Artificial intelligence in a crisis needs ethics with urgency

Artificial intelligence tools can help save lives in a pandemic. However, the need to implement technological solutions rapidly raises challenging ethical issues. We need new approaches for ethics with urgency, to ensure AI can be safely and beneficially used in the COVID-19 response and beyond.

Asaf Tzachor, Jess Whittlestone, Lalitha Sundaram and Seán Ó hÉigeartaigh

The novel coronavirus pandemic (COVID-19) is the largest global crisis in a generation, hitting the world at a time when artificial intelligence (AI) is showing potential for widespread real-world application. We are currently seeing a rapid increase in proposals for how AI can be used in many stages of pandemic prevention and response. AI can aid in detecting, understanding and predicting the spread of disease, which can provide early warning signs and inform effective interventions[1]. AI may improve the medical response to the pandemic in several ways: supporting physicians by automating aspects of diagnosis[2], prioritizing healthcare resources[3], and improving vaccine and drug development[4]. AI also has potential applications beyond immediate response, such as in combating online misinformation about COVID-19[5].

The current crisis presents an unprecedented opportunity to leverage AI for societal benefit. However, the urgency with which new technologies must be deployed raises particularly challenging ethical issues and risks. There is growing concern that the use of AI and data in response to COVID-19 may compromise privacy and civil liberties by incentivizing the collection and processing of large amounts of data, which may often be private or personal[6]. More broadly, although AI clearly has a great deal to offer, we must be careful not to overestimate its potential. Its efficacy will heavily depend on the reliability and relevance of the data available. With the worldwide spread of COVID-19 occurring so quickly, obtaining sufficient data for accurate AI forecasting and diagnosis is challenging. Even where AI models are strictly speaking accurate, they may have differential impacts across subpopulations, with harmful consequences that are difficult to predict in advance[7]. A further concern is that the lack of transparency in AI systems used to aid decision-making around COVID-19 may make it near impossible for the decisions of governments and public officials to be subject to public scrutiny and legitimation[8]. Finally, the current crisis may have longer-term impacts on public trust and norms around the use of AI in society. How these develop will depend on perceptions of how successful and responsible use of AI to address COVID-19 is.

The challenge of ethics in a crisis

Robust ethics and risk assessment processes are needed to ensure AI is used responsibly in response to COVID-19. However, implementing these at a time of crisis is far from straightforward, especially where new technologies need to be deployed at unprecedented speed and scale. For example, forecasting models have to be available at the early stages of disease spread and make use of all possible data to productively inform policy interventions. Current processes for ethics and risk assessment around uses of AI are still relatively immature, and the urgency of a crisis highlights their limitations.

Much work in AI ethics in recent years has focused on developing high-level principles, but these principles say nothing about what to do when principles come into conflict with one another[9]. For example, principles do not tell us how to balance the potential of AI to save lives (the principle of 'beneficence') against other important values such as privacy or fairness. One common suggestion for navigating such tensions is through engagement with diverse stakeholder groups, but this may be difficult to enact with sufficient speed at times of crisis.

When new technologies may pose unknown risks, we would ordinarily try to introduce them in gradual, iterative ways, allowing time for issues to be identified and addressed. In the context of a crisis, however, there is a stark trade-off between a cautious approach and the need to deploy technological solutions at scale. For example, there may be pressure to rely on systems with less human oversight and potential for override due to staff shortages and time pressures, but this must be carefully balanced against the risk of failing to notice or override crucial failures.

This does not mean that ethics should be neglected at times of crisis. It only emphasizes that we must find ways to conduct ethical review and risk assessment with the same urgency that motivates the development of AI-based solutions.

Doing ethics with urgency

We suggest that ethics with urgency must at a minimum incorporate the following components: (1) the ability to think ahead rather than dealing with problems reactively, (2) more robust procedures for assuring the behaviour and safety of AI systems, and (3) building public trust through independent oversight.

First, ethics with urgency must involve thinking through possible issues and risks as thoroughly as possible before systems are developed and deployed in the world. This need to think ahead is reflected in the notion of 'ethics by design': making ethical considerations part of the process of developing new applications of AI, not an afterthought[10]. For example, questions such as 'what data do we need and what issues might this raise?' and 'how do we build this model so that it is possible to interrogate key assumptions?' need to be considered throughout the development process. This means that experts in ethics and risk assessment need to be involved in teams developing AI-based solutions from the beginning, and much clearer guidelines are needed for engineers and developers to think through these issues. An ethics by design approach should also be supplemented with more extensive foresight work, looking beyond the more obvious and immediate ethical issues, and considering a wider range of longer-term and more systemic impacts. By synthesizing diverse sources of expertise, established foresight methodologies can be used to identify new risks and key uncertainties likely to shape the future, and use this to make better informed decisions today[11].

Second, where applications of AI are used at scale in safety-critical domains such as healthcare, ensuring the safety and reliability of those systems across a range of scenarios is of crucial importance. Finding ways to rapidly conduct robust testing and verification of systems will therefore be central to doing ethics with urgency. We suggest that the application of AI in crisis scenarios should in particular be heavily informed by research on best practices for the verification and validation of autonomous systems[12]. It may also be worthwhile for governments to fund further work on methods for establishing the reliability of machine learning systems across a range of circumstances, particularly where those systems may be deployed in high-stakes crisis scenarios.

Third, an important aspect of ethics with urgency is building public trust in how AI is being used. If governments use AI systems in ways perceived to be either mistaken or problematically value-laden, this could result in a loss of public trust severe enough to drastically reduce support for beneficial uses of AI not just in this crisis, but also in the future. Building public trust around new uses of technology may be particularly challenging in crisis times, where the need to move fast makes it easier for governments to fall back on opaque and centralized forms of decision-making. Several analyses of past pandemics have argued that transparency and public scrutiny are essential for maintaining public trust[13]. An independent oversight body, responsible for reviewing any potential risks and ethical issues associated with new technologies and producing publicly available reports, could help ensure public transparency. This oversight body could, among other approaches, make use of techniques such as 'red teaming' to rigorously challenge systems and their assumptions, unearthing any limitations and biases in the applications being proposed[14]. Red teaming is widely used in security settings, but can be applied broadly: at its core, red teaming is a way of challenging the blind spots of a team by explicitly looking for flaws from an outsider or adversarial perspective. As well as allowing developers to identify and fix issues before deployment, such processes could help assure public stakeholders that the interests and values of different groups are being thoroughly considered, and that all eventualities are prepared for.

Conclusion

As the COVID-19 pandemic illustrates, times of crisis can necessitate rapid deployment of new technologies in order to save lives. However, this urgency both makes it more likely that ethical issues and risks will arise, and makes them more challenging to address. Rather than neglecting ethics, we must find ways to do ethics with urgency too. We strongly encourage technologists, ethicists, policymakers and healthcare professionals to consider how ethics can be implemented at speed in the ongoing response to the COVID-19 crisis. If ethical practices can be implemented with urgency, the current crisis could provide an opportunity to drive greater application of AI for societal benefit, and to build public trust in such applications. ❐

Asaf Tzachor[1]✉, Jess Whittlestone[2], Lalitha Sundaram[1] and Seán Ó hÉigeartaigh[2]

[1]Centre for the Study of Existential Risk, University of Cambridge, Cambridge, UK. [2]Leverhulme Centre for the Future of Intelligence, University of Cambridge, Cambridge, UK.

✉*e-mail: at875@cam.ac.uk*

Published online: 22 June 2020
https://doi.org/10.1038/s42256-020-0195-0

References

1. van der Schaar, M. et al. Preprint at http://www.vanderschaar-lab.com/NewWebsite/covid-19/post1/paper.pdf (2020).
2. Wang, S. et al. Preprint at https://doi.org/10.1101/2020.02.14.20023028 (2020).
3. Butt, C., Gill, J., Chun, D. & Babu, B. A. *Appl. Intell.* https://doi.org/10.1007/s10489-020-01714-3 (2020).
4. Zhang, H. et al. Preprint at https://doi.org/10.20944/preprints202002.0061.v1 (2020).
5. Infodemic management - infodemiology. *World Health Organization* https://www.who.int/teams/risk-communication/infodemic-management (2020).
6. Ienca, M. & Vayena, E. *Nat. Med.* **26**, 463–464 (2020).
7. Wynants, L. et al. *BMJ* **369**, m1328 (2020).
8. Nyrup, R., Whittlestone, J. & Cave, S. *Why Value Judgements Should Not Be Automated* (Leverhulme Centre for the Future of Intelligence, 2019); https://doi.org/10.17863/CAM.41552
9. Whittlestone, J., Nyrup, R., Alexandrova, A. & Cave, S. In *Proc. 2019 AAAI/ACM Conf. AI, Ethics, and Society* 195–200 (ACM, 2019).
10. d'Aquin, M. et al. In *Proc. 2018 AAAI/ACM Conf. AI, Ethics, and Society* 54–59 (ACM, 2018).
11. *The Futures Toolkit: Tools for Futures Thinking and Foresight Across UK Government* (Government Office for Science, 2017).
12. Lyons, J. B., Clark, M. A., Wagner, A. R. & Schuelke, M. J. *AI Mag.* **38**(3), 37–49 (2017).
13. O'Malley, P., Rainford, J. & Thompson, A. *Bull. World Health Organ.* **87**, 614–618 (2009).
14. Brundage, M. et al. Preprint at https://arxiv.org/abs/2004.07213 (2020).

Competing interests

The authors declare no competing interests.