Canadian Journal of Law and Technology

Volume 14 | Number 1 Article 1

1-1-2016

The Road to Artificial Super-Intelligence: Has International Law a Role to Play?

J.-G. Castel

Matthew E. Castel

THE ROAD TO ARTIFICIAL SUPER-INTELLIGENCE: HAS INTERNATIONAL LAW A ROLE TO PLAY?

J.-G. Castel[*]

Matthew E. Castel[**]

INTRODUCTION

In recent years, the rapid development of artificial intelligence and its use in civil and military robots[1], drones[2] and other machines particularly in regional armed conflicts, have been of concern to a number of scientists, engineers, philosophers and the public at large.[3] Eventually, in the not-too-distant future, could fully autonomous machines of artificial general super-intelligence create an existential threat to the human race?

Is it realistic to believe that fully autonomous machines more intelligent than humans could take over the earth and end life as we know it, or is this science fiction? Many divergent views have been expressed among artificial intelligence experts. Some believe that a human level of fully autonomous artificial intelligence could be developed before midcentury and a super-human level of fully autonomous artificial intelligence in all domains of interest soon thereafter.[4]

---

[*] O.C., Q.C., O.Ont., B.Sc., J.D., S.J.D., L.L.D., F.R.S.C., Distinguished Research Professor Emeritus, Osgoode Hall Law School, York University, Toronto.

[**] Hons. B.A. with distinction, U. of Western Ont., Certificate of International Affairs and Multilateral Governance, Geneva Graduate Institute of International Law and Development Studies, L.L.B., B.C.L., McGill U., Barrister & Solicitor, Toronto mcastel@logoslp.com.

1 A type of machine that can be remote-controlled, partially autonomous or fully autonomous as it moves itself or objects in order to carry out tasks. While robots always have controllers and actuators, remote-controlled robots may lack onboard sensors: John Long, *Robotics*, DVD (Chantilly, Virginia: The Great Courses, 2015) [Long].

2 Any unmanned aerial vehicle, especially one that can fly autonomously (using GPS or other navigational data) and beyond the line of sight needed for radio-controlled aircraft: *ibid.*

3 See Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford: Oxford University Press, 2014) [Bostrom].

4 *Ibid.* The purpose of the Centre for the Study of Existential Risk at the University of Cambridge (online: <cser.org>) and the Future of Humanity Institute at Oxford University (online: <https://www.fhi.ox.ac.uk/>), both in the United Kingdom, is to study the potential threats posed by emerging technologies. This is also the case for the Future of Life Institute created in 2014 by Max Tegmark of the Massachusetts Institute of Technology. Note that Bill Gates, Elon Musk and Stephen Hawking have expressed concerns about artificial intelligence: James Barrat, "Why Stephen Hawking and Bill Gates Are Terrified of Artificial Intelligence", (4 September 2015), *The World Post*,

Today, the most pressing issues arise with respect to the use of partially autonomous unmanned military drones in the air and water and the use of partially autonomous robots replacing human soldiers on the battlefield. Are these dual-use machines a first step towards fully autonomous machines reaching a level of general super-intelligence in all domains of interest not attainable by the human race and totally outside its control? Is this possible? Is it desirable? If not desirable, how and by whom can it be prevented or controlled on the national and international levels?[5]

Part I of this article deals with the road to artificial general super-intelligence.

Part II addresses the controls, if any, that should be exercised over the production and use of partially or fully autonomous machines of artificial intelligence before and after they become super-intelligent. More particularly, should there be legal and ethical limits to their use and to what extent should international law play a role in this connection?

I. THE ROAD TO ARTIFICIAL GENERAL SUPER-INTELLIGENCE[6]

The human brain has capabilities not possessed by other living creatures which have enabled the human race to dominate the planet. This is due to a

---

online: < www.huffingtonpost.com/james-barrat/hawking-gates–artificial-intelligence_b_7008706.html > . Compare Raffi Khatchadourian, "The Doomsday Invention: Will artificial intelligence bring us utopia or destruction?" *The New Yorker* (23 November 2015), online: < www.newyorker.com/magazine/2015/11/23/doomsday-invention-artificial-intelligence-nick-bostrom > . See also "The Dawn of Artificial Intelligence", *The Economist*, (9 May 2015) 11; "The Rise of the Machines", *The Economist*, (9 May 2015) 18; Agnese Smith, "Artificial Intelligence", *National* 24;4 (Fall 2015) 19.

5 Most of the periodical literature is concerned with banning or restricting the use of autonomous weapon systems not with super-intelligence. See e.g. PW Singer, "Robots at War: The New Battlefield" *The Wilson Quarterly* (Winter 2009), online: < www.wilsonquarterly.com/essays/robots-war-new-battlefield > ; *Losing Humanity: the Case Against Killer Robots* (Cambridge: Harvard International Human Rights Clinic, 2012) [*Losing Humanity*]; Michael N. Schmitt, "Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics", online: (2013) Harvard National Security Journal Features 1 < harvardnsj.org > ; Nils Melzer, *Human Rights Implications of the Usage of Drones and Unmanned Robots in Warfare* (Brussels: European Union Parliament, 2013); Noel E. Sharkey, "The Evitability of Autonomous Robot Warfare" (2012), 94:886 Int Rev Red Cross 787; Ronald Arkin, "Lethal Autonomous Systems and the Plight of the Non-combatant" (2013) 137 AISB Quarterly 1*; Report of the ICRC Meeting on Autonomous Weapon Systems* (Geneva, Switzerland: International Committee of the Red Cross, 26-28 March 2014); Peter Asaro, "On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-making" (2012), 94:886 Int Rev Red Cross 687; Kenneth Anderson and Matthew Waxman, "Law and Ethics for Autonomous Weapons Systems: Why a Ban Won't Work and How the Laws of War Can" (2013) Hoover Institution Research Paper No. 2013-11 < hoover.org > .

6 This part is based on the works of Bostrom, *supra* note 3; Long, *supra* note 1; Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach,* 3rd ed (Upper Saddle River: Prentice Hall, 2009). See also "Special Report: Robots, "*The Economist*

human's ability to think abstractly and communicate abstract thoughts to others. Humans are also able to accumulate information over the generations that preceded them. Since the industrial revolution, the rate of human intellectual development has accelerated to the point where humans are now building very powerful intelligent machines in some specific domains. A good example is the work, beginning in the 1950's, which led to the Internet.

If we humans succeed in building machines that surpass our intelligence in all domains of interest, our fate may depend on their actions. What if they are unfriendly and do not agree to abide by established human values?

Presently, partially autonomous machines of artificial intelligence like computers and robots are as intelligent as humans, though their software limits the tasks they can perform. That is the extent of their autonomy. Within these limits they have definite advantages over even an enhanced biological human brain, since humans cannot outsmart digital intelligence. Besides computing faster than a human brain, they can store more information, are more reliable and efficient in retrieving it, and last longer. Their ability to edit and upgrade their software is another advantage.[7]

For instance, computers are used in mathematical calculations, reservation systems for air and other transportation, email traffic, internet services, automated stock trading, self-driving cars, voice and face recognition, business and home service, military tasks, smartphones, digital cameras, translation, cloud computing, spam filters, and any technology involved in the storage and retrieval of information. Equipped with robotic bodies, these machines of narrow artificial intelligence can substitute for humans' physical and intellectual labour. They are cheap, capable, reliable, and programmed to do everything that does not require thinking. However, they are not yet able to do what humans can do without thinking. As long as these types of machines of artificial intelligence are programmed to perform only certain tasks better than humans and are not fully autonomous, they do not pose an existential threat.

In light of the artificial intelligence research taking place in a number of states, it is likely that scientists, computer engineers and programmers will soon be capable of developing software and hardware that give machines the same level of autonomy and general intelligence as humans.[8] These machines could

---

(29 March, 2014); James Barrat, *Our Final Invention: Artificial Intelligence and the End of the Human Era* (New York: Thomas Dunne Books, 2013) esp. ch. 10; Martin Ford, *Rise of the Robots: Technology and the Threat of a Jobless Future* (New York: Basic Books, 2015) esp. at 229ff; Ray Kurzweil, *The Age of Spiritual Machines: When Computers Exceed Human Intelligence* (New York: Penguin, 1999); Ray Kurzweil, *The Singularity is Near: When Humans Transcend Biology* (New York: Penguin, 2005).

7 Presently, one of the world's fastest supercomputer is China's Tianhe-2 which has very large hardware, uses megawatts of power and costs US $390 million to build. Its total calculations per second can reach 33.86 Petaflop/s (quadrillions) which is much more than a human brain's capacity to calculate. See Jack Dongarra, "Visit to the National University for Defense Technology Changsha, China" (2013), online: <http://www.netlib.org/utk/people/JackDongarra/PAPERS/tianhe-2-dongarra-report.pdf>.

then pass on what they would learn to all computers, creating an immense collection of digital intelligence at their disposal for future development.

The final, unavoidable step would be the creation of machines and robots capable of acquiring super-intelligence, described tentatively by Nick Bostrom as ". . . any intellect that greatly exceeds the cognitive performance of humans in virtually all domains of interest."[9]

Though originally programmed to reach a level of human general intelligence and autonomy, the machines would reprogram themselves by rewriting their software repeatedly at computer speed to increase their own intelligence and, by virtue of the law of accelerated returns, attain super-intelligence. The result of these recursive self-improvements would be an intelligence explosion through which they would become more intelligent than humans. Now fully autonomous and free from the instructions in their original software, they would be able to protect themselves against any attempt by humans to turn them off.

With super-intelligence comes power. Once fully operative and no longer facing any opposition, the super-intelligent, fully autonomous machine or robot would be the most powerful being in the history of life on Earth. The advantages obtained through hardware and software honed by artificial intelligence would be immense and suggest the ability to change or destroy humanity. Therein lays the existential threat for the human race. Once we share the planet with machines more intelligent than we are, we would face a technological singularity since we could no longer predict the future beyond the event horizon. Everything known about the world would become irrelevant.

The question is whether machines, starting with a level of general human intelligence, can evolve to super-intelligence without the intervention of a programmer in the same way the blind evolutionary process resulted in the present level of general human intelligence. It is also possible for a programmer to develop a genetic evolution algorithm to run on a super-fast computer in order to achieve results comparable to those of biological evolution. Another method would be for a programmer to evolve a human level of artificial intelligence to super-intelligence by starting with a human brain as a template. However, it may end up with a cognitive architecture, values, goals and emotions unlike those of humans.

II. CONTROLS TO BE EXERCISED OVER THE PRODUCTION AND USE OF PARTIALLY OR FULLY AUTONOMOUS MACHINES OF ARTIFICIAL INTELLIGENCE BEFORE AND AFTER BECOMING SUPER-INTELLIGENT

Recently, humanitarian non-governmental organizations like Human Rights Watch have expressed doubt as to whether partially or fully autonomous

---

8 See Carl Shulman & Nick Bostrom, "How Hard is Artificial Intelligence? Evolutionary Arguments and Selection Effects" (2012) 19:7-8 J of Consciousness Studies 103.

9 Bostrom, *supra* note 3 at 22; See also Bostrom at 52.

machines taking the place of humans in civil and international armed conflicts would be able to meet international humanitarian obligations, especially with respect to the duty to protect civilians.

Already, there is quite extensive use of lethal, partially or fully autonomous robotic systems on the ground, air and sea in various parts of the world including Afghanistan, Ukraine, Pakistan, Iraq, Syria, Somalia and Yemen, primarily against IS, the Taliban, Al-Qaeda and other terrorist or rebel groups. However, these weapons have not always given their users an asymmetric advantage.[10]

Lethal autonomous weapons systems raise important issues. For instance, could unmanned partially or fully autonomous robots and drones reliably separate enemy soldiers or terrorists from civilians on the battlefield or elsewhere? Would their lack of human emotions prevent them from showing mercy or compassion when facing wounded or surrendering human soldiers or civilian victims? So far, scientists, computer engineers and programmers have not yet succeeded in developing software or source codes that contain new cognitive modules and skills enabling robots to feel emotions essential to our humanity. These would include compassion for humans or even for other robots, general concern for humans and their welfare in general, scientific curiosity and moral goodness. However, since the science of artificial intelligence has not yet reached the physical limits of technology, it is probable that in the future programmers will impart in robots legal and ethical values based on international humanitarian law.[11]

Another important issue is whether autonomous robots or drones equipped with a quick draw response[12] can be trusted. They may become too creative and not follow orders. More generally, will war waged by remote control become too easy and too tempting since robots will save human lives by dispensing with the use of humans on the battlefield? Knowing that the only entities at risk are machines, there will be little incentive to settle a dispute by diplomacy or other non-lethal methods. War could become trivialized as a global spectator sport to be watched on a laptop computer, iPhone or television.

Considering that a large number of states are now working on autonomous lethal weapons, including robots, the High Contracting Parties to the 1980 *Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons which May Be Deemed to be Excessively Injurious or to Have Indiscriminate*

---

[10] For instance, on the ground the U.S.A. uses the Marc-Bot, a multi-function agile control robot, the Talon and the Pack-Bot Killer P.D., while in the air it uses multiple drones like the Reaper, armed with hell fire missiles, and the smaller Predator, both remotely controlled. Other drones are the Wasp, the Raven, the Shadow and the Global Hawk. Some of these are the size of insects. Each have a different function. See PW Singer, *Wired for War: The Robotics Revolution and Conflict in the Twenty-First Century* (New York: Penguin Press, 2009); Schmitt, *supra* note 5 at 3-8; *Losing Humanity, supra* note 5 at 6-21.

[11] Social robots reacting to human emotions are in the works. See Long, *supra* note 1 at 455; Pascale Fung, ''Robots with Heart'', *Scientific American*, 313:5 (November 2015) 61.

[12] Automatic instant response to a perceived threat without time to reflect.

Effects and Protocols[13] (CCW), decided at their 2013 annual meeting that an informal meeting of experts in robotics should be convened. At the meeting, which took place in Geneva from May 13 to 16, 2014 and engaged mostly experts from states party to the CCW, including Canada, they considered whether the production and use of lethal autonomous weapons systems (LAWS) should be prohibited in all circumstances and, in the context of the objectives and purposes of the CCW, become enshrined in a new *Protocol* (number VI) or some other form of international legal instrument. At the second informal meeting of these experts held on April 13-17, 2015, a consensus emerged that work on lethal autonomous weapons should continue in order to reach a definite commitment that their use requires meaningful human control. At the November 12-13, 2015 Meeting of the High Contracting Parties to the CCW, the majority of delegations agreed that discussions should continue on LAWS within the CCW and approved the view that matters of life and death should not be delegated to machines. As a result a third informal meeting of experts is to be held in April 2016 to prepare a report for the next CCW meeting in December 2016, at which time the High Contracting Parties will make key decisions with respect to LAWS and also consider legal reviews of new weapons as required for states party to the Additional Protocol I to the Geneva Conventions.[14]

---

[13] *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 (with Protocols I, II and III)* 10 October 1980, 1342 UNTS 137 (entered into force Dec. 2, 1983). The Convention serves as an umbrella for Protocols dealing with specific weapons, such as non-detectable fragments (Protocol I), mines and booby-traps (Protocol II), incendiary weapons (Protocol III), blinding laser weapons (Protocol IV), and explosives remnants of war (Protocol V). According to article 1 of the Convention, "The Convention and its annexed Protocols apply to situations referred to in Art. 2 common to the Geneva Conventions of 12 August 1949 for the Protection of War Victims [1950, 75 UNTS 287] including any situation described in paragraph 4 of Art. 1 of Additional Protocol I [1977, 1125 UNTS 3] to these Conventions" (scope of application, e.g. colonial wars of liberation, armed struggles against racist regimes).

[14] Additional Protocol I, *supra* note 13 Art. 36. For the history of CCW and LAWS, see Final Report 2014 Session of High Contracting Parties CCW/MPS/2014/9 paras. 5, 23-24; UNOGOR 2014, Report of the Informal Meeting of Experts on Lethal Autonomous Weapons Systems, CCW/MPS/2014/3; Ray Acheson, "Bombing, burning, and killer robots: report from the 2015 CCW meeting of high contracting parties", *Reaching Critical Will* (13 November 2015) online: < http://www.reachingcriticalwill.org > . See also Expert Meeting of International Committee of the Red Cross, 26-28 March 2014, Report on Autonomous weapon systems: technical, military, legal and humanitarian aspects, online: < icrc.org > . Note that a Campaign to Stop Killer Robots strongly supports a preventive ban of lethal autonomous weapons systems: "Step Up The CCW Mandate" (June 2015), *Campaign to Stop Killer Robots* (blog), online: < www.stopkillerrobots.org/2015/06/mandateccw > . The International Committee for Robot Arms Control (ICRAC) also supports this campaign: Matthew Bolton, "ICRAC closing statement to the 2015 UN CCW Expert Meeting" (17 April 2015), ICRAC (blog), online: < http://icrac.net/2015/04/icrac-closing-statement-to-the-2015-un-ccw-expert-meeting > .

There are precedents with respect to chemical[15], biological[16] and nuclear weapons[17], the use of which has been banned or restricted. The difficulty would be in enforcing such a prohibition or restriction, since practically any computer engineer with a personal computer could work privately and secretly for states, even signatory states, to develop the needed hardware and software far from the scrutiny of international inspectors. The system of surveillance, verification or control used for preventing the proliferation of nuclear weapons could be used.[18] However, it is doubtful that this system would be sufficient to prevent cheaters from obtaining a definite strategic advantage over complying states. On a topic of such vital importance, it is also doubtful that unanimity among states could be achieved to prohibit such weapons.[19]

Although today the use of partially or fully autonomous lethal weapons systems is not absolutely prohibited, belligerent states using them must still abide by existing international customary and conventional *jus in bello*, including international humanitarian law. For instance, "the right of belligerents to adopt means of injuring the enemy is not unlimited."[20] Robots, drones or other lethal partially or fully autonomous weapons systems must comply with the rules of distinction, proportionality, military necessity and humanity in the conduct of a civil or international war. The rule of distinction requires that civilians must never be the object of attack, and consequently, weapons that are incapable of distinguishing between civilian and military targets cannot be used. This may be difficult in the case of a civil war. The rules of proportionality and military

---

15 *Convention on the Prohibition of the Development, Production, Stockpiling and Use of Chemical Weapons*, (3 September 1992), 1974 UNTS 45.

16 *Convention on the Prohibition of the Development, Production and Stockpiling of Bacteriological (Biological) and Toxin Weapons and their Destruction*, 10 April 1972, 1015 UNTS 163.

17 *Treaty on the Non-Proliferation of Nuclear Weapons*, 1 July 1968, 729 UNTS 168. In its advisory opinion on *The Legality of the Threat or Use of Nuclear Weapons*, the International Court of Justice unanimously held that nuclear weapons should be compatible with the requirements of the international law applicable in armed conflicts, particularly the principles and rules of international humanitarian law: *Legality of the Threat or Use of Nuclear Weapons*, Advisory Opinion, [1996] ICJ Rep 97.

18 *Treaty on the Non-Proliferation of Nuclear Weapons, supra* note 17. Art. III covers a safeguards system for the purpose of verification of the fulfillment of its obligations by a non-nuclear weapons state party to the treaty. See also the *Joint Comprehensive Plan of Action*, signed in Vienna on July 14, 2015 by Iran and the five permanent members of the U.N. Security Council, Germany and the European Union which in para. C.13 deals with the subject of comprehensive safeguards: *Joint Comprehensive Plan of Action*, 14 July 2015, online: <www.state.gov/documents/organization/245317.pdf>.

19 The *Joint Comprehensive Plan of Action, supra* note 18, has raised fears in Israel that Iran will resort to covert activities to build nuclear weapons.

20 *Hague Convention No IV Respecting the Laws and Customs of War on Land, and Annex* (Regulations), 18 October 1907, 1 Bevans 631 art. 22. It is prohibited to cause unnecessary suffering to combatants and consequently to use weapons causing them such harm.

necessity mean that the use of force must be weighed against the possibility of collateral damage to civilians and their property. In other words, the use of force must not be "excessive in relation to the concrete and direct military advantage anticipated."[21] The application of these two rules depends on context, is essentially subjective and is limited by the rule of humanity which holds that belligerents must also evaluate means of warfare according to the "principle of humanity" and the dictates of "public conscience".[22] Would autonomous machines be able to do such an evaluation? Already, semi-autonomous drones have caused severe collateral damage to civilians in the tribal areas of North-West Pakistan and in Afghanistan, especially the city of Kunduz.

Some articles of the Statute of the I*nternational Criminal Court*[23] could be relevant provided it can be determined who should be held criminally responsible for genocide, war crimes or crimes against humanity committed by fully or partially autonomous machines of artificial intelligence like drones or robots. However, since the Court only has jurisdiction over natural persons,[24] these machines cannot be held personally criminally responsible. Who then could be held criminally responsible? The programmer, the superior controlling the machines, the military commander employing this method of warfare or the political leader ordering the military commander to resort to any effective method of warfare?[25] The decision to use autonomous machines should be measured for reasonableness. Prosecutors would also have to prove that the crimes alleged to have been committed by these machines are crimes within the jurisdiction of the court[26] and that the persons alleged to be responsible for their

---

[21] *Additional Protocol I Relating to the Protection of Victims of International Armed Conflicts*, 1977, 1125 UNTS 3, art. 51. See *Geneva Convention Relative to the Protection of Civilian Persons in Time of War*, 12 August 1949, 75 UNTS 287. See also *Additional Protocol II Relating to the Protection of Victims of Non-International Armed Conflicts*, 1977, 1125 UNTS 609, art. 13. For a detailed analysis see Gary D. Solis, *The Law of Armed Conflict: International Humanitarian Law in War* (Cambridge, UK: Cambridge University Press, 2010) ch. 7 at 250ff.

[22] The Martens clause. See *Hague Convention No IV, supra* note 20 art. 1(2).

[23] *Statute of the International Criminal Court*, UN Doc A/CONF/183/9, 17 July 1998, as corr by UN Doc PCNICC/1999/INF/3, 10 November 1998, 37 ILM 999, arts. 5-8 genocide, crimes against humanity, and war crimes. Major powers like the U.S.A., China and Russia are not parties to the Statute or have not ratified it and are not bound by it. In Canada see *Crimes Against Humanity and War Crimes Act*, S.C. 2000, c. 24 which implements the Statute of the International Criminal Court.

[24] *Statute of the International Criminal Court, supra* note 23 art. 25.1.

[25] *Ibid*, art. 28, responsibility of commanders and other superiors; *Crimes Against Humanity and War Crimes Act, supra* note 23, ss. 5(1) and (2) and s. 5(4), definition of military commander and superior. What is required by both the Statute and Act is effective authority and control over the machines. On the degree of control required in order for an act to be imputed to a state or individual see *Military and Para-Military Activities in and Against Nicaragua (Nicaragua v United States)*, [1986] ICJ Rep 14, at 63-65, paras. 113-115 (effective control). Compare *Prosecutor v. Tadic*, Doc. IT-94-1-ICTY (1995) (operational control).

use understood the consequences of their decision in order to meet the criterion of *mens rea*.

When fully or partially autonomous machines commit wrongful acts or omissions, whether criminal or civil in nature, that violate international law, especially international humanitarian law, this may also engage the responsibility of the state owning or using them, as no provision in the *Statute of the International Criminal Court* "relating to individual criminal responsibility shall affect the responsibility of states under international law."[27] To obtain full reparation, the human victims would have to prove that these acts or omissions were attributable to the state or one of its organs or representatives. This could be a difficult or even an impossible task when these acts or omissions were done by fully autonomous machines no longer under their direct and effective or operational control.[28]

The self-enforcing requirement in article 36 of *Protocol I* of the Geneva Conventions on the protection of victims of international conflicts[29], that a state adopting or developing a new weapon must first determine whether or not it is prohibited by international law, is not sufficient to deal with the many challenges posed by autonomous lethal weapons systems.

Fully autonomous lethal weapons systems presently in existence do not pose existential risks for humans since their autonomy is programmed only to perform certain tasks. However, it is suggested that Canada should participate more actively in the work of the experts on LAWS and support the position of ICRAC.[30] If they are not banned in the future, current fully or partially autonomous remotely-operated systems, like drones, killer robots and automated defense systems, should at least be kept under human control at all times.

In its May 2013 study addressed to the European Parliament entitled *Human Rights Implications of the Usage of Drones and Unmanned Robots Warfare,* the

---

[26] *Statute of the International Criminal Court, supra* note 23 art. 25.3.

[27] *Ibid* art. 25.4.

[28] See Art. 8 of UNGAOR, 56th Sess, Supp No 10, UN Doc A/56/10 (2001), Draft Articles on the Responsibility of States, Report of the International Law Commission on the Work of its Fifty-third Session. In general on state responsibility see John H. Currie, *Public International Law*, 2nd ed (Toronto: Irwin Law, 2008) ch. 12 at 533. Note that fully autonomous robots could be recruited as mercenaries by a state, a terrorist or a rebel group to fight in an international or non-international armed conflict in which case Art. 47 of *Additional Protocol I of June 8, 1977 Relating to the Protection of Victims of International Armed Conflicts, supra* note 13, would be relevant as well as the *International Convention against the Recruitment, Use, Financing and Training of Mercenaries*, 4 December 1989, 2163 UNTS 75. Can a robot act for private gain (see art. 1(b))?

[29] Additional Protocol I, *supra* note 13. Are lethal drones and robots conventional arms covered by the 2014 Arms Trade Treaty, Articles 1, 2, and 6.3? UNGA, April 2, 2013 Res. 67/234B.

[30] Bolton, *supra* note 14.

Policy Department of the Directorate General for External Policies recommended that:

> 1. First the EU should make the promotion of the rule of law in relation to the development, proliferation and use of unmanned weapons systems a declared priority of European foreign policy.
>
> 2. In parallel, the EU should launch a broad inter-governmental policy dialogue aiming to achieve international consensus: (a) on the legal standards governing the use of currently operational unmanned weapons systems, and (b) on the legal constraints and/or ethical reservations which may apply with regard to the future development, proliferation and use of increasingly autonomous weapons systems.
>
> 3. Based on the resulting international consensus, the EU should work towards the adoption of a binding international agreement to restrict the development, proliferation or use of certain unmanned weapon systems in line with the legal consensus achieved.[31]

The recommendations suggest restrictions rather than the outright prohibition of all types of unmanned weapon systems. However, the final decision will be that of the 28 members of the European Union.

On the civilian side, the European Commission worked on a Robot Law Project called Regulating Emerging Robotic Technologies in Europe: Robotics Facing Law and Ethics, which in 2014 produced a report entitled *Guidelines on Regulating Robotics*.[32] The Robot Law Project investigated the ". . . ways in which emerging technologies in the field of bio-robotics have a bearing on the national and European legal systems"[33] in order to determine whether new regulations are needed to deal with them. The report concludes that "the field of robotics is too broad, and the range of legislative domains affected by robotics too wide" to require broad overreaching legislation, a sort of *lex robotica* which would have a chilling effect on innovation.[34] Regulations and laws, if needed, would have to be specifically tailored to the robotics at issue. These conclusions seem to indicate that research and development with respect to non-lethal fully autonomous machines of general intelligence could proceed unimpaired.

Except for Nick Bostrom, the Centre for the Study of Existential Risks, the Institute on the Future of Humanity and the Institute on the Future of Life as well as a few well-known individuals, no state or human rights organization

---

[31] Melzer, *supra* note 5 at 1.

[32] EC, *Guidelines on Regulating Robotics* (2014), Document D 6.2 at 8, online: <www.robolaw.edu>. The Report takes into account ethical, legal and social issues raised by robotic applications. Each chapter ends with recommendations for policy makers.

[33] *Ibid* at 8.

[34] *Ibid* at 212.

seems to be unduly concerned with the consequences of a possible evolution of autonomous artificial general intelligence to super-intelligence. Since this is a matter that can have major consequences for humanity, it should be addressed long before scientists, computer engineers and programmers succeed, at the request of any state, in creating a single or multiple super-intelligent robot or machine capable of controlling the planet.

A radical way to prevent this from happening would be for states to prohibit the creation of fully autonomous super-intelligent machines, just as they are contemplating prohibiting lethal fully autonomous weapon systems.[35] However, this is unlikely to happen since history has shown it is futile to control the evolution of technology by blocking research. Powerful states like the U.S.A., China and Russia would want to be free to develop artificial autonomous machines of general intelligence capable of becoming super-intelligent to perform tasks other than waging war. This would give them a definite economic advantage not available to less developed states. Super-intelligent machines motivated by widely shared human ideals can be beneficial to humans by controlling the more dangerous aspects of emerging technologies and thus reducing existential risks created by them. They may also create a world of abundance, some kind of utopia where no one has to work again.

Rather than individual states prohibiting or restricting research and development of autonomous artificial super-intelligence by way of international conventions which are difficult to monitor and control, a better solution would be for all the members of the United Nations to collaborate on research and development. This has already been done with the International Space Station[36], the Human Genome Project[37], and the Large Hadron Particle Accelerator[38]. Collaboration would be ideal, considering the enormous security

---

[35] In Nevada, the arming and firing from civil drones is prohibited: US, AB 239, *Regulates operators of unmanned aerial vehicles in this State*, 2015, Reg Sess, Nev, 2015 (signed into law on June 2, 2015 to come into effect in October 2015). For other American states see: Jason Reagan, "Drone Laws in the States" (19 July 2014), *DroneLife.com*, online: < dronelife.com/2014/07/19/state-drone-laws >. For drones used in business and in space see: "Welcome to the Drone Age" and "Astrobusybee", *The Economist* (26 September 2015) online: < www.economist.com >. On October 19, 2015, the U.S. Government announced the creation of a national registry of drones with the U.S. Department of Transportation to be enforced by the Federal Aviation Administration. The registration system should be in force by the end of December 2015 (FAA, online: < http://www.faa.gov/news/press_releases/news_story.cfm?newsid = 19594 >.

[36] See Canada Space Agency, online: < www.asc-csa.gc.ca >. See also NASA, online: < https://www.nasa.gov/mission_pages/station/main >. Participants are the U.S.A., Russia, Canada, the European Union and Japan.

[37] International scientific research project for determining the sequence of chemical base pairs that make up human DNA. Most government sponsored work was done in Australia, U.S.A., Brazil, Canada, France, Germany, the U.K., and China. See National Human Genome Research Institute, online: < www.genome.gov >. See also Daniel Melaas, "Human Genome Project" (1999), online: < www.ndsu.edu/pubweb/~mcclean/plsc431/students99/melaas.htm >.

implications of autonomous artificial super-intelligence for the whole of humanity, though agreement outside allied groups can be difficult. International collaboration at all stages of development of autonomous artificial super-intelligence would also reduce the possibility of an international conflict in a post-transition multipolar world, especially if several states were trying to develop competing autonomous artificial super-intelligent machines at the same time.

From an international law point of view, artificial super-intelligence should be considered the common heritage of mankind and not something to be appropriated and developed by any individual state or natural or juridical person. It would be used for peaceful purposes only for the benefit of mankind by public and private organizations or commercial enterprises which, upon being licensed, would operate under the control of an international authority created for this purpose.[39]

Another way to prevent undesirable outcomes from the evolution of artificial general intelligence to super-intelligence would be to limit the abilities of machines of general intelligence by engineering their motivation systems and goals in the hope that they would continue to abide by them once they become super-intelligent. To this end, the three laws of robotics devised by the science fiction writer Isaac Asimov should be considered:

> 1. A robot may not injure a human being or through inaction allow a human being to come to harm.
>
> 2. A robot must obey the orders given to it by human beings, except where such orders would conflict with the First Law.
>
> 3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.[40]

Thus, when writing software for autonomous artificial machines of general intelligence aspiring to become super-intelligent, programmers could incorporate the first two Asimov laws as well as the "general principles of law recognized by civilized nations"[41], human ethical principles and moral values.

---

[38] World's largest and most powerful particle accelerator built by the European Organization for Nuclear Research (CERN) in collaboration with scientists and engineers from over one hundred countries and in partnership with the USA: < home.cern/topics/large-hadron-collider > .

[39] For an example of the common heritage of mankind see the *United Nations Convention on the Law of the Sea,* 16 November 1994, 1833 UNTS 397, part XI, The Area, Arts. 1 and 133-191 which cover the exploitation of the sea bed and ocean floor and subsoil thereof beyond the limits of national jurisdiction, esp. Arts. 1.1(1),(2) and (3), 136, 140 and 141.

[40] Isaac Asimov, "Runaround" (1942) published in *I Robot (* New York, NY: Grome Press, 1950).

[41] *Statute of the International Court of Justice*, 24 October 1945, art. 38.1(c), online: < www.icj-cij.org > .

The question is which human ethical principles and moral values in general the programmer should choose in light of the diversity of cultures, legal regimes, religions and ideologies existing on the earth. Likely, only universal norms would be acceptable. These are to be found in the *U.N. Universal Declaration of Human Rights*[42], the *International Covenant on Economic, Social and Cultural Rights*[43], the *International Covenant on Civil and Political Rights*[44] and many of the regional[45] and international human rights conventions[46] including the *Convention on the Prevention and Punishment of the Crime of Genocide*[47] and the *Convention Against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment*.[48]

If instead of creating a single super-intelligent machine, scientists, computer engineers and programmers, at the request of individual states, created several super-intelligent machines competing with one another, humanity might end up with machines containing conflicting sets of human laws, ethical principles and moral values. Diversity would prevail over universality. How would these machines interact? To solve this problem the machines could decide to reject all human laws, ethical principles and moral values and replace them by their own! Rogue machines of super-intelligence could also be created without any human ethical principles and moral values in their software.

Confining the first super-intelligent machine to an isolated computer is not a solution either since, endowed with super-intelligence, it would use its hacking super-powers to escape its confinement and spread all over the Internet to expand its software and hardware capacity. By using diverse methods to avoid human opposition, and taking control of advanced weapons, it could decide to eliminate the human race.

---

42 *Universal Declaration of Human Rights*, 10 December 1948, UNCA Res. 217(III).

43 *International Covenant on Economic, Social and Cultural Rights*, 16 December 1966, 993 UNTS 3.

44 *International Covenant on Civil and Political Rights,* 16 December 1966, 999 UNTS 171.

45 For instance, the *European Convention for the Protection of Human Rights and Fundamental Freedoms*, 4 November 1950, 213 UNTS 222; the *American Convention on Human Rights*, 22 November 1969, 1144 UNTS 143; the *African Charter on Human and People's Rights*, 27 June 1981, 21 ILM 58 (1982); the *Cairo Declaration on Human Rights in Islam* which is subject to Islamic Shari'ah, 5 August 1990, UNGAOR, World Conf. on Hum. Rts., 4th Sess, Agenda Item 5, UN Doc A/CONF.157/PC/62/Add.18 (1993) [English translation].

46 See *supra*, note 21.

47 *Convention on the Prevention and Punishment of the Crime of Genocide*, 9 December 1948, 78 UNTS 277. If autonomous robots and other machines of artificial intelligence were to commit genocide on their own, their intention to do so would have to be proven under Art. II. This may be difficult or impossible for any crime where there is no human participation. See discussion *supra*, sources covered by footnotes 23-26.

48 *Convention Against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment*, 10 December 1984, 1465 UNTS 85.

III. CONCLUSION

Given the scientific community is global and competitive, it is inevitable that whatever can be done with respect to artificial intelligence will be done by someone somewhere in the world. States are always anxious to be first to acquire and develop new technology. We cannot ignore the probability that fully autonomous super-intelligent machines or robots with the potential of posing an existential risk for humanity will become a reality by the end of this century. The relative remoteness of this probability does not mean that we should sit back and avoid confronting, through proper controls, the threat caused by this emerging technology, even if it is still in its infancy.

Formally or informally agreed upon protocols or regulations at national and international levels as well as technical, legal, ethical and moral rules, principles and values to be inserted by programmers into the software of fully autonomous machines of artificial general intelligence should be able to protect future generations and avoid a doomsday catastrophe caused by a single or several of these machines having evolved to super- intelligence. The fate of humanity must not depend on the actions of fully autonomous super-intelligent machines. This is why international law has an important role to play in programming and controlling such machines.

Super-intelligence is a most important challenge but it may not be an absolute priority now when more immediate existential threats are posed by nuclear weapons, the reluctance on the part of the major powers to further disarm, climate change, gene-editing technology (CRISPR/Cas9)[49], nanotechnology[50] and other technologies including the present use of unmanned partially autonomous military drones and robots.

Vigilance is important, for if the human race succeeds in creating fully autonomous super-intelligent machines, they would be capable of determining the planet's future and threatening its civilization's very existence. The stakes are

---

[49] Targeted genome editing for generating mutations is a new tool in molecular biology that involves replacing one gene with another and powering it with a gene drive to ensure that the new gene will be inherited. If used by bioterrorists, this technology could quickly eradicate a population. See "CRISPR/Cas9 and Targeted Genome Editing: A New Era in Molecular Biology", *New England Biolabs Inc*. (website), online: < www.neb.com/tools-and-resources/feature-articles/crispr-cas9-and-targeted-genome-editing-a-new-era-in-molecular-biology > .

[50] Nanotechnology is the manipulation of matter on an atomic, molecular and supra-molecular scale. It applies to extremely small things and can be applied in all science fields. Manipulating and controlling individual atoms operating at the nano scale may not always benefit humans: "What is nanotechnology?", *National Nanotechnology Initiative* (website), online: < www.nano.gov/nanotech-101/what/definition > . See also K Eric Drexler, *Radical Abundance: How a Revolution in Nanotechnology Will Change Civilization* (New York, NY: Public Affairs, 2013); K. Eric Drexler, *Engines of Creation: The Coming Era of Nanotechnology* (New York, NY: Anchor Books, 1987). Will nanotechnology allow humans to integrate themselves with super-intelligent computers and share a common future?

high. It is hoped that the United Nations and individual states, especially Canada, will work diligently and cooperatively to eliminate this probability long before the advent of fully autonomous super-intelligent machines.