THE I.A., THE NEW ARMS RACE AND THE ROLE OF SCIENTISTS

María Vanina Martínez and Ricardo Oscar Rodríguez
Computer Department. FCEyN-UBA
Institute of Computer Science. UBA-CONICET.

Summary

The development of new intelligent technologies for military use inaugurates a new stage in the arms race that can mean a sophisticated loss of liberties for global citizenship. Governments, international organizations, civil society and the scientific community face the challenge of promoting legal, ethical and moral parameters for the development of I.A for military purposes. This article explores concepts and definitions about the development of autonomous weapons and reflects on the importance of establishing legal and cultural bodies that guide the development of the A.I.

There is no doubt that the development of Artificial Intelligence (A.I.) will have a strong cultural and social impact in the daily life of humanity. In fact, it already has and there are many AI-based applications that facilitate or improve people's lives. There are examples as simple as anti-spam systems or digital assistants who speak to us and advise with total naturalness. Others are somewhat more sophisticated, such as the system of remote monitoring that allows to predict forest fires, or diagnostic systems of cancer or prediction of blindness, or semiautomatic assistants of large landings airports and even cars without driver. And they are all insignificant examples in front of the great auguries that are predicted for this technology. But, as with any other disruptive technology, not everything is profit and prosperity. Its improper and unethical use it also exists. There are systems that predict the preferences of different social groups and try to influence / guide / manipulate their opinions and actions. A recent case has been the scandal of the massive "leakage" of Facebook data in favor of Cambridge Analytica, and the use of these to define campaigns to attract votes for the Brexit referendum or Trump's candidacy in the 2016 presidential election in USA. But there are many more subtle and veiled forms of opinion control to which the society is exposed daily, as are the systems of recommendations guided by A.I.

Our friendships in social networks, our consumption or Internet searches, determine a profile that is recognized by these systems and we become vulnerable. However, it is important to make clear that beyond the objective of manipulation with which these technologies can be used, those same profiles can serve to detect pedophiles, depressants with suicidal intentions, identify facts of discrimination or bullying in social networks, etc.

But if the subtle methods of manipulation / persuasion / social domination mentioned previously, weren't enough for corporate voracity, it could always resort to the ancestral method of violence. And again, here the A.I. will also have a strong impact through the development of intelligent military technology. In fact, the armed forces of the developed countries have started a new career armament in order to maintain its military might. For the experts, we would be entering the third era of armament technology after the irruptions of the firearms and nuclear bombs. For ordinary mortals we would be entering a sophisticated stage of loss of freedoms.

To visualize this vulnerability, it is enough to consider the destructive potential of the use of A.I. techniques, such as advances in the area of robotics, in the development armament. Even the slightest possibility of endowing the simplest weapon with very basic autonomy skills such as mobility, perception and understanding of the environment (such as those that allow today to implement techniques based on automatic learning and neural networks), quickly arouses our concern in relationship to how and for what purpose they can be used. If we extend the analysis with the possibility of incorporating autonomous reasoning and decision-making capabilities, the scenario becomes even more complex and worrisome. Moreover, imagine one of the modern cars without a driver loaded with explosives and it will become clear what we are talking about.

For all that, making society aware of the vulnerability to which we are exposed with the new technologies of A.I., is a responsibility that we have, especially, the scientists who develop and perfect these techniques. Essentially to not repeat the mistake made with the use of atomic energy. Precisely for that reason, the A.I. scientific community has been struggling for several years for the non-development of lethal autonomous weapons. During the IJCAI2015 conference in Buenos Aires, thousands of researchers at A.I. from the international community appealed to the no proliferation of lethal autonomous weapons through an open letter that had repercussions worldwide (see https://futureoflife.org/open-letter-autonomous-weapons/). In line with the staging of the problem, during the development of the IJCAI2017 it was made public an open letter addressed to the United Nations signed by the presidents of 116 companies leaders in the use of A.I. (from 26 countries) warning again about the danger of weapons with A.I. and calling its ban. An interesting effect of these conscientization campaigns has been the refusal of Google employees to participate in the *Marven Project*, generated from a contract between the company and the Pentagon for the development of "killer drones". In fact, more than three thousand employees of Google requested not only that the project be abandoned, but also that a clear policy that establishes that neither the company nor its contractors will ever participate in the development of war technology.

In that same direction, Stuart J. Russell, professor at the University of California at Berkeley, and the Future of Life Institute, created a video that shows the destructive ability that these type of drones have (see https://www.youtube.com/watch?v=9Pn17-Mr7wc&t=17s). This video was shown during an event in the Assembly of the United Nations Convention on Conventional Weapons to call for action that prohibits the development of this type of weaponry.

But proposing guidelines and enacting legislation on the use of A.I. for the weapons construction is not the only solution to this problem that society faces. To difference from other technologies that human beings have developed along our history, the systems of A.I. can be built based on moral codes and / or ethics so that their behavior can be informed and / or restricted according to the values that those codes define, in the same way that happens with the human behavior. As scientists it is also part of our responsibility to lay the groundwork for both discussion and advancement around research areas that allow to understand and incorporate ethical behavior in the systems of A.I. from development, as well as to fight for policies of construction and development of systems that are going to behave according to certain codes and / or values and can't be easily manipulated by other agents to break those principles.

It is because of all this that the activities that have been developed are not limited to alert about the dangers of using A.I. for war purposes, it has also appeared in different forums and international organizations to generate consensus about policies and laws, not only against the proliferation of weaponry based on technique of A.I., but also to establish an adequate use this technology in all areas. In this sense, the main role of scientists has been, and is, to clarify the scope of new technologies developed from A.I. to disarm not only minimalist arguments, but also fatalistic in which political debates tend to fall. Laying the bases for a constructive debate that allows to generate precise control tools that don't impede the progress of the A.I.

Photo of International Campaign #stopkillerrobots (www.stopkillerrobots.org)

In this regard it seems important to clarify and / or establish a position on some aspects or concepts that appear repeatedly in different official documents (sometimes without a unified meaning):

1. **Lethal Autonomous Weapons Systems, SAAL, (Lethal Autonomous Weapons Systems).** Essentially refers to any war system that can perpetrate a lethal attack against human beings autonomously, that is, that can plan how to approach your goal and make the decision to murder without human supervision. In general terms, any device designed to kill people executing the action with their own criteria. This does not include remote-controlled missiles or remotely piloted drones for which humans make all the decisions of attack.

2. **International Humanitarian Law (IHL).** Set of rules that seek to limit the effects of armed conflicts. Rest on the basis of five principles: Humanity (priority to respect the person over military needs), Military necessity (prohibition of unnecessary military actions), distinction (determines the need to differentiate at all times between civilians and fighters, as well as between civilian goods and military

objectives), Limitation (prohibition of certain methods and combat weapons such as chemicals, bacteriological, nuclear and incendiary and antipersonnel mines), and Proportionality (rules to assess as licit or illicit damages caused to people and property that do not participate in hostilities because of an attack directed against a target military).

3. **"Autonomy"** is a term that should not be understood in a one-dimensional way. On the contrary, it is a concept that is derived from two greek words ("auto" -self- and "nomo" -governance-) and that has two own senses: on the one hand “self-sufficiency”, referred to the ability to take care of itself, or what is the same, to the condition or state of who is sufficient to itself. On the other is "self-directedness", understood as the attribute of be free of all external control. An essential feature of autonomy is the self-learning that allows you to adapt to changes and increase your unpredictability At present the total autonomy has not been implemented in no artificial system. There are yes, devices with partial autonomies such as cars without a driver or drones without a driver. In known cases, the autonomy is specific to a task. A SAAL requires self-sufficiency in tackling multiple tasks, combining and coordinating the necessary actions to achieve an objective. Autonomy is an orthogonal or additive characteristic of a weapon system. That is, all types of conventional weapons are potentially capable of achieving total autonomy with the technologies of A.I.
   To be more precise, a SAAL would be a type of weapon that can select (ie, search, detect, identify and locate) and attack (use force against, neutralize, damage or destroy) objectives without human intervention. This type of weapon would have the ability to learn and / or adapt their functioning in response to the changing circumstances of the environment in which they are deployed, so that their use could reflect a qualitative change in the paradigms in the conduct of hostilities. In spite of those expressed in the preceding paragraphs, it must be made clear that it is very difficult to define with precision the concept of autonomy. The pretense of doing it as a step prior to stipulating the prohibition of systems that seek to achieve it, even if it seems sensible, leads to inaction.

4. **International treaties.** There are treaties dedicated to specific types of weapons for cluster munitions, antipersonnel mines, blinding lasers, weapons chemical and biological weapons, and of course nuclear weapons.

5. **Artificial intelligence.** Computer systems that carry out tasks that usually humans perform making use of intelligence. These tasks essentially involve reasoning and learning. Which imply other cognitive skills such as representing information / knowledge, recognizing images and sounds, etc. Currently the systems of A.I. have many limitations.

6. **Ethics in a computer system.** Provide the system with the ability to distinguish / discriminate what is right / good from those that is not, either in a sense cultural, social or legal.

Given these definitional and conceptual questions, one could ask today: to what point can be measured the risks of something that we don't know if it can be created ?; will be it be possible that some day there are levels of "artificial intelligence" so sophisticated that they generate completely autonomous armament systems? and without going very far, how could a human program an autonomous system so that it manages to differentiate a civilian from a fighter?; What formulas can be applied to program in the weapon a standard of proportionality in the use of lethal force in war zones that by definition are rather unpredictable? If there is a design and programming error, who should be held accountable for the damage caused by that war device? the commander of the mission, the human operator that activated it, the programmer? Even more, if these systems are completely autonomous, could they disobey orders? or, failing that, can we provide these systems of ethical blockers that prevent them from committing misdeeds? But above all, in what situation would be the human dignity of a person who comes to feel terror, panic, desolation or impotence to be affected by an injury committed against him for a machine and product of a technical error?

All these valid questions highlight the fact that the main discussion of the appropriate use of a technology so powerful potentially, must be addressed right now. In modern times, technological development generally advances in a way much more accelerated than the analysis of the ethical controversies it produces. But in this particular case, we believe that the approach of the ethical implications of the use of A.I. should be performed a priori or at least in parallel. It is clear that we are not encouraging to return t medieval practices where society and the state arrogated themselves the right to prevent the development of astronomy. Instead, we fight for the establishment of legal and cultural bodies that guide the development of I.A.

In this regard, it seems necessary to promote the following actions:

1. The promulgation of ethical codes for the application of A.I. in any device. According to these principles, intelligent systems could not commit illicit, or attempt against the physical or psychological integrity of people.
2. Define criteria for binding and non-binding legislation (hard and soft law) in the use of A.I.
3. The prohibition of the development of the SAALs before they are technologically feasible to be built. At present, the scientific and technological knowledge has not reached the scope and sufficient development for the effective construction of this type of armaments, but it is estimated that they will be achieved in the coming

decades. The advance settlement of international regulations that prohibit its development seeks to avoid the conditions of the Treaty on the Non-Proliferation of Nuclear Weapons where five countries arrogated themselves the right to own nuclear weapons by simple fact of having carried out tests prior to signing it.

4. Promote more interdisciplinary research projects to study the socio-cultural effects product of the development of systems of A.I. and its ethical and moral implication in society.
5. Encourage the creation of multidisciplinary forums (including, among others, psychologists, sociologists, political scientists, philosophers, scientific computers, economists, legislators, etc.) to discuss and provide guidance on issues emerging issues related to the impact of the A.I. in society.
6. The teaching of ethical aspects in professional training courses that will develop these new technologies.
7. Promote an international non-development treaty of SAALs (A Treaty of NonDevelopment of Lethal Autonomous Weapons Systems), which should be based in at least two pillars: the non-development and the use of Artificial Intelligence only to peaceful purposes. For which an International Agency of Artificial Intelligence (within the framework of the UN) that is the entity of control of application of the rules arising from the treaty.
8. Develop a protocol of social impact that allows to assess the relevance or to not launch a new product on the market that uses A.I.

Some of these aspects are already being solved in isolation by companies like Google, whose employees have promoted seven principles that the systems of A.I. should follow: 1) Be socially beneficial; 2) Avoid creating or reinforcing unfair biases; 3) Build socially safe systems; 4) Be responsible to the people; 5) Incorporate privacy design principles; 6) Maintain high standards of scientific excellence; 7) Be available for uses that are in accordance with these principles. (Details in https: // www.blog.google/technology/ai/ai-principles/).

Even so, we believe that actions cannot be limited to private initiative or unilateral of a group, beyond its good will. On the contrary, we think that it should be the society that commits itself in this debate.

To close this presentation we would like to highlight that the ethical arguments and conflicts wielded so far in relation to the construction and use of smart weapons, clearly demonstrates the importance of the discussion and the need to start from now on in the definition of action plans that help mitigate related problems. However, some scientists and other stakeholders interested in the area point to a problematic that seems in principle to be more urgent than the long-term discussion about killer robots, and they are the implications of the use of A.I. systems, called "black box" for decision making in everyday or civil situations. The algorithms based on machine learning techniques infer patterns statistically relevant from the analysis of a large amount of data. Two problems exist related to these algorithms, one is that, if they are fed with biased data, the results and therefore the decisions that are taken based on them can be biased. This is a problem because in many cases these algorithms are used by people without the necessary training to provide input data free of bias. The second problem is the opacity of these algorithms in terms of their operation. Some of the techniques used are really complex, which makes it difficult to understand and audit the systems that use them.

To make the controversy clear, we will propose a couple of examples (but there are very many more):

1. In 1991, Dr. Diane F. Halpern (from California State University, San Bernardino) and Dr. Stanley Coren (from the University of British Columbia) published an article whose conclusion was alarming. These researchers took a sample of 987 individuals who had died and asked their closest relatives, if they were left-handed or right-handed. The finding was very disturbing, the lefties died nine years before the right-handers. The work was published in the New England Journal of Medicine, which is one of the most prestigious medical journals of the world. If the conclusions were correct, this meant that being left-handed was so bad like smoking 120 cigarettes per day. Subsequently, it was found that shows that it had been taken was biased and did not correspond with reality. However, suppose that an insurance company uses that information to fix the premium of your policies. That would mean that life insurance for left-handers would be significantly more expensive.

2. In any case, the issue of bias is also human. The system of conditional release for prisoners in the United States provides a striking example. "It has been shown that it is much more likely that convicts will be granted probation if they appear before the judge immediately after lunch instead of just before ", (see https: //www.ncbi.nlm.nih.gov/pmc/articles/PMC3084045/). Algorithms, which of course are immune to empty stomach syndrome, are easy to program because in the USA probation depends essentially on just one parameter: the risk of fleeing or re-offending based on the background. Is then possible a truly 'blind justice' based in completely objective facts? The answer does not seem to be conclusive. Today the courts of the United States use a "black box" to take these decisions. This system is based on historical data and reproduces the biases of the data with which they are trained.

Recently, several prominent scientists in the area of A.I. have declared their position facing these problems and point to how difficult it is to be able to identify behaviors biased in the most commonly used systems. The current algorithms simply are not designed to explain the decisions they make, which makes it impossible for a user of an application to understand, much less question, how the application uses his or her preferences, or why the information is presented in a certain order. And the solution is not achieved by requesting service providers to publish the details of the data with which the algorithms or algorithms themselves were trained.

Many of these tools are too complex and require control of many parameters, so that they can be examined meticulously in order to understand their functioning. Therefore, alternative and complementary models are being explored that allow, for example, that the system itself offers justifications on the determinations that it assumes during the decision-making processes.

Modern machine learning techniques have allowed A.I. systems leave the laboratories and overcome applications of "toys" to be immersed in the real world, allowing the environment to be recorded and understood. Problems of opacity and potential bias from the handling of data are two of the most urgent limitations of current systems of A.I. that should be addressed before they further compromise the privacy and other ethical issues of our society.

About the authors

María Vanina Martínez is PhD in Computer Science, National Scientific and Technical Research Council (CONICET) researcher, works on the Institute for Computer Science (CONICET- University of Buenos Aires (UBA)) and Professor in the Department of Computer Science at the UBA. Her interests are focus on the development of models to represent knowledge on support system on decision making based on Artificial Intelligence.

Ricardo Oscar Rodriguez is Phd in Computer Science specialize on Artificial Intelligence. He is Associate Professor In the Department of Computer Science, Faculty of Exact and Natural Science-UBA and Science Computer Institute member (UBA-CONICET) His scientific work are focus in the development of logical models for reasoning under uncertainty and incompleteness. He was been co-chair and financial chair fro IJCAI 2015.

About the organizers

SEHLAC (www.sehlac.org)

The Latin America and Caribbean Region Human Security network (SEHLAC) worked on Humanitarian Disarmament since 2009. SEHLAC was founded during the Oslo Process that Ban Cluster Munitions and since then became a crucial actor on the negotiation disarmament processes as the Arms Trade Treaty (ATT) Landmine Treaty, the Nuclear Ban Treaty, from which one SEHLAC received the 2017 Nobel Peace Prize as part of ICAN.

SEHLAC is also part of the Campaign to Stop Killer Robots and International Action on Small Arms (IANSA) and is part of most of the international campaign boards.

Campaign to Stop Killer Robots (www.stopkillerrobots.org)

Formed in October 2012, the Campaign to Stop Killer Robots is a coalition of non-governmental organizations (NGOs) that is working to ban fully autonomous weapons and thereby retain meaningful human control over the use of force.

The Campaign to Stop Killer Robots Steering Committee is comprised of six international non-governmental organizations (NGOs), a regional NGO network, and four national NGOs. It serves as the principal leadership and decision-making body for the Campaign to Stop Killer Robots. SEHLAC is one of them.