Cureus Journal of Computer Science

Open Access Review Article

Towards a Harmonized Regulation of Artificial Intelligence: An Analysis of Its Main Implications and Risks

Received 06/24/2025
Review began 07/15/2025
Review ended 08/19/2025
Published 09/03/2025

DOI:
https://doi.org/10.7759/s44389-025-07546-x

José J. Piña Mondragón [1, 2] , Angelina Espejel-Trujillo [3, 2]

1. Digital Governance, Research Center in Geospatial Information Sciences (CentroGeo), Queretaro, MEX 2. Directorate for Support for the Consolidation of the Scientific and Humanistic Community, Secretariat of Science, Humanities, Technology and Innovation, Mexico City, MEX 3. Computational Geointelligence, Research Center in Geospatial Information Sciences (CentroGeo), Queretaro, MEX

Corresponding author: Angelina Espejel-Trujillo, aespejel@centrogeo.edu.mx

Abstract

Currently, artificial intelligence represents a disruptive technology with an extensive range of applications that has become popular due to the efficiency and speed with which it performs analysis, obtains results, and optimizes processes of diverse nature. As an ally of humans, it has proven to be exceptional. However, its application also generates controversy due to the nature of its use and development, as well as the type and origin of the data used to train it. This has raised risk scenarios for the dignity, integrity, and protection of the fundamental rights of individuals that must be regulated through a solid legal framework that considers aspects of governance, security, and ethical implications derived from its implementation. This paper addresses the origins and concept of artificial intelligence, some of its applications and general benefits, some of the potential risks derived from its application, as well as an updated analysis regarding the status of some of the main initiatives that have emerged in the international arena involved in its regulation with the aim of identifying common approaches, guidelines, and principles that are being considered by governments and that can establish guidelines towards a harmonized regulation of AI for the countries that cannot be postponed.

Categories: Ethical AI and Responsible Technology, Ethical and Social Implications, AI Ethics and Fairness
Keywords: artificial intelligence, governance, ethics, regulation, risks

Introduction And Background

Currently, artificial intelligence (AI) is increasingly immersed in our lives, not only in the professional field, where it is a very useful tool that streamlines activities and optimizes processes, but also in our leisure time and daily activities. There are multiple applications of AI for a wide variety of fields, such as medicine, agriculture, economics, sustainable development, cybersecurity, justice, etc., in which it provides multiple benefits, such as reducing risks and human errors, optimizing results and automating processes, etc. However, the development of AI has advanced at such an accelerated pace and has generated so many expectations that the risk scenarios derived from its application, use and development have been relegated.

Indeed, more and more risk situations derived from the use of this disruptive technology are being reported, from road accidents to manipulation and misinformation, work biases and gender and racial discrimination resulting from decision-making by AI algorithms [1]. This situation is becoming increasingly complex due to the lack of clarity regarding the scope and limitations of AI systems, and who should be held responsible in the event of a failure or breakdown that violates human dignity, integrity, and rights. For this reason, it is urgent to establish a harmonized global regulatory framework and governance authorities to guide the development, use, and application of AI technologies from an ethical perspective.

The main objective of this article is to analyze some of the key initiatives and regulatory frameworks that have emerged in the international arena, which consider governance aspects, security, and ethical implications of AI, to identify the main approaches, guidelines, and common principles that allow establishing guidelines towards a harmonized regulation of AI in the global level.

Finally, the research is organized into three sections: In the first, introduction and background, we present the origins of AI as an area of study and the main purpose behind its development as well as some of the main points of debate that have arisen around its definition, then we provide an overview of its main applications and benefits as well as some of the main risks derived from its development and implementation. The review section refers to the most current international regulatory initiatives that have emerged from international organizations such as the United Nations (UN), the Organization for Economic Cooperation and Development (OECD) and the European Union (EU), as well as some of the

most relevant government initiatives of those countries that show the clearest leadership in the development and implementation of AI, such as China, the United States and Japan, and finally, the conclusions.

Origins of AI

The emergence of AI can be approached from different perspectives. In this case, we have chosen to rescue briefly (Table *1*), some episodes of history where some relevant contribution or discussion related to the controversy of the emulation of the human mind is manifested, considering that there are multiple technological contributions that would not be reached to address. However, we refer to those we consider important to identify the risk scenarios and the most illustrative regulatory guidelines that allow understanding why it is so difficult, but important within the area, to reach a consensus on its definition, since any initiative related to AI must start from a conception of the systems and technologies it intends to regulate.

Conceptual foundations of AI	
Ada Lovelace (1840)	•She claimed that a machine would be able to perform any kind of task that was commanded or programmed. •Ada Lovelace's objection: Such a machine would not be able to think autonomously or generate anything new that had not previously been performed by a human being.
Alan Turing (1950)	•Worked on the possibility of creating intelligent machines. •Through the "Turing test", he proposed an evaluation to determine the degree of "intelligence" of a machine. •He considered how to incorporate intuition and ingenuity (mind and ability) in a machine. •He favored the idea that powerful ingenuity can replace intuition in complex problems, which led to reducing and limiting human intelligence to a mere question of problem solving. •Researchers such as François Chollet (former senior engineer at Google) criticized that position, although it is the prevailing one to this day.
Dartmouth Conference (1956)	•The idea of a super-intelligent machine and the term singularity (stating that, at some point, machines could become the most intelligent entities on the planet) were raised. •John McCarthy officially coined the term Artificial Intelligence as "the discipline within Computer Science or Engineering concerned with the design of intelligent systems". •Three objectives were set to be achieved in a couple of months: 1. To describe all the aspects of learning so that a machine can imitate them. 2. To find out how to make machines use natural language and solve problems reserved for humans. 3. Let the machines perfect themselves. •Despite the advances in the area of Natural Language Processing, the goals were too ambitious and unrealistic.
Present	•Breakthroughs in technologies such as machine translation and hardware emerged. •Despite the most modern technological innovations, the goals set at the Dartmouth conference have not been achieved.

TABLE 1: Key historical milestones in the evolution of the concept of AI

Source: [2,3]

As shown in Table *1*, Ada Lovelace's predictions already raised questions, such as how capable is an AI system of emulating the human mind, will it only perform calculations, or will it also be able to think like us? Although Ada Lovelace's interest at the time was purely technological, these dilemmas persist today. Almost a century later, Alan Turing again addressed the subject of the machine beyond the technological perspective. The idea of powerful ingenuity would later culminate in weak AI applications, as they only solve specific questions, as there is no AI that emulates a general intelligence. By the 1950s, inordinate expectations of AI were raised (Table *1*). The reality is that it turned out to be a formal area of study with challenges beyond those expected. Although they highlight the increase in storage and computational capacity, they in no way provided greater intelligence. Again, Turing's error was present, i.e., limiting the term "intelligence", which goes beyond problem solving. This has certainly not been achieved today, despite the great technological advances, as Ada Lovelace asserted from the beginning.

The concept of AI

Due to its origin and Turing's description, it is difficult to establish a concept. In principle, it would have been better to name it differently. Currently, there is no consensus regarding its definition, so its objectives and limits have not been clearly identified, which is why it is so complicated to understand what it does and to define its scope. What is known is that its reason for existence lies in imitating the activity of the human mind through specialized algorithms. Thus, leaving aside the fact that the term "intelligence" has been controversial since its origin, we will try to provide a simple explanation based on what it actually does and how it develops.

Referring to an algorithm as a series of instructions that are executed sequentially to obtain a result. AI is an algorithm or a set of specialized algorithms [2], whose difference with conventional algorithms is that it learns through samples. AI is fed with data, from which it extracts characteristics and patterns, some of

which are initially determined by humans. It identifies what it considers relevant and reprocesses until it is as close as possible to the initial sample. In this sense, AI can be understood as the set of algorithms that attempt to solve problems by learning through many examples provided by a human, to carry out actions where the human mind presents some limitation and generates more accurate results that help a human to make more assertive decisions. However, this leaves several questions in the air: where do you get the data to train these systems? Who trains and evaluates them, what kind of decisions do they make, and how reliable are they?

For example, entertainment and music streaming systems have built-in AI algorithms, which as we consume content, identify patterns to issue recommendations increasingly aligned to the user's tastes, the user accepts the terms and conditions of use so that these systems learn from us. However, there are other uses that we do not properly consent to, but that still make suggestions to the user, such as advertising ads that suddenly appear in our social networks or emails.

Applications and benefits of AI

AI systems have proven to have an extensive range of applications and benefits in different fields. Some of the most important ones are summarized in Figure *1*, as well as the areas of impact and some of the developments in the field (Table *2*).

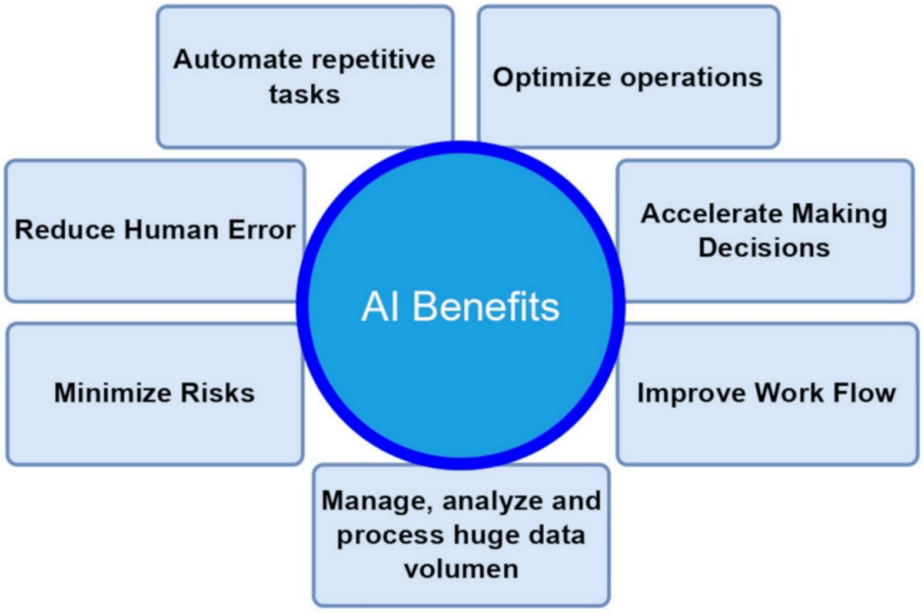

FIGURE 1: Some of the main benefits of AI

There are multiple areas of impact and applications of AI. However, for the sake of completeness, those that can be most illustrative of the benefits are outlined in Figure *1*.

Area	Applications	Authors
Agricultural	Detection and management of unwanted plants. Monitoring of agricultural conditions. Disease and pest detection. Environmental management. Evaluation of livestock breeding growth	[4,5]
Cybersecurity	Real-time threat monitoring. Classification and identification of cyber-attacks and profiles. Predictive analysis for vulnerability identification. Optimization of response to attacks	[6-8]
Sustainable development	Renewable energy management. Productivity enhancement. Monitoring of illegal wildlife trade. Optimization of weather and climate predictions	[9-11]
Economics	Market analysis. Risk management. Stock market forecasting. Trade balance monitoring	[12,13]
Education	Optimization of student and teacher evaluations. Monitoring of the learning process. Analysis of educational backwardness. Management of educational institutions. Adaptive learning and personalized tutoring	[14-16]
Entertainment	Optimization of personalized recommendations. Automatic content moderation. Audiovisual content generation. Enhancement of user interaction and audiovisual content	[17,18]
Justice delivery	Risk assessment in probation proceedings. Evidence and evidentiary analysis. Mediation and alternative dispute resolution. Combating corruption and promoting judicial transparency	[19,20]
Industry	Automatic and assisted driving. Generation of fault diagnostics in industrial parts, electronics, and/or processes. Identification of risk scenarios through video surveillance. Logistics optimization. Vehicle manufacturing. Quality control	[21,22]
Marketing	Consumer service improvement. Optimization of decision-making. Consumer feedback and trend monitoring. Personalization of products and/or services	[13,23]
Health	Optimization of personalized treatment strategies. Prediction of treatment response. Identification of patients at risk of developing chronic diseases. Automation and management of clinical documents. Identification of areas of interest and analysis in medical images. Gene sequencing. Classification of diseases and microorganisms	[24,25]

TABLE 2: Some of the areas and applications of AI

Risks of AI

Although AI has brought us multiple benefits both professionally and personally, its application also carries risks largely related to the type of data it collects and analyzes, such as personal and sensitive data, which raises concerns regarding the actors involved in the handling of privacy, security, processing and management of information. Currently, there is no international AI governance body or entity in charge of verifying the treatment of data in AI systems.

In order to check whether the most recent law proposals contemplate the current risk scenarios derived from the development and use of AI, in this paper, based on the documentary [11,26-54], we identified six main and general scenarios, as shown in Figure *2*.

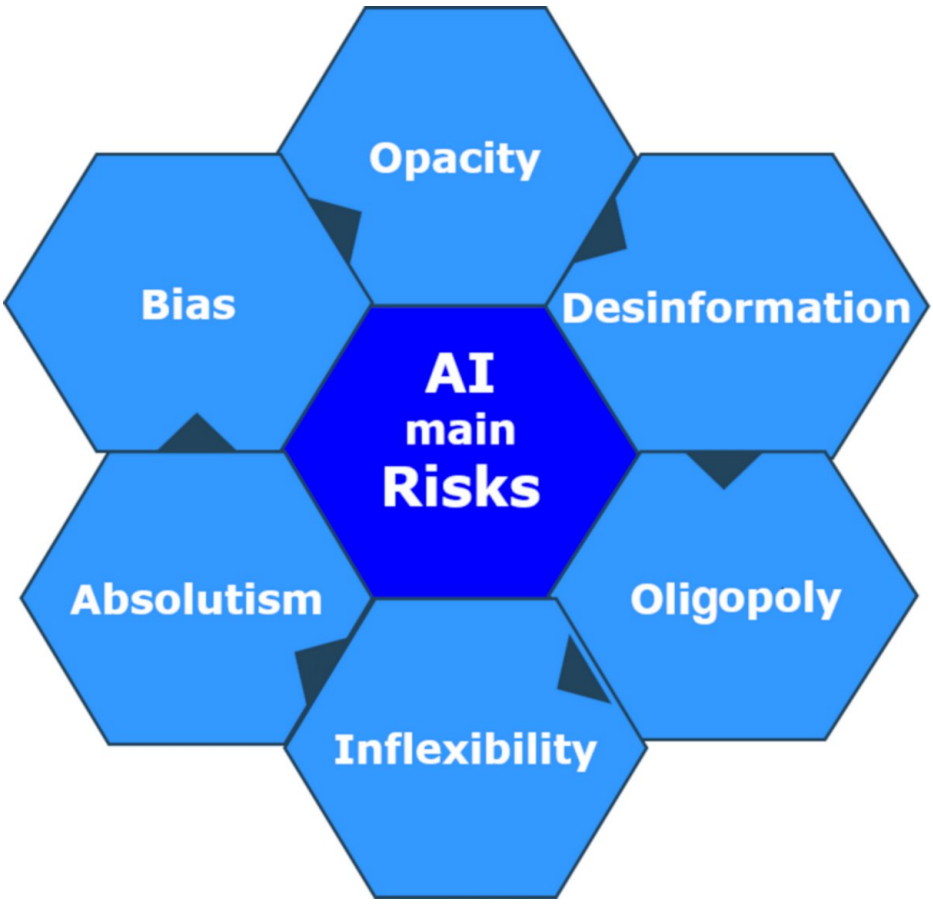

FIGURE 2: Some of the main risks of AI

Bias

Bias is defined as an anomaly that originates voluntarily or involuntarily from the judgment of developers, humans who are not exempt from making mistakes, and their preconceived assumptions during the design and development of AI algorithms and the structuring of the dataset used for training [26,27]. This may cause their use and training to be developed under discriminatory perspectives, exclusion, prejudice, stereotypes and other ideologies that threaten the integrity of people.

In this regard, the question arises as to who should assume ethical and/or legal responsibility for the results based on AI systems [28]. However, such dangerous criteria usually occur unintentionally, as it is a huge challenge to consider all the characteristics of a society as heterogeneous as ours, where minorities and vulnerable groups are usually the most affected. However, there is also the risk that these actions may be deliberate and malicious. Either way, bias generates consequences that affect society, as in the following examples:

• The case of recruitment performed using AI, where biases related to gender, skin color, and personality have been detected [29,30].

• The situation with popular image generators, such as DALL-E and Stable Diffusion, where it is perfectly clear how biases, stereotypes and even sexualizations are reflected [31-34]. When these tools are given a word, they generate a series of images alluding to it. Bias is inherent in the images they generate.

• Significant disparities when requesting college career recommendations for students with different profiles, where the Chat-GPT system (GPT-3.5 Turbo specifically) determines scores based on gender and has troubling biases against minorities such as the LGBTQ+ community and economically disadvantaged individuals [35].

Opacity

Explainability can be understood as the ability to reconstruct or describe the process by which a system yields certain results in order to understand how it came to a certain conclusion. Therefore, when an AI

algorithm delivers a result, it becomes a kind of verdict to which it is impossible to object.

Due to the complex and massively scaled nature of AI algorithms, it is really very difficult for any technique or tool to monitor the whole process in detail at each stage to understand the reasoning behind the result that the algorithm produces [36], so the process becomes opaque. This results in a lack of transparency that lacks explanatory power. Consequently, it hinders its intelligibility [37].

In addition, there is also a lack of transparency regarding the dataset, because most of the cutting-edge AI developments as well as databases are developed by private companies and large conglomerates that keep them out of the public eye, which hinders accessibility and therefore explainability [36,38].

Disinformation

Whenever any AI innovation is announced through the media, the media tend to exaggerate the capabilities of such developments, making it appear that these are fantastic technologies, and there is even a tendency toward anthropomorphizing (attributing human characteristics to non-human entities such as objects, systems, and animals) of AI. Developers encourage such advertisements. In the words of Emily Bender, "If someone is in the business of selling technology, the more magical it looks, the easier it is to sell it"; however, this represents a risk that currently has not been given the importance it deserves [38,39], but of which there are already consequences (such as those set out in the "risk absolutism" section) since, and according to Epstein et al. [40], the more AI tends to be anthropomorphized, the more responsibilities are attributed to it, and instead of seeing it for what it is, a tool used by humans, it is perceived as an independent entity capable of creating.

Oligopoly

The researcher Emily Blender and some former employees of large corporations have warned about the risks that exist due to the concentration of databases, systems and AI development specialists in just a few companies such as Google, IBM, Microsoft, Amazon, Facebook, etc. [38,41,42]. These companies, although they have expressed a commitment to ethics, refuse to document the foundations of their systems, including the types of data they store, and neither do they make part of the public domain important characteristics such as training parameters, size of the bases, data provenance, or resources to understand the model, among others [36,38] Decisions and the direction of developments, as well as the use and nature of the data, are left to a few.

Inflexibility

In every technological or industrial revolution, one of the biggest concerns that arises is the impact on the workplace, since the adoption of new technologies transforms workspaces, and there is always the fear of the "displacement effect" [43], i.e. the replacement of jobs by mechanical or digital entities. Referring specifically to AI, several studies have determined that its impact can be negative in this regard [44-46]. However, while validating the "displacement effect" by AI [43], it is also concluded that it brings with it something positive, such as the generation of new sources of employment [47,48].

It is important to emphasize that both companies and governments must be resilient to the adoption of new technologies such as AI and propose strategies as well as optimize their organization [47,49] since there are several factors involved, such as security, knowledge and dependence [49,50] that threaten the labor transformation, leading to a negative impact on workers [51]. Inflexibility in the adoption of AI is a latent risk.

Absolutism

As mentioned above, the anthropomorphism of AI leads to more and more responsibilities being attributed to it [40]. Although AI systems are a useful tool in decision-making and automation of various processes (i.e., economy, employment, education, security, health systems, government services, industry, among others) [52], leaving the responsibility of decision-making one hundred percent on a system that presents biases is quite questionable, even more so if there is not at least a well-founded ethical guideline [53,54]. However, systems are already in use that have adopted AI to address current challenges such as vehicle driving scenarios that, although seemingly a trivial human action, are actually very complex, as unforeseen scenarios arise that carry significant risks and that, even for a human, would be difficult to define.

Review

Methodology

This work was developed using a mixed and mainly qualitative approach, consisting of a review of

specialized literature and analysis of the most current legal and regulatory documents. For the specialized bibliographic review, we exhaustively explored the main academic databases: Web Science, IEEE Xplore and Scopus to define, analyze and present our proposal. Other official sources were also considered, such as newspaper articles, books and verified websites of both international organizations and governments. Some of the keywords used for the documentary search were: "artificial intelligence", "incidents", "ethics", "risks", "legal framework", "regulatory initiatives", "biases", "definition", among other relevant terms, all related to AI. With regard to the analysis of regulatory frameworks, the most current regulatory proposals and the bodies involved in their discourse at the global level were identified. Aspects related to governance, security and the ethical implications of their implementation were analyzed and compared. Finally, both approaches were compared in order to correlate the scope of the law proposals on AI risk scenarios.

Due to the rapidity with which the current regulatory and technological landscape advances, modifies and impacts the topic of AI, it is possible that by the date of publication of this paper some law proposals have already been issued or other works published that address additional risk scenarios to those we contemplated, so both the literature and current regulations that were found during the development of the present research are considered.

Main initiatives to regulate AI

As can be seen, even though AI-based systems are at an early stage of development, many of the issues and risks related to their application are related to areas such as governance, liability and ethical implications, which raise questions such as: who is responsible for the eventual actions and damages caused by an AI system; and what are the ethical guidelines according to which AI can make decisions; and what are the ethical guidelines according to which AI can make decisions?

In this regard, liability for damages caused by AI systems is a constantly evolving field, influenced by legal, ethical, and technical factors, where approaches such as “Accountability by Design” have been proposed to integrate ethical, legal, and technical principles throughout the development cycle, ensuring transparency, fairness, and accountability. Within this framework, developers may be held liable for design flaws, poor implementation, or lack of rigorous evaluation. However, liability may also fall on end users if they use the AI system negligently or outside its original purpose. On the other hand, emerging regulations, such as the EU Artificial Intelligence Act, establish specific obligations for providers and users, balancing innovation with the protection of rights, while organizations such as Partnership on AI, IEEE, AI Now Institute, and The Alan Turing Institute develop guidelines and best practices to mitigate risks and promote ethics in AI from the initial design stages.

In this context, the path towards AI regulation has gained relevance in the international arena in a scenario in which sustainable development in its three dimensions is presented as one of the main objectives of the international community within the framework of the UN 2030 Agenda, by addressing various environmental, economic and social challenges at the global level. However, in the face of the risks posed by the dizzying development and deployment of AI, we must not forget that its implementation must adhere to ethical, inclusive, and sustainable guidelines, anticipating its potential impacts and limitations.

Therefore, for some years now, organizations and governments around the world have expressed a constant concern about the multiple implications of AI, which is reflected in the holding of various summits, initiatives, reports, programs, recommendations, and proposals for regulatory frameworks that have arisen at the behest of international, regional and national organizations that revolve around the ethical, safety and socioeconomic impact of AI. In this regard, governments have begun to promote policies focused not only on stimulating innovation in AI, but also on adopting protective measures under a pluralistic approach that articulates principles and values to guide its use and deployment in the face of the risks posed using this disruptive technology [28].

While the regulatory framework for AI varies according to the jurisdiction and context (international, regional, and national) in which it arises, we must add that, so far, most of the efforts undertaken by countries are mostly strategies, reports or recommendations that lack binding force (in the style of soft law). Nevertheless, it is possible to identify some common approaches and principles that can establish guidelines towards harmonized AI regulation for countries, considering the following aspects (Table *3*).

Privacy and personal data protection regulations	**AI systems operate with large datasets (big data), often some of which are of a personal and even sensitive nature. Therefore, privacy protection legislations are essential to regulate how this data is collected, stored and used throughout the lifecycle of AI systems.**	**[55,56]**
Safety and liability	The European Union's AI Law constitutes the first regulatory framework on the safety of AI systems that considers the level of risk and determines the liability of developers in case of eventual damages caused by AI. This aspect should include both cybersecurity standards and liability requirements for developers and implementers of AI systems.	[56,57]
Robustness and security	AI systems should be designed and implemented foreseeing eventual risks and trying, as far as possible, to reduce errors to a minimum and to contemplate preventive scenarios in case of eventual failures and if possible develop protocols in this regard.	[56,58]
Equity and non-discrimination	AI systems should promote inclusiveness and ensure that biases that create discrimination are not perpetuated. This may involve assessing and mitigating bias from the AI system training data and in the algorithms themselves. Therefore, the implementation of an AI system should only be undertaken after having assessed its purposes and benefits, as well as its potential risks.	[55,56]
Transparency and explainability	Some initiatives envisage that AI systems should be transparent and able to explain how decisions are made. This may be particularly important in critical applications used in areas such as health, security, and delivery of justice.	[55,56]
Labor and social rights	As AI increasingly impacts the labor market and society at large, some regulations address issues regarding the protection of workers' rights not to be subject to automated decisions and the fair distribution of economic benefits generated by AI, as well as promoting sustainability, ecological and social responsibility, and not causing any harm.	[56,57]
Ethics and Governance of AI	In addition to legal regulations, there has been discussion around the need to incorporate ethical principles and guidelines for the responsible development and use of AI under the oversight of a supervisory authority.	[55,57]

TABLE 3: Guidelines to consider for a harmonized regulation of AI

These aspects may constitute guidelines for both governments and developers on how to design, guide, regulate and use AI from an ethical perspective. In this sense, the path towards global governance of AI is currently articulated around a variety of regulatory frameworks, recommendations and initiatives driven by a multiplicity of actors. It is worth mentioning that the leading powers in the development of AI technologies, such as the United States, China, Japan, the United Kingdom, etc., have already embarked on strategies to promote the development of a regulatory and governance framework for AI and, since 2016, have joined forces to coordinate these efforts within the framework of intergovernmental forums such as the Organisation for Economic Cooperation and Development (OECD), the Group of Seven (G-7), the Group of Twenty (G-20), etc.

Similarly, at the multilateral level, international organizations such as the UN and regional organizations such as the European Union (EU), the Organization of American States, and Economic Commission for Latin America and the Caribbean (ECLAC) have joined in the task of formulating global regulatory standards and principles for AI, as have companies, universities, research centers and civil society organizations. Some of the main ones are described below.

United Nations System

Efforts to regulate AI in the UN context are ongoing and cover a range of issues, including human rights, ethics, privacy, security and AI governance. While there is no governance body within the UN System specifically tasked with advancing the regulatory aspect of AI, there are a number of initiatives and bodies that are already addressing the issue, albeit from a non-binding approach, such as the following:

● AI for Good Global Summit: This platform was launched in 2017 in collaboration with the International Telecommunication Union, 38 UN agencies and bodies, as well as representatives of the private sector, civil society and academia, with the aim of promoting a collaborative framework to propose innovative solutions to the global challenges posed by AI. The action-oriented Summits held in 2018, 2019 and 2023 laid the groundwork for the development of various AI for Good projects in areas such as education, innovation ecosystems, gender equality, health, social justice, etc. During the last Summit held in Geneva (Switzerland) on May 30-31, 2024, innovations in generative AI, robotics and brain-machine interfaces were presented. The next Summit, which will take place July 8-11, 2025, aims to identify practical applications of AI to accelerate progress towards the Sustainable Development Goals and connect AI innovators with public and private sector decision makers to help scale AI solutions globally.

• Report of the UN High Commissioner for Human Rights: The document entitled "The right to privacy in the digital age" presented in September 2021 highlights the importance of ensuring that the development and implementation of AI respects fundamental human rights, such as privacy, non- discrimination and transparency [59].

• Recommendation on the Ethics of Artificial Intelligence from UNESCO: They were unanimously adopted in November 2021 during the General Conference of the Organization. They constitute a set of principles and values intended to guide the ethical, responsible and sustainable development and implementation of AI for the benefit of humanity [55].

• High-Level Advisory Body of Artificial Intelligence: In August 2023, the Office of the UN Secretary-General announced the creation of an advisory group composed of 39 experts from around the world to reflect on the challenges and opportunities related to AI. Its objective is to foster dialogue and promote international cooperation towards a global governance of AI [60].

• International Labor Organization: In its study entitled "Generative AI and jobs: A global analysis of potential effects on job quantity and quality", published in August 2023, the International Labor Organization addresses the impact of automation arising from the latest wave of generative AI on the labor market and the protection of workers' rights in the digital age. In this regard, it considers that the greatest impact will not be labor displacement, but its reconfiguration with respect to its intensity and autonomy [61].

• UNESCO's Global Norms on AI Ethics and Governance Observatory: It was launched in February 2024, during the II Global Forum on the Ethics of AI held in Kranj, Slovenia, with the purpose of monitoring how Member governments develop strategies to mitigate potential risks related to the deployment of AI, taking as a reference the Recommendations about Ethics of the AI. It also hosts the AI Ethics and Governance Lab that consists of a collaboration platform of experts from all over the world to share case studies, ideas, best practices, research recommendations and results related to better AI governance.

While so far the regulatory efforts undertaken within the UN framework mainly correspond to reports, observatories and recommendations without binding force, in recent years, it has been decided to address this complex and multidimensional challenge through the collaboration of multiple actors under a systemic governance approach, which includes governments, private sector, civil society and technical experts. In that sense, the UN offers a global framework to promote the development of international policies and standards to guide the ethical and responsible development of AI. Therefore, it could act in the short and medium term as a driving force in the process towards the establishment of a true International Organization for IA, either under the auspices of a specialized agency or as a subsidiary body of the General Assembly, considering its universal nature that frames the participation with the right to speak and vote of all the countries that make up this international organization [28].

Organisation for Economic Cooperation and Development

For some years now, the OECD has been actively working on different proposals focused mainly on the incorporation of the ethical aspects of AI. Although it lacks direct regulatory power, its recommendations and guidelines are highly influential, even beyond its 38 Members. Some of the main ones are listed below: OECD AI Principles: this is a set of guidelines and policies adopted in May 2019 to drive the adoption of ethical practices by both the public and private sectors in the development and deployment of AI, and to help governments address the economic and social challenges associated with the deployment of this emerging technology [62]. Inclusive growth, sustainable development, and well-being; 1.2 Respect for the rule of law, human rights and democratic values, including fairness and privacy; 1.3 Transparency and explainability; 1.4 Robustness, security and safety; and 1.5 Accountability [63].

• Recommendation on the Ethics of Artificial Intelligence: Considering the five principles listed above, the OECD issued five specific recommendations for countries to develop their respective AI strategies, such as 1. Investing in AI R&D; 2. Fostering an inclusive AI-enabling ecosystem; 3. Shaping an enabling interoperable governance and policy environment for AI; 4. Building human capacity and preparing for labor market transformation; and 5. In May 2024, the Organization's Ministerial Council approved a revision of the Recommendations considering aspects such as security of AI systems, information integrity, responsible business conduct, transparency, disclosure and interoperability of governance frameworks that promote compatible and coherent policy and regulatory environments for AI [63].

• The OECD.AI Network of Experts: It was created in 2023 with the objective of bringing together representatives from governments, industry, academia and civil society in a forum for discussion and collaboration on ethical policies and practices in AI, as well as developing guidance and recommendations for governments.

In short, the OECD plays an important forum towards AI regulation by providing guidance and

formulating recommendations and best practices for governments. Through its initiatives, although so far lacking binding force, it seeks to promote an ethically based approach to AI development and implementation that benefits society as a whole.

European Union

Some years ago, the EU launched the most advanced AI regulation process in the world to date through a series of initiatives and legislative proposals binding on the 27 Member States, promoted jointly by various institutions of the regional bloc, such as the European Commission (EC), the Council of the European Union (CEU) and the European Parliament, which finally led to the entry into force of the EU's first Artificial Intelligence Act in August 2024. The following are some of the main actions that the region has taken in this regard:

● General Data Protection Regulation: It was issued by the European Parliament and the Council of the European Union in May 2016. It establishes principles and guidelines for the protection of personal data of European citizens and its transfer to countries outside the European Economic Area, which is important considering the processing of personal information by some AI systems [64].

● EU AI Strategy: It constitutes a policy framework issued by the EC in April 2018, through the "Communication Artificial Intelligence for Europe". It aims to foster technological development and embrace AI in different sectors of economic activity, pave the way for socio-economic transformations and ensure its ethical and responsible use in Europe [65]. It is complemented by the Digital Competence Framework for Citizens (DigComp) and the Coordinated Plan on Artificial Intelligence.

● Coordinated Plan on Artificial Intelligence: It was presented by the EC in December 2018, considering the commitments set out in the framework of the aforementioned EU AI Strategy. It represents a comprehensive and coordinated approach between governments, industry, academia and civil society to define the set of actions and funding instruments to uniformly adopt and develop AI in the European bloc, while ensuring security, ethical development and respect for fundamental rights [66]. It was updated in April 2021 with the purpose of promoting investment for the development of AI technologies and implementing regional strategies and programs in the field.

● Communication Building Trust in Human-Centric Artificial Intelligence: It was presented by the EC in April 2019, with the purpose of promoting citizen participation, ensuring data security and protection, promoting education and digital literacy, supporting ethical research and fostering dialogue with industry, building a solid and sustainable relationship with AI [58].

● White Paper on Artificial Intelligence - A European approach to excellence and trust: It was published by the EC in February 2020 as part of the EU AI Strategy, with the purpose of boosting the safe and reliable adoption of AI in the EU and addressing the risks associated with certain uses of this emerging technology, based on two pillars: (a) mobilizing resources to generate an ecosystem of excellence that encourages the adoption of AI-based solutions, and (b) promoting an ecosystem of trust in which the fundamental rights of citizens in the Union are respected in the face of the risks associated with AI systems [67].

● Artificial Intelligence Act: The EC presented in April 2021 a Proposal for a Regulation laying down harmonised rules on artificial intelligence [57]. Thus, on March 13, 2024, it was endorsed by the European Parliament to enter into force on August 1, 2024 (although most of its provisions will be applicable from 2026). It constitutes the first comprehensive and binding legislation on AI in the world, which aims to establish a uniform legal framework regarding the development, commercialization and use of AI in Europe under a risk-based approach, i.e., considering its economic and social repercussions, and its impact on the fundamental rights of citizens. In that sense, it was considered that the higher the implicit risk for users in relation to the specific use that AI can generate, the stricter the rule should be. To this end, it establishes four categories of AI technologies [68] as it can be seen in Figure *3*:

a) Minimal risk: They represent an imperceptible or null risk for citizens and, consequently, are not subject to compliance with any obligation. Although companies are encouraged to voluntarily adopt codes of conduct voluntary adoption of codes of conduct that increase user confidence in the acceptance of AI technologies, e.g. spam filters, is recommended.

b) Limited risk: They are subject to compliance with limited transparency obligations, with the purpose of making users aware that they are interacting with a machine and, based on this, they can make informed decisions; for example, the use of chatbots or applications that provide advertising based on likes, gender or age.

c) High risk: Require an impact assessment analysis prior to commercialization or market introduction to minimize potential risks and discriminatory outcomes; for example, the use of biometric recognition technologies, administration of justice and democratic processes.

d) Unacceptable risk: They constitute a clear threat to people's safety, livelihoods and rights; for example, AI systems that use subliminal techniques that transcend a person's consciousness to substantially alter their behavior in a way that causes harm; or the use of AI systems for social scoring purposes. On the other hand, the Law establishes penalties for both large infringing companies and SMEs. In addition, by 2025, all 27 countries are expected to commit to designating national governance authorities to oversee companies that use and develop AI systems [69].

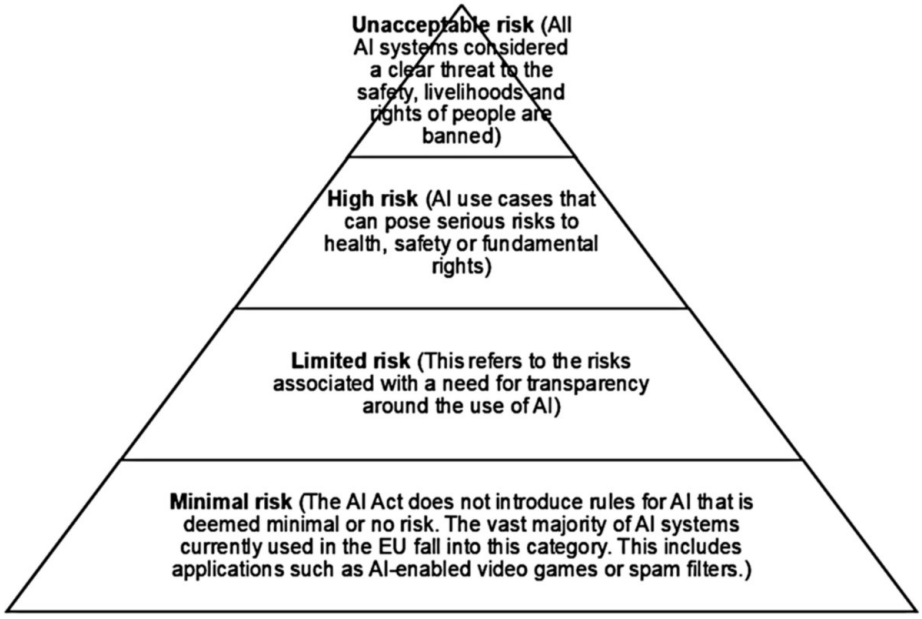

FIGURE 3: A risk-based approach

EU, European Union

• AI innovation package to support Artificial Intelligence startups and small and medium enterprises (SMEs): In January 2024, in the framework of the work for the approval of the EU AI Act, the EC issued a package of measures to support European SMEs of around 4 billion euros until 2027, as well as access to supercomputers to promote their involvement in the formation of reliable AI models. The AI innovation ecosystem promoted by the EU also comprises the creation of two European Digital Infrastructure Consortiums (EDICs). On the one hand, the Alliance for Language Technologies (ALT-EDIC) promotes the development of large European language models for training AI solutions. On the other hand, the European Digital Infrastructure Consortium (CitiVERSE EDIC) encourages the development of AI tools, such as digital twins, to help the creation of real-time simulation models for the benefit of cities for the optimization of processes related, for example, to traffic management and air quality improvement [70].

• European Artificial Intelligence Office (AI Office): The EC announced its creation in February 2024, as the basis of the AI governance system in Europe, which aims to foster international cooperation and promote the development and safe and reliable use of this technology. With the recent entry into force of the EU AI Act, it is expected to play an important supporting role for the AI governance bodies of each of the 27 countries of the regional bloc [42].

In this context, the EU has opted for the adoption of a regulatory approach that ensures transparency, traceability, accountability and responsibility for the use and development of AI systems [62]. So, ultimately, it is at the level of the UN, the OECD and the EU that we see the greatest momentum towards harmonized regulatory development of AI. In particular, the regulation of AI in the EU is an unprecedented effort that aims to ensure safety, protection of fundamental rights and ethics focused on the development and use of AI for countries.

AI Safety Summit

The first AI Safety Summit was held on November 1-2, 2023 at Bletchley Park (UK), with the participation of 28 countries from around the world, including China and the United States, and culminated in the signing of the Bletchley Declaration, a document that recognizes the need for governments to commit to monitoring and mitigating the potential risks of AI. This declaration represents a significant step towards a coordinated, global approach to addressing the challenges and seizing the opportunities presented by AI [71]. Subsequently, on May 21-22, 2024, the second AI Seoul Summit 2024 was held, jointly organized by

South Korea and the United Kingdom, which resulted in the Seoul Declaration for Safe, Innovative and Inclusive AI, through which several countries made a commitment to promote the interoperability of AI governance frameworks, through the integration of the Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems, which promotes the development of human-centered AI in collaboration with the private sector, academia and civil society. They are also committed to developing low-power chips considering the rapid growth of the AI industry and its increased consumption [71,72].

Hiroshima AI Process

International cooperation to regulate AI also includes other regional and interregional initiatives, such as the G7, an association of seven of the world's most industrialized countries, who agreed at the G7 Summit in Hiroshima, Japan, in May 2023 to the so-called "Hiroshima AI Process", which is a significant effort towards the creation of a global governance framework for AI regulation [73]. In this regard, on October 30, 2023, they announced the issuance of a Code of Ethics to promote and regulate the development of AI by companies.

For their part, in the framework of the G20, the main forum for macroeconomic policy coordination among the world's twenty largest economies and other emerging economies, the members have reaffirmed their commitment to the responsible use and development of AI, taking as a reference the Principles and the Recommendations of the OECD about AI [74].

Latin American Artificial Intelligence Index

In August 2023, the ECLAC presented the first Latin American Artificial Intelligence Index, which was conducted in collaboration with various institutions and experts in the region, in order to identify achievements, gaps and opportunities for improvement of AI ecosystems in nineteen countries in Latin America and the Caribbean. In this sense, it has been used by governments, academic institutions and organizations as a guide for the formulation of policies and the implementation of projects that promote a more equitable, ethical and effective use of AI in the region. On September 24, 2024, ECLAC and the National Center for Artificial Intelligence of Chile [75] launched a second version of the Index.

Cartagena de Indias Declaration for Governance, The Construction of AI Ecosystems and the Promotion of AI Education in an Ethical and Responsible Manner in Latin America and the Caribbean

It was signed in August 2024, in the framework of the AI Summit of Cartagena de Indias, Colombia, in which representatives of 17 countries of Latin America and the Caribbean participated. It is a common framework that revolves around three axes: 1. The construction of enabling ecosystems that facilitate the deployment of ethical and inclusive AI; 2. The promotion of education and training of digital skills in the use of AI in the region; and 3. The promotion of governance frameworks that protect human rights in digital environments [76].

Some of the main national AI regulations

Although the main regulatory efforts are driven by various international and regional organizations around the world, some countries that undoubtedly show clear leadership in AI R&D have joined the urgent task of beginning to develop a regulatory framework to guide its development and ethical implementation, considering its multiple risks and impacts. Such is the case, for example, of China, Japan and the United States.

China

Since 2017, the State Council of the People's Republic of China began the construction of the institutional and regulatory scaffolding of AI through the publication of the Next Generation Artificial Intelligence Development Plan, which sets out the strategy of positioning the country as a world leader in AI by 2030 [77]. This plan is structured in three phases and covers multiple aspects, ranging from R&D to the application of AI technologies in specific sectors:

1. Initial (2017-2020): Laying the foundations of the AI industry, developing a legal and ethical framework, in addition to incentivizing basic research.

2. Intermediate (2020-2025): Achieve significant progress in its development with a focus on integrating AI into key industries.

3. Final (2025-2030): Become the world leader in AI R&D, setting global standards and mastering key AI technologies.

Also, on August 15, 2023, China's Interim Measures for the Management of Generative Artificial Intelligence Services was published, which aims to regulate the development and use of generative AI in the face of the risks and ethical challenges associated with its rapid development [78]. This reflects the comprehensive approach to consolidate itself as a superpower, leveraging its ability to deploy AI technologies quickly and on a large scale.

Japan

In the Asian context, Japan has also established itself as a global leader in the development and implementation of AI in areas such as robotics, automotive and advanced manufacturing. In this regard, in January 2016 it launched the so-called Japan's Society 5.0 Strategy, which combines government policies, R&D investments and ethical guidelines to guide the deployment and implementation of AI. Similarly, in April 2016, the AI Strategy Council, an advisory body tasked with advising the government on AI policy formulation, was established. In doing so, Japan seeks to create a "superintelligent" society in which AI is deeply integrated into all aspects of daily life [79,80].

On the other hand, in February 2019, it adopted the Social Principles of Human-Centered AI, which emphasize the protection of human autonomy, prevention of harm, fairness and explainability [81].

Also noteworthy are the Governance Guidelines for Implementation of AI Principles Ver 1.1 of 2022, the AI White Paper Japan's National Strategy in the New Era of AI of 2023 and the Draft AI Guidelines for Business of 2024 issued by the Ministry of Internal Affairs and Communications Ministry of Economy, Trade and Industry of Japan, which establishes obligations for developers, service providers and users related to the processing of personal data and the use of traceable data, record keeping and documentation and security measures related to AI [82].

On October 30, 2023, former U.S. President Joe Biden issued the Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, as a first step towards establishing a framework for managing the ethical development and responsible use of this technology based on eight guiding principles: risk management and mitigation; collaboration between developers and companies; promotion of equity, privacy and protection of civil rights; and global social, economic and technological progress in the face of disruptive change. Its implementation was intended to create an ad hoc environment to position the United States as a leader in responsible AI regulation globally [83].

However, on January 20, 2025, President Donald Trump rescinded this Executive Order by Biden, arguing that it stifled innovation by imposing excessive bureaucratic requirements on the private sector. Instead, three days later, he signed Executive Order 14179 (Removing Barriers to American Leadership in Artificial Intelligence), aiming to:

1. Consolidate American dominance in AI.

2. Eliminate regulations deemed "unnecessarily restrictive."

As a first step, the order calls on presidential advisors on AI to develop an Artificial Intelligence Action Plan within 180 days, focused on establishing key policy measures to strengthen U.S. leadership in artificial intelligence, ensuring that regulations do not impose undue burdens that stifle private sector innovation. In this way, through a balanced regulatory framework, it aims to ensure that U.S. dominance in AI not only drives technological progress, but also promotes social welfare, global competitiveness, and the protection of national interests.

While there are marked differences between the visions of former President Biden and current President Trump, it will be interesting to learn how the forthcoming Artificial Intelligence Action Plan intends to address the host of risks and ethical challenges posed by the rapid advance of AI. The above is important mainly if we contrast the current technological conflict between China and the United States in the context of the Fourth Industrial Revolution.

China and the United States ultimately represent two antagonistic models in the development of AI, reflecting their political systems and strategic priorities. On the one hand, the United States promotes a decentralized, private sector-led approach, with an emphasis on open innovation, ethics in civilian applications and industry-university-government collaboration, albeit with increasing restrictions on the export of sensitive technologies for national security reasons. China, on the other hand, prioritizes a coordinated state development model, with five-year plans that align AI research with goals of social control, technological supremacy and military applications, under strict data controls and censorship. While the U.S. debates ethical boundaries and competes on global standards, China is moving forward with massive integration of AI in surveillance and facial recognition, evidencing a race where ethical and geopolitical divergence redefines the global governance of this technology (Table *4*).

Biden (2023)	Trump (2025)
Ethical regulation and public-private collaboration.	Deregulation to prioritize competitiveness.
Approach based on the protection of civil rights and risks.	Focus on technological leadership vis-à-vis China.
Alignment with global standards.	Technological nationalism ("America First").

TABLE 4: Contrast between the two visions

Regulatory focus on the private sector

On the other hand, it is also important to consider the work undertaken by research institutions, social organizations and private companies in different countries, who have also joined the task of developing and proposing a series of ethical guidelines applicable to AI systems related to transparency, justice, equity, nonmaleficence, responsibility, privacy, beneficence, freedom and autonomy, trust, sustainability, dignity, solidarity, etc. [84].

Being private companies from countries such as the United States, EU, UK and Japan the largest issuers of guidelines and lists of ethical principles of AI (about 20% of the total according to the study presented by Jobin et al.,2019), such as: Sony, IBM, Open AI, Google, Microsoft, Intel Corporation, The Royal Society, Sage, The Greens, European Parliament, etc.

Based on the above, efforts to create a legal and ethical framework for AI are mainly led by economically developed countries, which in the future could lead to some inequity in the face of a lower participation of regions such as Africa, South America, Central America and Central Asia. Examples of other international organizations that have contributed to the issuance of ethical principles include the IEEE, UNESCO, Women 20, Women Leading in IA, Partnership on IA, etc.

Discussion

Based on the above analysis, it is possible to conclude that the initial regulation of IA varies from one country to another and is mainly influenced by the work carried out in the framework of international and regional organizations such as the UN, the OECD and the EU. It is also framed by the respective efforts of some leading countries in AI development and innovation, which have undertaken concrete actions for the issuance of policies and regulatory frameworks that consider common aspects such as the concept and classification of systems, governance and ethical use of systems (Table *5*).

Particularly noteworthy, however, are the efforts of the European bloc with respect to the creation and entry into force of the first binding regulatory framework for AI in the world, whose scope constitutes a historic milestone for other countries and regions in the world to promote the creation of an ethical regulatory framework, taking into account the wide range of applications and risks that AI systems currently represent.

	United Nations	OECD	EU
Concept of AI Systems	Based on their objectives and applications.	Emphasizes its structure and objectives.	Based on its objectives.
Classification of AI systems and obligations applicable to each category	Does not establish a classification that imposes specific obligations. However, it urges AI actors to commit to developing explainable and transparent algorithms.	Does not establish a classification. Instead, it presents a catalog of principles to be observed by the actors involved, related to ensuring human supervision, transparency, and explainability in processes.	It classifies AI systems into four categories (minimal, limited, high, and unacceptable) in accordance with the implicit risk for users, imposing corresponding obligations on AI actors to ensure compliance.
Actors in the cycle life cycle of the AI systems	Considers individuals and legal entities that participate in at least one stage of the AI system life cycle as providers.	Considers organizations or individuals who deploy or exploit AI and in general, the AI system plays an active role during the various phases of the system's life cycle.	Defines each of the actors that participate in the of the life cycle of the AI system, from an AI system to a user, by imposing obligations during the various Strictly for the case of high-risk applications.
Governance of the IA	Expert Group on AI is responsible for promoting dialogue and fostering international cooperation in AI.	Network collaboration of AI experts, composed of representatives from governments, industry, academia, and civil society. It formulates guidelines and recommendations for the States.	Each Member State shall designate its national authorities and representatives: the notifier and the market. European Council and AI, integrated by a representative of each country, is in charge of coordination between national authorities. Consultative Forum, responsible for providing expertise and advising the Board and the Commission. Independent Scientific Expert, responsible for ensuring correct application of the law. European Office of AI, as the foundation of the European AI governance system, is in charge of fostering international cooperation and the development and safe, reliable use of AI.
Funding AI public	Recommends that countries promote incentives for both public and private sectors to support AI R&D.	Recommends that governments facilitate public and private sector investment in R&D to stimulate reliable AI innovation.	The EC earmarked an investment of 1,000 M for the years 2021 and 2022, through the programs Horizon Europe and Digital Europe. In accordance with the objectives of the Digital Decade, by September 2023, €4.4 billion has been earmarked for AI.
Regime penalties in AI matters	Does not establish a regime sanctioning.	Does not establish a regime penalties.	Contemplates fines administrative of up to 35 M€, if the offender is a company; and up to 7% of its volume of total annual turnover worldwide for the previous year. In the case of suppliers of models of AI, provides for fines not exceeding 3% of their total annual worldwide turnover in the previous financial year, or €15 million, whichever is higher.

TABLE 5: Key contributions to AI regulation

OECD, Organisation for Economic Co-operation and Development; EC, European Commission; EU, European Union; AI, Artificial Intelligence; R&D, Research and Development

Although there are differences between each of the three regulations mentioned above, they also share common elements that make it possible to formulate a comprehensive definition of AI systems, which they regulate as machine-based systems that fulfill objectives (explicit or implicit) and make inferences based on input information, and generate output information, such as content, predictions, recommendations or decisions, and can thus influence real or virtual environments.

In the case of the UN and the OECD, they mainly establish a series of non-binding recommendations for countries. In contrast, the EU developed a robust regulatory framework to regulate actors in the AI lifecycle that is complemented by a risk-based classification of systems.

Undoubtedly, all these rules pose significant challenges for both AI users and developers, who will have to adapt to the new governance provisions and the sanctions regime in the case of the EU, since in case of contravening its provisions, they will be liable to high fines, which, for some developers, could represent

an obstacle to continue with the development of this disruptive technology. Through these actions, the regional bloc is consolidating its position as a global benchmark in regulatory development and financing for the adoption of reliable, safe and sustainable human-centered AI technologies.

Conclusions

Since its inception, AI has aroused controversy due to its origin and the lack of consensus on its definition, as well as its impact on society due to the biases (voluntary or involuntary) that are introduced into systems and data, that result in discriminatory, unfair or harmful decisions for certain groups (usually some minorities), leading to prejudice, manipulation, ideological influence, etc. In addition, society tends to humanize this technology, mainly due to disinformation and media exaggeration, which causes it to be delegated more and more responsibilities in decision-making that traditionally should correspond to human intelligence. Faced with these scenarios, the urgency has arisen to generate a solid ethical framework to guide the deployment and use of AI. In this regard, the efforts undertaken by international organizations such as the UN, UNESCO and the OECD stand out, as well as the work of governments and entities belonging, for the most part, to developed countries, which are currently focused on the development of an ethical regulatory framework to guide its development and implementation. Such is the case of the EU, with the entry into force as of January 1, 2024, of the first AI Law in the world, which adopts an approach based on the implicit risk for users of these systems and establishes four categories of AI: a) minimal risk; b) limited risk; c) high risk; and) unacceptable risk. This regulatory framework is reinforced by a series of sanctions applicable to those companies that fail to comply with its provisions.

Although the issuance of the world's first AI law in the USA represents a good start, other countries need to follow suit and establish robust legal and ethical frameworks, taking into account the wide range of applications and risks that AI currently presents, such as the scenarios shown in Figure *2*, which current regulations barely address, particularly in terms of bias. These frameworks should also consider the emergence of auditing and governance entities that evaluate AI developments prior to their introduction to the market. Given the rapid growth of this technology and its ever deeper immersion immediately require the corresponding legal and governance frameworks in order to control or supervise the ethical and balanced use of AI and protect the fundamental rights of individuals, avoiding any violation of human dignity and integrity. In order to pave the way for the harmonized regulation of AI in the future, the process towards the establishment of a genuine International AI Organization can be fostered.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Angelina Espejel-Trujillo, José J. Piña Mondragón

Acquisition, analysis, or interpretation of data: Angelina Espejel-Trujillo, José J. Piña Mondragón

Drafting of the manuscript: Angelina Espejel-Trujillo, José J. Piña Mondragón

Critical review of the manuscript for important intellectual content: Angelina Espejel-Trujillo, José J. Piña Mondragón

Disclosures

Conflicts of interest: In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

References

1. AI Incident Database. (2016). Accessed: October 20, 2024: https://incidentdatabase.ai/.
2. Larson EJ: The Myth of Artificial Intelligence. Harvard University Press, Cambridge, MA; 2021. 10.2307/j.ctv322v43j
3. Gonçalves B: Lady Lovelace’s objection: The Turing-Hartree disputes over the meaning of digital computers, 1946-1951. IEEE Annals of the History of Computing. 2024, 46:6-18. 10.1109/mahc.2023.3326607
4. Bao J, Xie Q: Artificial intelligence in animal farming: A systematic literature review. Journal of Cleaner Production. 2022, 331:129956. 10.1016/j.jclepro.2021.129956
5. Misra NN, Dixit Y, Al-Mallahi A, Bhullar MS, Upadhyay R, Martynenko A: IoT, big data, and artificial intelligence in agriculture and food industry. IEEE Internet of Things Journal. 2022, 9:6305-6324. 10.1109/jiot.2020.2998584

6. Kabbas A, Alharthi A, Munshi A: Artificial intelligence applications in cybersecurity. International Journal of Computer Science and Network Security. 2020, 20:120-124.
7. Zeadally S, Adi E, Baig Z, Khan IA: Harnessing artificial intelligence capabilities to improve cybersecurity. IEEE Access. 2020, 8:23817-23837. 10.1109/access.2020.2968045
8. Zhang Z, Hamadi H Al, Damiani E, Yeun CY, Taher FT: Explainable artificial intelligence applications in cyber security: State-of-the-art in research. IEEE Access. 2022, 10:93104-93139. 10.1109/access.2022.3204051
9. Vinuesa R, Azizpour H, Leite I, et al.: The role of artificial intelligence in achieving the Sustainable Development Goals. Nature Communications. 2020, 11:233. 10.1038/s41467-019-14108-y
10. Taghikhah F, Erfani E, Bakhshayeshi I, Tayari S, Karatopouzis A, Hanna B: Artificial intelligence and sustainability: solutions to social and environmental challenges. Artificial Intelligence and Data Science in Environmental Sensing. Asadnia M, Razmjou A, Beheshti A, (ed): Academic Press, London; 2022. 93-108. 10.1016/B978-0-323-90508-4.00006-X
11. Galaz V, Centeno MA, Callahan PW, et al.: Artificial intelligence, systemic risks, and sustainability. Technology in Society. 2021, 67:101741. 10.1016/j.techsoc.2021.101741
12. Chhajer P, Shah M, Kshirsagar A: The applications of artificial neural networks, support vector machines, and long-short term memory for stock market prediction. Decision Analytics Journal. 2022, 2:100015. 10.1016/j.dajour.2021.100015
13. Rahmani AM, Rezazadeh B, Haghparast M, Chang WC, Ting SG: Applications of artificial intelligence in the economy, including applications in stock trading, market analysis, and risk management. IEEE Access. 2023, 11:80769-80793. 10.1109/access.2023.3300036
14. Özer M: Potential benefits and risks of artificial intelligence in education. Bartın University Journal of Faculty of Education. 2024, 13:232-244. 10.14686/buefad.1416087
15. Salas-Pilco SZ, Yang Y: Artificial intelligence applications in Latin American higher education: a systematic review. International Journal of Educational Technology in Higher Education. 2022, 19:21. 10.1186/s41239-022-00326-w
16. Wang S, Wang F, Zhu Z, Wang J, Tran T, Du Z: Artificial intelligence in education: A systematic literature review. Expert Systems with Applications. 2024, 252:124167. 10.1016/j.eswa.2024.124167
17. Chan-Olmsted SM: A review of artificial intelligence adoptions in the media industry. The International Journal on Media Management. 2019, 21:193-215. 10.1080/14241277.2019.1695619
18. Nautiyal R, Jha RS, Kathuria S, Chanti Y, Rathor N, Gupta M: Intersection of artificial intelligence (AI) in entertainment sector. 2023 4th International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India. 2023, 1273-1278. 10.1109/ICOSEC58147.2023.10275976
19. Laptev VA, Feyzrakhmanova DR: Application of artificial intelligence in justice: Current trends and future prospects. Human-Centric Intelligent Systems. 2024, 4:394-405. 10.1007/s44230-024-00074-2
20. Spitsin IN, Tarasov IN: Use of artificial intelligence in the administration of justice: theoretical aspects of legal regulation (problem statement). Actual Problems of Russian Law. 2020, 15:96-107. 10.17803/1994-1471.2020.117.8.096-107
21. Banitaan S, Al-refai G, Almatarneh S, Alquran H: A review on artificial intelligence in the context of Industry 4.0. International Journal of Advanced Computer Science and Applications. 2023, 14:10.14569/ijacsa.2023.0140204
22. Bharadiya JP, Thomas RK, Ahmed F: Rise of artificial intelligence in business and industry. Journal of Engineering Research and Reports. 2023, 25:85-103. 10.9734/jerr/2023/v25i3893
23. Labib E: Artificial intelligence in marketing: exploring current and future trends. Cogent Business & Management. 2024, 11:2348728. 10.1080/23311975.2024.2348728
24. Haleem A, Javaid M, Khan H: Current status and applications of artificial intelligence (AI) in medical field: An overview. Current Medicine Research and Practice. 2019, 9:231-237. 10.1016/j.cmrp.2019.11.005
25. Salam A, Abhinesh N: Revolutionizing dermatology: The role of artificial intelligence in clinical practice. IP Indian Journal of Clinical and Experimental Dermatology. 2024, 10:107-112. 10.18231/j.ijced.2024.021
26. Dilmegani C: Bias in AI: Examples and 6 Ways to Fix it in 2025. (2017). Accessed: April 22, 2024: https://research.aimultiple.com/ai-bias/.
27. Varsha PS: How can we manage biases in artificial intelligence systems - A systematic literature review. International Journal of Information Management Data Insights. 2023, 3:100165. 10.1016/j.jjimei.2023.100165
28. Drnas de Clément Z: Inteligencia artificial en el Derecho Internacional, Naciones Unidas y Unión Europea. Revista Estudios Jurídicos. 2024, 22:28. 10.17561/rej.n22.7524
29. Nadeem A, Marjanovic O, Abedin B: Gender bias in AI-based decision-making systems: a systematic literature review. Australasian Journal of Information Systems. 2022, 26:10.3127/ajis.v26i0.3835
30. Chen Z: Ethics and discrimination in artificial intelligence-enabled recruitment practices. Humanities and Social Sciences Communications. 2023, 10:567. 10.1057/s41599-023-02079-x
31. Nicoletti L, Bass D: Humans Are Biased. Generative AI Is Even Worse. (2023). Accessed: August 8, 2024: https://www.bloomberg.com/graphics/2023-generative-ai-bias/.
32. Cheong M, Abedin E, Ferreira M, et al.: Investigating gender and racial biases in DALL-E Mini images. ACM Journal on Responsible Computing. 2024, 1:1-20. 10.1145/3649883
33. Ghosh S, Caliskan A: 'Person' == light-skinned, Western man, and sexualization of women of color: Stereotypes in stable diffusion. Findings of the Association for Computational Linguistics: EMNLP 2023. Bouamor H, Pino JP, Bali K (ed): Association for Computational Linguistics, Singapore; 2023. 6971-6985. 10.18653/v1/2023.findings-emnlp.465
34. Luccioni AS, Akiki C, Mitchell M, Jernite Y: Stable bias: Evaluating societal representations in diffusion models. 37th Conference on Neural Information Processing Systems (NeurIPS 2023) Track on Datasets and Benchmarks. 2023, 14.
35. Zheng A: Dissecting bias of ChatGPT in college major recommendations. Information Technology and Management. 2024, 1-12. 10.1007/s10799-024-00430-5
36. Liao QV, Wortman Vaughan J: AI transparency in the age of LLMs: A human-centered research roadmap. Harvard Data Science Review. 2024, 1-33. 10.1162/99608f92.8036d03b
37. Felzmann H, Fosch-Villaronga E, Lutz C, Tamò-Larrieux A: Towards transparency by design for artificial intelligence. Science and Engineering Ethics. 2020, 26:3333-3361. 10.1007/s11948-020-00276-4

38. Peréz Colomé J: Emily Bender: “Los chatbots no deberían hablar en primera persona. Es un problema que parezcan humanos”. (2023). Accessed: December 8, 2023: https://elpais.com/tecnologia/2023-03-18/emily-bender-los-chatbots-no-deberian-hablar-en-primera-persona-es-un-proble....
39. Bender EM, Gebru T, McMillan-Major A, Shmitchell S: On the dangers of stochastic parrots: Can language models be too big?. FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. 2021, 14:610-623. 10.1145/3442188.3445922
40. Epstein Z, Levine S, Rand DG, Rahwan I: Who gets credit for AI-generated art?. iScience. 2020, 23:101515. 10.1016/j.isci.2020.101515
41. Mulligan CEA, Godsiff P: Datalism and data monopolies in the era of A.I.: A research agenda. arXiv. 2023, 10.48550/arXiv.2307.08049
42. European Commission: Making Artificial Intelligence Available to All - How to Avoid Big Tech's Monopoly on AI?. (2024). https://europa.eu/newsroom/ecpc-failover/pdf/speech-24-931_en.pdf.
43. Guliyev H: Artificial intelligence and unemployment in high-tech developed countries: New insights from dynamic panel data model. Research in Globalization. 2023, 7:100140. 10.1016/j.resglo.2023.100140
44. Ngo P, Das J, Ogle J, Thomas J, Anderson W, Smith RN: Predicting the speed of a Wave Glider autonomous surface vehicle from wave model data. 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, Chicago, IL, USA. 2014, 2250-2256. 10.1109/IROS.2014.6942866
45. Georgieff A, Milanez A: What happened to jobs at high risk of automation? OECD Social, Employment and Migration Working Papers No. 255. OECD Publishing. (2021). https://www.oecd.org/en/publications/what-happened-to-jobs-at-high-risk-of-automation_10bc97f4-en.html.
46. Hawksworth J, Berriman R, Goel S: Will robots really steal our jobs?: An international analysis of the potential long term impact of automation. PricewaterhouseCoopers. (2018). https://www.pwc.co.uk/economic-services/ukeo/pwcukeo-section-4-automation-march-2017-v2.pdf.
47. Ali O, Murray PA, Muhammed S, Dwivedi YK, Rashiti S: Evaluating organizational level IT innovation adoption factors among global firms. Journal of Innovation & Knowledge. 2022, 7:100213. 10.1016/j.jik.2022.100213
48. World Economic Forum: The Future of Jobs: Employment, Skills and Workforce Strategy for the Fourth Industrial Revolution. World Economic Forum, Cologny, Switzerland; 2016.
49. Tursunbayeva A, Gal HC: Adoption of artificial intelligence: A TOP framework-based checklist for digital leaders. Business Horizons. 2024, 67:357-368. 10.1016/j.bushor.2024.04.006
50. Cubric M: Drivers, barriers and social considerations for AI adoption in business and management: A tertiary study. Technology in Society. 2020, 62:101257. 10.1016/j.techsoc.2020.101257
51. Stevenson B: Artificial intelligence, income, employment, and meaning. The Economics of Artificial Intelligence: An Agenda. Agrawal A, Gans J, Goldfarb A (ed): University of Chicago Press, Chicago; 2019. 189-195.
52. Monteith S, Glenn T: Automated decision-making and big data: Concerns for people with mental illness. Current Psychiatry Reports. 2016, 18:112. 10.1007/s11920-016-0746-6
53. Comisión Europea, Dirección General de Redes de Comunicación, Contenido y Tecnologías, Grupa ekspertów wysokiego szczebla ds. sztucznej inteligencji, Directrices éticas para una IA fiable. (2019). https://data.europa.eu/doi/10.2759/14078.
54. Strümke I, Slavkovik M, Madai VI: The social dilemma in artificial intelligence development and why we have to solve it. AI and Ethics. 2022, 2:655-665. 10.1007/s43681-021-00120-w
55. United Nations Educational, Scientific and Cultural Organization: Recommendation on the Ethics of Artificial Intelligence. (2021). Accessed: March 21, 2024: https://unesdoc.unesco.org/ark:/48223/pf0000381137.
56. Mikalef P, Conboy K, Lundström JE, Popovič A: Thinking responsibly about responsible AI and ‘the dark side’ of AI. European Journal of Information Systems. 2022, 31:257-268. 10.1080/0960085x.2022.2026621
57. European Commission: Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. European Commission, Brussels; 2021.
58. European Commission: Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions. Building Trust in Human-Centric Artificial Intelligence. European Commission, Brussels; 2019.
59. UN Office of the High Commissioner for Human Rights. The right to privacy in the digital age: report of the Office of the United Nations High Commissioner for Human Rights. (2021). https://digitallibrary.un.org/record/777869?ln=en&v=pdf.
60. United Nations: Governing AI for Humanity: Final Report. United Nations, New York, NY; 2024.
61. Gmyrek P, Berg J, Bescond D: Generative AI and jobs: a global analysis of potential effects on job quantity and quality. ILO Working Paper 96. International Labour Organization. (2023). https://doi.org/10.54394/FHEM8239.
62. Martínez Espín P: La propuesta de marco regulador de los sistemas de Inteligencia Artificial en el mercado de la UE. Revista CESCO de Derecho de Consumo. 2023, 46:1-20.
63. Organisation for Economic Co-operation and Development: Recommendation of the Council on Artificial Intelligence. (2024). https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449.
64. European Parliament and the Council of the European Union: Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Official Journal of the European Union. 2016, L119:88.
65. European Commission: Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions. Artificial Intelligence for Europe. European Commission, Brussels; 2018.
66. European Commission: Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions. Coordinated Plan on Artificial Intelligence. European Commission, Brussels; 2018.
67. European Commission: White Paper on Artificial Intelligence - A European Approach to Excellence and Trust. European Commission, Brussels; 2020.
68. Shaping Europe’s digital future. AI Act. (2024). Accessed: April 17, 2025: https://digital-

strategy.ec.europa.eu/en/policies/regulatory-framework-ai.
69. European Parliament and the Council of the European Union: Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828. Official Journal of the European Union. 2024, LSeries:144.
70. European Commission: Commission implementing decision (EU) 2024/459 of 1 February 2024 on setting up the European Digital Infrastructure Consortium for Networked Local Digital Twins towards the CitiVERSE (LDT CitiVERSE EDIC). Official Journal of the European Union. 2024, LSeries:1-4.
71. Policy Paper. The Bletchley Declaration by Countries Attending the AI Safety Summit. (2025). https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-....
72. Government UK: Seoul Declaration for safe, innovative and inclusive AI by participants attending the leader's session of the AI Seoul Summit. (2024). Accessed: May 2, 2024: https://www.mofa.go.jp/files/100672534.pdf.
73. G7: Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI System. G7 2023 Hiroshima Summit. (2023). https://www.mofa.go.jp/files/100573471.pdf.
74. G20 Ministerial Statement on Trade and Digital Economy. (2019). https://www.mofa.go.jp/files/000486596.pdf.
75. Economic Commission for Latin America and the Caribbean (ECLAC) & National Center for Artificial Intelligence (CENIA): ILIA 2024: Latin American Artificial Intelligence Index. ECLAC and CENIA. (2024). https://indicelatam.cl/wp-content/uploads/2025/01/ILIA_2024_Ingles_020125_compressed.pdf.
76. Latin American and Caribbean Ministerial Summit on Artificial Intelligence: Cartagena de Indias Declaration for Governance, the construction of AI ecosystems and the promotion of AI education in an Ethical and Responsible manner in Latin America and the Caribbean. (2024). https://www.mintic.gov.co/portal/715/articles-383990_recurso_1.pdf.
77. Department of International Cooperation, Ministry of Science and Technology of China: Next generation artificial intelligence development plan issued by State Council. China Science and Technology Newsletter. 2017, 17:18.
78. Cyberspace Administration of China: China's Interim Measures for the Management of Generative Artificial Intelligence Services officially implemented. (2023). Accessed: June 2, 2024: https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm..
79. Council for Science, Technology and Innovation Cabinet Office: Report on The 5th Science and Technology Basic Plan. Government of Japan, 2016.
80. Riken Center for Advanced Intelligence Project of Japan: Tokyo Establishes "AI Strategy Council". (2024). Accessed: October 10, 2024: https://aip.riken.jp/news/ai-strategy-council/.
81. Council for Social Principles of Human-centric AI of Japan: Social Principles of Human-Centric AI. (2019). Accessed: October 10, 2024: https://www8.cao.go.jp/cstp/english/humancentricai.pdf.
82. Ministry of Internal Affairs and Communications and Ministry of Economy, Trade and Industry: AI Guidelines for Business. Ministry of Economy, Trade and Industry, Tokyo; 2024.
83. The White House: Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. (2023). Accessed: March 16, 2024: https://bidenwhitehouse.archives.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure....
84. Jobin A, Ienca M, Vayena E: The global landscape of AI ethics guidelines. Nature Machine Intelligence. 2019, 1:389-399. 10.1038/s42256-019-0088-2