Artificial Intelligence: Lethal Autonomous Weapons Systems and Peace Time Threats

Regina Surber, Advisor, ICT4Peace Foundation. Written for the Zurich Hub for Ethics and Technology (ZHET)

Research on Artificial Intelligence (AI) – the simulation of human intelligence processes through computer software – has enabled humanity to create software and software-systems which exhibit a level of intelligence that can make them perform tasks as well as to learn new tasks without human guidance, observance, or intervention. Such so-called *increasingly autonomous intelligent agents* can be purely software, or integrated into a physical system – a robot.[1]

Besides potentially promising applications of increasingly autonomous intelligent systems (e.g. self-driving cars, ISABEL in medical diagnostics), those software agents can be (and arguably already are) integrated into robots that can identify, select, track, and attack a (military) target (e.g. combatants and infrastructure) without a human operator.[2]

Often-called *Lethal Autonomous Weapons Systems (LAWS),* these systems have been taken up as an issue by the international arms control community in the framework of the United Nations Convention on Certain Conventional Weapons (CCW) in 2014.[3] After a series of annual informal discussions, a Group of Governmental Experts (GGE) has debated on the subject matter for the first time during a 5-day-gathering in the CCW framework in Geneva in November this year. The main points of discussion of the GGE were the potential legality under International Humanitarian Law (IHL) of such weapons systems, 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 challenge both existing law (IHL) as well as normative principles.[4]

[1] Guarino, Alessandro, 2013, Autonomous Intelligent Agents in Cyber Offence, in: Podins, K., Stinissen, J., and Maybaum, M. (Eds.), 5th International Conference on Cyber Conflict, NATO CCD COE Publications.
[2] ICRC, 2016, Convention on Certain Conventional Weapons, Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS), April 11 – 15, 2016, Geneva, Switzerland, 1.
[3] CCW/MSP/2014/3.
[4] CCW/GGE.1/2017/CRP.

However, to date, states neither agreed on a definition of LAWS, nor whether increasingly autonomous weapons systems or precursor technologies already exist. Moreover, national as well as international policy debates on LAWS have lacked precise terminology. Hence, there is a strong need to develop or have better technical understanding in the political debate.[5] This becomes even more imperative due to the rapid pace with which autonomy-enhancing technologies advance.[6]

Furthermore, the CCW's discussion on LAWS has focused on conventional (physical/ robotic) systems which interact in a 3D reality with other machines or humans. However, autonomous software agents which act entirely in the cyberspace are of tremendous military interest. The use of autonomy for intangible cyber operations (defensive or offensive) could be decisive and much more economic in current/future warfare.[7]

In addition, the CCW is a framework underpinned by IHL, which narrows the debate's focus on weapons on their use during *armed conflict*.[8] However increasingly autonomous weapons systems can be and are used during peace time in law enforcement operations (e.g. crowd control, hostage situations), where International Human Rights Law (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.[9]

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.

---

[5] Ibid.,1, 7, 9, 10.

[6] UNIDIR, 2017, The Weaponization of Increasingly Autonomous Technologies: Autonomous Weapons Systems and Cyber Operations, UNIDIR Resources No. 7, 1.

[7] Meissner, Christopher, 2016, The Most Military Decisive Use of Autonomy You Won't See, DefenseOne, November 7, 2016, available at http://www.defenseone.com/ideas/2016/11/most-militarily-decisive-use-autonomy-you-wont-see-cyberspace-ops/132964/ (accessed on November 25, 2017). See e.g. the United States' cyberwarfare program MonsterMind. This software could constantly be on the lookout for traffic patterns indicating known or suspected cyberattacks. When it detected an attack, it would automatically block it from entering the country. This is regarded as a "kill" in cyber terminology. See e.g. Zetter, Kim, 2014, Meet Monstermind, The NSA Bot That Could Wage Cyberwar Autonomously, Wired, August 13, 2014, available at https://www.wired.com/2014/08/nsa-monstermind-cyberwarfare/ (accessed on November 28, 2017).

[8] Art. 1 and 2 Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects as amended on 21 December 2001 (CCW).

[9] Heyns, Christof, 2016, Human Rights and the use of Autonomous Weapons Systems (AWS) During Domestic Law Enforcement, Human Rights Quarterly 38, 350-378; Heyns, Christof, 2014, Autonomous Weapons Systems and Human Rights Law, Presentation made at the informal expert meeting organized by the state parties to the Convention on Certain Conventional Weapons, May 13-14, 2017, Geneva, Switzerland.

However, the risk of the use of autonomous intelligent agents during peacetime is not limited to the lack of a legal review based on IHRL:

Mass disinformation generated by intelligent technology: For example, 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 entirely by autonomous intelligent agents.

Autonomously generated profiles: Computerized pattern and correlation recognition in order to identify and represent people, for example during criminal investigations, could be performed by autonomous intelligent agents. The detection and capture of potential (pre-emptive profiling) and actual criminals (e.g.) could be outsourced to increasingly autonomous machine calculation based on Big Data – uncontrollable for humans. Already today, so-called Deep Learning Mechanisms – a subfield of machine learning concerned with algorithms inspired by the structure and function of the brain – allow for ever-more 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 in Russia and China (e.g.) and used in order to continuously track potential criminals or public dissidents. [10] 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, noise and the like. Furthermore, categorizing potential criminals based on computational inferences somehow turns the presumption of innocence upside down, assuming a general potential for criminal conduct.[11]

Autonomous technology in light of emerging resource-scarcity on our planet: The current global social, economic (including financial and monetary) and environmental trends constitute a high risk to humanity and make our present global human coexistence potentially unsustainable. 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 a Citizen Score Card,[12] representing value of an individual citizen from a governmental perspective, become

[10] See e.g. Chin, Josh, and Lin, Lisa, 2017, China's All-Seeing Surveillance State Is Reading Its Citizen's Faces, The Wall Street Journal, June 26, 2017, available at https://www.wsj.com/articles/the-all-seeing-surveillance-state-feared-in-the-west-is-a-reality-in-china-1498493020 (accessed on November 27, 2017); Mezzofiore, Gianluca, 2017, Moscow's facial recognition CCTV network is the biggest example of surveillance society yet, Mashable, September 28, 2017, available at http://mashable.com/2017/09/28/moscow-facial-recognition-cctv-network-big-brother/#kF19SB72r8qA (accessed on November 27, 2017).

[11] Hildebrandt, Mireille, 2015, Smart Technologies and the End(s) of Law, Novel Entanglements of Law and Technology, Elgar Publishing, 97.

[12] Storm, Darlene, 2015, ACLU: Orwellian Citizen Score, China's credit score system, is a warning for Americans, Computerworld, October 7, 2015, available at https://www.computerworld.com/article/2990203/security/aclu-orwellian-citizen-score-chinas-credit-score-system-is-a-warning-for-americans.html (accessed on November 25,

the reference point of informing such decisions?[13] Who would take those decisions? Such decisions could potentially be outsourced to autonomous intelligent software - integrated into health insurance systems (e.g.) and feeding from their patients' data, they could determine who receives a potential treatment and who does not. The emergence of autonomous intelligent agents can force us even more to evaluate our current economic, social and environmental systems and trends in order not to put society at risk of being kept in quantitative borders set by algorithms and based on utilitarian calculations.

Besides potential risks of autonomous technology during peacetime and for society as a whole, the case of autonomous weapons systems trigger at least three further arguments for a rethinking of our society and of how we understand our human nature.

If Code is Law, it can be changed. By whom? Code is 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, which is currently still generally neutral regarding the content of the data and ignorant about the user. This feature of codes makes regulating behaviour in the cyberspace difficult. However, code is not fixed, but the architecture of the cyberspace can be changed by the ones 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 those who set it up are facing. In other words, there exists a choice whether to influence the 'regulability' of the cyberspace as well as a choice on how 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?[14]

Human decision vs. machine calculation: The fact that weapons systems are referred to as 'autonomous' due to their capacity to continuously interact with their environment over time, to generate an output without human intervention, and to supplant the human from a process

---

2017); see also India's mandatory biometric ID system 'Aadhar': Pahwa, Nikhil, 2017, How not to screw up your national ID, Medianama, November 21, 2017, available at https://www.medianama.com/2017/11/223-how-not-to-screw-up-your-national-id-india-aadhaar/ (accessed on November 27, 2017).

[13] Helbing, Dirk, Nagler, Jan, and Van den Hoven, Jeroen, 2017, Ethics for Times of Crisis: How not to use autonomous systems in an unsustainable world, available at https://www.researchgate.net/publication/320740872_Ethics_for_Times_of_Crisis_How_not_to_use_autonomous_systems_in_an_unsustainable_world (accessed on November 25, 2017).

[14] Lessing, Lawrence, 2000, Code Is Law, On Liberty in Cyberspace, Harvard Magazine, January-February 2000, available at http://socialmachines.media.mit.edu/wp-content/uploads/sites/27/2015/03/Code-is-Law-Harvard-Magazine-Jan-Feb-2000.pdf (accessed on November 25, 2017); see also Van den Hoven, Jeroen, Vermaas, Home Pieter, and Van de Poel, Ibo (Eds.), 2015, Handbooks of Ethics, Values and Technological Design: Sources, Theory, Values and Application Domains, Springer.

where he previously has taken decisions, bears the risk of prematurely regarding them as 'human-like', with a capacity to 'learn', 'understand', or 'decide'. However, if we compare a computer software with a human being based on the reference point of 'deciding' (e.g.), we must clarify what this term means both for software/machines and for a human. A software analyses, calculates and creates an output that from the outside may look like its 'deciding to act in a certain way'. However, a software is usually named by its purpose, and not by its structure. It is crucial to not prematurely underestimate human language or overestimate computer programs. The fact that a software substitutes 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 – a software performs a calculation. Language frames the way we think, understand, and compare. Using the same language for machines as for humans could lead us to overlook the risk that we could potentially live in a future world of calculations rather than of decisions. Therefore, we must ask: Do we need a new language for machines?

New (artificial) species threatening humanity and its ecosystem: 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. In this light, some experts claim that we should pre-emptively decide not to create an invasive artificial species of autonomous agents that could endanger the lives of human beings on the planet.[15]

Regina Surber (reginasurber@ict4peace.org)
ICT4Peace Foundation
Zurich, 27 November 2017

[15] Helbing, Dirk, 2017, Open Discussion on Presentation on Lethal Autonomous Weapons Systems, November 13, 2017, ETH Zurich, Switzerland; see also Cellan-Jones, Rory, 2014, Stephen Hawkings warns artificial intelligence could end mankind, BBC Online, December 2, 2014, available at http://www.bbc.com/news/technology-30290540 (accessed on November 27, 2017).