

# Solidarity should be a core ethical principle of AI

**To the Editor** — Over the past few years, dozens of organizations including governments, technology companies and think-tanks have released their ethical guidelines and principles in the development of artificial intelligence (AI)<sup>1</sup>. There is a consensus that AI is a dual-purpose technology and we must weigh potential risks and benefits<sup>2</sup>. However, while many efforts advocate for a human-centric approach to AI, few of the existing ethical AI guidelines consider the principle of solidarity upfront: the equal and just sharing of prosperity and burdens.

According to ref. <sup>1</sup>, just 6 out of 84 AI principles guidelines mention the concept of solidarity. This is surprising, because solidarity is one of the fundamental values at the heart of peaceful societies, present in more than 30% of world's constitutions<sup>3</sup> and a foundational principle of institutions such as the European Charter<sup>4</sup>. Furthermore, the understanding of the concept differs: for example, the Montreal Declaration<sup>5</sup> proposes the development of autonomous intelligent systems compatible with maintaining the bonds of solidarity among people.

Solidarity as an AI principle should imply the following: (1) sharing the prosperity created by AI, implementing mechanisms to redistribute the augmentation of productivity for all, and sharing the burdens, making sure that AI does not increase inequality and nobody is left behind; and (2) assessing the long-term implications before developing and deploying AI systems.

Solidarity should be a core ethical principle in the development of AI. Sharing prosperity would mean, for instance, that anyone whose actions provide data to train AI models will be paid — for example, though a royalties system where humans receive compensation each time an AI system trained with their data is used.

Human doctors teaching an AI model to diagnose a disease will be rewarded each time the model is used to diagnose, or humans producing text to feed an AI automatic text generator will get something back each time the robot writes an article<sup>6</sup>. At a public scale, tax on robots or automation are also options for financial solidarity.

Sharing prosperity implies creating and releasing digital public goods in the form of AI models and open sourcing algorithms — from climate models for smart agriculture to pandemics management. The global AI research agenda needs a compass to direct efforts towards a shared prosperity and achievement of the Sustainable Development Goals, and not to a narrow optimization of Internet businesses or academic benchmarks.

International cooperation in the regulation of digital technologies<sup>7</sup> is also needed. This should include a commitment to address bias, attacks and unexpected functioning states in AI models used at global scale. Similar to the mechanisms we have for declaring a global health emergency or investigating human rights violations, we may now need to prepare ourselves for the declaration of global AI emergencies where the international community will offer its support to address an AI crisis: for example, when thousands of deep fake videos with ethnic violence circulate on a day of elections in a country with a history of genocide.

Beyond trustworthy AI, solidarity in long-term thinking implies assessing risks and harms before embarking on new AI deployments. A clear example is the urgent need to understand the climate impact of computing resources used to train AI models — is it reasonable to generate tons of CO<sub>2</sub> emissions to teach a machine to discern photos of cats and dogs in the Internet? Shouldn't we establish sustainability policies

by which the expected benefits of the AI model should at least outpace its carbon footprint<sup>8</sup>? Furthermore, thinking long term implies developing international policy instruments such as extending human rights so they include its digital dimension, agreeing on bans for lethal autonomous weapons, working on regulation of global social media companies, or even preemptive bans until societal impacts and regulatory needs are clear — as could be the case for facial recognition technologies<sup>9</sup>.

One of the biggest long-term challenges of AI will be how to redistribute the augmentation of productivity so nobody is left behind. Solidarity as an AI principle can provide a framework and a narrative to face this challenge, so we do not create new and exacerbate existing inequalities. We need a global safety net for AI technologies, with solidarity as a core principle for AI development. □

Miguel Luengo-Oroz 

United Nations Global Pulse, New York, NY, USA.  
e-mail: miguel@unglobalpulse.org

Published online: 18 October 2019  
<https://doi.org/10.1038/s42256-019-0115-3>

## References

1. Jobin, A. et al. *Nat. Mach. Intell.* **1**, 389–399 (2019).
2. Brundage, M. et al. *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation* (Future of Humanity Institute, 2018).
3. Rutherford, et al. *Nat. Hum. Behav.* **2**, 592–599 (2018).
4. *The Charter of Fundamental Rights of the European Union* (European Parliament, 2001).
5. *Montreal Declaration for Responsible Development of Artificial Intelligence* (Université de Montréal, 2017).
6. Luengo-Oroz, M. *El País Retina* [https://retina.elpais.com/retina/2019/01/04/tendencias/1546604928\\_551805.html](https://retina.elpais.com/retina/2019/01/04/tendencias/1546604928_551805.html) (2019).
7. *The Age of Digital Interdependence* (UN, 2019).
8. Strubell, E., Ganesh, A. & McCallum, A. in *Proc. 57th Annual Meeting of the Association for Computational Linguistics* 3645–3650 (ACL, 2019).
9. Crawford, K. *Nature* **572**, 565 (2019).

## Competing interests

The author declares no competing interests.