

## 2.3 Theoretical notes on the risks, challenges and opportunities of AI at the societal level

*“By far, the greatest danger of Artificial Intelligence is that people conclude too early that they understand it”.*

**Eliezer Yudkowsky**

When the innumerable possibilities resulting from the extensive dissemination of AI throughout the world are discussed, at least two points of view on this phenomenon can quickly be examined:

The first point of view, refers to Artificial Intelligence as an imminently powerful **opportunity** for lifestyle improvement, as it facilitates the performance of activities, allows rapid progress in jobs that require more time and dedication, promotes agile and efficient decision-making, automates repetitive tasks, can diagnose and/or treat diseases effectively, drives innovation and creativity through the design of personalised and customised products and has the potential to increase competitive advantages for companies, among many other functions.

Without denying that this is a false dilemma, AI has shown itself to be categorically helpful and versatile in the management of almost all areas, fields and sectors of human life; and this could be established in previous sections, where the usefulness of this technological tool in areas such as medicine, education, systems engineering and robotics was mentioned. However, this list is much more extensive, as finance, recreation, housekeeping or even law, for example, are areas that have also benefitted from the introduction of intelligent systems. A few other benefits or professions could be listed, and even then, the list would be endless.

Against this backdrop, exploring the field of AI is therefore an element that is of interest to many and should be further investigated, as the fact that intelligent machines have potentially useful effects for large industries and society, and that their competitive advantages are becoming more and more decisive for everyday life, cannot be neglected.

When we speak of Artificial Intelligence as an opportunity, we are not only referring to the possibility of convergence between physical, biological or digital technologies; above all, we are alluding to the fact that AI as it is known today is much more than automating and mechanising human thought; ‘it is, in general, a structure that visualises the usefulness and versatility of technological advances in modern society, which takes on a prominent role in the development, evaluation and promotion of the goals of sustainable development’ [50]. However, it cannot be forgotten that, while there are significant contributions from AI, there can be errors in its processes, as happened, for example, with an AI system that, instead of distinguishing the types of tanks, learned to distinguish the landscapes in which these tanks were located.

The second point of view, however, is identified by those who, without the need to engage in anachronisms, foresee AI as a negative force that alters the historical course of humanity in the current century. Undoubtedly, Artificial Intelligence has ethical implications, as researchers and experts have pointed out that its evolution is not only due to the way in which it changes or improves the way people live, but also that ‘it is an entity that evolves dialectically, has the capacity to act autonomously, to learn and to model itself on the basis of its own individual and collective experience, at a pace that exceeds human capacities’ [2].

This overview shows that the unstoppable technological transformation has been subject to growing global concern, especially because it is a product that, according to experts, is becoming less and less controlled. The great importance that humans have given to Industry 4.0 has also made it necessary to discuss its **risks** or potential threats to the living conditions of humanity; risks that are directly affecting the role played by humans and that have a deeper and broader significance if the changes resulting from digitalisation are taken into account. In any case, these risks:

“are of a very unique nature in that they do not physically compromise our survival in the way that environmental, health or public safety risks do. On the contrary, digital risks affect rights and freedoms and even our political system, since, in addition to the privacy of individuals, they involve freedom of expression, political freedoms and the very functioning of democracy, the principle of equality and, ultimately, human dignity.” [51]

Such an argument envisages a human rights model that is rarely taken into account by companies that create AI systems. This model, although diffuse and superficial, crystallises from a paradigm of protection that is created through international covenants and which, if linked to the development of new technologies, presupposes the acceptance of two elements: The first is that the idea of inclusive innovation for sustainable development must be strengthened and, the second is that, in the human rights model, disruptive innovation must incorporate the principles of algorithmic identity, algorithmic vulnerability and algorithmic dignity; principles aimed at inhibiting the risk to which people are exposed by the digital world.

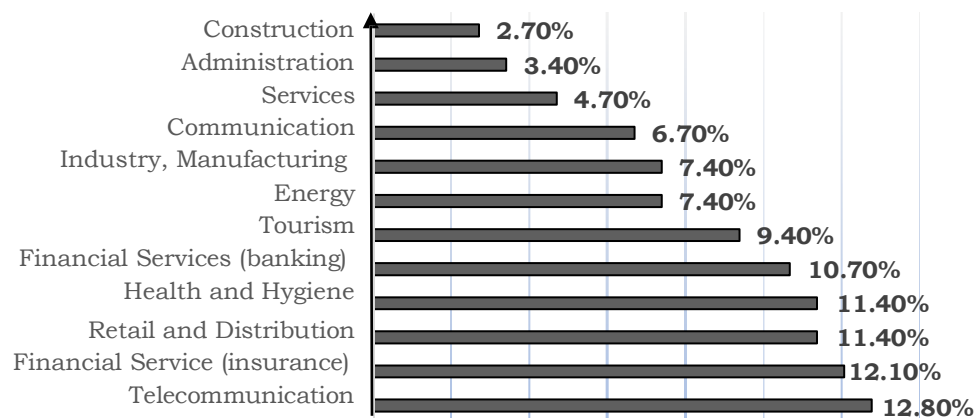
This not only makes the understanding of Artificial Intelligence as a whole more complex, but also calls for specific intervention by public authorities. It could even be said that laws sometimes seem insubstantial to solve the problems arising from the misuse of intelligent systems, which is why real governance of digital risks is also required, as Vida Fernández proclaims in his article ‘The governance of digital risks: challenges and advances in the regulation of artificial intelligence’ [51]. Through governance, the author intends to give meaning to the new initiatives that governments have adopted to manage the risks that new technologies entail and that we human beings are facing. Although it is a broad and expressive concept, his interest is focused on European public power, which is radically different from those established in countries such as China, the United States or Russia.

When the risks of AI are mentioned, they implicitly allude to the possibility of future harm or damage to individuals, and in this sense, we can highlight a few examples:

- Individuals may see the security and privacy of their personal data threatened.
- There is the possibility that in the future there will be a significant loss of interpersonal interaction. The development of this technology is likely to increase isolation and loneliness.
- Large industries foresee problems inherent in technology and the implementation of intelligent machines in everyday life.
- The ease of access to such systems (e.g. Open AI's GPT Chat in all its versions) has allowed malicious entities and individuals to misuse them. So much so that, ‘the European Union's police cooperation body EUROPOL, detailed how Artificial Intelligence language models, which are set to revolutionise technology and the global economy in the near future, can also fuel fraud, cybercrime and terrorism’ [52]. Just a few days after the

launch of this system, it was possible to anticipate the impact that its use could have on society, as it was demonstrated that, although the system refuses to comply with certain orders, some people have found a way to avoid the chat's content filters; therefore, the greater the evolution of the functions and services of AI, the greater the likelihood that it will be used for illegal activities.

- Currently, there is a concentration of wealth in multinationals that employ AI systems in their manufacturing processes. With this, the potential risk is the limitation of other companies or states regarding access to the technological advances they implement; exposing a monopolised environment that inhibits the free market.
- One of the most well-known inherent risks of disruptive innovation is the loss of jobs, which calls into question the fact that machines could replace humans at any time. In fact, the Ministry of Economic Affairs and Digital Transformation publicised a list of sectors with the greatest expected impact of AI in the short and medium term, as shown in the graph below:



*Diagram 7. Sectors with Highest Expected Impact of AI*  
*Source: Adapted from Ministry of Economic Affairs and Digital Transformation [49]*

Faced with this scenario, people are being forced to re-educate themselves professionally, given that the professional and educational backgrounds demanded by the market will be different, and with this, 'analysts predict that, within 20 years, 4 out of every 10 jobs will be affected by the robotisation and automation of certain tasks' [17].

- While one of the greatest limitations of AI is the absence of common sense, the machine-human interrelationship, as well as its boundaries, need to be carefully developed from a foundation that incorporates the ethical component, in order to prevent machines from making the same or more serious mistakes as people.

"we must rethink the unstoppable digitalisation process we are undergoing, the intensification of which in recent years has exponentially increased the level of risks involved. It is an irreversible process that is set to increase as all states, companies and other private subjects make their growth and development dependent on digitalisation." [51]

However, despite the fact that these risks are constantly highlighted by governments and other entities, people still do not assimilate them or perceive them as real dangers. 'In this sense, it should be remembered that, in the digital sphere, there have been disasters that have revealed very serious damage on a global scale, such as the NSA Prism Surveillance system<sup>40</sup> revealed by Edward Snowden or the Cambridge Analytica case with Facebook'

[51] and it is hoped that, in the face of these cases, there will be at least a similar awareness to that of environmental protection and care.

In this vein, it is important to note that, given the presence of Artificial Intelligence in our lives, **challenges** relating to the labour and gender gap that currently exists are also postulated, as it has been arduously recognised that both Industry 4.0 and AI are driven to a greater extent by men. Women, in this particular scenario, account for no more than 35% of enrolments in careers such as mathematics, technology, science or engineering (STEM careers<sup>41</sup>), according to data supported by UN Women [53]; drastically reducing the contributions and possibilities for future innovation that this gender can bring.

Another challenge that is also worth highlighting is related to the strong possibility that, with the constant use of AI, geopolitical inequality gaps in the world will become more and more accentuated. There is no doubt that different countries have placed bets on cutting-edge technology in the field of AI, as well as on the constant training of their professional workers in the field; a situation that has led to competition between great superpowers (China, USA) for leadership in the development and research of Artificial Intelligence. The European Union, although in the background, has also made efforts to gain a place and recognition in this field.

While these challenges appear to be a dark omen, prominent celebrities and experts such as Bill Gates, Stephen Hawking and Elon Musk have requested a cautious approach to AI. From their perspectives, this technology inexorably represents an inevitable risk for years to come, as they have seen AI labs venture to realise, deploy and develop 'digital brains' that are increasingly more powerful than anyone else, and which 'not even their creators have been able to understand, predict or reliably control' [54]. Among their demands, primarily, is the intention of a pause until strong regulatory frameworks and authorities are established worldwide. They also expressly request developments that help distinguish the real from the artificial.

Elon Musk, for his part, has constantly stressed that, although AI will continue to advance in leaps and bounds over the next 20 years, it is humanity that must decide in which direction to steer its development, because although it may seem harmless, it can also be manipulated and used maliciously.

Therefore, it is important to consider two interrelated aspects: on the one hand, if experts on the subject are currently reflecting on human intervention in algorithms, how much human intervention is necessary or relevant for AI systems to be respectful, legitimate and to promote Human Rights? And, on the other hand, how could companies or States guarantee human intervention in the decisions that an intelligent machine decides to take? Prospectively, these questions invite us to reflect on the challenges that Artificial Intelligence

---

<sup>40</sup> National Security Agency /USA.

<sup>41</sup> Acronym: Science, Technology, Engineering, Mathematics.

is currently presenting and that also have to do with our identity as a species; since, as Corvalán rightly points out:

“If human beings are characterised by diversity, randomness and imperfection, we are entering an era of automation that could bring these traits into crisis. As unlikely as it may sound, in the not too distant future, it will be essential to think seriously about guaranteeing a fundamental right, which could be the cornerstone of the age of Artificial Intelligence: the right to the random and imperfect diversity inherent in human beings.” [55]

Undoubtedly, humanity is in the presence of technologies that disrupt its tranquillity. Today, issues such as discrimination in employment, housing and credit, as well as the high dependence of people on technology, have gone beyond the limits, which has undoubtedly generated great social inequalities around the world. It is not only about the elimination of jobs, limited access to AI technologies or the concentration of power in some companies; this picture is much broader and also has to do, for example, with biases in the selection of personnel or in the decisions that areas such as human resources make about their workers. Although it seems straightforward and lacking significant meaning or repercussion, this scenario has perpetuated discrimination in various fields and has markedly increased unemployment and social inequality around the world.

In this case, and wishing to go deeper into the proposed research question: Does the use of Artificial Intelligence by human beings today represent a complementary device or an antagonistic technology, it is highlighted that, although there are opposing positions regarding the use of AI, this type of technology will continue to be a truly useful tool for society as a whole and although this question deserves serious study, Artificial Intelligence is the cornerstone in the next stage of evolution.



*Illustration 9. AI: A New Way of Thinking and Evolving.  
Source: Taken from Martínez [56]*

The global concern that exists regarding the potential substitution of human beings is, although unwanted, a development that must be faced due to the simple fact that the tasks and activities that people carry out must be modernised, must be facilitated and must come to question our existence, because only in this way will human beings have the capacity and need to develop new technologies that mitigate or limit the functions that have

been damaging their role. It is also important to highlight that:

Today, machines are already beginning to learn from their own mistakes and from human behaviour in order to make decisions, and their 'rational' processes increasingly resemble the way the human brain does when faced with a common problem. However, the human brain is still far ahead of computers. A person's behaviour or the way they solve problems often does not follow logic, and the decision-making path they follow can be highly flexible, unlike an algorithm-based programme, no matter how complex. [10]

This being so, there are fundamental reasons to recognise the benefits that Artificial Intelligence has been bringing to society as a whole, but there are also plenty of reasons to fear the progress of machines, so it is necessary to take into account that:

“nothing is certain and can only be speculated. In any scenario, the transformations that will materialise in the coming decades will not be easy and will bring with them changes in people's identities that could transform their genetic integrity and, therefore, significantly alter their way of being, doing and living together with the way they feel and live their worlds of life, which is fixed at all times by the limitations and subjectivities inherent to the human condition at this stage of its historical development.” [2]

Finally, both positions will not only continue to challenge the functionality of AI, but are also valid, and it can be clearly stated that Artificial Intelligence is both a complementary device and an antagonistic technology, as there is no single established position to date.

## 5 Bibliography

### Books:

- [2] Ortiz Ríos, A., Moncada Niño, Á., Contreras Fernández, A. B., Hoyos, C. A., Álvarez Calderón, C. E., López Cortés, D. E., Puentes Becerra, J. A., Cerón Rincón, L. A., Parra, O. I., and Ramírez Pedraza, Y. E. *Enfoques y Gestión en Seguridad Integral*. Comercializadora Comsila SAS, 2020.
- [10] Jones, T. *Artificial Intelligence: A Systems Approach*. Jones and Bartlett Publishers LLC Canada, 2009.
- [17] Villena Román, J., Crespo García, R., and García Rueda, J. J. *Historia de la Inteligencia Artificial*. Universidad Carlos III de Madrid, Madrid, 2023.

### Journal Articles:

- [50] Vida Fernández, J. 'La Gobernanza de los Riesgos Digitales: Desafíos y Avances en la Regulación de la Inteligencia Artificial'. *Cuadernos de Derecho Transnacional*, 14(1), 2022, pp. 489-503.
- [54] Corvalán, J. G. 'Artificial Intelligence: Challenges and Opportunities - Prometea: The First Artificial Intelligence of Latin America at the Service of the Justice System'. *Revista de Investigación Constitucional*, 5(1), 2017, pp. 295-316.
- [56] Vadillo Bueno, G. 'Futuros de la Inteligencia Artificial'. *Revista Digital Universitaria de la Universidad Nacional Autónoma de México*, 21(1), 2020, p. 12.

### Reports:

- [49] European Economic and Social Committee. 'Dictamen del Comité Económico y Social Europeo sobre la "Inteligencia Artificial: Las Consecuencias de la Inteligencia Artificial para el Mercado Único (Digital), la Producción, el Consumo, el Empleo y la Sociedad"' [Artificial Intelligence: The Consequences of Artificial Intelligence on the (Digital) Single Market, Production, Consumption, Employment and Society]. Official Journal of the European Union, 2017.

### Online Sources:

- [51] Perfil. 'Europol Advierte que ChatGPT ya Está Ayudando a Cometer Delitos'. *Perfil*, 2 April 2023. [Online]. Available at: <https://www.perfil.com/noticias/economia/europol-advierte-que-chatgpt-ya-esta-ayudando-a-cometer-delitos.phtml>. [Accessed 11th April 2023].
- [52] ONU Mujeres. 'Necesitamos Más Mujeres en Carreras STEM'. ONU, 11th February 2022. [Online]. Available at: <https://lac.unwomen.org/es/stories/noticia/2022/02/necesitamos-mas-mujeres-en-carreras-stem>. [Accessed 12th April 2023].
- [53] Revista Semana. 'Elon Musk y Cientos de Expertos Están Preocupados por la Inteligencia Artificial, Piden una Pausa'. *Revista Semana*, 29 March 2023. [Online]. Available at: <https://www.semana.com/tecnologia/articulo/elon-musk-y-cientos-de-expertos-estan-preocupados-por-la-inteligencia-artificial-piden-una-pausa/202331/>. [Accessed 29th April 2023].
- [55] Martínez, D. 'Ventajas y Desventajas de la Inteligencia Artificial | Beneficios vs. Riesgos'. *Futuro Eléctrico*. [Online]. Available at: <https://futuroelectrico.com/ventajas-y-desventajas-de-la-inteligencia-artificial/>. [Accessed 15th April 2023].