rac.2025.113.1.org04
RACSG.2025;113(01):54-65
Recibido: 25/09/2024 | Aceptado: 11/11/2024

ORIGINAL

NAVIGATING THE ETHICAL AND SOCIETAL CHALLENGES OF ARTIFICIAL INTELLIGENCE: ENSURING RESPONSIBLE INNOVATION

ABORDANDO LOS RETOS ÉTICOS Y SOCIALES DE LA INTELIGENCIA ARTIFICIAL: HACIA UNA INNOVACIÓN RESPONSABLE.

Amparo Alonso Betanzos[1]
1. CITIC (Research Center in Information and Communication Technologies). University of A Coruña. SPAIN

ABSTRACT

Artificial intelligence (AI) is a new revolution, with rapid advances driven by its increasingly widespread adoption across multiple sectors, from healthcare and education to finance and transportation. AI has many benefits, including improved decision-making, increased efficiency, and the creation of new business models. However, these advances entail seismic changes in the economy, employment, and education that need to be addressed to ensure responsible and ethical AI use. Key concerns include bias, discrimination, privacy, transparency, accountability, surveillance, and misuse. Bias in AI algorithms, for instance, can lead to unfair treatment that exacerbates social inequalities; privacy concerns arise from the vast amounts of personal data mined by AI systems; a lack of transparency in some complex algorithms poses challenges for accountability, making it difficult to understand and question output; and potential misuse or malicious use of AI, through deepfakes or autonomous weapons, raises security and ethical issues. Automation of tasks in many areas is also likely to lead to significant job displacement and will require adapted educational systems. Balancing the transformative potential of AI with the need to responsibly and sustainably address social and economic challenges, while promoting ethical use of AI that maximizes its benefits for humanity, requires a multidisciplinary effort involving governments, researchers, technologists, and society at large.

Keywords: Artificial Intelligence; Ethical and Trustworthy AI.

RESUMEN

La inteligencia artificial (IA) es una nueva revolución, con rápidos avances impulsados por su adopción cada vez más generalizada en múltiples sectores, desde la atención sanitaria y la educación hasta las finanzas y el transporte. La IA tiene muchos beneficios, como son una mejor toma de decisiones, una mayor eficiencia y la creación de nuevos modelos de negocio. Sin embargo, estos avances implican enormes cambios en la economía, el empleo y la educación que deben abordarse para garantizar un uso responsable y ético de la IA. Las principales preocupaciones incluyen el sesgo, la discriminación, la privacidad, la transparencia, la rendición de cuentas, la vigilancia y el uso indebido. El sesgo en los algoritmos de IA, por ejemplo, puede conducir a un trato injusto que exacerba las desigualdades sociales; las preocupaciones sobre la privacidad surgen de las grandes cantidades de datos personales extraídos por los sistemas de inteligencia artificial; la falta de transparencia en algunos algoritmos complejos plantea desafíos para la rendición de cuentas, lo que dificulta comprender y cuestionar los resultados; y el posible uso indebido o malicioso de la IA, a través de deepfakes o armas autónomas, plantea cuestiones éticas y de seguridad. También es probable que la automatización de tareas en muchas áreas conduzca a un importante desplazamiento de puestos de trabajo y requerirá sistemas educativos adaptados. Equilibrar el potencial transformador de la IA con la necesidad de abordar de manera responsable y sostenible los desafíos sociales y económicos, y al mismo tiempo promover un uso ético de la IA que maximice sus beneficios para la humanidad, requiere un esfuerzo multidisciplinar que involucre a gobiernos, investigadores, tecnólogos y a la sociedad en general.

Palabras clave: Inteligencia artificial; Retos éticos para IA confiables.

Correspondencia
Amparo Alonso Betanzos
Academician of the Royal Academy of Exact,
Physical, and Natural Sciences of Spain
E-mail: amparo.alonso.betanzos@udc.es

1. A BRIEF HISTORY OF ARTIFICIAL INTELLIGENCE

Artificial intelligence (AI), currently one of the most frequently mentioned terms in the media, refers to a present and future where intelligent programs and machines perform tasks that previously required human intelligence. However, although this term may seem to correspond to the 21st century, its origins date back to 1956, when, at a summer seminar at Dartmouth College (USA), a handful of scientists gathered together to ponder the question"Can machines think?" This question had been posed in 1950 by one of the fathers of AI, Alan Turing, in an article published in the journal ***Mind*** . Since then, the discipline has alternated between periods of commercial success and decades confined to academic environments (known as "AI springs" and "AI winters", respectively), due to issues related to with software maintenance, scalability, and other challenges that hindered AI applications to real-world problems. However, these cyclical phases of slow and mostly incremental progress have, by now, given way to AI as the leading technology driving the transition to a new society 5.0. This is partly due to several disruptions of recent years, brought about by factors such as the vast availability of data due to massive digitalization processes, the availability of the necessary computational power, and key advances in software. In this new context, companies find they can monetize previously non-exploited data. The convergence of all these factors has positioned AI as a mature technology with significant economic and social impact.

So, what is AI? Drawing on many different definitions, we can briefly define AI for our purposes as a scientific discipline aimed at creating computer systems capable of performing, with a degree of autonomy, tasks or processes that would require certain levels of intelligence in a human. This includes research into perception and visual recognition, natural language understanding and speech recognition, and automated learning and reasoning. In the recently published EU AI Act (the world's first such regulation)[1], an AI system is defined (in Article 3) as ***"a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments"***.

Until about a decade ago, advances in AI occurred slowly and incrementally. However, after the convergence of the key factors mentioned above, the field began to experience further disruptions that significantly altered the way AI was developed and how it was perceived by society. Notable early disruptions occurred in the 2010s, led by the system known as AlphaGo , which became the first AI system to defeat a professional human champion in Go, a game regarded as the most challenging of classic games (including chess) due to its complexity. Employing a different approach to the brute-force predecessors used for chess, AlphaGo successfully deployed new machine learning techniques such as deep neural networks and reinforcement learning . Those innovations opened up the way to rapid progress in other areas like natural language processing and computer vision, enabling AI systems to match and even surpass human-level performance in tasks such as object recognition, automatic translation, recommendation systems, and more . Advancing scientific discovery was exemplified in the 2010s by AlphaFold , a revolutionary system that accurately predicted the 3D structures of proteins for a large percentage of the human proteome (98.5%) , as well as for around 20 other organisms. By 2022, this tool had determined the structures of approximately 200 million proteins. More recent major disruptions have been the emergence of generative AI (which creates new content) and large language models (LLMs, trained on vast amounts of text); these have led to the develop-

[1] ***https://artificialintelligenceact.eu/ai-act-explorer/***

ment of systems like ChatGPT, DALL-E[2], Mistral[3], and similar. In quickly generating text, images, and voice outputs, such tools demonstrate an astonishing level of natural interaction between humans and machines. Interestingly, until the arrival of newer versions that integrate text, image, and more recently voice, there was little debate about AI's ability to achieve the original vision of its pioneering researchers: creating human-like general AI, i.e., systems as capable as humans in adapting, learning, and applying knowledge across a broad range of tasks and domains[4] . Advances until recently have been largely confined to narrow applications using intelligent systems developed to solve very specific problems within limited domains .

2. THE PRESENT CONTEXT AND REGULATORY EFFORTS

In the last two decades, AI has emerged as a transformative technology with the potential to significantly alter various aspects of our everyday lives and reshape the global socioeconomic landscape. From automating routine tasks to revolutionizing healthcare through personalized treatments, AI is redefining industries and presenting both new opportunities and challenges. The AI market is growing exponentially, with the compound annual growth rate estimated to be 35.7% for the 2024-2030 period . Research and innovation, mainly driven by major US technology companies, are leading to the rapid adoption of AI technology across industries and public administrations, resulting in significant changes in the automotive, healthcare, finance, and retail sectors, among others. As mentioned above, contributing to this acceleration is the availability of vast sets of historical data, accessible through advanced technologies like deep learning and generative AI techniques. Given the growing strategic importance of AI, the sector is currently characterized by a high level of merger and acquisition activity among industry leaders, driven by the desire to access new technologies and AI talent and establish a foothold in a rapidly growing market.

AI faces many challenges, and, as with all technological advances, those challenges are not limited to scientific progress or developments in various fields, but also imply significant repercussions for society in general. The transformation resulting from rapidly advancing technological change is profoundly social, and concerns raised include increased inequality and the impact on employment. To equip future generations with the skills necessary to thrive in an AI-driven world, the rapid pace of technological progress requires reforms at all education system levels and highlights the need for reskilling, upskilling, and lifelong learning.

Returning briefly to a more technical aspect, researchers who have experimented with GPT-4 (at the time, the latest version of ChatGPT) claim that it is approaching a level of intelligence comparable to that of humans, and even suggest that it could be viewed as an early-stage version of general AI . However, this remains highly debatable, as significant limitations have yet to be overcome, including the lack of cohesion between learned components, insufficient deep reasoning, limited common sense, and limited generalization abilities.

What is clear, however, is that beyond technical challenges, society needs to tackle the ethical dilemmas posed by AI systems. These include the large-scale use of data, decision-making biases, privacy concerns, security risks, the potential for widespread persuasion and influence, and questions of accountability. To address these issues, the call for AI regulation has become more pressing, with countries currently taking different approaches. In the EU, the focus is on creating ethical, trustworthy, and human-centric AI systems. Its AI Act[5], which entered into force on 1 August 2024, is broadly based on a risk-level approach; it also includes content regarding generative AI models that was not on the table when the regulation was originally proposed in 2021, highlighting the rapid evolution of technology. While the USA as yet has no binding federal law that specifically regulates AI, the White House (borrowing many foundational ideas from the EU framework) issued an Executive Order in October 2023 aimed at ensuring safe and reliable AI deployment. The approach in the USA to date has been decentralized, allowing states to draft their own regulations. Colorado, in May 2024, became the first US state to enact a regulatory framework governing AI development and use, while California, home to the

[2] ***https://platform.openai.com/docs/models***

[3] ***https://mistral.ai/***

[4] ***https://lamaletadeportbou.com/articulos/sobre-la-inteligencia-de-la-inteligencia-artificial/***

[5] ***https://artificialintelligenceact.eu/***

main US technology companies, has recently passed bills regulating AI (testing for threats to critical infrastructure, curbing children's exposure to algorithms, limiting the use of deepfakes, and more). Finally, although with a different focus than the EU and USA, China has also implemented legislation (in 2022) that regulates recommendation systems and generative AI.

The AI regulatory landscape is evidently far from uniform, and concern is growing regarding international coordination in this area. As AI technologies rapidly advance, international bodies are collaborating to develop multilateral AI governance frameworks. Intergovernmental bodies (such as UNESCO and ISO) and several regional entities are all engaged in efforts to create cohesive global frameworks that address AI's challenges and ensure its ethical use. Taking steps to guide this process, the UN has published a resolution[6] that underscores the need to ground AI development in principles aligned with international human rights law, reinforces the importance of ethical standards, and urges nations to prioritize transparency, accountability, and fairness in their AI policies. Maintaining trust and sustainable growth in the AI industry requires that the private sector companies playing a crucial role in shaping AI's future develop and deploy AI technologies that comply with emerging regulations. Ensuring their compliance is essential not only for ethical reasons but also to ensure the prioritization of data privacy, transparency, and legal liability. However, the challenge is that compliance will vary significantly across different jurisdictions. This fragmentation risks the division of the global AI ecosystem into separate regulatory regions, each with its own set of rules and standards. Despite these challenges, the push for AI regulation is a truly global effort, with laws and policies being developed on six continents. The AI Safety Summit[7], the G7 agreement on Guiding Principles and a Code of Conduct on Artificial Intelligence[8], regulation in China, the US Executive Order, and the EU AI Act are all examples of wide-ranging efforts to uphold responsible AI governance, demonstrating that nations are starting to align on key principles.

Ultimately, for AI to benefit society, its governance must be rooted in international cooperation. While national regulations may differ, the collective effort of governments, international organizations, and the private sector will be critical in ensuring that AI serves humanity responsibly and equitably. While international cooperation is crucial to long-term solutions, in the short term we need to address specific high-risk areas, including data privacy, ownership, access, bias, explainability, transparency, and sustainability. The following sections delve deeper into these challenges.

3. THE CHALLENGES AHEAD

AI undoubtedly represents a technological opportunity that we must learn to harness effectively. As highlighted by the authors of one of the most widely used AI textbooks : ***"Our civilization is the product of our human intelligence. If we gain access to greater intelligence in machines, the ceiling of our ambitions is substantially raised. The potential for AI to free humanity from repetitive work and dramatically increase the production of goods and services could herald an era of peace and prosperity. The ability to accelerate scientific research could result in the cure of diseases and provide solutions for climate change and resource scarcity"***.

New areas of AI application are continuously emerging, with the fintech, education, and healthcare sectors set to undergo significant transformations due to AI's influence. Figure 1, for instance, shows how the number of AI medical devices approved by the US Federal Drug Administration has grown apace in just a decade, suggesting that the potential to develop new drugs, create more personalized treatments, and provide individualized health input throughout our lives is becoming a reality. Virtual doctors and educators can potentially provide low-cost access to high-quality services, reaching areas where these services are unavailable, and thereby helping restore our welfare society in an economically viable way. AI is also being explored as a tool to assist populations at risk of exclusion and in underdeveloped geographic areas, and, importantly, as a critical component in the fight against climate change. Also attracting increasing attention is how AI can enhance civic participation by enabling citizens to engage more actively in the governance of their institutions.

[6] ***https://documents.un.org/doc/undoc/ltd/n24/065/92/pdf/n2406592.pdf***

[7] ***https://www.aisafetysummit.gov.uk/***

[8] ***https://digital-strategy.ec.europa.eu/en/library/hiroshima-process-international-guiding-principles-advanced-ai-system***

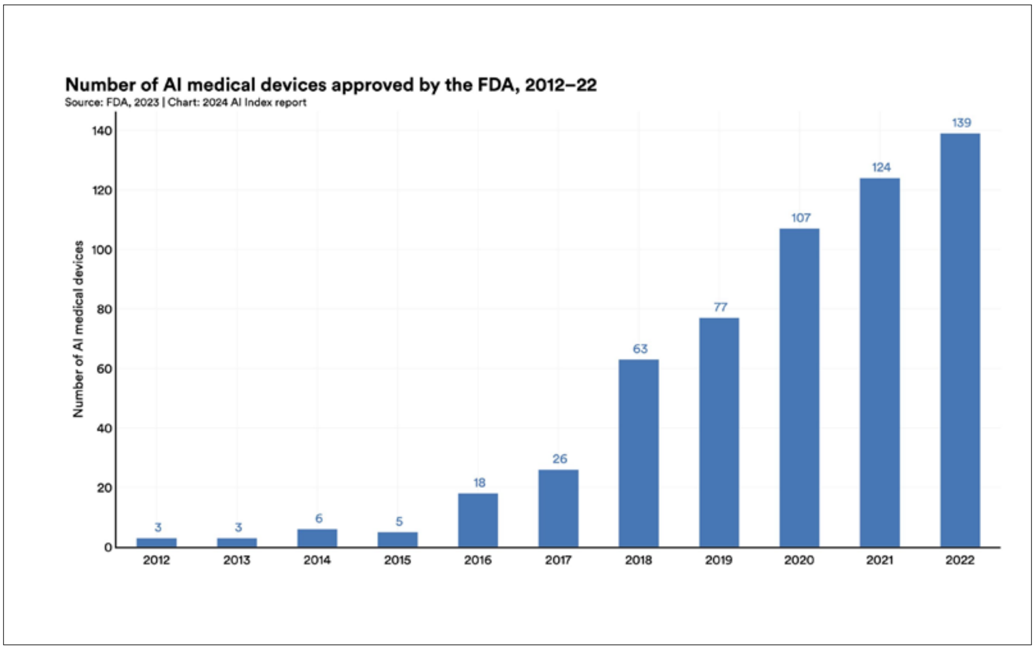

Fig. 1. Artificial intelligence is increasingly being used for real-world healthcare, with the number of medical devices using AI multiplying more than 40-fold since 2012.

However, we face considerable risks if AI technologies are not developed and used responsibly. Some of these challenges are briefly explored below (see figure 2).

3.1. Ethical challenges

Key factors driving the rapid advancement of AI are the availability of vast amounts of data across virtually all sectors and the fact that it is a transversal discipline applicable to nearly every field. Adherence to fundamental ethical principles, as outlined in the European Commission's 2019 publication on Ethics Guidelines for Trustworthy AI[9], is crucial. Six fundamental aspects, discussed below, are data privacy and security; bias, discrimination, and the perpetuation of prejudices; transparency and explainability; responsibility and accountability; environmental sustainability; and social impact and justice.

Data privacy and security. The volume of data we generate daily is enormous. In 2023 alone, around 120 ZB (zettabytes) of data were created, and the trend is for the volume to double every two years. As previously noted, virtually all sectors and human activities are undergoing digitalization, with data being collected from sensors, mobile devices, web environments, etc, in various formats (text, video, images, voice, etc). These large volumes of data are used to train AI algorithms. However, since some of this data may be sensitive personal or confidential information (e.g., medical or financial records), it is crucial to implement a governance model that ensures data privacy and security of access. Ensuring data privacy means protecting individuals' identities and guaranteeing that their data is only used with their consent for clearly specified purposes. Data access security is essential to prevent unauthorized access, breaches, or misuse. For instance, if sensitive information on a cured cancer survivor is accessed without authorization and revealed, it could affect the decision on whether this person is granted medical insurance. Guarantees must be put in place that ensure that data is protected, individual privacy is respected, identity is not revealed, and data is used ethically and only for the specified purpose.

Bias, discrimination, and the perpetuation of prejudices. Many routine applications and tools rely on AI. For instance, AI algorithms perform part or all of the tasks related to obtaining directions to a specific location or receiving content recommendations on a website. AI also assists in decision-making in more critical contexts, such as selecting job can-

[9] ***https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai***

didates, approving bank loans, and diagnosing diseases, although some of the regulations mentioned earlier mandate that final decisions should be made by human experts. AI algorithms can more rapidly and more accurately automate decision-making processes involving the analysis of large datasets than humans. However, since these algorithms learn from real-world data which is inherently imperfect, detecting bias is challenging, as the data might reflect historical inequalities or social stereotypes. If not addressed, such biases can inadvertently replicate and even exacerbate existing discrimination, producing unequal outcomes for different groups. This issue is particularly of concern when AI is used for critical decision-making in areas like hiring, loans, law enforcement, and healthcare. For instance, in 2019 in the USA, a hospital algorithm designed to predict which patients would require additional medical attention was significantly biased in favor of white patients. The bias, reduced by 80% once identified, was discovered when it was determined that using healthcare expenditure as an input parameter was not appropriate because economic disparities meant that healthcare expenditure by black patients with similar health conditions was frequently lower than that of white people . Another example is facial recognition systems, which are less accurate for people with darker skin and for females . Although error rates have been reduced in recent years, the accuracy rates of above 90% reported for most systems cannot be considered universal, as their training datasets are frequently unbalanced. AI-based facial recognition, for instance, has caused several people of color to be falsely identified as suspects in criminal investigations, wrongfully arrested and charged, or denied employment[10]. AI algorithm design itself can be discriminatory. Early virtual assistants, typically designed with female voices and submissive characteristics, reinforced traditional gender roles[11] by subtly perpetuating the stereotype of the subservient and accommodating woman. AI aimed at predicting potential repeat offenders can result in unfair treatment of certain demographic groups, wrongly denied access to parole programs[12]. To build ethical AI systems that do not reinforce harmful prejudices or contribute to social inequality, biases need to be actively identified and mitigated in both datasets and algorithms, which can be done by auditing AI algorithms and systems for fairness, using more diverse datasets, and ensuring that AI decision-making is transparent and explainable.

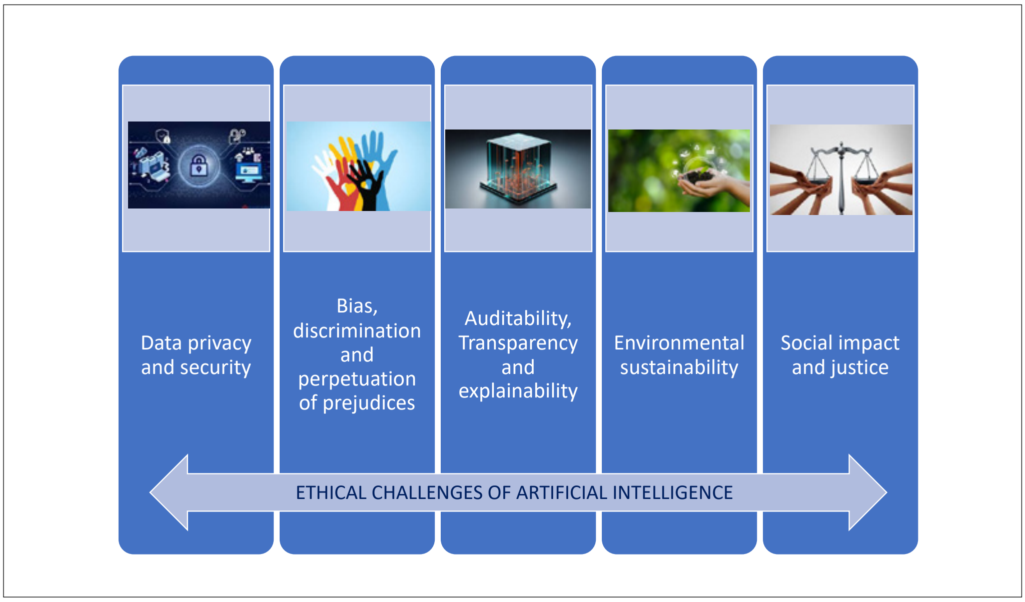
Data privacy and security
Bias, discrimination and perpetuation of prejudices
Auditability, Transparency and explainability
Environmental sustainability
Social impact and justice
ETHICAL CHALLENGES OF ARTIFICIAL INTELLIGENCE

Fig. 2. Some ethical considerations to be taken into account in AI.

[10] https://www.irwinmitchell.com/news-and-insights/expert-comment/post/102j5zu/uber-eats-compensates-driver-after-its-ai-facial-recognition-tool-discriminated-a

[11] https://unesdoc.unesco.org/ark:/48223/pf0000367416

[12] https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Transparency and explainability. AI algorithms are becoming increasingly accurate, but this precision often implies greater complexity (numerous parameters and processing layers), particularly in models based on deep learning . Consequently, although inputs and outputs are known, the black-box nature of AI systems is becoming increasingly opaque, which makes it challenging to understand or interpret exactly how the output was produced. Transparency refers to the ability to understand how an AI system operates (data sources, algorithms, and processes), whereas explainability refers to a system's decisions being understandable to humans in terms of clear and accessible reasons for why a particular outcome was obtained. Transparency is crucial in areas like healthcare, criminal justice, and finance, where the impact of AI decision-making may be critical. An AI system that does not provide clear reasons or justifications for its decisions makes it difficult to evaluate whether a decision is fair, unbiased, or even correct; people should be able to understand, for instance, on what basis an AI algorithm denies a loan or diagnoses a medical condition. Explainability when non-existent can lead to mistrust in AI systems, especially when outcomes seem unjust or biased, and also makes it hard to hold them accountable or challenge their decisions. Transparency and explainability are both key to bolstering trust and ensuring fairness in automated decision-making. Promoting both requires designing systems that can articulate decision-making processes in a way that is understandable to humans, e.g., simplifying models, using algorithms that are inherently more interpretable, or developing techniques that provide insights into complex models. In high-stakes environments in particular, AI systems need to be transparent and explainable so that they can be fully audited to ensure that users, regulators, and stakeholders can trust their operations and their outputs.

Responsibility and accountability. These aspects address the need for someone to be held responsible or accountable for discrimination, harm, or error resulting from increasingly complex AI decisions and interactions between humans and machines. The many different parties involved in the design, commercialization, and use of AI systems, however, complicate the identification of a possible culprit for poor or harmful outcomes. Responsibility, referring to the obligations of all those involved in the design, development, deployment, and use of AI systems, concerns AI companies in general, policymakers, and users, each playing a different role. Designers and developers are responsible for creating systems that are free from bias, safe, transparent, and reliable; deployers must ensure that their AI systems comply with ethical standards and the relevant regulations; and users need to understand the capabilities and limitations of AI to avoid misuse. Accountability addresses the need to determine who is liable when the system fails, harms, or errs. This is a particularly complex issue, first because multiple actors (developers, companies, organizations, users, etc) are typically implicated, and second, because both ethical and legal considerations apply; for an accident involving a self-driving vehicle, for instance, should the programmer, the distributor, the maintenance technician, or the driver be held responsible? In the USA in March 2018, an Uber driver was found guilty of negligent homicide because a built-in emergency braking system had been disabled while the car was in autonomous mode[13]. However, who might have been taken to court if that system had been activated?

Environmental sustainability. Sustainability focuses on minimizing the environmental impact of AI technologies while ensuring that development and deployment contribute positively to long-term societal goals. While AI advances have led to the development of increasingly accurate models, they have also significantly increased the computational resources required to train and run those models[14], raising concerns about the carbon footprint and overall environmental sustainability of AI systems. A major sustainability challenge is the energy consumption associated with training and running complex models, and especially deep learning algorithms. Training ChatGPT-3, for instance, consumed approximately 1,300 MWh of energy, for carbon emissions of close to 550 tons equivalent to 3 round trip flights between New York and San Francisco , or driving 123 gasoline vehicles during a year. Consuming even more energy than model training are real-time AI applications like speech recognition and recommendation systems. Energy is also consumed in our use of these systems, in which the energy demands appear to be considerably higher

[13] *https://www.nytimes.com/2018/05/24/technology/uber-autonomous-car-ntsb-investigation.html*

[14] *https://theconversation.com/is-generative-ai-bad-for-the-environment-a-computer-scientist-explains-the-carbon-footprint-of-chatgpt-and-its-cousins-204096*

[15],[16]: Approximately half a million kilowatt-hours of electricity are needed daily to handle around 200 million requests. For perspective, this amount of energy could power an average U.S. household, which consumes about 29 kWh per day, for over 17,000 days. Another sustainability concern is the growing need for extensive data storage infrastructure as AI systems come to rely on increasingly large datasets for training. The data centers housing the servers that process and store data consume vast amounts of electricity and water for cooling. In 2019, MIT reported that cloud computing had a higher carbon footprint than the airline industry and that a single data center was capable of consuming as much electricity as 50,000 homes[17]. These resource needs continue to rise, with new resource-guzzling AI tools, such as the recent multimodal ChatGPT that combines text, images, and voice, suggesting that this trend is unsustainable. As for the substantial water usage required to cool servers, data centers use water from 90% of the U.S. watersheds, ranking them among the top 10 commercial water users in that country [18]. As an example, in 2021, to maintain the appropriate temperature in its data centers, Google needed around 1,7 million liters of water per day [19], [20]. It is clear that we urgently need to move towards adopting a more sustainable approach to developing and deploying AI models that enhance accuracy while also reducing the corresponding carbon footprint—ideally aiming for net zero emissions to prevent AI from becoming an environmental time-bomb. More sustainable, or green, AI is characterized by a low carbon footprint, smaller models, lower computational complexity, and greater transparency, achievable through strategies such as leveraging edge computing capabilities and employing modular deep learning models. Green AI has two dimensions, referred to as "green-in" AI and "green-by" AI. Green-in AI is focused on designing more energy-efficient systems, optimizing algorithms, and adopting sustainable data storage and processing practices, whereas green-by AI is concerned with enhancing eco-friendly practices and applying AI to environmental and social challenges like climate change, waste management, energy efficiency, conservation efforts, etc . AI models, for instance, can predict extreme weather patterns, optimize energy grids, and help monitor deforestation and biodiversity loss. Finally, an emerging concept is circular AI focused on sustainability-by-design, which ensures that systems can be reused, repurposed, or recycled at the end of their useful life. This involves building AI systems that are modular, transparent, and easy to update and maintain, reducing the need for frequent replacements and the retraining of new models.

Social impact and justice. AI can have a significant positive impact on society, but it also poses certain challenges. One issue is the exacerbation of inequalities between countries. While AI has the capacity to drive economic growth, improve healthcare, and enhance education, uneven development and deployment across different world regions will widen existing gaps between technologically advanced developed countries and less developed countries. North America, Europe, and certain parts of Asia have the technical, financial, and educational resources needed to invest in AI research and to develop sophisticated AI systems, but a divide is growing with most developing countries, which lack the necessary infrastructure, expertise, and capital to keep pace with AI advances. Thus, while some nations can leverage AI-driven innovation to their benefit, other nations are left behind, unable to leverage AI even for such critical areas as healthcare, education, and economic growth. Another relevant issue is the concentration of talent. Research at the frontier of knowledge tends to be clustered in a reduced number of elite universities, private companies, and government institutions, leading to unequal opportunities for innovation. Talent is also becoming concentrated where better opportunities are available, which means a brain-drain of key human resources from poorer countries that, in turn, leads to their reliance on technology imports with the consequent dependence on more advanced economies. Another source of concern regarding many developing countries, deri-

[15] ***https://towardsdatascience.com/chatgpts-electricity-consumption-7873483feac4/***

[16] ***https://alltechmagazine.com/chatgpt-uses-17000-times-more-energy-us-family/***

[17] ***https://www.scientificamerican.com/article/science-needs-to-shrink-its-carbon-footprint/***

[18] ***https://www.hivenet.com/post/tech-needs-to-reduce-its-water-consumption-in-a-thirsty-world***

[19] ***https://blog.google/outreach-initiatives/sustainability/our-commitment-to-climate-conscious-data-center-cooling/***

[20] ***https://www.techtarget.com/searchdatacenter/tip/How-to-manage-data-center-water-usage-sustainably***

ved from the fact that AI systems need to be trained with large, high-quality datasets to work effectively, is the lack of local data and difficulties in generating and collecting data, due to an underdeveloped digital infrastructure. Using AI models trained on datasets from the developed world is not a viable option, as such models perform poorly or have unintended consequences when applied in different cultural, social, or economic contexts. This misalignment can perpetuate inequality by failing to meet the unique needs of underrepresented populations. The potential for AI to widen economic disparities between countries is also a matter of concern, as AI-driven productivity and efficiency gains in advanced economies will lead to greater economic growth and increased competitiveness, leaving developing nations at risk of being left behind in the global economy. The scenario is one in which the benefits of AI will be disproportionately captured by a few wealthy countries, thereby deepening the economic divide and reducing opportunities for equitable development. Efforts to close the gap between advanced and developing countries require international cooperation in the form of global partnerships that facilitate knowledge transfers, share AI resources, and support initiatives that prioritize the needs of developing countries. Additionally, investment in AI education and skills training in underdeveloped regions can help cultivate the local talent needed for these countries to build their own AI capabilities. By fostering inclusive AI development, the global community can work to narrow the digital divide and ensure that AI contributes to a more just and equitable world for all.

3.2. Economy, employment, and education

The influence of AI in the economy is indisputable, with generative AI in particular expected to have a substantial impact on the global economy. In 2023, the five largest companies in the world by market capitalization were US technology companies (Apple, Microsoft, Nvidia, Alphabet, and Amazon), all involved in AI. According to the European Parliament, the global AI market, valued at over €130 billion in 2023, is projected to expand to nearly €1.9 trillion in value by 2030[21]. Most investment in AI now comes from the private sector, with the USA leading investment in 2023 (€62.5 billion), followed by China (€7.3 billion) . Private AI funding in the EU and UK combined was €9 billion in 2023; between 2018 and the first three quarters of 2023, the EU received €32.5 billion in AI investments, compared to €120 billion secured by the USA, with firms like OpenAI and Anthropic particularly contributing to widening the gap[22]. AI use has grown significantly in organizations and is being adopted in various operational areas, most particularly in the finance, pharmaceuticals, healthcare, and education sectors, highlighting how AI, as a strategic priority, is significantly transforming business operations across a wide range of industries. Emerging AI startups at the forefront in developing cutting-edge robotics, biotechnology, and personalized education solutions are attracting substantial venture capital and becoming significant contributors to the growing technology industry. While they challenge established firms, startups also broaden business opportunities and foster a more dynamic and varied economy.

The job market is also undergoing profound transformation by AI, with a recent IMF study concluding that AI will affect around 40% of jobs worldwide, replacing some and complementing others . While employees performing routine data processing and assembly line work are clear candidates to be replaced by AI systems, a differentiating aspect from previous technological change is that AI is also affecting highly skilled and qualified jobs. The significant changes to employment patterns will create both opportunities (new and better jobs) and challenges (downgraded and disappearing jobs) for the workforce in several sectors, including human resources, production, risk management, and strategic planning. One of the most notable impacts of AI is the automation of routine and repetitive tasks, leading to increased efficiency and cost savings for companies, but also resulting in job displacement for roles that are highly susceptible to automation, such as in the administrative and legal fields[23]. Some jobs will remain broadly the same, except that the human will require new skills to be able to effectively interact with technology. Leading to potentially higher wages due to increased skills are jobs where human intelligence and AI work in cooperation. According to the IMF, AI's impact on

[21] *https://www.eca.europa.eu/ECAPublications/SR-2024-08/SR-2024-08_EN.pdf*

[22] *https://stateofeuropeantech.com/*

[23] *https://www.gspublishing.com/content/research/en/reports/2023/03/27/d64e052b-0f6e-45d7-967b-d7be-35fabd16.html*

employment will vary depending on a country's economic development level, ranging from up to 60% in advanced economies, to 40% in emerging economies and 26% in low-income countries[24]. The ILO (International Labour Organization) has also warned that AI could disproportionately affect women, particularly in administrative roles where female employment rates are high[25].

According to the OECD, jobs in manufacturing and finance are likely to be the most negatively impacted by rapid advances in technology , while jobs in agriculture, farming, fishing, associative activities, extractive industries, and construction are expected to remain largely unaffected. How work is structured will need to evolve, especially in terms of upskilling to enable effective human-machine collaboration. AI adoption could potentially exacerbate inequality by opening a wider divide between individuals with and without sought-after skills—a disparity that is likely to be further exacerbated in individual economies and across the global economic landscape. This scenario is compounded by the fact that competition is already intensified by workforce globalization and remote work, as local companies now have to face off against international firms that offer fully remote positions, with better salaries, more attractive projects, and more dynamic career paths. This issue poses a significant challenge for companies and nations alike, as rapid technological progress (including in generative AI), decreasing costs, and the increasing availability of AI-savvy workers suggest that OECD countries may witness dramatic shifts in the coming decade.

New job profiles are likely to emerge in areas such as programming, consultancy, scientific and technical fields, telecommunications, media, and publishing, and evidently, the demand will grow, and wages will increase, for professionals with advanced technical skills and expertise in AI, machine learning, data science, and robotics. Contributing to job creation and economic growth will be entirely new AI-related industries and business models, such as those centered around AI ethics, AI-driven healthcare solutions, and smart technologies. Moreover, AI has the potential to enhance human work rather than replace it, since by automating routine tasks, it enables workers to focus on more complex and creative aspects of their jobs. In healthcare, for instance, AI can assist doctors by analyzing medical data, enabling them to spend more time on patient care and complex diagnostics, while in the creative industries, AI tools can implement repetitive tasks and so free up professionals to engage in higher-level creative processes.

In relation to education, integrating AI in the workforce poses immediate and significant challenges regarding the future. There is a growing need for reskilling and upskilling initiatives to help workers transition to new roles and adapt to the evolving job market. Educational institutions and organizations need to develop training programs that equip individuals with the skills needed for AI-related jobs and that ensure a suitably skilled workforce for the future. Educational levels from early childhood to pre-tertiary stages will need to focus on the development of basic skills such as communication abilities, critical thinking, and creativity, while third-level education will need to ensure deeper scientific and technological knowledge and to offer flexible and interdisciplinary curricula that include ethics and responsibility issues.

In sum, given that AI is reshaping the employment landscape in negative as well as positive ways, to develop the talent that will be needed in the coming years and to ensure that workers are equipped to thrive in an AI-driven economy, proactive measures are needed that address potential job displacement, reorient the workforce, and adapt education systems. Crucial to navigating the impact of this transformative technology on society is balancing the economic benefits of AI with labor- and education-oriented strategies.

4. CONCLUSIONS

A balanced and mindful approach is urgently needed regarding the ethical aspects of AI and its potential impact on the economy, employment, and education. Regarding the economy, AI offers great opportunities for efficiency and innovation, but also raises challenges related to the concentration of power and inequalities between countries and regions. It is therefore crucial to

[24] *https://www.imf.org/en/Blogs/Articles/2024/01/14/ai-will-transform-the-global-economy-lets-make-sure-it-benefits-humanity*

[25] *https://webapps.ilo.org/static/english/intserv/working-papers/wp096/index.html*

promote policies that ensure an equitable distribution of the benefits of AI. In terms of employment, automation and the transformation of production processes are reshaping the labor landscape, creating new jobs while rendering others obsolete. This calls for a commitment to reskilling, upskilling, and lifelong learning to prevent segments of the population from being left behind. It is also important to ensure that AI's impact on workers is positive and that effective human-machine cooperation is fostered. Finally, the educational sector will need to adapt, adopting a more interdisciplinary approach that promotes not only technical skills, but also critical thinking, creativity, and ethical responsibility. Curricula in particular need to include the development of the kind of human skills that complement AI, thereby ensuring that future generations are equipped to thrive in a constantly evolving environment. Finally, AI must be designed, developed, deployed, and used transparently, fairly, and responsibly, respecting fundamental rights and avoiding the perpetuation of biases and inequalities. Only through ethical governance can we ensure that AI has a positive and sustainable impact on society.

TRANSPARENCY STATEMENT

The author of this article declares that they have no conflicts of interest regarding the content presented in this work.

BIBLIOGRAPHY

1. Bolón-Canedo, V., Morán-Fernández L., B. Cancela, and A. Alonso-Betanzos. 2024. "A Review of Green Artificial Intelligence: Towards a More Sustainable Future." Neurocomputing 599. hhttps://doi.org/10.1016/j.neucom.2024.128096.

2. Bubeck, S., Chandrasekaran V., Eldan R., and et col. 2023. "Sparks of Artificial General Intelligence: Early Experiments with GPT-4." arXiv. https//doi.org/10.48550/arXiv.2303.12712.

3. Cazzaniga, M., Jaumotte F., Li L., Melina G., Panton A. J., Pizzinelli C., Rockall E. J., and M. Mendes Tavares. 2024. Gen-AI: Artificial Intelligence and the Future of Work. International Monetary Fund, Staff Discussion Notes No. 2024/001.

4. Goodfellow, I., Bengio Y., and A. Courville. 2016. Deep Learning. MIT Press.

5. Grand View Research. 2024. Artificial Intelligence Market Size, Share & Trends Analysis Report by Solution, by Technology (Deep Learning, Machine Learning, NLP, Machine Vision, Generative AI), by Function, by End-Use, by Region, and Segment Forecasts, 2024 - 2030.

6. Jumper, J., Evans R., Pritzel A., and et col. 2021. "Highly Accurate Protein Structure Prediction with AlphaFold." Nature 596: 583–89.

7. Lane, M., Williams M., and S. Broecke. 2023. "The Impact of AI on the Workplace: Main Findings from the OECD AI Surveys of Employers and Workers." Organization for Economic Cooperation; Development (OECD). Technical report 288. https://doi.org/https://doi.org/https://doi.org/10.1787/ea0a-0fe1-en.

8. Larson, E. K. 2021. The Myth of Artificial Intelligence: Why Computers Can't Think the Way We Do. Shackleton Books.

9. Leslie, D. 2020. "Understanding Bias in Facial Recognition Technologies: An Explainer." The Alan Turing Institute. https://doi.org/https://doi.org/10.5281/zenodo.4050457.

10. López de Mántaras, R., and P. Meseguer González. 2017. Inteligencia Artificial. Los libros de la Catarata.

11. Maslej, N., Fattorini L., Perrault R., Parli V., Reuel A., Brynjolfsson E., Etchemendy J., et al. 2024. "The AI Index 2024 Annual Report." AI Index Steering Committee, Institute for Human-Centered AI, Stanford University.

12. Mittermaier, M., M. M. Raza, and J. C. Kvedar. 2023. "Bias in AI-Based Models for Medical Applications: Challenges and Mitigation Strategies." NPJ Digit. Med. 6 (113). https://doi.org/10.1038/s41746-023-00858-z.

13. Patterson, D., Gonzalez J., and Le Q. et col. 2021. "Carbon Emissions and Large Neural Network Training." https://arxiv.org/abs/2104.10350. https://arxiv.org/abs/2104.10350.

14. Russel, S., and P. Norvig. 2022. Artificial Intelligence: A Modern Approach, 4th Ed. Pearson.

15. Silver, A.& Maddison, D. & Huang. 2016. "Mastering the Game of Go with Deep Neural Networks and Tree Search." Nature 529 (7587): 484–89.

16. Sutton, R. S., and A. G. Barto. 1998. Reinforcement Learning: An Introduction. Bradford Books.

17. Tunyasuvunakool, K., Adler J., Wu Z., and et col. 2021. "Highly Accurate Protein Structure Prediction for the Human Proteome." Nature 596: 590–96.

18. Turing, A. M. 1950. "Computing Machinery and Intelligence." Mind 59 (236): 433–60.

19. Vries, A. de. 2023. "The Growing Energy Footprint of Artificial Intelligence." Joule 7: 2191–94.

If you want to quote our article:

Alonso Betanzos A. Navigating the Ethical and Societal Challenges of Artificia Intelligence: Ensuring Responsible Innovation. RACSG.2025;113(01):54-65. rac.2025.113.1.org04