Artificial intelligence: promises, risks and regulation. Something New Under the Sun[1]

Luís Roberto Barroso

President of the Supreme Federal Court. Full professor at the State University of Rio de Janeiro (Uerj) and the University Center of Brasilia (Uniceub). PhD and associate professor at Uerj. Master's degree from Yale Law School, USA. *Visiting scholar* at Harvard Law School. *Senior fellow* at Harvard Kennedy School, USA.

Patricia Perrone Campos Mello

Secretary of Advanced Studies of the Brazilian Supreme Court. Professor of constitutional law at UERJ and UNICEUB. PhD and master's degree in law from UERJ. Post-doctoral studies as *a visiting scholar* at the Max Planck Institute for Comparative and International Public Law, Germany, and at Harvard Kennedy School, USA. Attorney for the State of Rio de Janeiro.

Abstract: This article deals with the benefits and risks of artificial intelligence (AI). It places AI at the dawn of the Fourth Industrial Revolution, explaining its essential categories and way of operating. It addresses the benefits brought by this new technology: expansion of human decision-making capacity, automation, research and innovation, medical school and education, among others. It examines the risks it presents concerning impacts on the labor market, use for military purposes, dissemination of misinformation and violation of fundamental rights. Finally, it proposes principles for the regulation of AI. The paper demonstrates that AI is a technology with great potential, whose real effects depend on the use we make of it. In such conditions, the law and its operators must look for an institutional design that encourages its positive use and contains its distortion.

Keywords: artificial intelligence; fundamental rights; democracy; labor market; risks; regulation.

---

[1] Article published in the Journal Law and Practice, v. 15, n. 4, 2024.

Initial note[2]

There is something new under the sun[3]. Many of our beliefs and certainties may have their days numbered. Just as ancient navigators contemplated the immensity of the oceans, full of promises, mysteries and dangers, we are once again facing an unknown world. There is a feeling in the air that a profound transformation is coming. A revolution, perhaps. Something grand like the invention of the printing press with movable type, which exponentially spread human knowledge, or the Enlightenment, which reshaped social life, culture and politics (Kissinger; Schmidt; Huttenlocher, 2023). The future has never seemed so close and unpredictable[4].

Faced with the seemingly infinite possibilities of technology, there is only one safe navigation chart: the values that have long since guided the advancement of civilization and the evolution of the human condition on Earth. Whether secular or mythical, they come from Greece, and include the Torah, the Gospels, Buddha, Thomas Aquinas, Kant and many others who have built the ethical heritage of humanity. But there is a dramatic point here: the dizzying scientific progress that we have witnessed, cumulatively, over the centuries, has not been accompanied by a corresponding ethical – and even spiritual – evolution of the human condition. Goodness, real justice and solidarity are often neglected in a world of extreme poverty in many places, unjust inequalities, wars and a domestic and international order in which some win everything and others always lose. It is in this scenario that the issue of artificial intelligence (hereinafter also referred to as AI) and its potential to make the world better. Or worse. Or even to annihilate it[5].

Perhaps no topic in the history of civilization has sparked so much simultaneous reflection. In the media, in bars, in universities, at major international events, and at meetings of experts, one subject has become omnipresent: artificial intelligence. There is no aspect of its implications that has not been explored by the most brilliant minds and the most ordinary citizens. The following text is part of this profusion of writings that seek to capture the spirit of the times, chart routes, and push history in the right direction. Avoiding the abysses that would put at risk, if not our lives, at least our humanity as we know it. Faith in science, like all faith in this world, cannot lead to fanaticism. We need to define directions and limits. Here is just one more attempt to do so.

This article is structured as follows. An introduction presents some basic notions about the topic. Part I explores the positive potential of AI. Part II seeks to catalogue the main risks that accompany

---

2 The authors are grateful to Pedro Henrique Ribeiro Morais e Silva for his help with research and suggestions in preparing the text, and also to Frederico Alvim for important comments and bibliographic recommendations.

3 There is a well-known passage in the Bible, found in Ecclesiastes 1:9, which reads: "What has been what will be; what has been done will be done again. There is nothing new under the sun." The meaning of this phrase is that as the ages pass, human beings struggle with the same existential questions. Perhaps some new questions are emerging, however.

4 This unpredictability can be explained, in part, by the progressive autonomy acquired by solutions based on machine learning and, in part, by the accelerated nature of innovations in this field, which follow one another (or are renewed) at a pace that hinders their full understanding. Nina Schick (2020, p. 11) notes that there were four centuries between the invention of the printing press and the development of photography, for example, but in just three decades, we went from the emergence of the internet to smartphones, and from there to the platformization of lives on social networks, with serious implications for the information regime. According to the author, such rapid changes in vital segments entail a high component of uncertainty, which must be considered by society as a whole.

5 AI research avoids alarmism, which does not mean that potentially catastrophic developments are not considered serious hypotheses. See Stuart Russell's work on the subject, *Artificial intelligence on our behalf. How to keep control over technology*. The most relevant concerns revolve around general artificial intelligence, also referred to as artificial superintelligence, which defines a state in which computers would surpass human capabilities to a prominent extent, leading to, according to Kai-Fu Lee (2019, p. 159-173), "control problems" and "alignment problems".

it. Part III identifies some principles that should govern the regulation of the matter. And, at the end, a conclusion seeks to assuage our concerns about the future.

Introduction
The dawn of the fourth industrial revolution

1. A Brave New World

A new industrial revolution is on the horizon. The first occurred in the mid-18th century and is represented by the use of steam as a source of energy. The second industrial revolution, at the turn of the 19th century to the 20th century, is symbolized by electricity and the internal combustion engine. The third took place in the final decades of the 20th century and culminated in the replacement of analog technology by digital technology. Known as the Technological Revolution or Digital Revolution, it allowed the universalization of personal computers and smartphones and is symbolized by the internet, connecting billions of people around the world (Barroso, 2019, p. 1262). The fourth industrial revolution, which is beginning to invade our lives, comes with the combination of artificial intelligence, biotechnology and the expansion of the use of the internet, creating an interconnected ecosystem that encompasses people, objects and even pets, in an internet of things and senses.

In this challenging new world that is unfolding, new technologies can free us from the simplest daily activities, as well as perform highly complex tasks. They can clean environments, regulate the temperature, and soon, they will drive cars autonomously (Manyika, 2022, p. 12). They promise to restore lost body movements (Caczan, 2023), provide more accurate medical diagnoses (Dilsizian; Siegel, 2014, p. 441), remedy neurological deficiencies, enhance cognitive abilities (Schmidt, 2017, p. 6-10), create someone's "virtual twin"[6], reproduce a person who has already died[7], allow reunions with loved ones who have passed away (Here. After [20--]), care for the elderly (Horowitz, 2023), find the ideal friend or romantic partner (Inner Circle, [20--]; Tinder, [20--]), write texts in the most diverse languages (ChatGPT, [20--]), distribute welfare aid to the most vulnerable, direct essential public services to the neediest places (Katyal, 2022, p. 327; Urueña, 2023). They also aim to predict the practice or recurrence of crimes (Eubanks apud Eubanks, 2015), improve environmental monitoring, promote the planning of smart cities (Galaz et al, 2021, p. 2), estimate the performance of job candidates, the probability of payment of financing, as well as the development of serious diseases (Silberg; Manyika, 2019), among other issues[8].

There is more. It is estimated that the same technologies can reveal a person's sexual orientation (Morrison, 2021), foresee and report the intention to have an abortion (Cox, 2022), replace hundreds

---

6 The site invites users to "virtually duplicate themselves" in order to "boost their productivity, mental health and longevity" (Mindbank, 2024).

7 The system's objective is to replicate the users' personality, their way of thinking, speaking and other characteristics, so that it can even interact with loved ones after the duplicate person's death (Ramirez, 2023).

8 The positive outcomes are indeed impressive, which leads some currents to consider the use of new technologies to transform governance mechanisms, in favor of establishing an "algorithmic democracy", supposedly neutral and effective. However, algorithmic neutrality does not exist, and democratic legitimacy is necessarily related to representation based on the will of the people. In this line, the Parliamentary Assembly of the Council of Europe understands that the definition of political and social objectives cannot be left to algorithms. On the contrary, it must remain in the hands of human beings who submit to a system of political and legal accountability (European Union, 2022).

of extras and actors in Hollywood (Beckett, 2023), create or eliminate thousands of mechanical or creative jobs (Manyika, 2022, p. 20), manipulate or falsify information, sounds, images, beliefs and desires (Hacker; Engel; Mauer, 2023, p. 1 and 2), generate addictions (Mohammad; Jan; Alsaedi, 2023; Becket; Paul, 2024), interfere with consumer behaviors (Makhnoumi, 2024), influence the outcome of electoral processes (Heawood, 2018, p. 429-434; Berghel, 2018, p. 84-89), provoke violent behavior (Pauwels, 2020), strengthen extremist agendas (Vlachos, 2022), aggravate inequality and discrimination against minority groups (Angwin et al., 2016; Eubanks apud Eubanks, 2015), alter and acquire free will (Hutson, 2023)[9], activate weapons of mass destruction, put people's lives, health and safety at risk (Manyika, 2022, p. 21 and 27).

The list is endless and can lead us to the sublime or to horror, to freedom or to slavery. To the broad affirmation of human rights or to their suppression. As intuitively stated, the problem is not in the technology itself, but in the use we will make of it and, above all, in how we intend to distribute the benefits it will generate. The challenge, therefore, is to produce an institutional design that encourages the good use of Artificial Intelligence and that contains its distortion, preventing the automation of the production of injustices (Degli-Esposti, 2023, p. 10) and the multiplication of existing risks (Coeckelbergh, 2023, p. 167).

2. What artificial intelligence is

In a simple definition, it is possible to state that artificial intelligence consists of programs (software) that transfer human capabilities to computers. These capabilities include cognitive tasks and decision-making, usually based on the data, instructions and objectives with which they are fed[10]. However, there is no full convergence on the technical concept of AI and its scope[11]. Numerous entities and institutions, such as the OECD[12] and UNESCO[13], seek to delimit its contours. It is possible to point out some common features in these attempts at definition: they are systems with the capacity to process data and information in a way similar to human intelligence, which includes learning, reasoning, perception and communication through language. When consulted, ChatGPT4 provided the following definition:

> Artificial intelligence (AI) is a branch of computer science devoted to creating systems capable of performing tasks that traditionally require human intelligence. These tasks include learning (the ability to improve performance with experience), reasoning (the ability to solve problems

---

9 There are fears that AI's ability to learn autonomously could lead it to acquire superhuman intelligence, making it uncontrollable. This phenomenon is called "singularity."

10 The expression "artificial intelligence" is attributed to a workshop held in 1956, in Dartmouth, with the aim of seeking to develop machines capable of solving problems solved by humans and improving themselves (MacCarthy et al., 1955; Manyika, 2022, p. 15).

11 Organizations representing AI companies advocate for the formulation of a more restrictive concept of artificial intelligence, while human rights organizations advocate for the expansion of the concept to other technologies, which can also produce adverse effects on human rights. In such a context, the scope of the AI context itself depends, in part, on how much one intends to regulate it (Madiega, 2023, p. 6-8).

12 "An AI system is a machine-based system that, for explicit or implicit purposes, infers from the information it receives how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptability after deployment" (OECD, 2019; Russel; Perset, Marko, 2023).

13 "Therefore, this Recommendation approaches AI systems as systems which have the capacity to process data and information in a way that resembles intelligent behaviour, and typically includes aspects of reasoning, learning, perception, prediction, planning or control" (Unesco 2021).

through logical methods), perception (the ability to interpret sensory data to understand aspects of the world), and linguistic interaction (the ability to understand and produce natural language).

At its current stage (Degli-Esposti, 2023, p. 10; Rebollo Delgado, 2023, p. 24)[14], artificial intelligence is not self-aware, has no discernment of right or wrong, nor does it have emotions, feelings, morality, or even common sense. In other words, it is entirely dependent on human intelligence to feed it, including ethical values. Computers have no will of their own (Winston, 2018; Lenharo, 2023). Although this is conventional wisdom on the subject, some experiments reveal a surprising capacity for learning, raising new concerns. One of them was Alpha Zero, an AI program developed by Google that defeated Stockfish, until then the most powerful chess program in the world. Unlike previous programs, Alpha Zero was not fed with moves previously designed by man. In other words, it was not based on human knowledge, experience, or strategies. It was given only the rules of the game. Alpha Zero trained by playing with himself, developed his own movements and strategies, original and unorthodox, with his own logic (Kissinger; Schmidt; Huttenlocher, 2021, p. 7 et seqs. and 26).

Two views have competed for primacy in artificial intelligence research over the years. The first was inspired by the way the human mind works, seeking to mimic the way we formulate questions and develop logical reasoning. This first perspective dominated AI experiments until the 1980s. The second view was inspired by the way the structures of the human brain work. It proposed connecting information processing units, equivalent to neurons, in order to simulate how they work (Dreyfus; Dreyfus, 1988, p. 15-44). This is the view that has become dominant in the AI scenario, called the "connectionist approach". approach). It does not seek to reproduce the way the human mind rationalizes. On the contrary, it seeks to establish correlations and patterns between thousands of data and certain results. Its main support points are statistics and neuroscience.

Artificial Intelligence systems are based on data and algorithms. The larger the data set they have access to, the greater the number of confirmed and discarded correlations and, naturally, the more accurate the results tend to be (Dreyfus; Dreyfus, 1988, p. 15-44). A given universe of data or correlated characteristics leads AI to identify a dog or a cat, a good or a bad debtor, a person with depressive tendencies, a child at risk. Establishing correlations between such elements may seem random or irrational to the human mind's way of knowing. But remember, the model is based on statistics, not logic.

The concept of algorithm, in turn, is a fundamental one in computer science. The term identifies the set of instructions, rules and parameters that guide computers to perform the tasks assigned to them. They are formulas, codes and scripts that select, process and store data, with the aim of obtaining a certain result. The selected data (inputs) and their correlations allow us to lead to the results targeted by the program (outputs), which can be the most varied. For example: if the result gives rise to the differentiation between objects and living beings, we speak of discriminative AI; if the result is the prediction of behaviors – consumer, financial or political – we have predictive AI; if it is the generation of content – texts, images or sounds –, we say it is generative AI (Hacker; Engel; Mauer, 2023, p. 1-3 and 13)[15].

---

[14] This caveat is necessary given that it cannot be ruled out that the AI of the future will grant machines intense doses of autonomy and consciousness, in a scenario in which intelligent applications acquire their own rationality, pursuing unforeseen objectives.

[15] The above qualifications are not exhaustive. Reference is made, for example, to adversarial AI, intended to prevent the operation of another AI for data protection purposes, among others (Hacker; Engel; Mauer, 2023, p. 13).

3. Machine learning, foundational models and other relevant concepts

In terms of operating mode, the most advanced artificial intelligence systems currently are those capable of developing machine learning. Machine learning refers to the ability of a model to acquire knowledge autonomously, without prior explicit programming, based on the identification of correlations between large volumes of data, as described above. It is also worth noting that, for more restricted concepts of AI, the capacity for machine learning is what differentiates artificial intelligence from mere automation, which would be a broader phenomenon (Nunes; Andrade, 2023, p. 4; Brown, 2021). Machine learning is the process that underpins most of the AI services we use today, such as content recommendation systems on platforms like Netflix, YouTube and Spotify, results selection and ranking models on search engines like Google, Bing and Baidu, as well as feeds and contact recommendation systems on social media like Facebook and X (formerly Twitter) (Hao, 2018; Nunes; Andrade, 2023, p. 5).

Machine learning relies on algorithms and artificial neural networks. Artificial "neural networks" (neural networks) are inspired by human neuron networks. They are mathematical models that imitate our nervous system (Porto; Araújo; Gabriel, 2024, p. 37). Through them, different data processors work together to process such data. Machine "deep learning" (machine deep learning) is the technique that expands machine learning capacity by establishing several layers of artificial neural networks, simulating the complex functioning of the human brain, so that such multiple networks and processors manage information and establish correlations simultaneously (Hao, 2018; Re; Solow-Niederman, 2019, p. 244-246). It has applications in image and speech recognition, automatic translation and text processing. There are three types of machine learning: supervised, *unsupervised and reinforcement* (Brown, 2021)[16].

In terms of use or purpose, AI systems come under numerous classifications, which sometimes overlap. Foundational models are trained with large amounts of data, prepared to adapt to multiple tasks. ChatGPT (Chat Generative Pre-Trained Transformer), a generative AI, is an example of a foundational model (also called general-purpose) and large language model, with the ability to generate art, images, texts and sounds (Jones, 2023)[17]. Note well: generative AI programs can create new content, not just analyze or classify existing content. If someone searches on Google how an electric car works, they will be sent to a link that leads to a third-party website. ChatGPT, however, will explain how an electric car works in its own words (Ariyaratne, 2023, p. 4 and 5).

Foundational models are considered a step towards so-called general-purpose AI. purpose AI), which has not yet been fully achieved, but is already capable of performing a large number of tasks, not being limited to specific objectives (Bommasani et al., 2022, p. 4-12; Hacker; Engel; Mauer, 2023, p. 2-4; Uuk; Gutierrez; Tamkin, 2023). In turn, fixed-purpose AI systems, or narrow AI, as the name suggests, are intended for a specific purpose and are therefore defined in a more restrictive way. They are trained with a more targeted database. This category includes most of the AI systems

---

16 Supervised learning, which is the main type currently used, is the type in which it is possible to define what the appropriate output is. This is the case of AI systems focused on categorizations (in cats, dogs, means of transport). Unsupervised learning, less common, is the type that seeks to identify patterns, correlations and groupings without previously defining the appropriate result. Finally, reinforcement learning aims to train and improve machines through a trial-and-error system, in order to teach them to make the best decisions (Brown, 2021).

17 Not all generative AI is a foundational model. It can be modeled for very specific purposes. Generative capabilities can include text, image, and video manipulation and analysis, and speech production. Generative applications include chatbots, photo and video filters, and virtual assistants.

currently in use, which include voice assistants such as Siri, Alexa and Google Assistant, which are designed to understand and carry out voice commands, content recommendation systems on streaming platforms, spam filters, weather prediction systems, facial recognition software, among others. All are intended for a very specific purpose.

Finally, the expression strong AI, also referred to as artificial general intelligence (AGI), designates systems with the capacity for understanding, learning and concrete application equivalent to that of human beings. It would thus be capable of reasoning, problem-solving and decision-making on its own. This type of AI constitutes a theoretical concept, not yet achieved, but possibly achieved in a few years, according to some researchers (Taulli, 2020, p. 218).

It would be possible to continue exploring the multiple technicalities of the topic. For example, the AI value chain, with its different phases: design, development, implementation and maintenance of the system (Hacker; Engel; Mauer, 2023, p. 8-11). Each of these phases has its own challenges and risks, involving different actors, which include the developer, the[18] deployer, the user[19] and[20] the final recipient. These[21] agents have different expertise and skills, and can generate different contributions, damages and responsibilities (Hacker; Engel; Mauer, 2023, p. 8-11). This variety of roles, as intuitive, adds some degrees of difficulty to the regulation of the matter.

It is time, however, to move forward, exploring implications of artificial intelligence that transcend strictly technical issues.

PART 1
ARTIFICIAL INTELLIGENCE AND ITS BENEFITS

Artificial intelligence has been increasingly incorporated into our daily lives, sometimes so naturally that we don't even associate certain utilities with it. The truth is that analog life is being left behind. It is true that there will always be those who prefer objects manufactured in the old style, such as some watches from very expensive brands, which celebrate a handcrafted past, although they are behind, ahead and work much worse than their digital equivalents. But they continue to attract buyers, proving that the human species is not entirely driven by reason and pragmatism. However, barring extravagances and idiosyncrasies, the fact is that today we research any topic using search algorithms. We choose products, reading material, trips and accommodations using recommendation algorithms. We opt for faster commutes or decide what to wear based on intelligent systems for measuring vehicle traffic and temperature. In short, AI brings many positive things that make our lives better and easier.

Considering the large-scale impact of artificial intelligence, there are so many uses and potentialities that it is not even easy to select and systematize them. Below are some significant examples.

---

18 This would be the case with OpenAI regarding GPT-4 (foundational model).

19 In the case of the products referred to in the previous note, the user is the one who generates the text with ChatGPT or consults the voice application, via Virtual Volunteer.

20 In the case of ChatGPT (a special-purpose model), OpenIA is both a developer and an implementer. Be My Eyes, in turn, is only an implementer of Virtual Volunteer (also a special-purpose model), adapted from GPT-4.

21 Who uses text and voice guidance? Note that the news produced may or may not be true. The voice may be good guidance or a deep fake, produced to deceive the recipient.

1. Better decision-making in many areas

In many domains, AI will have a better decision-making capacity than humans, for a variety of reasons. First, because it can store a much larger amount of information than the human brain. Second, because it is able to process it much faster. Third, because it is able to make correlations within a massive volume of data, beyond the capabilities of a single person or even a team. Such correlations can reveal associations between factors that we are not aware of, due to their complexity or subtlety. As already noted, however, the efficiency of AI will depend on the quantity and quality of the data it is fed. Furthermore, in the current state of the art, generative AI tools can produce invented or absurd information, in a deviation known as "hallucination"[22]. It should be noted that in areas that depend on emotional intelligence, ethical values or understanding the nuances of people's behavior, human intervention will be indispensable and their decision-making capacity superior.

2. Automation

AI enables the automation of numerous tasks, both routine and more complex, increasing productivity and efficiency in various areas of activity. Tasks that are repetitive, exhausting or strenuous for humans can be performed by machines, such as in industrial production lines. In addition, the margin for error is reduced and it is possible to eliminate risks in jobs such as mining, disarming bombs, repairing cables on the ocean floor or space travel. In addition, AI can work continuously for 24 hours, every day of the week, producing on a larger scale, with greater precision and at a lower cost. It does not tire, does not get sick, does not change moods and there is no risk of filing a labor claim. The negative impact that all this can have on the labor market will be examined later.

3. Language

The impact of AI on the field of language has been profound and multifaceted, especially through the use of Natural Language Processing. The quality of translations made by Google Translate, ChatGPT and DeepL, to name a few examples, has been significantly improved, making them quite accurate and fluent. This has broken down many of the language barriers in human communication. Tools like Siri, Alexa and Google Assistant respond to voice commands. Other tools transform text into speech. Chatbots help resolve questions and problems for consumers and customers. And generative AI, which has been taking the world by storm, communicates with users through text, sounds and images. The advances in this area are extraordinary.

4. Research and innovation

AI has expanded the frontiers of research and innovation in almost every area of human activity, from physics and chemistry to the automotive and space industries. The volume of science that has been produced based on AI has grown exponentially. AI can simplify and shorten clinical research and testing of new drugs, materials and products. The analysis of vast amounts of data accelerates the process of scientific discovery. It is important to highlight the reduction in cost and time in the

[22] It should be noted that AI hallucinations can have a dramatic impact on relevant social processes. For example, according to a report by the AI Forensics lab, Microsoft's Bing chatbot provided incorrect information in 30% of basic queries about topics related to elections in Germany and Switzerland. It also found that the problem was exacerbated when questions were asked in languages other than English (Guadián, 2024).

development of new drugs, as well as self-driving cars, with the promise of reducing the number of accidents. The expectation is that AI, in its relationship with research and innovation, can help in facing numerous challenges facing humanity, such as climate change, combating hunger, controlling pandemics, and the sustainability of cities and diseases such as cancer and Alzheimer's (European Union, 2023a). The movement that promotes the beneficial exploration of the potential of AI is known as Data for good (Muñoz Vela, 2022, p. 65).

5. Applications in medicine

Medicine is one of the areas where artificial intelligence will have the greatest impact on people's lives and health. Technologies such as machine learning and natural language processing will improve the quality of patient care and reduce costs. Improved diagnostics, image analysis, robotic surgeries, treatment planning and personalization, telemedicine, prediction of future diseases, and management of patient data are some of the many benefits that may arise. AI will not make doctors' work dispensable, but it may change some of the roles they play, transferring responsibilities from the technical plane to the human plane of empathy and motivation. There will also be ethical and legal implications, such as errors made by AI equipment (Davenport; Kalacota, 2019).

6. Applications in the justice system

AI brings the prospect of profound transformations in the practice of law and the provision of jurisdiction. In an environment where precedents are becoming more important, its value for efficient research into case law is enormous. The possibility of lawyers drafting briefs, the Public Prosecutor's Office issuing opinions and judges making decisions based on drafts researched and prepared by AI will simplify life and shorten processing times. Of course, everything will be under strict human supervision, as each of these professionals remains responsible. In courts, AI programs that group cases by subject matter, as well as those that can summarize large cases, optimize judges' time and energy. Likewise, the digitalization of cases – in Brazil, today, almost all cases and their processing are electronic –, the automation of certain procedures and online dispute resolution have the potential to make justice more agile and efficient. In Brazil, within the scope of the country's various courts, there are more than a hundred projects for the use of AI in the provision of jurisdiction.

There is a controversial and particularly interesting point here: the use of AI to support the drafting of judicial decisions. Many fear, not without reason, the risks of bias, discrimination, lack of transparency and explainability. Not to mention the lack of social sensitivity, empathy and compassion. But it is important not to forget that human judges are also subject to these same risks. For this reason, there is another side to this coin: the prospect that AI may actually be more prepared, impartial and less subject to personal interests, political influences or intimidation. This can happen anywhere, but especially in less developed countries, with a lower degree of judicial independence or a higher degree of corruption (Ariel Gustavo, 2021; Sustein, 2022). In any case, at the current stage of civilization and technology, the supervision of a human judge is essential, although an increased burden of argumentation may be imposed on him in cases where he intends to produce a result different from that proposed by the AI.

7. Education and culture

Artificial intelligence will transform the educational landscape in the world, both in teaching methods and in learning possibilities. Initially, the Internet, powered by AI, has exponentially expanded access to knowledge and information, expanding the horizons of all people who have access to the World Wide Web. In addition, distance learning has broken down the barriers of time and space, allowing learning at any time, from anywhere in the world. Digital libraries eliminate the need for physical travel and allow consultation of repertoires located anywhere in the world. From the perspective of teachers, it can help in preparing classes, writing questions and even correcting papers, in addition to performing administrative tasks that free up teachers for more time for academic activities. Everything always, should be reiterated, under human supervision.

From the students' perspective, AI, especially generative AI, facilitates research, can summarize long texts, correct grammatical errors and suggest improvements to writing, in addition to helping to overcome language barriers, as seen above. It also allows for personalized teaching, tailored to the needs of students, including for people with disabilities, as it can, for example, transform text into voice or vice versa[23]. Naturally, to bring benefits that are distributed equally across the population, the use of AI presupposes quality connectivity for all (digital inclusion). Here, we should not rule out some dysfunctions that may arise from the use of AI in education, ranging from plagiarism to limited creativity and critical thinking. For this very reason, international organizations such as the OECD (2023) and UNESCO (2021) have produced relevant documents, with principles and guidelines for the use of AI in education.

The impact of AI on culture will also be immense. On the positive side, it will open up avenues for creativity, in synergy with musicians, painters, writers, architects, graphic designers, and countless other creative agents. Generative AI can assist in the composition of symphonies, literary works, poems, storytelling, etc., increasing creativity and the aesthetic universe, but also raising countless ethical questions about intellectual property and copyright (Beiguelman, 2021, p. 59). Yuval Noah Harari's statement that AI has already hacked the operating system of human culture, which is language, is worth reflecting on. And he asks: what will it mean for human beings to live in a world in which a percentage of novels, music, images, and laws, among many other creations, are generated by non-human intelligence? (Harari, 2023)

8. Other useful applications of AI

8.1 Practical everyday utilities

AI technology is present in personal computers and smartphones in multiple applications, such as Google Maps, Waze, Uber, Spotify, Zoom, Facebook, Instagram. And also in personal assistants, such as Siri and Alexa. AI also plays an important role in the entertainment industry via streaming (Netflix, Amazon Prime, HBO Max) and gaming. Not to mention applications that allow banking transactions and credit card payments, among countless other utilities.

---

[23] In the educational field, AI can also benefit gifted students, given that perception, recognition and recommendation skills allow for the supervision, understanding and adaptation of each student's learning process, in addition to freeing up teachers to spend more time on individual instruction (Lee, 2019, p. 149).

8.2 Environmental protection

AI will play an increasingly critical role in environmental protection, data analysis, forecasting phenomena and monitoring situations. Examples are numerous and include: examining data on climate change, using satellite and drone images, monitoring air, water and soil pollution levels, rationalizing the distribution and consumption of energy and water, predicting natural disasters (such as hurricanes, earthquakes and floods), assisting in sustainable agriculture through soil sensors and other instruments, reducing the use of pesticides, guiding irrigation and helping in planning reforestation[24].

8.3 Personalization of commercial and other relationships

AI enables industry, commerce, services, media and digital platforms to direct information, news and advertisements to their consumers that match their interests. This naturally optimizes people's time and makes it easier to purchase products, books, plan trips and countless other choices and decisions that need to be made. Recommendations for films, music or other forms of entertainment come from this use of artificial intelligence. However, one should not disregard here the negative aspects associated with a certain tribalization of life, due to the confirmation bias resulting from the sending of materials that, in general, reiterate preferences and convictions. This phenomenon reduces the plurality of views, generates new forms of social control[25] and can lead to polarization and radicalism (Barroso; Barroso, 2023). In terms of personal relationships, research shows that marriages resulting from relationships initiated online, with the help of algorithms, have proven to be slightly more satisfactory than marriages in which the partners met through conventional, offline methods. (Tropiano, 2023; Harms, 2013).

It is not the case that we can endlessly list all the uses and benefits of artificial intelligence, which, moreover, are expanding every day. These include the development of autonomous vehicles, monitoring equipment to detect possible infrastructure failures, detecting fraud, especially of a financial nature, improving cybersecurity, aviation controls, etc. It is now time to turn our attention to the problems, risks and threats that may arise from the large-scale use of artificial intelligence.

PART II
ARTIFICIAL INTELLIGENCE AND ITS RISKS

Every new technology has a disruptive effect on production and consumption relations, and on the labor market, impacting social life. Furthermore, like many things in life, innovations can have a negative side or be appropriated by bad social actors. The weaving loom put seamstresses and

---

[24] However, the energy expenditure resulting from the feeding, operation and maintenance of AI, as well as its systemic impacts on different ecosystems, should not be ignored either, as some researchers have already observed (García-Martín et al., 2019, p. 75-88; Centeno et aal. 2021, p. 1-10).

[25] On the use of AI as a form of social control: "Search engines present another challenge: ten years ago, when they were driven by data mining (rather than machine learning), if a person searched for a "gourmet restaurant" and then for "clothes," their last search would be independent of the first. In both cases, a search engine would aggregate as much information as possible and give them options [...]. Contemporary tools, in contrast, are guided by observed human behavior. [...] A person may be looking for designer clothes. However, there is a difference between choosing from a variety of options and taking an action – in this case, making a purchase; in other cases, adopting a political position or ideology [...] – without ever having seen the initial range of possibilities or implications, just trusting a machine to configure the options in advance" (Kissinger; Schmidt; Huttenlocher, 2023, p. 20).

artisans out of work; offset printing eliminated linotype jobs. Computerization reduced the need for bank employees in the financial system. Digital platforms paved the way for extremist polarization (Fisher, 2023, p. 20), disinformation (Kakutani, 2018, p. 17), and hate speech (Campos Mello, 2020; Williams, 2021, p. 207). Even more serious: the invention of the caravel allowed transoceanic trade, but also the slave trade (Acemoglu; Simon; 2023, p. 4-5)[26].

For these reasons, it is necessary to pay attention to the adverse effects of the use of artificial intelligence, seeking to neutralize or mitigate them. Such negative impacts of AI can have social, economic, political implications or even undermine world peace. Below, we list some consequences, risks and threats brought about by artificial intelligence.

1. Impact on the job Market

This is the most obvious and predictable effect, resulting from what normally happens when new technology disrupts the previous mode of production. With the advance of automation, the labor market landscape will change profoundly, requiring workers from different areas of the economy to adapt to new jobs. This transition is not always easy. It should be noted that in the case of AI, the impact will not only be on more mechanical jobs but will also affect more skilled and creative roles[27]-[28]. It is true that new technologies also tend to generate new markets and, consequently, new jobs. However, there is a problem of timing and scale in this consideration. It is unlikely that new jobs will be spontaneously generated at the same pace and volume (Keynes, 1930)[29]. This is an important challenge, which will require governments to invest in social protection and worker training. It is worth remembering that the expansion of economic vulnerability tends to impact on the sphere of democratic protection, given that it has historically emerged as a potential factor of destabilization.

2. Use for military purposes

There is relatively little literature on the use of AI for military purposes, partly because of the secrecy that is normally imposed on the subject for security reasons. However, throughout history, new technologies either originate from research for military purposes or are quickly directed to that end. It is not difficult to imagine countries like the United States and China competing to employ AI for military purposes, using new technologies and robots. In fact, remotely-operated automated drones have been used for some time for this purpose, with reconnaissance, surveillance, equipment delivery or even air strike missions. A topic that has raised great concern is that of autonomous lethal weapons, which can engage in combat and attack targets on their own accord, without human control. There are ongoing debates about strict control of their use by international acts (Klare,

[26] The authors point out some inventions that, in the last thousand years, have not necessarily brought prosperity to everyone.

[27] It is estimated that banks and some technology companies spend 60% to 80% of their payrolls or more on workers who are highly likely to be affected by new technology (Lohr, 2024). In a similar vein, see Maheshwari (2024) regarding the advertising and marketing market.

[28] Other studies indicate that roles that require social intelligence (public relations), creativity (biologists and designers), perception and fine manipulation (surgeons) tend to be more spared (Frey; Ousborne, 2017). They also assess that analysis, forecasting and strategy work will be the most affected (Webb, 2019).

[29] As observed by Keynes (1930) almost a century ago, technological advances tend to generate at least a temporary imbalance in terms of work, until new work opportunities are identified.

2023). The ethical implications of this type of weaponry are dramatic and strict regulation of their use or, perhaps preferably, their banning is imperative.

Furthermore, communication and information technologies have been mobilizing military efforts for some time now, featuring recurring tactics in the context of hybrid warfare. These are new forms of aggression that involve, in addition to destruction by physical means, influence and disinformation campaigns (cognitive warfare), in addition to cyberattacks with the purpose of compromising vital computer systems, such as energy supply structures (Alvim; Zilio; Carvalho, 2023, p. 69).

3. Mass dissemination of disinformation

Since at least 2016, the dissemination of information through digital platforms and messaging apps has represented a serious problem for the democratic and electoral process. Studies document that the circulation of falsehoods and online radicalism occurs at a faster rate and with greater engagement than the dissemination of truthful and moderate discourse. What is emotional, improbable, and alarming produces more engagement and mobilization. Deep fake news makes things even worse, as it simulates people saying things they never said, adulterating content and realities in a way that is imperceptible to citizens (Barroso; Barroso, 2023; Campos Mello; Rudolf, 2023, p. 53-78). This scenario is not hypothetical, and the precedents are worrying. The influence that the dissemination of disinformation had on historical events such as the United Kingdom's exit from the European Union (Brexit), the elections in the United States, both in 2016, and the Brazilian elections of 2018 has become notorious. Democracy presupposes the informed participation of citizens and, naturally, is seriously compromised by the widespread circulation of deliberate lies, destruction of reputations and conspiracy theories.

4. Violation of privacy

The business model of platforms that use AI is based on collecting as much personal data as possible from individuals, which turns privacy into a commodity (Morozov, 2018, p. 36). Based on these data, complex algorithms and multiple neural layers establish deep correlations, which allow obtaining their genetic data, their psychic systems, vulnerabilities, consumer, political, financial, sexual and religious behaviors (Huq, 2020, p. 37). With such data and correlations, AI is able to make predictions, recommendations, manipulate interests and produce the results desired by the algorithm. Therefore, access to private data, from people and companies, is central to the AI business model as currently established (Zuboff, 2022, p. 1-79)[30]. It is no coincidence that in academia, data has been treated as the oil of the current century (Rebollo Delgado, 2023, p. 17).

There are at least three aspects that require attention regarding the topic of privacy. The first is the collection of data from Internet users without their consent by digital platforms and websites. Such information is used for commercial sales, for targeting information and advertising, or even for manipulating the will of users, as neuroscience research shows. A second aspect concerns surveillance and tracking by the government and police authorities, using facial recognition technologies and location tools. Although the legitimate purpose is to combat crime, the risks of abuse are very high.

[30] Many of the other restrictions on rights derive from restrictions on privacy, such as those related to physical, reputational, relational, psychological (emotional), economic, discriminatory and human autonomy-related damages (coercion, manipulation, misinformation, distortion of expectations, loss of control, among others) (Citron; Solove, 2022, p. 793-863; Huq, 2020).

These risks are aggravated in the case of authoritarian governments. Finally, a third point is that AI systems require the collection of vast amounts of data to train their models, with the risk of leaks and cyberattacks by malicious actors, for example, in spear phishing activities (Muñoz Vela, 2022, p. 64)[31] and doxxing (Prado, 2023, p. 162)[32] which often fuel harassment, political violence, malinformation and disinformation practices.

5. Algorithmic discrimination

Algorithms are trained on existing data, which in turn express past and present human behaviors, full of biases and prejudices, deeply determined by historical, cultural and social circumstances (Prado, 2023, p. 162; Horta, 2019, p. 85-122). For this reason, they tend to reproduce current and past social structures of inclusion and exclusion. To this extent, data on employability show a lower hiring of women, blacks and indigenous people, a tendency that is unrelated to their capacity and productivity, but which can lead to the reproduction of future behaviors (Dastin, 2018)[33]; data on public safety record a greater propensity for recidivism and violence involving black people, not necessarily because they are more violent, but possibly because they live in more adverse social contexts (Larson, 2016); data on health costs tend to overestimate the spending of some groups and minimize the spending of others, for reasons not necessarily related to their physical conditions[34]; data on credit risk will increase the risks and, consequently, the financing costs of those with lower economic and social status, even when they have managed to improve their conditions, depending on the circumstances of data collection (Pasquale, 2016). To this extent, it is found that some hiring algorithms may tend to discard women, criminalize black men and make it difficult for the poorest to access credit. In such conditions, the way AI works can be profoundly reinforcing of existing inequalities, to the detriment of the most vulnerable groups in society (Huq, 2020, p. 29-34; Silberg; Manyika, 2019, p. 3).

6. Issues regarding intellectual property and copyright

The AI business model raises important questions about copyright and intellectual property. Who owns the copyright to the vast universe of songs, films, news reports and content collected by big media? tech companies with the purpose of feeding their AIs? To their authors and creators or to those who began to employ and exploit them through algorithms? Generative AI is fed with an incredible amount of data. However, the answers to the questions asked to it come without identifying the source and the author. Discussions on this topic have been getting more heated and have reached the courts. Take the example of the press. The content produced by news companies is collected by AI companies, which use it to train applications that compete with the press themselves

---

31 These are malicious emails or messages, customized for a specific recipient, with an appearance of credibility, aimed at obtaining sensitive information (passwords, for example) or installing malware, which are malicious programs with serious effects on the affected systems.

32 Doxxing means the malicious removal of information about someone, either from public files or by hacking. computers, for the purpose of harassing, intimidating or extorting, among others.

33 In fact, Amazon stopped using a job screening system after it found that the system discriminated against women. The company found that the discrimination stemmed from the fact that the AI system had been trained on a hiring data set collected over the past 10 years, when women were less present in the job market. The system interpreted the reduced presence of women as meaning that hiring men was preferable and discarded female candidates.

34 In this case, the data used to train the algorithm was incomplete. Data on costs for white and black patients were used to estimate the scope of their health needs. The resources used for black patients were lower than those used for white patients, not because their needs were less, but because they had greater difficulty accessing the service. As a result, the needs of black patients were wrongly underestimated by the AI (Obermeyer, 2019).

in the production of information[35]. The issue is the subject of a lawsuit filed by the newspaper The New York Times in the face of OpenAI and Microsoft (Grynbaum; Mac, 2023). A similar lawsuit involves Getty Images, a visual media company and image provider, and Stability AI, an artificial intelligence company (Vincent, 2023)[36].

There is no way to exhaustively explore the risks involved in the development of AI, as there are countless possibilities to be considered, not to mention those that we are not even capable of imagining or anticipating. But there is one last concern that deserves special consideration. It concerns what is called "singularity," a term used to identify the risk of computers becoming conscious, acquiring a will of their own, and becoming dominant over the human condition. This is because, being able to process a much larger volume of data at an equally much greater speed, if they have consciousness and a will, they will become superior to all of us. The fear arises from the fact that AI systems can self-improve, reaching "superintelligence," mastering scientific knowledge, general culture, and social skills that would place them above the best human brains.

Someone skeptical of human potential might even assume that an extra-human superintelligence would have a greater capacity to address some of humanity's great unresolved issues, such as poverty, inequality, or environmental degradation. But one could never know whether this out-of-control intelligence would serve the cause and values of humanity. For this very reason, AI governance, both domestic and international, needs to establish safety protocols and ethical parameters designed to manage and mitigate this risk. If technology can reach this point, many scientists doubt—the very future of civilization and humanity will be at stake.

Yuval Noah Harari (2023) makes an interesting comment on the subject. According to him, in 2022, around 700 of the most important AI scientists and researchers were asked about the dangers of this technology impacting human existence itself or causing significant disempowerment. Half of them responded that the risk would be 10% or more. In light of this, he asks the fateful question: would you get on a plane if the engineers who built it told you there was a 10% risk of it crashing? If that's true, you won't be able to sleep peacefully.

PART III
SOME PRINCIPLES FOR REGULATING ARTIFICIAL INTELLIGENCE

1. Complexities of regulation

From everything that has been explained so far, it is clear that regulating artificial intelligence has become essential. However, the task is not simple and faces challenges and complexities. Below, we seek to identify some of them.

Regulation needs to be done while the train is moving. In March 2023, more than a thousand scientists, researchers, and entrepreneurs signed an open letter calling for a pause in the

[35] On the crisis of the press business model and the impact it has on democracy, see: Minow, 2021, p. 35; Jackson, 2022 p. 280 et seqs; Barroso; Barroso, 2023; Campos Mello; Rudolf apud Cunha França; Casimiro, 2023.

[36] Getty Images claims that Stability used the images it produced to train an image-generating AI system called Stable Diffusion, without authorization, violating its intellectual property rights and the copyrights of its collaborators, with the purpose of offering services similar to its own.

development of the most advanced AI systems, given the "profound risks to society and humanity" they posed. The pause, for at least six months, would be intended to introduce "a set of shared safety protocols" (Future for Life Institute, 2023). The concerns were fully justified, but the suspension of research did not happen. Development continued at a rapid pace. Not least because advances in this area have become the subject of disputes between nations, researchers, and entrepreneurs. The letter, however, reinforced the demands for governance, regulation, monitoring, and attention to the social, economic, and political impacts of new technologies.

The speed of change is staggering. This makes it extremely difficult to predict what is to come and to understand new realities in legal norms, which run the risk of becoming obsolete in a short space of time. It is not difficult to illustrate the point. It took the traditional landline telephone 75 years to reach 100 million users. The mobile phone took 16 years. The Internet, 7 years. Well, ChatGPT reached 100 million users in two months (The Feed, 2023). It is not easy for legislation and regulation to keep up with the pace of innovation.

Risks of overregulation. Regulation has become essential, as noted above, but it also involves risks. Two of them are worth highlighting. The first is that restrictions and civil liability cannot be so onerous as to inhibit the impetus for innovation. Second, disproportionate regulation can create a market reserve for established companies, creating a gap between them and the competition, worsening economic concentration in the hands of a few enterprises. The prevailing conventional wisdom is that regulation should focus on results, not on research itself.

Asymmetry of information and power between companies and regulators. AI technology is controlled, above all, by the companies involved in its development, which have greater knowledge than potential regulators. Added to this is the fact that Big Techs are some of the most valuable companies in the world, enjoying economic power that can easily be transformed into political power. This power was evident when the Brazilian National Congress voted on a bill regulating disinformation on social media. Some technology companies launched an intense campaign against the measure, on their own platforms and in lobbying in the National Congress, managing to have the bill removed from the agenda (Rezende, 2024; Poder 360, 2023a; Poder 360, 2023b)[37].

Need for global harmonization of regulation. AI is a predominantly private technology that does not observe national borders. Companies operate globally and do not usually even have their headquarters in the main centers of their business. Data can be collected and fed into the training of systems in different parts of the world. In such conditions, the way AI operates calls into question some essential elements of law, as we practice it. These elements are: the enforceability of fundamental and human rights against States (and not specifically against private agents) and the scope of national jurisdictions, which are limited by the sovereignty of other countries. In addition, the heterogeneous regulatory treatment of the subject in different countries can lead to investment flight and obstacles to technological development in restrictive States and represent an invitation to widespread violation of rights in more permissive places.

2. Some regulatory efforts

At the international level, some initiatives involving non-binding propositions (soft law) were notable. Among them, the following stand out: a) the Recommendation of the Council on Artificial

[37] It is estimated that there is no great interest in containing fake news or in moderating content. The more fake news there is, the greater the user engagement and the greater the interaction on networks, therefore, the greater the production of data, the raw material for Big Techs.

Intelligence, of the OECD (Organization for Economic Cooperation and Development), of 2019 (OECD, 2019)[38]; and b) the Recommendation on Ethics in Artificial Intelligence, of UNESCO, of 2021 (UNESCO, 2023)[39-40]. Both documents seek to respond to the risks already indicated above, are convergent and complementary and bring together very general principles on AI, to be detailed by the domestic regulations of the respective countries.

At the national level, the United States of America published, at the end of 2023, a long Executive Order (EO) on AI[41]. This is a broad regulation that covers multiple areas of technology risk, through which the US president essentially addressed federal agencies, according to their expertise, ordering them to establish standards and measures to test, ensure the security and reliability of the technology, prevent fraud, prevent algorithmic discrimination, and the violation of fundamental rights of citizens, consumers, competitors, and students. The EO also provided for the identification of content produced by AI with watermarks. It established the definition of good practices and the carrying out of studies on the impacts of AI on labor relations, with measures to mitigate them. It included funding for research and support for small businesses in accessing technical assistance, resources, and the market in AI, as well as the attraction of new talent through immigration measures. It is determined that developers of foundational models that may present risks to national security, the national economy and public health must notify the Public Authorities when training their systems and share with them the results of their security tests (red-team safety tests). And he called on Congress to pass a law protecting the right to privacy and protecting citizens' data.

The European Union (EU) approved the Artificial Intelligence Act (EU AI Act) in March 2024. The regulation proposed within the EU, unlike what occurs with the Executive Order in the US, is characterized by directly establishing rules and sanctions regarding the development, implementation and operation of AI. It also provides for the concentrated action of certain bodies in its monitoring and implementation. Such rules are, however, proportional to the risk posed by the technology (risk-based approach) for people and goods (European Union, 2021, 2023). In this line, systems are classified into three levels: a) systems subject to unacceptable risks, whose implementation is prohibited[42]; b) high-risk systems, whose implementation is permitted, as long as they meet mandatory standards[43]; and c) AI systems that do not pose high risk, for which incentives

---

38 This recommendation was also adopted by the G-20.

39 The document lists the following 10 principles: 1 – Proportionality and Do No Harm; 2 – Safety and Security; 3 – Fairness and Non-discrimination; 4 – Environmental and Ecosystem Protection; 5 – Right to Privacy and Data Protection; 6 – Human Oversight and Determination; 7 – Transparency and Explainability; 8 – Awareness and Literacy; 9 – Responsibility and Accountability; 10 – Multi-stakeholder and Adaptive Governance & Collaboration.

40 The United Nations also adopted Principles for the Ethical Use of AI in the United Nations System, which are quite similar to those included in the UNESCO recommendation (Ceb, 2022).

41 An Executive Order is a normative document, a type of directive, issued by the President of the United States of America, aimed at the management of the federal government. It is similar to a decree in Brazil. It has limits regarding its possibility of standardization, as it is not a law produced by the Legislature and can be changed by decision of the next president.

42 This category includes social scoring technologies, biometric identification in public places for law enforcement purposes (with specific exceptions), as well as subliminal practices of manipulating people and/or exploiting the vulnerabilities of vulnerable groups.

43 This category includes social scoring technologies, biometric identification in public places for law enforcement purposes (with specific exceptions), as well as subliminal practices of manipulating people and/or exploiting the vulnerabilities of vulnerable groups.

are provided for the voluntary adoption of codes of conduct, a type of self-regulation (European Union, 2024).

In Brazil, Bill No. 21/2020 and Bill No. 2,338/2023 are currently being processed in the National Congress, with a tendency to move closer to the standards set out in the European Union's proposed regulations (Brazil, 2023). In general terms, the proposals seek to: a) guarantee rights to people directly affected by AI systems; b) establish responsibilities according to the levels of risk imposed by systems and algorithms guided by this type of technology; and c) establish governance measures applicable to companies and organizations that explore this field.

3. Some guidelines

In light of everything that has been explained so far, it is possible to extract some values, principles and objectives that should guide the regulation of AI, so that these technologies serve the cause of humanity, enhancing their benefits and minimizing their risks. Such regulation should focus on defending fundamental rights, protecting democracy and promoting good governance. Below are some elements and aspects related to each of these purposes.

3.1 Defense of fundamental rights

the) Privacy. The use of AI must respect the individual data of individuals and legal entities, without being able to use them without consent. Invasive surveillance, such as facial recognition, biometrics and location monitoring, must be used in a restricted and controlled manner. And, given the vast amount of data used to feed AI, there must be adequate security mechanisms against leaks.

b) Equality (non-discrimination). The equality of all people, in its formal, material and recognition dimensions, is one of the most valuable pillars of contemporary civilization. The dangers of algorithmic discrimination have already been warned about here. AI regulation must prevent people from being unequally treated based on suspicious categories that exacerbate vulnerabilities, such as gender, race, sexual orientation, religion, age and other characteristics. There is a bad record in this regard (Dastin, 2018; Larson et al., 2016; Obermeyer et al., 2019; Heaven, 2021).

c) Freedoms. With regard to individual autonomy, the use of neuroscience and targeted advertising (microtargeting) has the power to manipulate people's behavior and will, through feelings of fear, prejudice, euphoria and other cognitive biases, inducing them to buy goods, hire services or adopt behaviors that are contrary to their interests, violating their cognitive freedom or mental self-determination. Furthermore, the right to information, pluralism of ideas and freedom of expression can be compromised by recommendation or moderation algorithms, which filter, direct and delete content, in conduct equivalent to private censorship.

3.2 Protection of democracy

the) Combating disinformation. Democracy is a regime of collective self-government, which presupposes the enlightened and well-informed participation of citizens. For this very reason, the circulation of disinformation and conspiracy theories deceives or generates unfounded fears in people, compromising their discernment and choices. As already noted, all of this is aggravated by deep fakes, which simulate non-existent videos and speeches, with the appearance of reality. We are

all educated to believe what we see and hear. Manipulations of this nature break the paradigms of experience (Filimowicz apud Filimowicz, 2022, p. x and xi)[44] and are destructive of democracy.

b) Combating hate speech. Since universal suffrage was historically enshrined, democracy has involved the equal participation of all people. Hate speech consists of attacks on vulnerable groups, racist, discriminatory or ableist statements regarding black people, gays, people with disabilities and indigenous people, among others. By seeking to disqualify, weaken or silence certain social groups, hate speech undermines the protection of human dignity and weakens democracy.

c) Combating attacks on democratic institutions. Social media, aided by AI, has been instrumental in orchestrating attacks on democratic institutions, aiming to destabilize them. Insurrectionary acts such as those on January 6, 2021, in the United States, or January 8, in Brazil, with coup attempts to disrespect the election results, put democracy at risk and cannot be tolerated (Ramonet, 2022).

3.3 Promoting good governance

In light of the aforementioned international, regional and domestic recommendations and normative acts, and the broad public debate currently underway in academia, civil society and the press, it is possible to extract some overlapping consensuses regarding AI governance, outlined in the five guidelines set out below.

the) Centrality of the common good. AI must be developed and oriented towards the well-being of people, countries and the planet. Its benefits must be distributed fairly among all, and its negative impacts must be mitigated through legislation and regulation[45].

b) Plural governance. AI governance must consider, in its different stages, with the appropriate proportionality, the participation of a diverse set of stakeholders, including the government, scientists and researchers, civil society, academia, companies and human rights organizations. The diversity of perspectives and the weighing of values and interests are very important for the legitimacy of decisions and appropriate regulations.

w) Transparency and explainability. Transparency identifies the user's minimum knowledge about how the system works and the information that they are interacting with an AI system. Explainability means making the reasons for the decisions taken intelligible, including allowing possible questioning of the results. Both requirements combine to mitigate concerns about the accuracy and impartiality of algorithms, as well as to encourage the responsible use of automation technologies (Bender, 2022, p. 12).

d) Security. AI systems must be internally secure to prevent errors that produce undesirable results, and they must also be protected against external attacks. Security in the use of AI includes impact analysis, data quality and cybersecurity, and mapping the processes and decisions that make up the AI lifecycle (traceability).

---

44 Accumulated experience leads us to understand videos and photographs as elements that attest to the veracity of information. However, AI based on deep learning technologies allows the production of fraudulent but ultra-realistic media that confuses minds and allows the synthetic production of false representations of reality.

45 Regulation should be seen as a necessary but insufficient condition. In this sense, addressing the risks associated with artificial intelligence goes beyond the legal dimension to also reach other fields, among which ethics applied to economics and programming stand out. For Lucrecio Rebollo: "Conceiving law as the only way to order and equalize digital society is a serious mistake. Law must be, as it has always been, a way to resolve social conflicts with a common good perspective, but in all cases, it requires the collaboration of other areas of knowledge, of all the elements that make up the social structure" (Rebollo Delgado, 2023, p. 52).

e) Control and responsibility. Human supervision or control is essential for AI to operate within the boundaries of legality, ethics and justice. Despite the relative autonomy in its decision-making processes, responsibility will always lie with an individual or legal entity. In the event of improper or malicious use, one or both of them will be subject to civil, administrative and criminal liability.

Conclusion

The role of knowledge is to comfort the afflicted and afflict the comforted (Shedden, 2014)[46]. This article aims to fulfill this role. Artificial intelligence, as demonstrated here, presents potential and risks in almost all areas in which it can be applied. In the political sphere, it can help to improve the representative system and to better capture the feelings and will of citizens. But it can also spread disinformation, hate speech and conspiracy theories, misleading voters, weakening vulnerable groups or spreading unfounded fears, bringing out the worst in people.

On the economic front, AI can help increase productivity in a variety of areas, from agribusiness to industry, and significantly improve the service sector. However, it can also concentrate wealth in the most favored sectors and in the richest nations, increasing inequality in the world. On the social front, it can be an important tool in solving problems related to poverty and unfair inequalities, but it can also lead to unemployment for masses of workers. There are also ethical dualities. A greater understanding of human nature can raise the humanistic or spiritual level in the world, but the loss of the centrality of the human person cannot be ruled out.

In short, we live in an era of ambiguities and decisive choices. In the authors' view, the history of the world has been a constant – although non-linear – flow towards goodness, justice and the advancement of civilization. We have come from times of hardship, human sacrifice and despotism, until we reached the era of human rights. For this reason, it is possible to have a constructive vision and attitude towards artificial intelligence. Without paralyzing fears, but also without naivety or fantasies. We will need legislation, regulation and, above all, education and awareness among scientists, companies and citizens so that we do not get lost along the way. And, as already mentioned, the compass, the direction indicated by the stars, are the values that lead to a good life: virtue, practical reason and moral courage. If we lose the references of goodness, justice and human dignity, then it would be time to let the machines take over and bet that they can do better.

But it doesn't have to be that way. Perhaps, paradoxically, artificial intelligence can help us rescue and deepen our own humanity, valuing empathy, fraternity, solidarity, joy, the ability to love and other attributes that will always differentiate us from machines.

[46] This phrase is a paraphrase of The job of the newspaper is to comfort the afflicted and afflict the comfortable, attributed to a fictional character – Mr. Dooley –, created by journalist Finley Peter Dunne, from the Chicago Evening Post.