

Using AI ethically to tackle covid-19

Taking a principled approach is crucial to the successful use of AI in pandemic management, say **Stephen Cave and colleagues**

In a crisis such as the covid-19 pandemic, governments and health services must act quickly and decisively to stop the spread of the disease. Artificial intelligence (AI), which in this context largely means increasingly powerful data driven algorithms, can be an important part of that action—for example, by helping to track the progress of a virus or to prioritise scarce resources.¹ To save lives it might be tempting to deploy these technologies at speed and scale. Deployment of AI can affect a wide range of fundamental values, however, such as autonomy, privacy, and fairness. AI is much more likely to be beneficial, even in urgent situations, if those commissioning, designing, and deploying it take a systematically ethical approach from the start.

Ethics is about considering the potential harms and benefits of an action in a principled way. For a widely deployed technology, this will lay a foundation of trustworthiness on which to build. Ethical deployment requires consulting widely and openly; thinking deeply and broadly about potential impacts; and being transparent

about goals being pursued, trade-offs being made, and values guiding these decisions. In a pandemic, such processes should be accelerated, but not abandoned. Otherwise, two main dangers arise: firstly, the benefits of the technology could be outweighed by harmful side effects, and secondly, public trust could be lost.²

The first danger is that the potential benefits increase the incentive to deploy AI systems rapidly and at scale, but also increase the importance of an ethical approach. The speed of development limits the time available to test and assess a new technology, while the scale of deployment increases any negative consequences. Without forethought, this can lead to problems, such as a one-size-fits-all approach that harms already disadvantaged groups.³

Secondly, public trust in AI is crucial. For example, contact tracing apps rely on widespread adoption for their success.⁴ Both technology companies and governments, however, struggle to convince the public that they will use AI and data responsibly. After controversy over, for example, the partnership between the AI firm DeepMind and the Royal Free London NHS Foundation Trust, privacy groups have warned against plans to allow increased access to NHS data.⁵ Similarly, concerns have been raised in China over the Health QR code system's distribution of data and control to private companies.⁶ Overpromising on the benefits of technology or relaxing ethical requirements, as has sometimes happened during this crisis,⁵ both risk undermining long term trust in the reputation of the entire sector. Whether potential harms become obvious immediately or only much later, adopting a consistently ethical approach from the outset will put us in a much better position to reap the full benefits of AI, both now and in the future.

Bringing together AI ethics and health ethics

AI can broadly be defined as digital systems that can make complex decisions or recommendations on the basis of data inputs. This simple definition highlights three reasons why ethical challenges arise from such systems.

Firstly, AI applications, particularly in healthcare, often require a lot of personal data, and so invoke all the concerns about responsible data management, such as privacy, consent, security, and ownership.⁷

Secondly, AI systems are often used to automate decision making processes that were previously made by humans. This automation gives rise to ethical challenges, such as who is to be held accountable for these decisions, or how stakeholders can know which value judgments are guiding them.⁸ For example, is the system optimising a commercial value, the interests of a government, or the health of the individual? These concerns can arise even when an AI system is only recommending a course of action because of automation bias—the propensity for people to suspend their own judgement and over-rely on automated systems.⁹

Thirdly, the operations of AI systems are often unclear, owing to the complexity of the data or the algorithm (especially many powerful and popular algorithms used in machine learning).¹⁰ This lack of clarity, as well as compounding problems of accountability, can make it hard to assess ethically relevant factors, such as unintended biases in the system or the robustness of results across different populations.¹¹

Ethical decision making is, of course, already an integral part of healthcare practice, where it is often structured according to the four pillars of biomedical ethics: beneficence, non-maleficence, autonomy, and justice.¹² When considering the use of AI in a public health setting, such as a pandemic, it might therefore be useful to consider how the distinctive challenges posed by AI pertain to these four well established principles.⁸

Beneficence

It might seem obvious that the use of AI in managing a pandemic is beneficent: it is intended to save lives. A risk exists, however, that the vague promise that a new technology will “save lives” can be used as a blanket justification for interventions we might not otherwise consider appropriate, such as widespread deployment of facial recognition software.¹³ Those developing or deploying

KEY MESSAGES

- AI based technologies promise benefits for tackling a pandemic like covid-19, but also raise ethical challenges for developers and decision makers
- If an ethical approach is not taken, the risks increase of unintended harmful consequences and a loss of stakeholder trust
- Ethical challenges from the use of AI systems arise because they often require large amounts of personal data; automate decisions previously made by humans; and can be highly complex and opaque
- The four pillars of biomedical ethics—beneficence, non-maleficence, autonomy, and justice—are a helpful way of seeing how these challenges can arise in public health
- Open and transparent communication with diverse stakeholder groups during development of AI systems is the best way of tackling these challenges

such a system must be clear about whom their intervention will benefit and how. Only by making this explicit can one ensure that the intervention is proportionate to its benefit.¹⁴ For example, if a data driven contact tracing app does not require large amounts of location data to be collected and stored indefinitely, it would not be proportionate to engage in large scale data gathering that we would normally find excessive. Even if some additional benefit could be had, one needs to consider whether this benefit is sufficient to justify creating such a database.

Non-maleficence

To avoid unintended harms from the use of AI in the management of a pandemic, it is important to carefully consider the potential consequences of proposed interventions. Some interventions—for example, imposing self-isolation—may cause mental health problems for those who are already vulnerable (eg, elderly people) or carry high economic costs for individuals. AI systems seek to optimise a particular objective function—that is, a mathematical function representing the goals it has been designed to achieve. Any potential harms not represented by this function will not be considered in the system's predictions. For example, some systems designed to inform the prioritisation of hospital resources are optimised to predict death from covid-19,¹⁵ but not other possible harms for patients (eg, “long covid”). If these other harms do not correlate with the risk of fatality, deciding how to prioritise health resources purely based on this system might considerably aggregate harm (depending on the incidence and severity of the other harms). Additionally, as these systems will be widely deployed, they must reliably perform as expected across different populations and potentially changing conditions. Trying to rapidly develop AI systems while our understanding of the virus is still limited, and with less time than usual to ensure the quality and representativeness of the data used, risks creating systems based on simplifying assumptions and datasets that do not cover all real world cases. For instance, a recent systematic review of 145 prediction models for diagnosis of covid-19 (including 57 using AI for image analysis) found all to have a high risk of statistical bias.¹¹ Inaccurate diagnoses or inappropriate interventions arising from such models could cost more lives than they save.

Autonomy

The benefits of new technologies almost always depend on how they affect peoples'

behaviour and decision making: from the precautions an individual chooses to take, to treatment decisions by healthcare professionals, and politicians' prioritisation of different policy responses. Respecting peoples' autonomy is therefore crucial. Evidence from across cultures and age groups shows that people feel the need to be in control and endorse the use of technology, or its influence on their behaviour is likely to be limited.¹⁶ A particular challenge for AI systems is that they might affect patients, healthcare professionals, and other stakeholders in more subtle and individualised ways than, for example, a mask or vaccine, where the desired behaviours are obvious. Designers can help users to understand and trust AI systems so that they feel able to use them with autonomy.¹⁷ For example, diagnostic support systems used by healthcare professionals in a pandemic should provide sufficient information about the assumptions behind, and uncertainty surrounding, a recommendation, so that it can be incorporated into their professional judgment.

Justice

Data driven AI systems can differentially affect different groups, as is well documented.¹⁸ When data of sufficient quality for some groups are lacking, AI systems can become biased, often in ways which discriminate against already disadvantaged groups, such as racial and ethnic minorities.¹⁹ For example, smartphone apps are increasingly heralded as tools for monitoring and diagnosis, such as the MIT-Harvard model for diagnosing covid-19 through the sound of coughs.²⁰ But access to smartphones is unevenly distributed between countries and demographics, with global smartphone penetration estimated in 2019 to be 41.5%.²¹ This limits both whose data are used to develop such apps and who has access to the service. If care is not taken to detect and counteract any biases, using AI for pandemic management could worsen health inequalities.²² Again, the speed and scale at which systems might be deployed in response to the pandemic exacerbate these risks, making foresight and vigilance all the more crucial.

More broadly, when AI systems are proposed for a response to a pandemic, difficult trade-offs between values could be introduced. For example, leading UK public health officials argued for a centralised approach to data collection in the design of the NHS digital contact tracing app, arguing that machine learning could be applied to the resultant dataset to aid in disease prediction. Legal and

security experts, however, argued for a decentralised approach, citing concerns about privacy and data security.²³ The UK chose to use a decentralised app. These are inherently value laden choices, about which reasonable people might disagree, and particular groups might have much greater reasons for concern (eg, owing to worry about surveillance or historic discrimination). Involving diverse communities in decisions, and being open about the values and trade-offs, will help to reduce these risks.²⁴

Ethics in practice: a participatory approach

Politicians and public health officials are charged with final decisions about deployment of AI, and thus are responsible for ensuring these ethical challenges are met. But this requires that they draw on the expertise of designers, engineers, and healthcare experts, as well as the views of affected groups. No single checklist is available that these decision makers can mechanically follow to ensure that AI is used ethically and responsibly. Especially during a crisis, there will be some trade-off between the use of AI for good and the need to mitigate its harms.

Public decision makers should not tackle these trade-offs alone but must communicate with diverse stakeholder groups to ensure decisions about the use of AI are fair. To ensure that this can be done rapidly and effectively, even during a fast moving crisis, it is essential that processes are put in place in advance, detailing who should be consulted and how to do so if a public health crisis arises. Decision makers can then be held accountable for following those processes and for making transparent the reasoning behind their decisions to deploy AI.

Broad stakeholder engagement means consulting with both a wide range of experts and diverse groups from across society, to better understand potential trade-offs involved in deploying a system and acceptable ways to resolve them. Consulting with experts might, for example, include talking to the engineers building AI systems to develop a fuller understanding of their weaknesses, limitations, and risks; experts in domains such as human centred design or value sensitive design to understand how the envisaged benefits of a system might depend on human behaviours and how to support adherence; and ethicists to understand where the use of AI systems might introduce value judgments into decision making processes.¹⁸ Consultation with diverse public groups can highlight

blind spots, identify previously ignored harms or benefits to different groups, and help decision makers to understand how trade-offs are perceived by different communities.²⁴

AI has the potential to help us solve increasingly important global problems, but deploying powerful new technologies for the first time in times of crisis always comes with risks. The better placed we are to deal with ethical challenges in advance, the easier it will be to secure public trust and quickly roll out technology in support of the public good.

Contributors and sources: All five authors conduct research in the ethics of technology, in particular AI. SC and JW contributed significantly to planning and drafting; RN and SOhE contributed to planning and drafting; RAC contributed to planning and editing. SC and JW are joint first authors and guarantors.

Competing interests: We have read and understood BMJ policy on declaration of interests and have no relevant interests to declare.

Provenance and peer review: Commissioned; externally peer reviewed.

This collection of articles was proposed by the WHO Department of Digital Health and Innovation and commissioned by *The BMJ*. *The BMJ* retained full editorial control over external peer review, editing, and publication of these articles. Open access fees were funded by WHO.

Stephen Cave, executive director¹

Jess Whittlestone, senior research fellow¹

Rune Nystrup, senior research fellow¹

Sean O hEigeartaigh, programme director¹

Rafael A Calvo, professor²

¹Leverhulme Centre for the Future of Intelligence, University of Cambridge, Cambridge, UK

²Dyson School of Design Engineering, Imperial College London, UK

Correspondence to: S Cave
sjc53@cam.ac.uk



OPEN ACCESS

This is an Open Access article distributed under the terms of the Creative Commons Attribution IGO

License (<https://creativecommons.org/licenses/by-nc/3.0/igo/>), which permits use, distribution, and reproduction for non-commercial purposes in any medium, provided the original work is properly cited.



Check for updates

- van der Schaar M, Alaa AM, Floto A, et al. How artificial intelligence and machine learning can help healthcare systems respond to COVID-19. *Mach Learn* 2020;110:1-14. doi:10.1007/s10994-020-05928-x
- Tzachor A, Whittlestone J, Sundaram L, O hEigeartaigh S. Artificial intelligence in a crisis needs ethics with urgency. *Nat Mach Intell* 2020;2:365-6. doi:10.1038/s42256-020-0195-0
- Veinot TC, Mitchell H, Ancker JS. Good intentions are not enough: how informatics interventions can worsen inequality. *J Am Med Inform Assoc* 2018;25:1080-8. doi:10.1093/jamia/ocy052
- Parker MJ, Fraser C, Abeler-Dörner L, Bonsall D. Ethics of instantaneous contact tracing using mobile phone apps in the control of the COVID-19 pandemic. *J Med Ethics* 2020;46:427-31. doi:10.1136/medethics-2020-106314
- Wright M. (2020). NHS can save lives by giving data to tech giants, says Matt Hancock. *Telegraph* 2020 Sep 8. <https://www.telegraph.co.uk/news/2020/09/08/nhs-can-save-live-giving-data-tech-giants-says-matt-hancock/>
- Xingdong F, Feng Y. Research on the challenges of digital social governance behind the "Health QR code". *Frontiers (Boulder)* 2020;(16):78-91. doi:10.16619/j.cnki.mltxsqy.2020.30.026
- Kroll JA. Data science data governance [AI ethics]. *IEEE Secur Priv* 2018;16:61-70. doi:10.1109/MSEC.2018.2875329
- Peters D, Vold K, Robinson D, Calvo R. Responsible AI—two frameworks for ethical design practice. *IEEE Transactions on Technology and Society* 2020;1:34-47. doi:10.1109/TTS.2020.2974991
- Strauch B. Ironies of automation: still unresolved after all these years. *IEEE Trans Hum Mach Syst* 2018;48:419-33. doi:10.1109/THMS.2017.2732506
- Watson DS, Krutzinna J, Bruce IN, et al. Clinical applications of machine learning algorithms: beyond the black box. *BMJ* 2019;364:l886. doi:10.1136/bmj.l886
- Wynants L, Van Calster B, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. *BMJ* 2020;369:m1328. doi:10.1136/bmj.m1328
- Beauchamp TL, Childress JF. *Principles of biomedical ethics*. 7th ed. Oxford University Press, 2013.
- Shachar C, Gerke S, Adashi EY. AI surveillance during pandemics: ethical implementation imperatives. *Hastings Cent Rep* 2020;50:18-21. doi:10.1002/hast.1125
- Hick JL, Hanfling D, Matthew K, Wynia MK, Pavia AT. Duty to plan: health care, crisis standards of care, and novel coronavirus SARS-CoV-2. NAM Perspectives. Discussion paper. National Academy of Medicine, 2020. <https://nam.edu/duty-to-plan-health-care-crisis-standards-of-care-and-novel-coronavirus-sars-cov-2/>
- Li Y, Horowitz MA, Liu J, et al. Individual-level fatality prediction of covid-19 patients using AI methods. *Front Public Health* 2020;8:587937. doi:10.3389/fpubh.2020.587937
- Ryan RM, Deci EL. *Self-determination theory: basic psychological needs in motivation, development, and wellness*. Guilford Publications, 2017. doi:10.1521/978.14625/28806
- Calvo RA, Peters D, Vold K, Ryan RM. Supporting human autonomy in AI systems: a framework for ethical enquiry. In: Burr C, Floridi L, eds. *Ethics of digital well-being. Philosophical Studies Series*. Springer, 2020:31-54. doi:10.1007/978-3-030-50585-1_2
- Rovatsos M, Mittelstadt B, Koene A. Landscape summary: bias in algorithmic decision-making. London: Centre for Data Ethics and Innovation, 2019. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/819055/Landscape_Summary_-_Bias_in_Algorithmic_Decision-Making.pdf
- Noor P. Can we trust AI not to further embed racial bias and prejudice? *BMJ* 2020;368:m363. doi:10.1136/bmj.m363
- Laguarta J, Hueto F, Subirana B. COVID-19 artificial intelligence diagnosis using only cough recordings. *IEEE Open Journal of Engineering in Medicine and Biology*, 2020. doi:10.1109/OJEMB.2020.3026928
- Statista. Global smartphone penetration rate as share of population from 2016 to 2020. 2020. <https://www.statista.com/statistics/203734/global-smartphone-penetration-per-capita-since-2005/>
- Röösli E, Rice B, Hernandez-Boussard T. Bias at warp speed: how AI may contribute to the disparities gap in the time of COVID-19. *J Am Med Inform Assoc* . 2021;28:190-2. doi:10.1093/jamia/ocaa210
- Burgess M. There's a big row brewing over the NHS Covid-19 contact tracing app. *Wired* 2020 May 5. <https://www.wired.co.uk/article/nhs-contact-tracing-app-data-privacy>
- Thompson AK, Faith K, Gibson JL, Upshur RE. Pandemic influenza preparedness: an ethical framework to guide decision-making. *BMC Med Ethics* 2006;7:E12. doi:10.1186/1472-6939-7-12

Cite this as: *BMJ* 2021;372:n364
<http://dx.doi.org/10.1136/bmj.n364>