thirds majority of the members present and voting. These questions shall include: recommendations with respect to the maintenance of international peace and security, the election of the non-permanent members of the Security Council, the election of the members of the Economic and Social Council, the election of members of the Trusteeship Council in accordance with paragraph 1 (c) of Article 86, the admission of new Members to the United Nations, the suspension of the rights and privileges of membership, the expulsion of Members, questions relating to the operation of the trusteeship system, and budgetary questions.
- 3) Decisions on other questions, including the determination of additional categories of questions to be decided by a two-thirds majority, shall be made by a majority of the members present and voting.

Article 19

A Member of the United Nations which is in arrears in the payment of its financial contributions to the Organization shall have no vote in the General Assembly if the amount of its arrears equals or exceeds the amount of the contributions due from it for the preceding two full years. The General Assembly may, nevertheless, permit such a Member to vote if it is satisfied that the failure to pay is due to conditions beyond the control of the Member.

UNITED NATIONS GENERAL ASSEMBLY

Procedure

Article 20

The General Assembly shall meet in regular annual sessions and in such special sessions as occasion may require. Special sessions shall be convoked by the Secretary-general at the request of the Security Council or of a majority of the Members of the United Nations.

Article 21

The General Assembly shall adopt its own rules of procedure. It shall elect its President for each session.

Article 22

The General Assembly may establish such subsidiary organs as it deems necessary for the performance of its functions.

INTRODUCTION

Autonomous Technology (AT) can be said to be the essence of Lethal Autonomous Weapons Systems (LAWS), which have triggered a legal and policy debate within the international arms control framework of the United Nations Convention on Certain Conventional Weapons (UNCCW) that is now entering its fifth year.

Since LAWS highly challenge existing International Humanitarian Law (IHL) due to their capacity of replacing a human operator on a weapons platform, the CCW's tasks of, i.a., ensuring that the concepts of legal accountability and human responsibility do not become void, and assessing whether LAWS are legal under IHL, are of utmost importance.

However, LAWS are not the only manifestation of the security risks of AT. AT not only poses risks to global society if weaponized, but can pose tremendous systemic risks to global society and humanity also when not weaponized. This potentially dangerous transformative power of AT, which is beyond the scope of the CCW's mandate, will be the thematic core of this paper. Based on a risk assessment of not-weaponized AT, the paper will present thought-provoking impulses that can shape an international interdisciplinary debate on the risks of AT specifically and of emerging technologies more generally.

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. 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 by 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.

AUTONOMOUS TECHNOLOGY

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 automation 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.

Some experts use a narrower understanding and limit the use of the attribute ‘autonomous’ to more complex technological processes. They argue that AT extends beyond conventional automation and can solve application problems by using materially different algorithms and software system architectures. This perspective is narrower and clearly locates the emergence of AT within the research of modern AI.

LETHAL AUTONOMOUS WEAPONS SYSTEMS

An autonomous weapon system is defined as follows: "Any weapon system with autonomy in its critical functions – that is, a weapon system that can select (search for, detect, identify, track or select) and attack (use force against, neutralize, damage or destroy) targets without human intervention."

After initial launch or activation by a human operator, it is the weapon system itself – using its sensors, computer programming (software) and weaponry that takes on the targeting functions that would otherwise be controlled by humans. This working definition encompasses any weapons system that can independently select and attack targets, including some existing weapons and potential future systems.

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.

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. 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. 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; 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.

Compliance with International Humanitarian Law

Autonomous weapon systems, as defined, are not specifically regulated by International Humanitarian Law (IHL) treaties. However, it is undisputed that any autonomous weapon system must be capable of being used, and must be used in accordance with the IHL.

DEBATE AT THE UNCCW

The main points of discussion of the GGE were LAWS's potential legality under IHL, questions of accountability and responsibility for the use of LAWS during armed conflict, potential (working) definitions of LAWS, as well as the need for emerging norms, since LAWS highly challenges both existing IHL as well as normative principles. However, this first GGE on LAWS brought no agreement on a political declaration and also no path toward a new regulator international treaty. The only common denominator was the general will of states to continue conversations in 2018. The UN CCW's debate highlights at least five severe challenges to a comprehensive understanding of the risks of LAWS and AT.

(1) To date, states have not agreed on a definition of LAWS or the concept of autonomy, or on whether increasingly autonomous weapons systems, or precursor technologies, already exist. Moreover, national as well as international policy debates on LAWS have lacked precise terminology.

Bearing in mind the above-described thoughts on the technological concept of 'autonomy', this is no surprise. However, it is claimed that definitions will most likely play a key role in the international deliberation on the issue of LAWS.

In order to comprehensively discuss, and reach agreement on, a topic, it is crucial to base the debate on a common understanding of the issue. It is also important in light of the fact that there exists a not negligible movement of some states and NGOs to ban LAWS.

However, since AT and the concept of 'autonomy' for technological artefacts may be a proxy term for an ongoing trend in human technological endeavors to give away control to technological agents thereby relinquishing human responsibility for outcomes of autonomous systems, a premature agreement on a definition of 'autonomy' in weapons systems by the GGE on LAWS would probably hide this trend.

Therefore, instead of pressing for a definition of LAWS and 'autonomy' within the GGE, it would be advisable to locate these challenges within a bigger picture of the general relationship between humans and technology, and focus on the question whether we want to continue to regard technology as a controllable tool. In this sense, the GGE framework could be deemed as unfitting. Surely, principles for responsible AI research are both a first reflection of this underlying and ongoing paradigm change, as well as a first step in the direction of responsibly addressing the seriousness of this risk. A list of existing principles is found in the ANNEX of this paper.

DEBATE AT THE UNCCW

(2) States are generally unwilling to share information on their capacity to develop LAWS.

However, in order to gain a better understanding of the lessons learned from already existing weapons with certain levels of autonomy, the sharing of information is vital.

(3) The GGE's mandate comprises the discussion of emerging technologies in the area of lethal autonomous weapons systems (LAWS) in the context of the objectives and the purposes of the convention.

However, the misuse of technology, e.g., by non-state actors, does not fall within the scope of this mandate.

Certainly, though, a holistic analysis and discussion of the peace and security implications of AT and new technologies requires the international community to address also the use of such by non-state actors.

LAWS represent a new category of weapons, in that their novelty lies in a formless technological capacity to recognize patterns from a continuous inflow of data. The difference between a currently existing remotely controlled drone and a 'fully autonomous' drone does not lie in the casing, but in the fact that the latter is controlled by a software with autonomous capacities. The UN CCW, established in 1983, seeks to prohibit the use of certain conventional weapons. Its protocols currently prohibit the use of weapons whose primary effect is to injure by fragments that, once within the human body, escape X-Ray detection, as well as the use of mines, booby-traps and incendiary weapons against civilians. One may argue that the UN CCW's GGE on LAWS is not capable to fully understand the technological complexity of current (not to mention future) AT.

(5) In addition, the CCW is a framework underpinned by IHL, which narrows the debate's focus on weapons and their use during armed conflict.

However, increasingly autonomous weapons systems can be and are used during peace time in law enforcement operations (e.g. crowd control, hostage situations), where IHRL represents the legal benchmark.

Compared to IHL, IHRL is much more restrictive on the use of force. Military technology often finds its way into law enforcement. One may assume that once the advantages of increasingly autonomous systems have been proven in the military context, they might be considered for use during domestic law enforcement, although IHRL, regulating the latter, would prohibit their use. Therefore, the CCW's/ GGE's approach could be criticized as not being legally comprehensive enough due to its limited focus on the use of a weapons during times of war.

ARTIFICIAL INTELLIGENCE

(1) First, software with autonomous capacities can be used to act and interact entirely in cyberspace. Those sometimes-called autonomous intelligent agents are of tremendous military interest for ‘conventional’ military operations: Autonomous intelligent agents acting in cyberspace can support the decision-making process, they can identify an adversary’s vulnerabilities and they can enable an ever-greater speed of response. Hence, the use of autonomy for intangible cyber operations (defensive or offensive) could be decisive and much more economic in current/future warfare. Five UN GGE discussions on cyber security have taken place since 2004/ 2005, and have confirmed that international law applies to the cyber space. Moreover, those GGEs have decided on a variety of confidence building measures (CBMs), and recommended norms for responsible State behaviour in the domain. Since both autonomous cyber weapons as well as LAWS are characterized by AT, both the GGE on LAWS as well as those on cyber security share the thematic technological basis. Nevertheless, those international policy discussions have nearly no overlap.

(2) Second, it is highly necessary to consider potentially malicious linkages of AT and other emerging technologies. Theoretically, it may be possible to create autonomous systems that control processes with the core aim of harming humans, e.g. the malicious use of biotechnology, 5G radiation, or products of molecular nanotechnology. Current examples of such linkages do not exist. However, it is crucial to raise this concern early enough in order to trigger both research in this field as well as a comprehensive debate of peace and security implications of both AT and other emerging technologies.

(3) Moreover, based on the above-mentioned understanding of ‘autonomy’ for artefacts as any result of a technological process for which the human cannot or does not want to bear responsibility, any intentionally harmful use of a technology whose causes for harm cannot be traced back to a human ‘trigger’ may be deemed an autonomous weapon.

Both fake news (deliberate misinformation via traditional or online media with the intent to mislead the readers) and internet trolls (the posting of erroneous, extraneous and off-topic messages in order to manipulate public opinion) could potentially be generated by autonomous intelligent agents, which could lead to mass disinformation guided by AT. Not only news portals that deliberately and automatically spread fake information, but also social bots on twitter have an immense potential for mass manipulation. Moreover, bots that deceive us are currently already more numerous than those that tell us the truth, and they hardly cost anything.

ARTIFICIAL INTELLIGENCE

In addition to general mass manipulation through widely spread disinformation by bots, research on AI makes it possible to generate individualized information. In this case, people do not share a common reference point for information anymore. The borders between reality and artificial creation with regards to knowledge through individual research would blur.

Further, AI research is able to create so-called ‘generative adversarial networks’ (GAN) that can currently generate fake images and videos whose quality is such that humans are incapable of telling that they are not real shots. Moreover, it is said that GANs could soon generate speech, language and behaviour.

With Adobe’s application ‘Project Voco’ it is also possible to rapidly alter an existing voice recording to include words and phrases that the original speaker has never said. One may assume that an altering of a recording by a machine or software instead of a human may soon be possible too. When real videos, images and voice recordings become indistinguishable from fake ones, fake news will become even more prevalent, and video, image, and voice evidence could become inadmissible in court.

AUTONOMOUSLY GENERATED PROFILES

Computerized pattern and correlation recognition in order to identify and represent people, for example during criminal investigations, could be performed by AT. The detection and capture of potential (pre-emptive profiling) and actual criminals could be outsourced to increasingly autonomous machine calculation based on Big Data – uncontrollable for humans. Already today, deep learning software allows for increasingly perfected facial recognition. Facial recognition technology is a computer application capable of identifying and verifying a person from a digital image or video. It is currently installed in public surveillance cameras, i.a., in Russia and China and used in order to continuously track potential criminals or public dissidents.

Through increasingly autonomous criminal profiling the border between a criminal and a legally innocent person would be drawn exclusively by an algorithm, and vulnerable to incorrect data due to bad sensor-technologies, incompleteness or noise. Furthermore, categorizing potential criminals based on computational inferences somehow turns the presumption of innocence upside down, assuming a general potential for criminal conduct.

Often, AI systems are claimed to be more ‘objective’ in their ‘behavior’ than humans, because they are not influenced by human feelings and prejudices. However, as ‘intelligent’ software and machines need to be ‘fed’ by a huge amount of data in order to ‘learn’ (a trait that we deem ‘intelligent’), there exists the risk that they learn human prejudices from biased data. And so-called machine biases constitute a danger for AI-controlled or autonomous systems that some experts regard as far more acute than LAWS. Based on the data a bot is fed in order to learn, it could learn, e.g., to discriminate people of colour or minorities, or gain a strict political attitude.

AUTONOMOUS TECHNOLOGY AND RESOURCE SCARCITY

The current global social, economic (including financial and monetary) and environmental trends render the planet's resources scarce. This constitutes a risk to humanity and makes our present global human coexistence potentially unsustainable. Hence, some experts ask the question: In an increasingly unsustainable society in critical times, what kind of citizens should be protected, and whose lives could be sacrificed? Should the worth of people's lives be weighed according to a certain benchmark, so that we can more easily decide who could stay alive? And does a human being have the guts to decide – or should we outsource this decision to autonomous software.

For example, autonomous intelligent agents could be integrated into health insurance systems and feeding from patients' data, they could determine who receives a potential treatment and who does not. This may yet be a dystopian idea. However, the idea of a rating system for citizens is already tested in China with the so-called Citizen Score Card, which represents the value of an individual citizen from a governmental perspective. A rating system like this could potentially become a reference point for informing decisions that aims at limiting population figures.

The emergence of AT forces us to evaluate our current economic, social and environmental systems and trends, to ensure that we do not put society at risk of being kept in quantitative borders set by algorithms and based on utilitarian calculations.

Code can be regarded as the regulator of the cyberspace, the way a constitution can be regarded as a regulator of society. Code enables the exchange of data among networks and is currently still generally neutral regarding the content of the data and ignorant about the user. This makes regulating behaviour in the cyberspace difficult. However, code is not fixed, but the architecture of the cyberspace can be changed by the people who code. The fact that it is hard to know who someone is in the Net and what the character of the content is that is delivered, can be changed. New architecture can facilitate identification and rate data content. This architecture can either be privacy-enhancing or not. This depends on the incentives that are being faced by those who set it up.

In other words, there exists a choice whether to influence the 'regulability' of the cyberspace as well as a choice about what this regulation should look like. Moreover, the way a constitution represents the normative values of a society through codifying them by law, code can be said to reflect a 'choice' of values that should guide actions and inactions in the cyberspace. If code represents the law of cyberspace, and computer software potentially interferes with citizens' privacy and maybe physical integrity (LAWS), should their use be restricted and regulated by a democratic process?

AUTONOMOUS TECHNOLOGY AND RESOURCE SCARCITY

This argument for a value-sensitive design of any software code that does or could interfere with citizen's privacy and physical integrity approved by a democratic political process would, as a first step, require a constant and very strong interaction between technological experts and both national and international policy-makers. Only thereby could the current policy discussions on technology lose their theoretical aspect and become more practical, which is crucial in order to potentially introduce the necessary aspects into a legislative process. Creating a fixed national and international policy-technology interface would require an architectural change of national and international political institutions, similar to the United Arab Emirates new state minister for AI.

Furthermore, source codes of AT and AI-controlled systems need to be open source in order to be accessible for a political discussion and potential introduction into a legislative process. This condition will require deeply considered answers on the question of property rights of source codes of autonomous systems.

The international debate on AT and LAWS contains the unexamined assumption that humans and artificially intelligent systems are different only to a degree, and that human qualities can be reproduced in a machine. This underlying belief is the reason why the international debate uses anthropomorphic language – machine 'decision-making', machine 'learning', machine 'intelligence' or 'autonomy' – to describe current technological artefacts.

On this subject it is crucial to highlight two points: First, the human-machine analogy grew out of the initial wish and claim of AI research to understand the human brain by modelling it. However, this analogy still has a mere hypothetical character. Science could not yet fully reveal what happens in the human brain when, e.g., a decision is taken,¹¹¹ or how and if 'consciousness' can be linked to a physical process.

And second, a software is usually named by its purpose, and not by its structure. If the purpose of, e.g., an 'autonomous' software is to supplant the human in an area where the latter used to take a human decision in no way implies that the software 'takes a decision' as well. Hence, by comparing humans and machines or software at a common reference point (e.g. capacity to 'decide', 'learn' or 'behave morally') we may risk falling into a linguistic trap and prematurely overestimate technological artefacts and underestimate human capacities, let alone human language.

Language frames the way we think, understand and compare. Using the same language for machines and software as for humans could lead us to make potentially false comparisons – 'machines decide better than humans'. Keeping in mind also the above-discussed risk for terminological confusion through the term 'autonomy' or 'intelligence', the question whether we need a new language for technological artefacts may be legitimate.

MORAL ARGUMENT FOR SUSTAINABLE ENVIRONMENT

We are on the threshold of a paradigm shift where the human being will not be the only existing ‘intelligent system’ on the planet with the capacity for autonomous action anymore. Depending on the features that are encoded in increasingly autonomous systems and the existing risks of unpredictable outcomes and vulnerabilities to hacking (e.g.), these systems may challenge the structure of current human society and might even become a risk for humanity as a species. Some experts also argue that organic human life is merely a short precursor in the evolutionary history of intelligent ‘life’ in the universe, which might soon be represented by inorganic machines with a far more powerful intellect than humans.

Some are already preparing for a potential emergence of general AI through the enhancement of human brain power through AI itself: Elon Musk’s recently launched company ‘Neuralink’ is exploring ‘neural lace’ technology – the implanting of tiny electrodes into the human brain to give us direct computing capabilities. He argues that a ‘[...] merger of biological intelligence and machine intelligence [...]’ would be necessary for humans to stay economically valuable in a future of general AI. Another way to keep up with AI and AT systems in a potential future world could also be paved by a genetic upgrade of humans through gene editing, which can nowadays already be used to alter the DNA of embryos. In other words, research is focused on technology that would not only help us do, but that has the potential to help us be.

A recent survey with the aim of clarifying expert opinions on the possibility and risks of human- like machine intelligence, based on 550 AI expert opinions, revealed a view among experts that AI systems will probably (over 50%) reach overall human ability by 2040-2050, and very likely (with 90% probability) by 2075. From reaching human-level-intelligence, experts assume that artificial superintelligence will be reached within 30 years after with a probability of 75%. Moreover, the respondents say that the probability that this development may be ‘bad’ or ‘extremely bad’ for humanity is 31%.

Some experts claim that there exists a moral duty to pre-emptively decide not to create an invasive artificial species of autonomous agents that could endanger the lives of human beings on the planet.

CONCLUSION

The international community should not get lost in attempts to define the term ‘autonomy’ for technological artefacts. Years of research and four years of discussions of LAWS within the UN CCW have not lead to terminological clarification, but opinions on the scope and content of the term ‘autonomy’ or AT have become more diverse.

If the term ‘autonomy’ for technological artefacts was defined to include a set of clearly delineated characteristics (e.g. ‘learning’, ‘creating or pursuing of a goal’, ‘independent of human control, operation or intervention’), future technological research might reveal further potential characteristics which then would be excluded from this definition.

Yet a fixed definition of ‘autonomy’ for technological artefacts could lead to a clear definition of LAWS within the GGE. On the one hand, this could encourage a potential outcome of the UN discussions (e.g. Code of Conduct or norms for responsible State behaviour). On the other hand, again, new and yet unknown technological developments interesting for military use might be beyond the scope of this definition of LAWS. Hence, the pressure of defining ‘autonomy’ in order to proceed with the GGE debate would most possibly lead to a definition that reflects the current and maybe also conceivable future technological potentials. However, the exponential pace with which AI research advances must alert us to yet unknown potentials and risks.