Artificial Intelligence and Autonomous Weapons: The Pandora's Box of our days?

Bianca Ferro Scarano
Heloisa Alves de Souza
Marcelo Chiavassa de Mello Paula Lima
Viviane Vaz Di Rossi Arantes Curvello
Universidade Presbiteriana Mackenzie
Observatório Mackenzie de Inteligência Artificial
Campinas/SP, Brazil

***Abstract* — Artificial Intelligence, especially deep learning, is a path of no return. It is necessary, however, to reflect on the risks of this new technology. The article aims, therefore, to analyze some of these risks. For that, the most extreme situation was chosen: autonomous weapons, capable of thinking and deciding for themselves who, when and how to kill a human being**

.

Keywords – Artificial Intelligence; Autonomous weapons; Ethical programmation

I. INTRODUCTION

The advancement of Artificial Intelligence has allowed the development of fully autonomous military weapons capable of identifying and destroying human and inanimate targets. But this technology comes up against questions of a moral nature, such as the legitimacy of machines to hurt and / or kill human beings. Several scholars have been worried and warned of the risks of this venture, especially when there is no safeguard clause (population protection) at the time of programming the machine. UNESCO and the European Parliament are trying to curb this technological race, with very detailed reports on the problems of authorizing such technology.

It is a current problem, whose decisions and studies today will mirror future legal solutions.

II. ASIMOV RULES AS A SAFEGUARD OF ARTIFICIAL INTELLIGENCE

In the 70s Isaac Asimov consolidated himself as one of the greatest, if not the greatest, science fiction writer of all time; his books whare set in a future where robots were not only a reality but also a threat. For this reason, they should be programmed according to three basic laws:

1st Law: A robot may not injure a human being or, through inaction, allow a human being to come to harm.

2nd Law: A robot must obey orders given it by human beings except where such orders would conflict with the First Law.

3rd Law: A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.

Asimov later added the "Zero Law" that was above all others: a robot may not harm humanity or, through inaction, allow humanity to come to harm.

Today, after 50 years AI's are no longer protagonists of fiction, but a daily reality and present in almost every aspect of our life. Although Asimov has been successful in predicting various aspects of our current reality in his works it is certain that he never imagined that his laws would be used until this day as a safeguard of humanity in relation to machines[1].

Although such guidelines have so far more or less effectively addressed moral discussions about the limits given to algorithms, robots, and AIs, they need to be updated for variables that even Asimov himself cannot predict:

- This information and how profitable the sharing of data can be[2], mainly because the information that society makes available about herself on the internet can be used as patrimony of companies like Google and Facebook.
- The use of autonomous weapons in wars
- What is the public duty of large Internet companies to society?
- The biased algorithms that hurt the ZERO rule.

[1] PASQUALE, Frank. Toward a Fourth Law of Robotics: Preserving Attribution, Responsibility, and Explainability in an Algorithmic Society

[2] BALKIN, Jack M.. Information Fiduciaries and the First Amendment
https://lawreview.law.ucdavis.edu/issues/49/4/Lecture/49-4_Balkin.pdf

We need to break the myth that robots are neutral intelligences that do not have any kind of prejudices. Algorithms are highly bias regardless of how they were programmed, either because their programmers reproduce in code their way of viewing the world (best approached in the sequence of this work) or by processing data loaded with prejudice during the deep learning / self-learning.

Currently, there is no way to teach algorithms to be truly neutral, robots are a mirror of the data they were exposed, so if we live in a bias society there is no possibility of AI's also not be.

It is already common having AI's present at the moment of making decisions; there are algorithms that help judges make court decisions, that help companies to choose the best candidate in a selection process but, increasingly these mechanisms operate autonomously combining this with the fact, already established, that they are biased this can be the beginning of a new type of segregation, a much more dangerous one because it cannot be predicted or visible in the short term[3].

III. The philosophical Discussion on Autonomous Weapons killing human beings (is this what we want?).

Every few decades humanity is faced with a dilemma, with a technological breakthrough in the war industry so great that we always have a question: Although we have the technical capacity to produce such a weapon, should we? In the overwhelming majority of the time the answer to this question leads to the death of thousands of people. It was so in the creation of firearms and in the use of nuclear weapons. Human beings have an almost fundamental need to test their creations and take them to the limit of their potential. It is precisely this distinction we have today with lethal autonomous weapons systems (LAWS), weapons that run from algorithms (self-learning??) that allow their total autonomy in the battlefield[4]

For over a decade we have drones routine functioning as weapons in war zones and their use, although controversial, is no longer so questioned. The difference between them and the LAWS comes from the fact that the drone alone does not kill anyone; the decision to shoot or not is always tied to a soldier, that is, the risk control is human, the decision is human, which technically disqualifies its purpose as an affront to Asimov's laws.

So, if in this article it has become very clear that no type of program is exempt from containing a value judgment, what prevents them from turning against us? This discussion may seem to have come straight from science fiction books but delegating the power to take a human life to a machine creates serious risks to our security, liberty, and dignity principles to which we struggle every day to keep intact and that may be beyond our control in no time.

The way AIs, that run through deep learning, process the data delegated to them and the way they arrive at a result are no longer understood by the scientists who designed it, making it a serious threat to humanity. The algorithms surpassed human intelligence, that's a fact, we can no longer predict if they, after their analysis, will reach the conclusion that the most efficient way to achieve their goal is to eliminate an entire city.

Such a discussion is important because we are used to living in a world where it is common and normal for nations to exercise power over each other through new technologies[5], it is understandable that the use of autonomous weapons is extremely tempting and to some even justifiable port that this would take soldiers of the battle field but, few nations have the resources to make such fight merely fair, entire populations would be left defenseless, the fear of the collapse of society as we live today is not unfounded and its risk should not be given as frivolous.

Legalizing the LAWS and putting them on a battlefield is to allow a non-human intelligence, with a high war potential and a judgment of value and that we do not know very well how they think to decide who lives and who dies, what isn't a good idea[6]. Allowing human beings to govern one another in ways never seen before.

IV. Ethical criteria for the AI to accomplish its war goals

New technologies are based on values. Like any innovative technology with transformative capability, some Artificial Intelligence applications can raise new ethical and legal issues linked, for example, to liability or potentially biased or harmful decisions. But what, after all, would be the limits and ethical criteria for a machine to fulfill its purpose?

It is observed that the potential reinforcement of powers through the use of robotics and AI contrasts with a set of tensions or risks and must therefore be seriously assessed from the point of view of safety, health and human protection, freedom, privacy, integrity and dignity, self-determination and non-discrimination, and the protection of personal data; in order to maintain these principles, new technologies must follow ethical standards, which will limit the performance of autonomous machines in certain contexts.

However, this delimitation of what is or isn't correct is not that simple. Considering that the use of autonomous lethal weapon systems raises fundamental ethical and legal issues, concerning key functions such as targeting and launching attacks that would be carried out exclusively by

[3] ZEYNEP Tufekci, Machine intelligence makes human morals more important.
https://www.youtube.com/watch?v=hSSmmlridUM&feature=youtu.be

[4] RUSSELL, Stuart. Take a stand on AI weapons. Nature

[5] BALKIN, Jack M.. The Three Laws of Robotics in the Age of Big Data
https://ssrn.com/abstract=2890965

[6] PARLAMENTO EUROPEU. Sistemas de armamento autónomos. Resolução do Parlamento Europeu (2018/2752(RSP)).

machines, and considering also that this use could compromise fundamental principles, it is necessary to define minimum standards and criteria in the execution of its objectives. That's because machines and robots, unlike humans, are incapable of making decisions to which the legal principles of distinction, proportionality and precaution apply, even the autonomous ones.

It is fundamental to create a climate of trust and responsibility around the development and use of AI and, therefore, its ethical limit is defined based on respect for the abovementioned principles and values set out in Article 2 of the Treaty on European Union, which together constitute the basis of the rights enjoyed by persons living in the Union.

There is much to debate on the subject, since it is an entire society living with a plurality of opinions and values that are constantly in conflict to each other. However, it is necessary to search for the moral criterion that best serves the interests of the population, based on fundamental and not expendable human principles, even in the name of social technological development.

V. Programmer Ethics

All the way through the potential destruction of the machines reveals the need to pay attention to the ethics of those who have programmed the artificial intelligences. There is still a great mistrust about how they work and what they can do, whether in relation to information collected that affects people's privacy and freedom, as well as the need for transparency of their algorithms or even the fear about their security (the possibility of cyberattacks, for example).

The programmer must map in advance situations of risk and make safeguard criteria regarding the possibilities of errors of the machines, as well as create mechanisms that make it harder as possible to be invaded by crackers. The difficulty lies precisely in avoiding to the maximum the fallibility of such systems.

The concern regards on international order, given that the information collected may affect even the sovereignty of States, which, to a greater extent, may end up creating an environment of hostility and mistrust between nations.

In this matter, the European Commission[7] has sought to implement in advance the trust and ethical responsibility for the use of AI to encompass both the values of the Treaty on European Union, the General Regulation on Data Protection and the Charter of Fundamental Rights of the European Union.

It is imperative that the programmer builds on the foundation of historically conquered legal institutions as a basis for the development of AI in order to support their conduct in international human rights treaties. From this it follows that values such as the dignity of the human person and the search for the common good are guiding sieves of the possible interactions for new technologies with the human being and with the society.

VI. Conclusion

Artificial intelligence is a reality and can not be refuted. Its use will bring multiple advantages to society. However, it is necessary to think about how to mitigate the problems stemming from deep learning, noting that data are not neutral, an equivocated decision making, as AI can not, in theory, decide on the basis of proportionality and reasonableness, and still, biased programming of the AI.

In relation to the autonomous weapons, the present observations are relevant. Are they really necessary? Would we be safe giving it the power to eliminate human beings, and thereby counteracting the safeguard clauses known as Asimov's rules?

It seems that these answers must be negative, as has been defended by the European Union and by several engineers and researchers around the world. The AI is still a black box: if opened can bring several breakthroughs, but can also be the pandora box of modern times.

VII. References

BALKIN, Jack M., Information Fiduciaries and the First Amendment (February 3, 2016). UC Davis Law Review, Vol. 49, No. 4, 2016, Forthcoming; Yale Law School, Public Law Research Paper No. 553. Available at SSRN: https://ssrn.com/abstract=2675270

BALKIN, Jack M., The Three Laws of Robotics in the Age of Big Data (August 27, 2017). Ohio State Law Journal, Vol. 78, (2017), Forthcoming; Yale Law School, Public Law Research Paper No. 592. Available at SSRN: https://ssrn.com/abstract=2890965

EUROPEAN COMISSION. COMMUNICATION ARTIFICIAL INTELLIGENCE FOR EUROPE. https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe

PASQUALE, Frank. Toward a Fourth Law of Robotics: Preserving Attribution, Responsibility, and Explainability in an Algorithmic Society. Ohio State Law Journal, Ohaio jul. 2017. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3002546

PARLAMENTO EUROPEU. Sistemas de armamento autónomos. Resolução do Parlamento Europeu (2018/2752(RSP))

RUSSELL, Stuart. Take a stand on AI weapons. Nature, Reino Unido, v. 521, p.415-416, maio 2015.

ZEYNEP Tufekci, Machine intelligence makes human morals more important. https://www.youtube.com/watch?v=hSSmmlridUM&feature=youtu.be

[7] EUROPEAN COMISSION. COMMUNICATION ARTIFICIAL INTELLIGENCE FOR EUROPE. https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe.