

Artificial Intelligence for a Fair, Just and Equitable World

Published at IEEE Technology and Society, 2021

<https://technologyandsociety.org/artificial-intelligence-for-a-fair-just-and-equitable-world/>

Ángeles Manjarrés, Celia Fernández-Aller, Maite López-Sánchez, Juan Antonio Rodríguez, Manuel Sierra

Already in the 70s we began 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... 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 US 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 e.g. [2]). In 2019, the world's 2,153 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].

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, strong inequalities in access to resources and power, alienation, exclusion. In this scenario, an elite monopolizes and uses 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][6][7], professional bodies [8], academia [9][10][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 marginalised populations are not treated with the importance they deserve in this debate.

Universal-ethics considerations gave rise to the 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 lower 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 organisations, such as UN Global Pulse [12], UNHCR [13], the 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 on potentialities, not only on risk mitigation, and from such a view arises the ethical imperative to harness AI technologies to the benefit of humanity in order to improve quality-of-life for all rather than contribute to perpetuating systemic injustices. To this end, more multi and inter-disciplinary 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].

Firstly, 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, in order 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, socio-economic, etc.) and the particular impact AI-based technological innovation can have in each context.

Secondly, 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, moreover, 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 papers mainly selected from the submissions to the workshop "Advancing Towards 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 the advances of AI augur, there is the option of an AI that catalyses that necessary transformation towards a fair, just and equitable world.

Here we present a series of papers where these issues are discussed, as well as papers describing real experience with SDG-oriented AI applications, and 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.

We present first “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, multi-level and multi-actor alliances that imply the private sector and civil society.

Next we reflect on initiatives addressing these issues from three different organizations: the IEEE, the European Commission, and the Latin American fAIr LAC alliance.

Elizabeth D. Gibbons in “Towards 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 be ‘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 analysing 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.

Finally, the authors of “To be fAIr or not to be. Using AI for the good of citizens” present the fAIr LAC initiative, that brings together a multidisciplinary group of Latin American experts from different governments, academic institutions, private companies, nongovernmental organisations, ethics experts, innovation centres 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. For this purpose, it promotes standards, methodologies and tools that guarantee the development of a responsible, human-centric and trustworthy AI. The authors introduce a local hub of the fAIr initiative implemented in Jalisco, Mexico, and the experience with a first pilot AI-based application for the Healthcare public sector.

Concerning the experience with SDG-oriented AI applications we next include three papers in which applications in the fields of mental health, humanitarian emergency, and social impact measurement, respectively, are discussed.

“Persuasive Technology for Mental Health: One Step Closer to (Mental Health Care) Equality?”, by Kolenik and Gams, shows 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 Sustainable Development Goals. The paper 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 paper 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 health care solutions to reduce inequalities in access as well as inequalities resulting from the lack of it.

In “From Artificial Intelligence Bias to Inequality in the Time of COVID-19”, Luengo et al. illustrate 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 shown 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.

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.

Finally, the other articles included illustrate legislative 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 the European Commission’s White Paper on

Artificial Intelligence is focused mainly on risk and some individual rights, such as privacy, while 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 debate. A description of the main principles of the rights approach is offered. Another key element of the paper 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.

The paper "AI ethics for Sustainable Development Goals", by Monasterio et al., shows how AI technologies can be used to meet the 17 Sustainable Development Goals and its 169 targets. This paper clarifies what people really mean by "ethics" in AI ethics and elucidate a roadmap 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.

The authors of "Bias and Discrimination in AI: a cross-disciplinary perspective" 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.

Finally, the paper "Explaining the Principles to Practices Gap in AI", by Schiff et al., reviews the principles-to-practices gap. 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.

References

- [1] Lord, B. (2020). Inequality in America: Far Beyond Extreme. <https://inequality.org/great-divide/inequality-in-america-far-beyond-extreme/>
- [2] International Federation of Red Cross and Red Crescent Societies. (2013). *Think differently: Humanitarian impacts of the economic crisis in Europe*. Internat. Federation of Red Cross and Red Crescent Societies.
- [3] Max Lawson et al. (2020). Time to care. Unpaid and underpaid care work and the global inequality crisis. Oxfam Report 2020.
- [4] World Bank Group (2000). Poverty and Shared Prosperity 2020. Reversals of Fortune. International Bank for Reconstruction and Development / The World Bank, Oct.

- [5] European Commission, "Draft ethics guidelines for trustworthy AI. digital single market report," European Commission, Tech. Rep., 12 2018. [Online]. Available: <https://ec.europa.eu/digital-single-market/en/news/draft-ethics-guidelines-trustworthy-ai/>
- [6] OECD, "Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449," 2020, (Date last accessed May 15, 2020). [Online]. Available: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>
- [7] House of Lords Select Committee on Artificial Intelligence, "Report of Session 2017–19 HL Paper 100. AI in the UK: ready, willing and able?" April 2018, (Date last accessed May 15, 2020). [Online]. Available: <https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf>
- [8] IEEE Standards Association et al., "The IEEE global initiative on ethics of autonomous and intelligent systems," 2019, (Date last accessed May 15, 2020). [Online]. Available: <https://ethicsinaction.ieee.org/>
- [9] Montréal: Université de Montréal et Fonds de recherche du Québec, "Montréal Declaration for Responsible Development of Artificial Intelligence," 2018, (Date last accessed May 15, 2020). [Online]. Available: <https://www.montrealdeclaration-responsibleai.com/>
- [10] P. Stone, R. Brooks, E. Brynjolfsson, R. Calo, O. Etzioni, G. Hager, J. Hirschberg, S. Kalyanakrishnan, E. Kamar, S. Kraus et al., "Artificial intelligence and life in 2030," One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel, vol. 52, 2016, (Date last accessed May 15, 2020). [Online]. Available: <http://ai100.stanford.edu/2016-report>
- [11] D. Leslie, "Understanding artificial intelligence ethics and safety," arXiv preprint arXiv:1906.05684, 2019.
- [12] UN Global Pulse, "Big data for development: Challenges and opportunities," 2012. [Online]. Available: <http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseMay2012.pdf>
- [13] UN, "Human Rights Based Approach to development," 2003, (Date last accessed May 15, 2020). [Online]. Available: [https://www.unicef.org/rightsresults/files/HRBDP Urban Jonsson April 2003.pdf](https://www.unicef.org/rightsresults/files/HRBDP%20Urban%20Jonsson%20April%202003.pdf)
- [14] UNICEF Global Innovation Centre, "Generation AI," 2003, (Date last 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 Foundation, Tech. Rep., 2017. [Online]. Available: [http://webfoundation.org/docs/2017/07/AI Report WF.pdf](http://webfoundation.org/docs/2017/07/AI%20Report%20WF.pdf)
- [16] International Telecommunications Union, "AI for good global summit 2020," Geneva, 2020, (Date last accessed May 15, 2020). [Online]. Available: <https://aiforgood.itu.int/>

[17] World Economic Forum, "Harnessing artificial intelligence for the earth," 2018. [Online]. Available: [http://www3.weforum.org/docs/](http://www3.weforum.org/docs/Harnessing%20Artificial%20Intelligence%20for%20the%20Earth%20report%202018.pdf) Harnessing Artificial Intelligence for the Earth report 2018.pdf

[18] R. Ochigame, "The invention of "Ethical AI"," 2019, (Date last accessed May 15, 2020). [Online]. Available: [https://theintercept.com/](https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence) 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] U. 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 New York: United Nations, 2014.

[21] European Commission. "A European strategy for data". In: available at European Commission: <https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-datastrategy> (01/10/2020) (2020).