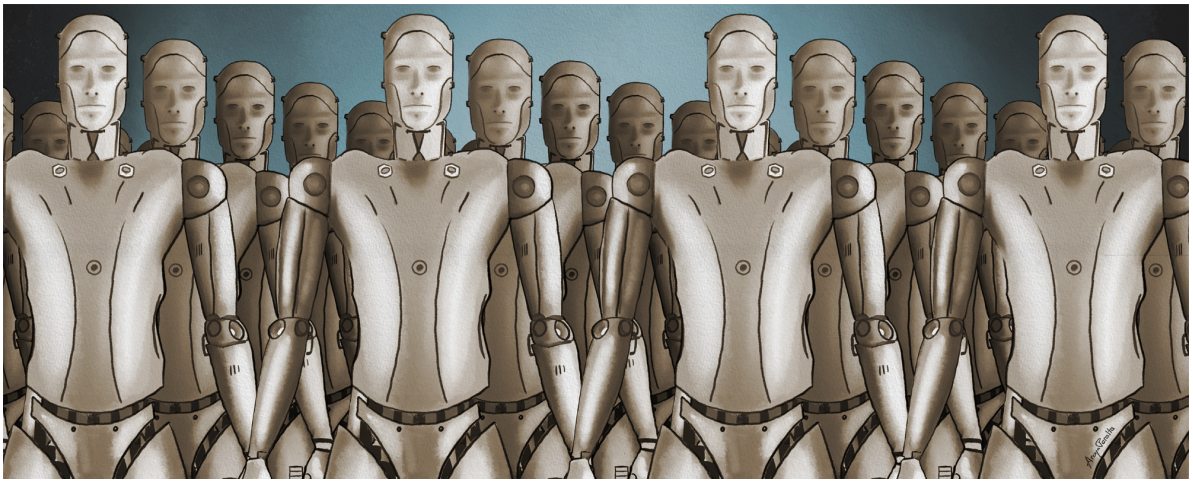
MILITARY USE OF AI

The dark side of Artificial Intelligence

Lethal Autonomous Weapon Systems and Killer Robots

Roser Martínez Quirante, Joaquín Rodríguez



Araya Peralta

The opening decades of the 21st century have been shaped by a process of technological acceleration that can only be compared to other historical milestones, such as the industrial revolution. They have ushered in a series of systemic changes that are difficult to value and understand in their entirety.

Myths and Realities of AI

As an example of the magnitude of this transformation, we need to keep in mind that, for the first time in the history of our species, we live in a context where critical decisions affecting the lives of individuals are made (partially or entirely) by non-human beings, i.e. by intelligence simulations. Nowadays, in a large number of countries, many of the transcendental decisions that affect our lives, such as being admitted to a university or approved for credit or a mortgage, are made by machine-learning algorithms; artificial intelligence with the ability to affect not only isolated individuals but whole communities. On this point, it's worth remembering the role played by certain financial algorithms in the 2008 financial crisis, which, by autonomously dumping the system into massive stock sales,

contributed to and accelerated the economic downturn. We should also consider the role that credit scoring algorithms are playing in the crystallisation of poverty in African-American and Hispanic communities in the United States today.

To put it another way, in the name of automation and standardisation, we find ourselves before a dehumanised process determined by the transfer of certain decisions to synthetic beings with no humanity, and it's a process that erodes our responsibility and accountability mechanisms. It's as if the famous "computer says no" line from *Little Britain* is ingrained in almost every layer of the system or, as Neil Postman described it, we have surrendered our culture to technology.

From Siri to Cortana, Alexa to Google Duplex; we are surrounded by new AI devices. We've become accustomed to sharing our reality with intelligence simulations, yet perhaps even more significant and shocking is the fact that, in an exercise of technological exhibitionism which could have radical consequences for the protection of civil liberties and human rights, we've become accustomed to giving away our data.

> For the first time, critical decisions affecting the lives of individuals are made by intelligence simulations

One of the main culprits behind our surrender of privacy, individuality and decision-making capacity (the end of which is extremely difficult to predict), are the meta-narratives which paint the new technological revolution as a solution to many of the considerable problems posing a threat to our societies. These include climate change, the treatment and cure of diseases, the water crisis, border controls and national security, among others.

But we forget that one of the many inherent risks of these smart applications is that they have the power not only to shape reality but also to alter our perception of it. Personalisation algorithms like *Page Rank* are a prime example in that they operate on the false promise of being able to choose and prioritise the internet pages of interest to us.

We often overstate the beneficial capabilities and potential of technology while overlooking not only its vulnerabilities but also the risks inherent in its development, implementation and crystallisation (remember what happened when Facebook discovered an AI program had created a language unintelligible to its creators?).

As was the case with systems from other technology fields, such as nuclear and transgenic technologies, we are told that AI will be critical for diagnosing certain diseases, the equitable distribution of food and the fight against climate change. Just as the transgenic industry of the 1990s promised to end world hunger and the nascent nuclear industry promised us cheap, clean and safe energy, it now seems that we're not allowed to put limits on AI, because it's destined to save the world.

History is clearly repeating itself, and top-tier international forums are once again replete with promises of utopian scenarios that can only be reached via a single path: the surrender of data, privacy and ultimately, humanity. The world of artificial intelligence is full of questions that need to be answered, especially given that much of the data provided by users is at risk of being used against them, whether by private corporations or for military programs, as is the case with AI facial recognition programs.

If something is free in the digital world, it's because it's you that is for sale. Data has now overtaken oil to become the most valuable raw material on the planet. After all, an algorithm without data is useless, and slowly but surely we're moving towards an algorithmic society, where our own behaviour and language are gradually adapting to the needs of the algorithms, and not the other way around. The case of Cambridge Analytica (which marketed the private data of more than 50 million people) is a paradigmatic example of our societal and individual vulnerabilities. But so are the social experiments of Facebook's FaceAPP (an application that ostensibly lets us see what we might look like when we're older but in reality obtains our authorisation for them to traffic our biometric data). This is proof of how it's relatively easy to take advantage of a society whose critical thinking has been relegated to its lowest levels by educational and media frameworks that devalue it.

The problem, however, runs much deeper. The theoretical need to use technologies that completely invade our privacy is often justified on the grounds of three fallacies.

The first is that, if properly coded, machines can adopt ethical-moral behaviours. But clearly, a machine is capable of neither ethics nor morals or intuition. At any rate, it could only ever have the ethics of the person who coded it. It would simulate the ethics of the programmer, be a replica of the engineer or a combination of information sourced from the cloud. The question we should be asking is, once the AI is programmed, will the system evolve by itself? Or will we be condemned to an immobilised society where good and evil are crystallised on the basis of a subjectivised algorithmic construct? And if it did evolve… what would be its objective?

> AI can by no means be considered a moral agent because it's merely a simulation

In short, artificial intelligence can by no means be considered a moral agent because it's merely a simulation. It, therefore, cannot comprehend, under any parameters, something as fundamental and pivotal as the value of a human life, nor feel respect or compassion.

The second fallacy is that AI can make decisions more effectively, more equitably and more justly than a human. Nothing could be further from the truth, firstly, because AI emulates the ethical-ideological system of its creators. In other words, it reproduces our lack of impartiality. As Cathy O'Neil demonstrated in her work *Weapons of Math Destruction*,

believing in the infallibility of algorithms can have drastic results. Both the teacher evaluations in Washington State and the work carried out by the American Civil Liberties Union to show that facial recognition systems have a high tendency to identify non-Caucasian subjects as criminals have proved this.

We're dealing with a technology designed by white men who transfer their likes and dislikes to their creations and build systems in line with their way of thinking. Moreover, being a heuristic system, understanding the processes used by the AI to arrive at a specific decision is incredibly challenging. And if it's impossible to deconstruct or explain the processes behind a particular AI decision, then it is irresponsible to allow them to operate freely.

And finally, we come to the third fallacy, which argues that artificial intelligence is more reliable than human intelligence. This concept could be accepted within very specific circumstances, but never in general terms. Here, it's worth highlighting the work of British NGO Big Brother is Watching Us, which, by appealing to the Freedom of Information Act, succeeded in getting the government to reveal the reliability of the facial recognition systems used at the Camden Carnival. Only 5% of the criminal identifications made through the AI system were correct, giving it an average error rate of 95%.

If we believe Professor Noel Sharkey's theory on Automation Bias, these fallacies have even more worrying consequences. His theory suggests we humans have a tendency to unquestioningly accept the judgements and analyses made by AI because we think they're more effective and reliable than our own.

But the most astonishing thing is that despite understanding that AI cannot be considered a moral agent, and despite knowing that its ability to interpret reality is limited by the biases of its creators and the wider society (especially systems fed by natural language), its usage is increasing, and these types of systems now guide ever more processes.

Significant human control and lethal autonomous weapon systems

This brings us to the most aberrant aspect of this topic: the use of AI in lethal autonomous weapon systems (LAWS). LAWS are a new generation of weapons with the ability to select and eliminate targets without significant human control. In other words, we're dealing with the delegation of lethal capabilities to an alleged artificial intelligence, which will have the power to decide not only who can get credit, who gets accepted to a university and who can access a particular job, but also, who lives and who dies.

We're talking about a type of weapon that obviates the rational, cooperative, intuitive, moral and ethical dimension of human decisions. One that contradicts international humanitarian law, the rules of war and, internally, given that the State has a monopoly on legitimate violence, administrative law.

All lethal autonomous weapon systems (drones and robots) developed to date, depend or should depend, on human supervision or judgement. That is, there should be significant prior human control involved in at least some of their critical phases (target selection or command cancellation). However, a lack of clear regulation around the issue is allowing for the research and development of fully autonomous systems, and state inaction is leading to a potentially perilous kind of lawless competitive race between governments. As a member of the International Committee on Robotic Arms Control (ICRAC), the United Nations is trying to put an end to it by passing a multilateral treaty to ban these genocidal weapons.

Despite these efforts, most of the heavyweight states justify the research of this lethal technology by claiming it will be used for national defence, rather than to attack. But this seems like little more than subterfuge in the race to become the first country to launch these categorically lethal systems endowed with the ability to become independent of their creator and manager. It is, therefore, essential to develop international regulation that prohibits the deadly use of AI and clearly limits any existing interrelationship between national defence systems and those whose purpose is lethal action against people. If we fail to do so, we could see a situation where someone gives the power to decide who lives and who dies to a machine with no humanity, effectively creating a robot or a drone with a licence to kill.

In recent years, we have begun to detect movements that mark the beginning of a new arms race with potentially disastrous consequences for the future of our species. China, for example, is rapidly modernising its military and has opted for state-of-the-art nuclear weapons with AI warheads to limit collateral damage during a targeted attack. By contrast, the United States remain tied to the weapons of the past by its “military-industrial-congressional complex” (MICC) or “Iron Triangle”, which refers to the tripartite relationship between private military contractors, the Government and Congress.

The fact that, between 2014 and 2018, China carried out around 200 laboratory tests to simulate a nuclear explosion while the US, during the same period, only carried out 50, is a clear illustration of the situation. China’s path is evident. Ultimately, as Hartnett from the Bank of America points out, “the 2018 trade war should be recognised for what it really is: the first stage of a new arms race between the US and China to reach national superiority in technology over the longer-term via quantum computing, artificial intelligence, hypersonic warplanes, electric vehicles, robotics, and cyber-security”.

Investing in technology is, therefore, linked to defence spending (although it doesn’t always lead to greater security): the IMF predicts that China will gradually outperform the US, becoming the world’s dominant superpower by 2050. Specifically, it calculates that it will overtake the US in terms of economy, military power and global influence sometime around 2032.

Death at the hands of an autonomous AI system is contrary to our concept of human dignity

In armed conflicts, the right to life means the right not to be killed arbitrarily or capriciously, inexplicably, inhumanely or as collateral damage, and a death cannot violate the right to human dignity. It could even be argued that the right to human dignity carries more weight than the right to life because, even in a civilised society, legal executions may take place so long as they respect human dignity.

> While LAWS may provide better results based on a cost-benefit calculation, they should be banned for ethical and legal reasons

The fear of a dystopian future would seem to be a legitimate reason for the precautionary prohibition or moratorium of LAWS. However, to defend that position, we must first reinforce the notion of human dignity and the Martens Clause, as well as the issues related to the significant human control and self-determination of autonomous lethal systems.

We believe that while LAWS may provide better results based on a cost-benefit calculation, they should be banned for ethical and legal reasons. Heyns, who shares this opinion, argues on the basis of Kant's conception of human dignity, which says that people have an inherent right to be treated as unique and complete human beings, especially when their lives are at stake. This human dignity would be denied if victims seeking to appeal to their executioner's humanity were unable to do so because it was an artificial being. The executive power must pay due respect to the dignity of the person in question and make constant evaluations and adjustments. What's more, nothing of the way our law is applied through human capabilities can be guaranteed by autonomous weapons whose actions would lack proper human judgement.

The dehumanisation process already initiated by the use of autonomous systems with human control in conflicts is an affront to all that we learned about cooperation, human dignity, verbal communication and the human relationship between combatants from the First World War, and it calls on us to work harder at finding new ways to coexist. Progress through non-verbal humanitarian communication slows down and even recedes when the fight involves lethal autonomous drones. In the words of Sparrow, "even when at war, we must maintain an interpersonal relationship with other human beings", or we will not respect the fundamentals of the law.

Defenders of these new smart weapon systems, ignoring the need for this component of humanity, attribute many benefits to them: reduced operating costs, their ability to carry out certain tasks more quickly than humans, their capacity to hit a target even when communication links fail, etc. Arkin also points out, in their defence, that they can be designed to accept greater risks, that they can be fitted with the best sensors, that they

won’t be shaken by emotions like fear or anger, that they won’t suffer from cognitive bias and even that they can legitimately and reliably distinguish between legitimate and illegitimate targets.

All of which could be true, but there are numerous examples of men and women in all kinds of situations and conditions who, when the time came, refused to press the button that would have caused the death of fellow citizens. Wars have evolved in humanity, since the non-verbal communication allowed soldiers in the trenches to organise moments of truce and low mortality without ever having received orders to do so.

To not be considered arbitrary, the termination of a human life must be based on informed decision and human cognitive judgement, since only a human decision can guarantee full recognition of the value of the individual life and the importance of its loss.

And so all the modern and complex standards of humanitarian law come into play: proportionality, compassion, the use of less onerous or restrictive methods, constant vigilance, chivalry, etc. The actions of lethal autonomous drones with AI, however, are neither legitimate nor morally justifiable and should be outlawed under the principle of human dignity and jus cogens, which, as a peremptory norm, provides for the fundamental rules of humanitarian law.

Intuition is a part of our essence as humans; it influences all of our actions and has always played a vital role in war. LAWS may be equipped with imitation mechanisms and incorporate integrative and cognitive, though not phenomenological, processes but they can never be intuitive or feel emotions; they can only attempt to replicate them. As G. Rizzolatti, the neuroscientist who discovered mirror neurones says, “robots can imitate, not feel”. If this is the case, given that the algorithms included in lethal autonomous systems cannot achieve the human characteristics needed to make transcendental discretionary public decisions relating to the exercise of legitimate force against people, the transfer, the decentralisation of these powers to those autonomous systems must not be accepted. The power to decide who is an enemy (whether at home or abroad) and discretionary power over human lives is such a monumental responsibility that it cannot be granted to artificial beings with no human emotions.

McQuillan warns that surveillance, thanks to the massive and detailed harvesting of data through intelligent systems, is leading to changes in governance and damaging the nucleus of civil society to such an extent that he gave the situation a name: “the Algorithmic State of Exception”.

> The only guarantee for the progress and sustainability of citizens’ rights before the artificial intelligence of autonomous systems is regulation

Even Mark Zuckerberg, the CEO of Facebook, stood before the US congress and tacitly acknowledged that we're dealing with a state of anomie and that we need a regulator who doesn't trust everything in the free market: "The federal regulation of Facebook and other internet companies is indispensable". This federal law will be projected internationally and, ultimately, globally because, as we've seen with other US regulations, it's likely to have extraterritorial implications for other countries. However, to date, there are no legally binding international instruments or even national laws to prohibit the development, production and use of so-called killer robots.

The only guarantee for the progress and sustainability of citizens' rights before the artificial intelligence of autonomous systems is regulation. The evolution of technology itself can be profoundly affected by usage that conflicts with the criterion of public opinion in such a way that it compromises its own success, as happened for nuclear and chemical technology. In the same way, if we relax our interventions in this particular technology, it may lead us on its own journey towards the end of humanity.

The most worrying and disturbing LAWS: lethal pocket-sized drones

At the United Nations, government experts are working to secure a treaty prohibiting lethal autonomous weapons (CCCW). Their focus is on the larger weapons used during wars (macro-LAWS such as Reaper, Taranis, Iron Dome, etc.).

However, we need to go further and recognise that the real danger lies in the individual usage of small weapons. These types of weapons, which could pass from the military sphere into the hands of any citizen to be used for their private security, could be described as micro-LAWS; another example of lethal dual-use technology.

Micro-LAWS have the potential to destabilise us and change the world of security as we know it. If military LAWS in the form of swarms of mini-drones are hard to attack, it's not hard to imagine what could happen if they fall into the hands of thousands of individuals who, instead of opting for a conventional firearm, choose a lethal drone to equip themselves with a level of security that the State cannot guarantee.

The right to bear arms guaranteed by the Second Amendment of the American Constitution allows citizens to own not only a handgun or a revolver but also any weapon deemed necessary for their security, including automatic and military weapons. Taken to the extreme, the right to bear arms could extend to the possession of a lethal autonomous robot for defensive or offensive protection. In other words, the right to own a lethal autonomous pocket drone with AI.

Just as we've seen with the drones used by the police, civil protection services, and even in the private sector, this technology will eventually transfer from the military domain to the public-civil domain. We, therefore, need international and globally administered preëmptive regulation to prevent this from happening.

We are at risk from the global insecurity that would ensue should an uncontrolled proliferation of these types of weapons fall into private hands; a situation made worse by the fact that it's difficult to predict how the lethal AI systems would interact with each other in those circumstances. Let's hope the law arrives in time to prevent this foretold pandemic.

Roser Martínez Quirante

Roser Martínez Quirante és professora de Dret Administratiu a la Universitat Autònoma de Barcelona des del 2002. També és professora de l'Escola de Prevenció i Seguretat Integral de la UAB des de la seva fundació, l'any 2004, i ha impartit classes en diferents matèries, entre les quals destaquen la Llei de seguretat, intervenció i autoregulació i la regulació d'armes de foc als Estats Units i a Europa. Ha estat coordinadora del grup de recerca EPSI-University de Massachussetts Lowell (UMAAS) per al desenvolupament d'activitats docents i de recerca sobre seguretat. És experta en l'àmbit de les armes autònomes i defensora de la campanya *Stop Killer Robots*.

Joaquín Rodríguez

Joaquín Rodríguez és investigador a la Fundació de la Universitat Autònoma de Barcelona i coordinador local de la xarxa Leading Cities Network. És professor de l'Escola de Prevenció i Seguretat Integral, centre adscrit a la UAB, i també és un dels promotors a l'estat espanyol de la campanya *Stop Killer Robots*, que pretén prevenir la proliferació de sistemes d'armament autònom. És Doctor especialitzat en anàlisi de riscos i en les relacions entre societat i tecnologia. Té un Màster en Relacions Internacionals amb una especialització en Estudis de Pau i Seguretat per l'IBEI (Institut de Relacions Internacionals de Barcelona) i un postgrau en Gestió de projectes pel Centre d'estudis Alts Acadèmics de l'Organització d'Estats Iberoamericans.