Special Issue Introduction

Artificial Intelligence for a Fair, Just, and Equitable World

Ángeles Manjarrés
Dept. Inteligencia Artificial, ETSI, Universidad Nacional de Educación a Distancia (UNED), 28040 Madrid, Spain

Celia Fernández-Aller
ETSISI, Universidad Politécnica de Madrid (UPM), 28031 Madrid, Spain

Maite López-Sánchez
Dept. Mathematics and Computer Science, Universitat de Barcelona, 08007 Barcelona, Spain

Juan Antonio Rodríguez-Aguilar
Artificial Intelligence Research Institute (IIIA), Spanish Council for Scientific Research (CSIC), 08193 Bellaterra, Spain

Manuel Sierra Castañer
Dept. Señales, sistemas y radiocomunicaciones, ETSIT, Universidad Politécnica, 28040 Madrid, Spain

From the 1970s onward, we started to dream of the leisure society in which, thanks to technological progress and consequent increase in productivity, working hours would be minimized and we would all live in abundance. We all could devote our time almost exclusively to personal relationships, contact with nature, sciences, the arts, playful activities, and so on. Today, this utopia seems more unattainable than it did then. Since the 21st century, we have seen inequalities increasingly accentuated: of the increase in wealth in the United States between 2006 and 2018, adjusted for inflation and population growth, more than 87% went to the richest 10% of the population, and the poorest 50% lost wealth [1]. Following the crisis of 2008, social inequalities, rights violations, planetary degradation, and the climate emergency worsened and increased (see [2]). In 2019, the world's 2153 billionaires had more wealth than 4.6 billion people [3]. The World Bank estimates that COVID-19 will push up to 150 million people into extreme poverty [4].

Digital Object Identifier 10.1109/MTS.2021.3056292
Date of current version: 16 March 2021.

The future brought to us by technological advances and, in particular, by the spectacular development of data science and artificial intelligence (AI) evokes the dystopian future painted by numerous science-fiction stories. These stories speak to us through powerful allegories of human existence in the AI era: automated, dehumanized, and depressed societies, solitude in the company of machines, predation of the planet, ecological degradation, totalitarian governments, and strong inequalities in access to resources and power, alienation, and exclusion. In this scenario, the elite people monopolize and use sophisticated intelligent technology as an instrument of commodification, repression, exploitation, manipulation, and control of the dispossessed.

The rise of AI has fueled the debate on the potential contribution of new technologies to the creation of a prosperous and equitable world, as against the countless ethical, moral, legal, humanitarian, and political–social risks, as well as physical and mental health risks. The ethical questions raised by intelligent systems are currently being addressed by diverse national and international governmental bodies [5]–[7], professional bodies [8], academia [9]–[11], and

the industry (initiatives on AI ethical codes such as those of Google, IBM, Microsoft, and Intel). In essence, these initiatives aim to identify the potential benefits and risks, and issue recommendations on the principles to be followed by the different actors involved.

However, this ethical debate is taking place mostly in high-income countries so that much of it is of little relevance to the more than 700 million people living in extreme poverty. Reciprocally, ethical questions that greatly affect marginalized populations are not treated with the importance they deserve in this debate.

Universal-ethics considerations gave rise to the United Nations (UN) Agenda for sustainable development, to be reached by 2030. Eradicating poverty is a central objective of the sustainable development goals (SDGs), and though the emphasis is on low- and middle-income countries (LMICs), they also target the growing pockets of underdevelopment in high-income countries. There is a growing interest in the role that AI can play in achieving these objectives on the part of international organizations, such as UN Global Pulse [12], UN High Commissioner for Refugees (UNHCR) [13], the UN International Children's Emergency Fund (UNICEF) Global Innovation Centre [14], the World Wide Web Foundation [15], the International Telecommunications Union [16], and even the World Economic Forum [17].

A wide view of ethics focuses not only on risk mitigation but also on potentialities, and from such a view arises the ethical imperative to harness AI technologies to the benefit of humanity to improve quality of life for all rather than contributing to perpetuating systemic injustices. To this end, more multi- and interdisciplinary R&D in the potential of AI to contribute to the SDGs is urgently needed; a practical research that goes beyond cataloging risks and potentialities, in part as a counterweight to the heavily plugged corporate sector view on AI ethics, which is often little more than "ethicswash" for a program in which the effect of AI/S development and deployment will most likely be to increase inequality [18], [19].

First, there is a need to study the current panorama of AI applications in sectors crucial to the UN SDGs, to share the lessons learned in applying them, to identify strengths and weaknesses, and to document and disseminate the development and deployment of the most significant innovative applications. Attention should be drawn to the idiosyncrasy of each application context (cultural, climatic, environmental, organizational, infrastructural, socioeconomic, etc.) and the particular impact AI-based technological innovation can have on each.

Second, progress in standards and R&D methodological and technical tools that guide the development of ethical AI is also essential. Ethical AI should be respectful of and even actively committed to fundamental human rights and of the particular values of the culture where it is implemented and should take into account the idiosyncrasy of each context. Additionally, these methodological and technical tools could ensure compliance with regulations, laws, and policies, particularly those focusing on protecting and empowering the most vulnerable and marginalized. Although manuals of good business practices are also necessary, in the academic field, there is a need for independent and scientifically rigorous research, with an empirical dimension which, so far, is mostly lacking. Academic research, private sector self-regulation, and legislation are necessary and complementary actions.

In this special issue, we aim to illustrate this R&D path that would confer a decisive role to AI in achieving the SDGs, by presenting a set of articles mainly selected from the submissions to the workshop "Advancing Toward the SDGs Artificial Intelligence for a Fair, Just, and Equitable World," held in conjunction with the "European Conference on Artificial Intelligence" in September 2020. The spirit of the 2030 Agenda, as reflected in [20], is expressed as an "inescapable transformation," that is, a profound change in the systems and structures in which all organizations and individuals in society must participate. In the face of the dystopian futures of the advances of AI augur, there is the option of an AI that catalyzes that necessary transformation toward a fair, just, and equitable world.

This special issue first presents "AI4Eq: For a True Global Village not for Global Pillage," by Manjarrés *et al*. as a call for action on researchers to participate and promote an interdisciplinary research field "AI for Equity" dealing with the distinctive challenges posed by AI technologies in the context of a human rights-based approach to sustainable development. The authors show how AI4Eq occupies a particular area within ICT4D due to the very significant ethical and philosophical problems and dilemmas that it gives rise to, and to the fact that many of the risks associated with ICT, in general, are magnified in the

case of AI. They present a first exploration of the way forward for AI4Eq and discuss the relevance of multidisciplinary, multilevel, and multifactor alliances that imply the private sector and civil society.

The rest of the special issue is then divided (conceptually) into three parts: first, a set of organizational initiative addressing these issues; second, a set of papers discussing real experience with SDG-oriented AI applications; and third, a set of papers describing tools (legislative, methodological, and technical) to support design, development, and deployment of SDG-oriented AI, reflecting on their strengths and weaknesses with an emphasis on reducing inequalities.

In the first part, the reflection on initiatives addressing the issue of AI4Eq is from three different organizations: 1) the IEEE; 2) the European Commission; and 3) the Latin America and the Caribbean (LAC) alliance.

In an Opinion piece, Elizabeth D. Gibbons in "Toward a More Equal World: The Human Rights Approach to Extending the Benefits of Artificial Intelligence," emphasizes the dangers of AI driving inequality, concentrating wealth, resources, and decision-making power in the hands of a few countries, companies, or citizens. She stresses the need for adopting a human rights framework in AI design, development, and deployment and introduces the work of the Sustainable Development Committee of the IEEE's ethically aligned design project [8]. This committee (whose multidisciplinary members included academics, lawyers, robotics engineers, businessmen and women, and international development experts) was concerned that there is "equal availability" of access to AI's benefits that would, to use the SDG's driving principle, "leave no one behind."

In "An Inclusive and Sustainable Artificial Intelligence Strategy for Europe Based on Human Rights," Fernández *et al*. summarize the reply that a group of professionals and experts drafted in response to the European Commission public consultation process on the "White Paper on Artificial Intelligence: A European Approach Oriented to Excellence and Trust" [21]. The authors highlight how the position expressed in the white paper is technologically reductionist, in contradiction with the European commitment to the UN Agenda 2030, which is not given its due centrality and, indeed, is hardly mentioned. There is an under-representation of the importance of human rights when analyzing AI impacts, and notions of regulation, self-regulation, and ethics are used in an imprecise and interchangeable way: proposed policies on AI appear to be exclusively conceived to improve the competitiveness of European companies in AI.

In "To be fAIr or not to be: Using AI for the Good of Citizens," the authors present the fAIr LAC initiative, which brings together a multidisciplinary group of Latin American experts from different governments, academic institutions, private companies, nongovernmental organizations, and innovation centers, as well as ethics experts and specialists from different areas of the Inter-American Development Bank. This initiative seeks to harness the potential of AI to create more efficient, fair, and personalized social services for Latin America and the Caribbean. To this end, it promotes standards, methodologies, and tools that guarantee the development of a responsible, human-centric, and trustworthy AI. The authors also introduce a local hub of the fAIr initiative implemented in Jalisco, Mexico, and the experience with a pilot AI-based application for the healthcare public sector.

In the second part of the Special Issue, concerning the experience with SDG-oriented AI applications, there are three articles in which applications in the fields of humanitarian emergency, mental health, and social impact measurement, respectively, are discussed.

In an Opinion piece, "From Artificial Intelligence Bias to Inequality in the Time of COVID-19," Luengo et al. illustrate the potential of AI to make a positive impact in the fight against the COVID-19 pandemic, while warning that AI applications, in practice, may suffer from problems of bias and interpretability which can result in systems that amplify health, economic, and social inequalities already exacerbated by the pandemic. The examples of bias that increase inequality range from systems for diagnosis and treatment trained with data from populations with very narrow demographics, to epidemiological models which cannot be adapted to different cultural and social settings, to AI algorithms driving the spread of mis- and disinformation targeting the attention of particularly vulnerable groups.

The article "Persuasive Technology for Mental Health: One Step Closer to (Mental Health Care) Equality?," Kolenik and Gams show how persuasive technology, which tries to influence people's

behavior or attitudes for their own goals without coercion, can be used to improve mental health, a part of the SDGs. This article focuses on stress, anxiety, and depression and examines why mental health is a considerable barrier to equality and why people with mental health issues have problems accessing health care. This article presents such systems with a brief overview of the field and offers general, technical, and critical thoughts on the implementation as well as impact. The authors think that such technology can complement existing mental healthcare solutions to reduce inequalities in access as well as inequalities resulting from the lack of it.

In "SIAMES: Social Impact Advisor and MEasurement System," Daniel Hernández and Marta Solórzano present the third SDG-oriented application included in our compilation. The authors highlight the importance of social impact measurement and the lack of generally agreed-upon indicators for such measurement and illustrate the potential contributions of AI to creating objective and empirically based measures that capture the social impact of an organization, with a goal of increasing standardization, verifiability, and accountability. They briefly describe SIAMES, a prototype recommender system of social impact indicators that extracts structured information from a corpus of impact measurement, reports through ontology-based semantic text mining and retrieves appropriate indicators by applying case-based reasoning.

In the third part of the Special Issue, which will be appear in a subsequent issue of the Magazine, the articles illustrate legislative, methodological, and technological proposals for the promotion and support of an inclusive AI and equitable access to its benefits.

In "A Wide Human-Rights Approach to Artificial Intelligence Regulation in Europe," Jesús Salgado-Criado and Celia Fernández Aller propose human rights as the basic framework for a future AI regulation. The authors argue that three European Commission's White Paper on AI is focused mainly on risk and some individual rights, such as privacy, whereas the collective dimension of society as a whole is overlooked. They highlight the importance of following a human rights-based approach in the regulatory efforts as it is necessary to establish a universal governance model and a general normative framework for AI. Human rights should replace ethics as the dominant framework for a debate. A description of the main principles of the rights approach is offered. Another key element of the article is the need to develop a sound technical framework within the regulation, as any regulation on a technical matter should encompass an architectural model on how the overall system functions and interacts. Finally, the authors point out that an auditing system is also required to allow accountability in the algorithmic process.

In the article "AI Ethics for Sustainable Development Goals," Monasterio Astobiza *et al*. show how AI technologies can be used to meet the 17 SDGs and its 169 targets. This article clarifies what people really mean by "ethics" in AI ethics and elucidates a road map to implement "ethics by design" standards to establish satisfactory measures of fairness, transparency, and explainability of algorithms when used for social good as, for example, in the promotion of the SDGs.

In "Bias and Discrimination in AI: A Cross-Disciplinary Perspective," the authors critically survey relevant literature about bias and discrimination in AI from an interdisciplinary perspective that embeds technical, legal, social, and ethical dimensions. The authors show that finding solutions for attesting and avoiding discrimination in AI requires robust cross-disciplinary collaborations and highlight a number of interdisciplinary challenges to address in this area.

Finally, in the article "Explaining the Principles to Practices Gap in AI," Schiff *et al*. review the gap between high-level principles for responsible uses of AI and the effective application of those principles in practice. The authors outline five explanations for this gap ranging from a disciplinary divide to an overabundance of tools and argue that an impact-assessment framework which is broad, operationalizable, flexible, iterative, guided, and participatory is a promising approach to closing the principles-to-practices gap.

We hope that this selection of articles will illustrate the path toward an AI for a fair, just, and equitable world, and motivate researchers and practitioners to travel along it. ■

References

[1] B. Lord. (2020). *Inequality in America: Far Beyond Extreme*. [Online]. Available: https://inequality.org/great-divide/inequality-in-america-far-beyond-extreme/

[2] *Think Differently: Humanitarian Impacts of the Economic Crisis in Europe. Internat. Federation of Red Cross and Red Crescent Societies*, Int. Fed. Red Cross Red Crescent Societies, Geneva, Switzerland, 2013.

[3] M. Lawson, A. P. Butt, R. Harvey, D. Sarosi, C. Coffey, K. Piaget, and J. Thekkudan, "Time to care: Unpaid and underpaid care work and the global inequality crisis," Oxfam GB, Oxford, U.K., Tech. Rep., 2020.

[4] *Poverty and Shared Prosperity 2020. Reversals of Fortune. International Bank for Reconstruction and Development/The World Bank*, World Bank Group, Washington, DC, USA, Oct. 2000.

[5] European Commission, "Draft ethics guidelines for trustworthy AI. Digital single market report," Eur. Commission, Brussels, Belgium, Tech. Rep., Dec. 2018. [Online]. Available: https://ec.europa.eu/digital-single-market/en/news/draft-ethics-guidelines-trustworthy-ai

[6] OECD. (2020). *Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449*. Accessed: May 15, 2020. [Online]. Available: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

[7] House of Lords Select Committee on Artificial Intelligence. (Apr. 2018). *Report of Session 2017-19 HL Paper 100. AI in the UK: Ready, Willing and Able?* Accessed: May 15, 2020. [Online]. Available: https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf

[8] IEEE Standards Association et al. (2019). *The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems*. Accessed: May 15, 2020. [Online]. Available: https://ethicsinaction.ieee.org/

[9] Université de Montréal et Fonds de recherche du Québec, Montreal, QC, Canada. (2018). *Montréal Declaration for Responsible Development of Artificial Intelligence*. Accessed: May 15, 2020. [Online]. Available: https://www.montrealdeclaration-responsibleai.com/

[10] P. Stone et al., "Artificial intelligence and life in 2030: One hundred year study on artificial intelligence: Report of the 2015–2016 study panel," Stanford Univ., Stanford, CA, USA, Tech. Rep., 2016, vol. 52. Accessed: May 15, 2020. [Online]. Available: http://ai100.stanford.edu/2016-report

[11] D. Leslie, "Understanding artificial intelligence ethics and safety," 2019, *arXiv:1906.05684*. [Online]. Available: http://arxiv.org/abs/1906.05684

[12] UN Global Pulse. (2012). *Big Data for Development: Challenges and Opportunities*. [Online]. Available: http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseMay2012.pdf

[13] UN. (2003). *Human Rights Based Approach to Development*. Accessed: Feb. 26, 2021. [Online]. Available: https://unsdg.un.org/2030-agenda/universal-values/human-rights-based-approach

[14] UNICEF Global Innovation Centre. (2003). *Generation AI*. Accessed: May 15, 2020. [Online]. Available: https://www.unicef.org/innovation/stories/generation-ai

[15] World Wide Web Foundation, "Artificial intelligence: The road ahead in low and middle-income countries," World Wide Web Found., Geneva, Switzerland, Tech. Rep., 2017. [Online]. Available: http://webfoundation.org/docs/2017/07/AI_Report_WF.pdf

[16] International Telecommunications Union, Geneva, Switzerland. (2020). *AI for Good Global Summit 2020*. Accessed: May 15, 2020. [Online]. Available: https://aiforgood.itu.int/

[17] World Economic Forum. (2018). *Harnessing Artificial Intelligence for the Earth*. [Online]. Available: http://www3.weforum.org/docs/HarnessingArtificial Intelligence for the Earth report 2018.pdf

[18] R. Ochigame. (2019). *The Invention of 'Ethical AI.'* Accessed: May 15, 2020. [Online]. Available: https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence

[19] Y. Benkler, "Don't let industry write the rules for AI," *Nature*, vol. 569, no. 7754, pp. 161–162, 2019.

[20] UN Secretary-General, "The road to dignity by 2030: Ending poverty, transforming all lives and protecting the planet-synthesis report of the secretary-general on the post-2015 agenda," United Nations, New York, NY, USA, Tech. Rep., 2014.

[21] European Commission. (2020). *A European Strategy for Data*. Accessed: Oct. 1, 2020. [Online]. Available: https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en

Ángeles Manjarrés is a Lecturer with the Department of Artificial Intelligence, Universidad Nacional de Educación a Distancia (UNED), Madrid, Spain, the Spanish national distance-learning university. Her research is focused on the field of ontologies and educational recommender systems, in e-learning and educational robotics. She has participated in national, European, and international research projects, and in educational innovation projects integrating service-learning methodology into artificial intelligence studies.

Ms. Manjarrés is a member of the UNED Service-Learning Department management team, of the UNED Research Ethics Committee, and of the

Education for Development Group EDETIC, affiliated with the Innovation and Technology for Development Center of the Technical University of Madrid.

Celia Fernández-Aller received the Ph.D. degree in law and technology from the Universidad Nacional de Educación a Distancia (UNED), Madrid, Spain, in 1998.

She is a Lecturer of Ethical and Legal Aspects with the Department of Computer Systems, Technical University of Madrid, Madrid, Spain. She has been a Visiting Scholar at José Simeón Cañas Central American University (UCA), San Salvador, El Salvador, and Bristol University, Bristol, U.K. She is one of the experts who will draw up the charter of digital rights for the Spanish government. She has several papers and books related to her area of interest. Her research interests focus on the human rights approach to technology, mainly privacy.

Maite López-Sánchez received the Ph.D. degree in artificial intelligence from the Artificial Intelligence Research Institute (IIIA), Spain.

She was the Research Manager with the Innovation Department of a iSOCO: Intelligent Software Components company, and a Visitor Researcher with the University of Southern California (USC), Los Angeles, CA, USA. She is an Associate Professor (TU) with the University of Barcelona (UB), Barcelona, Spain, and Adjunct Scientist with the Artificial Intelligence Research Institute (IIIA-CSIC), Bellaterra, Spain. Moreover, she is the Coordinator at UB of the interuniversitary master on Artificial Intelligence (UPC-UB-URV), member of a consolidated research group, and one of the board directors of the European Association for Multi-Agent Systems (EURAMAS). During her academic career, she has published nearly 150 scientific publications in indexed journals, books, and ranked international conferences. Her current interests focus on the inclusion of moral values within autonomous systems, social deliberation, and DSGs.

Juan Antonio Rodríguez-Aguilar received the Ph.D. degree in computer science from the Autonomous University of Barcelona, Barcelona, Spain, in 2001.

He is a Professor of Artificial Intelligence with the Artificial Intelligence Research Institute (IIIA-CSIC), Bellaterra, Spain. His research interests encompass multiagent optimization, artificial intelligence and ethics, industrial applications of artificial intelligence, and artificial social systems.

Dr. Rodríguez-Aguilar is also a fellow of the European Association for Artificial Intelligence.

Manuel Sierra Castañer was born in Zaragoza, Spain, in 1970. He received the degree of telecommunication engineering in 1994 and the Ph.D. degree in 2000, both from the Technical University of Madrid (UPM), Madrid, Spain.

He has been a Full Professor at UPM since 2017. He has been a Visitor Researcher at Tokyo Tech, Tokyo, Japan (September–December 1998) and École polytechnique fédérale de Lausanne (EPFL), Lausanne, Switzerland (September–December 1999) during the Ph.D. degree and Visitor Professor in Tokyo Tech during the summers of 2012 and 2013.

Dr. Sierra Castañer is currently a Senior Member of IEEE and Fellow of the AMTA Society. He has been the UPM Director for International Cooperation from 2010 to 2020. Since January 2016, he has been a member of the European Association on Antennas and Propagation (EurAAP) board of directors, where he is currently the vice-chair.