The EU Trustworthy Artificial Intelligence: How the EU Wants to Regulate the Artificial Intelligence Practices

Miguel Adolfo Rodríguez Cuadros
migrodcuadros@gmail.com
Universidad de Lima, Perú
Johnson & Johnson Switzerland

Recibido: 11/8/2021 Aceptado: 20/9/2021
doi: https://doi.org/10.26439/ciis2021.5585

Abstract. The European Union (EU) recently proposed in April 2021 an Artificial Intelligence legal framework. The EU highlights the advantages of using Artificial Intelligence (AI) especially in prediction, optimizing operations, resource allocation and personalizing service delivery. Nevertheless, the EU considers that AI might bring new risks and negative consequences for individuals and society when an AI system violates the EU values and fundamental rights.

This paper explains how the EU is proposing a legal framework for trustworthy AI. Tech or nontech companies and professionals should know what kind of AI applications and practices might be prohibited or restricted within the EU. More specifically, this paper focuses on what type of AI system can be considered unacceptable risk, high or low risk AI.

KEYWORDS: European Union / EU / Artificial Intelligence /AI / trustworthy AI / fundamental rights

LA CONFIABLE INTELIGENCIA ARTIFICIAL DE LA UE: CÓMO LA UE QUIERE REGULAR LAS PRÁCTICAS DE INTELIGENCIA ARTIFICIAL

RESUMEN. La Unión Europea (UE) ha propuesto recientemente, en abril del 2021, un marco jurídico para la inteligencia artificial (IA). La UE destaca los beneficios de usarla, especialmente en áreas de predicción, optimización de operaciones, asignación de recursos y servicios personalizados de entrega. Sin embargo, la UE también considera que la inteligencia artificial puede conllevar nuevos riesgos y consecuencias negativas a los individuos y a la sociedad cuando, por ejemplo, esta viola los valores y derechos fundamentales de la UE.

El propósito de este coloquio es explicar cómo la UE está proponiendo un marco jurídico para una IA confiable. Las compañías o profesionales en tecnología deben tener conocimiento sobre qué clase de aplicaciones y prácticas de IA puedan estar prohibidas o restringidas en la Unión Europea. Este coloquio les proporcionará información acerca de cuándo una IA puede ser considerada: (a) un riesgo inaceptable; (b) un riesgo alto; o (c) aplicación o herramienta de bajo o mínimo riesgo.

PALABRAS CLAVE: Unión Europea/ UE / inteligencia artificial / IA / IA de confianza / derechos fundamentales

INTRODUCTION

AI system can be defined as software with the particular ability to set human-defined objectives, to generate outputs such as content, predictions, recommendations, or decisions with influence to the environment with which the system interacts, be it in a physical or digital dimension.

There are three ways in how AI will reshape global technology. One way is that AI will exceed human intelligence and escape human control. The second way relates to the new industrial revolution where machines will disrupt and replace the human workforce in every or almost every area of society, e.g., transportation, military, healthcare. The third way focuses on surveillance, namely how the government will monitor, understand and control their citizens and foreigners (Wright, 2018).

China and other countries like Thailand, Vietnam, Sri Lanka, Russia, Zambia and Zimbabwe have started to use AI as tool to monitor their citizens. The Chinese government (directly or through stated own-private companies), for example had begun to build a digital authoritarian state, by using surveillance and machine learning tools and creating a so-called "social credit system".

The Chinese AI company Yitu sold in 2018 wearable cameras with artificial intelligence-powered facial-recognition technology to Malaysian police department (Wright, 2018). Another example is the company BGI, a Chinese genome-sequence company (started as a government funded research group), that cooperated with 50 foreign laboratories to support foreign governments for the COVID-19 virus tests. Although this can be a legitimate way to support and fight against the COVID-19 pandemic, BGI runs China's national library of genomics data. Some foreign DNA might end up in the BGI repository, especially as China has been collecting the DNA data from foreign minorities groups, such as Tibetans and Uighurs (Darby & Sewall, 2021)

By following the current Chinese approach, the AI is being using as a social control tool, that might draw data from several devices (computers, smartphones, smart tv, emails, video-cameras) and different sources e.g., tax returns, criminal records, medical records, bank statements, sexual health clinics, genetic screenings, physical information (location, biometrics, CCTV with facial recognition software) and biotechnology (Wright, 2018).

Biotechnology is an area that developed countries and their multinationals are increasingly investing in. Along with big data and powerful computers, researchers have mastered the gene editing tool CRISPR[1] with the support of AI, allowing them for example to grow

1 "CRISPR" stands for "clusters of regularly interspaced short palindromic repeats". CRISPR is a technological tool for editing genomes. It allows researchers to easily alter DNA sequences and modify gene function, including correcting genetic defects, treating and preventing the spread of diseases and improving crops. However, its promise also raises ethical concerns.

wheat that resists plagues or encode the DNA of bacteria and viruses to produce drugs and vaccines. Despite of these positive results, there is a sensitive and controversial side of CRISPR technology, which relates to human DNA manipulation. Indeed, He Jiankui, a Chinese scientist, produced in 2018 genetically altered babies in a way to develop resistance to HIV and "spare the babies the possibility of becoming infected with HIV later in life" (Normile, 2019).

Regarding the United States AI approach, in 1983 the U.S. military's research and development (R&D) department started a ten-year USD 1 Billion machine intelligence program, which focused to keep the U.S. ahead from their technological rivals (at that time Japan and Germany). The program failed and the AI – machine learning research was put on ice.

In 2020, there is growing sense of urgency around AI in the United States, especially in the Defence sector. Indeed, the U.S. government wants to incorporate AI into the U.S. military, however it will require disruptive changes and investment in new AI applications and algorithms to the existing processes (Ciocca *et al*, 2021).

For example, the U.S. government will need to manage expectations about what AI can do for the military sector, particularly by making clear that AI would not replace humans but rather enhance their capabilities. They need to create algorithms more like computing power and less than science-fiction substitute weapons. More important, the U.S. government needs to modernize their defence infrastructure (hardware and software) and improve AI literacy among policymakers and government officials. They are lacking common programming languages used in the private sector and making it difficult to update or make them suitable for AI tools. There is also a deficit of technical and legal framework understanding about AI (Ciocca *et al.*, 2021).

The approach of the EU with regard AI is to work on a legal framework to become a global leader in the development of secure, trustworthy, and ethical artificial intelligence, that ensure the protection of ethical principles and fundamental rights (European Commission, 2021).

The EU AI legal framework sets harmonized rules for the development, placement on the market and use of AI systems in the European Union by following a risk-based approach. In other words, the EU wants to prevent harmful AI practices by regulating "prohibited artificial intelligence practices", "high-risk AI systems" and "transparency obligation for certain AI systems". (See figure 1).

Figure 1. The current AI industry landscape (China, EU, United States)

In the next section, we are explaining the scope of the EU AI legal framework (the so-called Artificial Intelligence Act), in particularly which kind of AI tools, application or system might be prohibited, restricted in accordance with the EU fundamental rights.

2. AI BASED ON FUNDAMENTAL RIGHTS

The proposal of the EU to regulate the use of AI focuses on protecting fundamental rights. The use of AI has specific characteristics, such as opacity, complexity, dependency on data, autonomous behaviour, that can negatively affect fundamental rights.

Without deep diving into the legal analysis of these fundamental rights, the EU is proposing to regulate the use of AI based on the following rights:

- Human dignity
- Respect for private life and protection of personal data
- Non-discrimination
- Equality between woman and men
- Freedom of expression and assembly
- Right of defense and presumption of innocence
- Worker's right to fair and just working conditions
- High level of consumer protection and environmental protection and improving of the quality of environment.

Although the EU proposal relies on the protection and promotion of the above fundamental rights, it is important to mention that the same proposal is imposing restrictions to

other fundamental rights, for example on the freedom to conduct business and freedom of art and science. This is to ensure compliance with overriding reasons of public interest, such as health, safety, consumer protection and protection of other fundamental rights, when high-risk AI technology is developed and used.

To protect and promote fundamental rights, tech or non-tech companies that use or develop AI tools or systems will need to comply with transparency obligations. These obligations will not affect the right to the protection of intellectual property, because they will be limited only to minimum information for individuals to exercise their right to a remedy, and to the necessary transparency towards supervision and enforcement authorities. An important point of this EU proposal is that when public authorities must give access to confidential information or source codes, they are placed under binding confidentiality obligations.

In general, the development and use of a trustworthy AI tool, software or system should not violate the EU fundamental rights. Tech or non-tech companies, data science or programming specialists, tech and non-tech professionals should be aware about what kind of tools or software will be prohibited, restricted, or permitted. For this, the EU proposal follows a risk-based approach to categorize AI practices. These risk categories are unacceptable risk AI, high-risk AI or low-minimal risk AI. (See figure 2)

Figure 2. The EU trustworthy artificial intelligence legal framework

3. DISCUSSIONS

Prohibited AI or unacceptable risk AI

Under this category, the EU is prohibiting the development of AI practices that have the potential to manipulate persons through subliminal techniques beyond their consciousness

or exploit vulnerabilities of vulnerable groups such as children, persons with disabilities or elderly persons.

AI tools, applications or software that can cause psychological or physical harm to persons, environment or animals will be also considered as prohibited AI.

AI-based social scoring for general purposes performed by public authorities is prohibited. There is currently no official or standard definition of social scoring, but it can be understood as the analysis of massive behaviour, physical or health characteristics or actions of persons through algorithms and rate them for awarding, accessing, controlling, clustering and/or penalizing purposes.

Likewise, the EU is prohibiting the use of real time remote biometric identification systems in public spaces for law enforcement purposes, unless specific and limited exception apply, such as, targeting potential victims of crime, including missing children, prevention of a specific, substantial, and imminent threat to the life or physical safety of natural persons or of a terrorist attack, detention, localization, identification, prosecution of criminals.

High-risk AI

In general, high-risk AI are those that create high-risks to health, safety, or fundamental rights of natural persons. High-risk AI will be permitted in the EU if they meet mandatory requirements and a previous conformity assessment.

The EU classifies high-risk AI in two main categories: AI systems to be used as safety component of products and stand-alone AI systems. The first one refers to AI tools that are physically integrated into the product (embedded), while the second one works independently without being integrated in the product (non-embedded).

This classification is important, because the EU wants to apply the regulation to high-risk AI systems regardless their location. In other words, the EU regulation would be applicable to providers and users of high-risk AI systems that are established in a third country (e.g. China, U.S), provided that the output produced by those AI systems is used within the EU.

Examples of high-risk AI systems are autonomous robots for manufacturing or personal assistance and care. AI systems used in the health sector to perform diagnostics, operations or support human decisions (e.g., cancer treatment, radiography) are also considered as high-risk AI systems. Other examples of high-risk AI systems are:

- Real time and post remote biometric identification systems.
- AI systems used in education or vocational training to determine access or assigning persons to education or vocational trainings or to evaluate persons on tests as a precondition for their education.

- AI systems used in employment, employee management and access to self-employment for the recruitment and selection of persons for making decisions on promotion and termination, monitoring, or evaluation of persons in work-related contractual relationship are also considered as high-risk AI.
- AI systems used to monitor the performance and behavior of workers
- AI systems to evaluate the credit score or creditworthiness of natural person.
- AI systems used in migration, asylum, and border control management.
- AI systems intended for the administration of justice and democratic processes.
- AI systems intended to assist judicial authorities or support on the application or interpretation of law or facts

Low-minimal risk AI

Low-minimal risk AI are those systems that do not have a significant risk to health, safety, or fundamental rights of natural persons. These systems are mostly related to purely ancillary administrative activities tasks, translations, data analytics, anonymization of documents, communication between persons, resource allocation, process optimization, chatbots (when the user knows that a machine is answering).

In general, low-minimal risk AI systems relates to tools, applications or software used by companies and their processes, provided that these systems do not affect the health, safety, or fundamental rights of natural persons. (See figure 3)

Figure 3. The EU trustworthy artificial intelligence legal framework: Prohibited AI, High risk AI, Low minimal risk AI

4. CONCLUSIONS

AI is reshaping the global order. On one hand, China and other countries are having an authoritarian AI approach that focuses currently on citizen surveillance and big data control. On the other hand, the United States is focusing on implementing AI systems in their military and defence infrastructures, but they are lacking a suitable AI framework and infrastructure.

Unlike the U.S. and China, the EU is working on legal AI framework aimed to protect the interest and fundamental rights of EU citizens and companies. This approach promotes the development of AI technology and systems by humans and for humans and the environment. Although, this proposal is still pending to becoming a binding regulation within the EU, data science or software developers professionals should be ready to understand and include it in their IT projects. AI and algorithm compliance regulations and non-tech professionals should start learning and understanding more about AI technology.

REFERENCIAS

Ciocca, J., Horowitz, M. C., & Kahn, L. (2021). The Perils of Overhyping Artificial Intelligence: For AI to Succeed, It First Must Be Able to Fail. *Foreign Affairs.* https://www.foreignaffairs.com/articles/united-states/2021-04-06/perils-overhyping-artificial-intelligence

European Commission. (2021). Proposal for a Regulation of the European Parliament and the Council Laying Down Harmonized Rules on Artificial Intelligence (artificial intelligence act): (COM/2021/206 final). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

Darby, C., & Sewall, S. (2021). The Innovation Wars: America's Eroding Technological Advantage. *Foreign Affairs.* https://www.foreignaffairs.com/articles/ unitedstates /2021-02-10/technology-innovation-wars

Normile, D. (2019). Chinese Scientist Who Produced Genetically Altered Babies Sentenced to 3 Years in Jail. *Science.* https://www.sciencemag.org/news/2019/12/chinese-scientist-whoproduced-genetically-altered-babies-sentenced-3-years-jail

Wright, N. (2018). How Artificial Intelligence Will Reshape the Global Order. *Foreign Affairs.* https://www.foreignaffairs.com/articles/world/2018-07-10/how-artificial-intelligence-willreshape-global-order