NHS AI Lab: why we need to be ethically mindful about AI for healthcare

Jessica Morley[1], Luciano Floridi[1,2]

[1] Oxford Internet Institute, University of Oxford, 1 St. Giles', Oxford, OX1 3JS

[2] Alan Turing Institute, 96 Euston Road, London, NW1 2DM

Abstract

On 8th August 2019, Secretary of State for Health and Social Care, Matt Hancock, announced the creation of a £250 million NHS AI Lab. This significant investment is justified on the belief that transforming the UK's National Health Service (NHS) into a more informationally mature and heterogeneous organisation, reliant on data-based and algorithmically-driven interactions, will offer significant benefit to patients, clinicians, and the overall system. These opportunities are realistic and should not be wasted. However, they may be missed (one may recall the troubled Care.data programme) if the ethical challenges posed by this transformation are not carefully considered from the start, and then addressed thoroughly, systematically, and in a socially participatory way. To deal with this serious risk, the NHS AI Lab should create an Ethics Advisory Board and monitor, analyse, and address the normative and overarching ethical issues that arise at the individual, interpersonal, group, institutional and societal levels in AI for healthcare.

NHS AI Lab: why we need to be ethically mindful about AI for healthcare

Introduction

On 8th August 2019, the UK's Secretary of State for Health and Care announced the creation of a £250 million National Health Service (NHS) AI Lab (1). The creation of the Lab shows the UK government's commitment to deliver on the plans made in the NHS Long-Term Plan (2) with regards to data-driven technology, and the Grand Challenge to "use data, AI and innovation to transform the prevention, early diagnosis and treatment of chronic diseases by 2030 (3)."

There are several reasons to see this as a positive investment in the NHS. As (4) notes "cooperation between doctors and machines could represent a turning point with regards to our ability to tackle diseases and improve our wellbeing." Today, AI systems can estimate bone age, predict which patients will not show up to an appointment (5), diagnose retinal disease, or quantify cardiac risk with greater consistency, speed, and reproducibility than humans (6). These and many other opportunities are to be valued, but capitalising on them is challenging. This is because the opportunities are not created by the technologies *per se* but by their ability to transform fundamentally the intrinsic nature of how healthcare is delivered in the NHS (7).

This fundamental transformation is happening at a pace with which the system's governance mechanisms are struggling to keep up. To ensure fairness, efficiency, efficacy and patient safety, a rigorous and robust regulatory system will be required (8) and it is being developed (9). However, compliance is necessary but still insufficient to steer the development of AI for healthcare in the right direction. Regulations, even when they are not in need of interpretation or extension, will only indicate what *may* and *may not* be done, but not what *ought* or *ought not*. They specify what is socially unacceptable but leave undetermined what is socially preferable. With an analogy, they are the necessary rules of the game, not the best strategy to win it. For this, we need *soft ethics* to help guide considerations of what ought and ought not to be done post-compliance (10).

By 'ethics' we mean to refer to the need to evaluate not only the intentions and responsibilities of different agents in the AI for healthcare system (for example, clinicians, developers, policymakers), but also the impacts that the actions of these agents will have on the 'receiver' (individuals, groups, systems or indeed whole populations), their expectations, demands, needs, and rights. By taking this 'patient-centric' approach, we can collectively design a pro-ethical blueprint for algorithmic healthcare that ensures the "right actions are facilitated, promoted,

amplified, and rewarded and the wrong actions hindered, prevented, mitigated, or punished in reparation" (p8 (11)).

The challenge lies in creating a framework that enables a consistent approach to this kind of analysis. Accepted ethical principles of health research are well established, issues related to privacy, effectiveness, accessibility and utility are clear or becoming clearer (13), but other issues (see next section) are still uncertain (14). Therefore, the Lab should establish an Ethics Advisory Board in order to monitor, analyse, and address the normative and overarching ethical issues that arise at the individual, interpersonal, group, institutional and societal levels in AI for healthcare (15).

Mapping the Issues

Building on the map developed by (12) of the normative, epistemic and overarching ethical issues associated with algorithmic decision making, Table 1 shows how ethical concerns can lead to harm when algorithms are used in the context of healthcare. They are related to (a) inconclusive, inscrutable or misguided evidence; (b) unfair outcomes or transformative effects; or (c) traceability.

	Ethical Concern	Explanation	Medical Example
Epistemic concerns	Inconclusive Evidence	Algorithmic conclusions are probabilities that are not infallible. They are rarely sufficient to posit the existence of a causal relationship.	*EKG readers in smartwatches may 'diagnose' a patient as suffering from arrhythmia when it may be due to a fault with the watch not being able to accurately read that user's heartbeat (for example due to the colour of their skin) or the 'norm' is inappropriately calibrated for that individual (16)*
	Inscrutable Evidence	Receivers of an algorithmic decision very rarely have full oversight of the data used to train or test an algorithm or the datapoints used to reach a specific decision.	*A clinical decision support system deployed in a hospital may make a treatment recommendation, but it may not be clear on what basis it has made that 'decision' raising the risk that it has used data that are inappropriate for the individual in question or that there is a bug in the system leading to issues*

			with over or under prescribing (17).
	Misguided Evidence	Conclusions can only be as reliable (but also as neutral) as the data they are based on.	*Watson for Oncology is in widespread use in China for 'diagnosis' via image recognition but has primarily been trained on a Western data set leading to issues with concordance and poorer results for Chinese patients than their Western counterparts (18).*
Normative Concerns	Unfair outcomes	An action can be found to having more of an impact (positive or negative) on one group of people	An algorithm *'learns' to prioritise patients it predicts to have better outcomes for a particular disease. This turns out to have a discriminatory effect on people within the Black and minority ethnic communities (19).*
	Transformative effects	Algorithmic activities, like profiling, re-conceptualise reality in unexpected ways.	*An individual using personal health app has limited oversight over what passive data it is collecting and how that is being transformed into a recommendation to improve, limiting their ability to challenge any recommendations made and a loss of personal autonomy and data privacy (20).*
Overarching	Traceability	Harm caused by algorithmic activity is hard to debug (to detect the harm and find its cause), and it is hard to identify who should be held responsible for the harm caused.	*If a decision made by clinical decision support software leads to a negative outcome for the individual, it is unclear who to assign the responsibility and or liability to and therefore to prevent it from happening again (21)..*

Table 1: A summary of the epistemic, normative and overarching ethical concerns related to algorithmic use in healthcare based on (12).

The examples given in Table 1 provide a broad overview of the ethical challenges that need to be considered if the benefits of data-driven healthcare are to be achieved (14). However, they primarily focus on potential risks and harms at the individual level. This is typical of the current literature on AI for healthcare (22). The concern is that this dominance in the literature has prompted policy responses that also focus solely on *individual* level impacts. For example, the NHS's Code of Conduct for Data-Driven Health and Care Technology (23) asks developers to consider their 'specific' user when carrying out tasks, such as a data protection impact assessment, data minimisation processes, and evaluation of evidence of effectiveness, but gives no guidance to commissioners on how to assess the impact of introducing an algorithmic service at a group level (24).

Although vital, this exclusive focus on an individual level fails to recognise that the ethical risks highlighted in Table 1 can impact relationships between people, groups of people, whole populations, and even institutions, not just individuals. Data are now circulating outside of the boundaries of formal healthcare systems, shared with third parties for research and commercial purposes (19), connecting personal, providers' and population's health information in complex feedback loops that exist at many levels (25). Due to these complexities, existing institutional review boards are struggling to evaluate increasingly technical research proposals (26). Thus, in order to design and manage an appropriate pro-ethical blueprint that would effectively deal with these new risks, an Ethics Advisory Board of ethicists, policymakers, clinicians, patients, developers, academics, regulators and information governance experts (as a minimum) seems necessary. This Board will be able to take into account a broader set of variables than those already

recognised or protected in the above outlined policies and examples, by considering these risks at a number of different Levels of Abstraction (LoA)[1](27).

A recent systematised thematic review (28) of the ethics of AI for healthcare literature identified five main LoA at which ethical concerns with regards to AI for healthcare arise: (i) individual, (ii) interpersonal, (iii) group (e.g. family or population), (iv) institutional, and (v) societal. Take, for example, the epistemic and ethical concern of misguided evidence and AI triaging systems, which use patient-reported symptomatic and demographic data to help patients identify what they should do next (e.g. stay home, see a GP, go to hospital), or help clinicians identify which patients should be prioritised in a hospital setting. We have five possible LoA:

a. **Individual LoA:** the algorithm could mis-judge the severity of an individual's symptoms. For example, nausea and back-pain are symptoms of a heart-attack in women but not typically in men (29), if the triaging algorithm has been trained to recognise a heart-attack on a dataset biased towards men, it may not identify these as potentially severe symptoms in a female patient.
b. **Interpersonal LoA:** there could be a negative impact on the relationship between clinician and patient, if the patient trusts the recommendation or diagnosis of the triaging system more than that of their clinician, but the clinician disagrees with it (30).
c. **Group LoA:** biased training datasets could lead to disproportionately better or worse health outcomes for different groups of people. For example, approximately 80% of participants in genome-wide association studies are of European descent (31)—this is true of the majority of medical datasets—meaning that the triaging

1 A level of abstraction can be imagined as an interface that enables one to observe some aspects of a system analysed, while making other aspects opaque or indeed invisible. For example, one may analyse a house at the LoA of a buyer, of an architect, of a city planner, of a plumber, and so on. LoAs are common in computer science, where systems are described at different LoAs (computational, hardware, user-centred etc.). LoAs can be combined in more complex sets, and can be, but are not necessarily always, hierarchical (27).

algorithm could be consistently less accurate for individuals from Black and Ethnic Minority groups leading to overall poorer health outcomes for these individuals than others.

d. **Institutional LoA:** there is a considerable reduction in the ability of the system to protect patient safety and therefore maintain patient trust, as data used to train and test healthcare algorithms flow from data collectors, to data aggregators, to analysers, (19) some of whom may be under no obligation to tell regulators or healthcare providers how the data were collected, aggregated, stored, or processed. This can significantly limit the ability of an Institution to respond if any of the harms associated with the other LoAs come to pass.

e. **Societal LoA:** the triaging algorithm is just one node in a learning healthcare system (14). The feedback it provides to policymakers on population health—for example where certain diseases are more likely to occur and therefore where resources should be allocated—could lead to greatly inequal provisions of care. With no ability to audit the algorithm itself, people living in negatively affected areas may have no mechanisms to evidence calls for redress (32).

This is just an example, but it makes clear the number of issues that could be missed if ethical considerations focus solely on individual level impacts or are reduced exclusively to a matter of legal compliance.

Clearly, a pro-ethical blueprint for AI for Healthcare must consider epistemic, normative and traceability ethical concerns at five different LoAs, but there is one more element that needs to be considered: the different stages of the life-cycle of an algorithm: design, development, deployment, use, and possibly re-design, and so forth. It is relatively common for those commenting on the ethical implications of algorithms to use the phrase 'rubbish in, rubbish out' to stress that if, for example, an algorithm trained on a biased dataset will likely produce

discriminatory outcomes. However, the ethical impact of an algorithm can be altered in either direction at each stage of development, from intention setting (business case development) to design phase, to training and test data procurement, to building phase, to testing phase, and finally to deployment (33), sometimes recursively. For example, an algorithm designed to recognise breast cancer in mammography scans more accurately—if it is built on an ethically procured and representative dataset, but deployed in a healthcare system that is unable to cope with the potential increase in volume of diagnosed patients—could lead to individuals living with a potentially worrying diagnosis, with no help, for significantly longer than before it was deployed. This scenario could result in a loss of autonomy for the individuals in question due to the very negative psychological impacts of the associated anxiety; an example of a negative transformative effect playing out at both the individual and institutional LoAs. Thus, unless all those involved in the life-cycle of healthcare algorithms are encouraged to consider the ethical impacts of the decisions they make at each stage, there is a risk that pro-ethical design principles written into the business case may be coded out by the time a system gets to deployment (34).

Conclusion

Recent developments in the use of AI for diagnostics, drug discovery, epidemiology, system efficiency and 'P4' Medicine make it clear that the opportunities presented by the increasing use of algorithms in national and international healthcare systems are significant and should be welcomed (4). However, it is also clear that the ethical risks are broad, serious, and complex and need to be considered at both different LoAs and stages of algorithm development. This could make the challenge of pro-ethically designing AI for healthcare seem overwhelming. This is not our intention. We do not wish to give the impression that healthcare systems need to be timid about, and afraid of, adopting algorithmic solutions. On the contrary, we are recommending a bold and systematic approach, based on the recognition of the challenges and the need to address them as early as possible. and not as an afterthought, or a mere formality, by an Ethics Advisory Board.

Those designing, developing, deploying, and using algorithms in healthcare should anticipate and address relevant ethical concerns so that they can make better, pro-ethical, design decisions and be ready to redress and change direction if a mistake becomes apparent (35). The failure of care.data should not be repeated and a backlash can and must be avoided. If this bold approach will be pursued, the significant benefits of AI for health will become a concrete possibility for all (14).

BIBLIOGRAPHY

1. Department of Health and Social Care. Health Secretary announces £250 million investment in artificial intelligence [Internet]. [cited 2019 Aug 8]. Available from: https://www.gov.uk/government/news/health-secretary-announces-250-million-investment-in-artificial-intelligence

2. NHS England. The NHS Long Term Plan [Internet]. NHS; 2019 Jan [cited 2019 Apr 15]. Available from: https://www.longtermplan.nhs.uk/wp-content/uploads/2019/01/nhs-long-term-plan.pdf

3. Department for Business, Energy & Industrial Strategy. Artifical Intelligence and Data Grand Challenge. Mission: Use data, Artificial Intelligence and innovation to transform the prevention, early diagnosis and treatment of chronic diseases by 2030 [Internet]. 2018 [cited 2019 Aug 8]. Available from: https://www.gov.uk/government/publications/industrial-strategy-the-grand-challenges/missions#artificial-intelligence-and-data

4. Bartoletti I. AI in healthcare: Ethical and privacy challenges. Lect Notes Comput Sci Subser Lect Notes Artif Intell Lect Notes Bioinforma [Internet]. 2019;11526 LNAI:7–10.

5. Nelson A, Herron D, Rees G, Nachev P. Predicting scheduled hospital attendance with artificial intelligence. Npj Digit Med. 2019 Apr 12;2(1):26.

6. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med [Internet]. 2019 Jan [cited 2019 Apr 18];25(1):30–6. Available from: http://www.nature.com/articles/s41591-018-0307-0

7. Morley J, Floridi L. How to design a governable digital health ecosystem - Preprint. 2019 [cited 2019 Jul 22]; Available from: http://rgdoi.net/10.13140/RG.2.2.28320.74244/1

8. Macrae C. Governing the safety of artificial intelligence in healthcare. BMJ Qual Saf [Internet]. 2019 Apr 12 [cited 2019 Apr 22];bmjqs-2019-009484. Available from: http://qualitysafety.bmj.com/lookup/doi/10.1136/bmjqs-2019-009484

9. Morley J, Joshi I. Developing effective policy to support Artificial Intelligence in Health and Care. Eurohealth [Internet]. 2019;25(2):11–4. Available from: https://apps.who.int/iris/bitstream/handle/10665/326127/Eurohealth-V25-N2-2019-eng.pdf?sequence=1&isAllowed=y

10. Floridi L. Soft ethics, the governance of the digital and the General Data Protection Regulation. Philos Transact A Math Phys Eng Sci. 2018 Oct 15;376(2133).

11. Floridi L. Faultless responsibility: on the nature and allocation of moral responsibility for distributed moral actions. Philos Trans R Soc Math Phys Eng Sci [Internet]. 2016 Dec 28 [cited 2019 Mar 9];374(2083):20160112. Available from: http://rsta.royalsocietypublishing.org/lookup/doi/10.1098/rsta.2016.0112

12. Floridi L. The Logic of Design as a Conceptual Logic of Information. Minds Mach [Internet]. 2017 Sep [cited 2019 Mar 9];27(3):495–519. Available from: http://link.springer.com/10.1007/s11023-017-9438-1

13. Nebeker C, Torous J, Bartlett Ellis RJ. Building the case for actionable ethics in digital health research supported by artificial intelligence. BMC Med [Internet]. 2019;17(1).

14. Char DS, Shah NH, Magnus D. Implementing Machine Learning in Health Care - Addressing Ethical Challenges. N Engl J Med. 2018 Mar 15;378(11):981–3.

15. Mittelstadt BD, Allo P, Taddeo M, Wachter S, Floridi L. The ethics of algorithms: Mapping the debate. Big Data Soc [Internet]. 2016 Dec [cited 2019 Mar 9];3(2):205395171667967. Available from: http://journals.sagepub.com/doi/10.1177/2053951716679679

16. Hailu R. Fitbits and other wearables may not accurately track heart rates in people of color. STAT [Internet]. 2019 [cited 2019 Aug 8]; Available from: https://www.statnews.com/2019/07/24/fitbit-accuracy-dark-skin/

17. Wachter RM. The digital doctor: hope, hype, and harm at the dawn of medicine's computer age. New York: McGraw-Hill Education; 2015. 330 p.

18. Liu C, Liu X, Wu F, Xie M, Feng Y, Hu C. Using Artificial Intelligence (Watson for Oncology) for Treatment Recommendations Amongst Chinese Patients with Lung Cancer: Feasibility Study. J Med Internet Res [Internet]. 2018 Sep 25 [cited 2019 Aug 8];20(9):e11087. Available from: http://www.jmir.org/2018/9/e11087/

19. Garattini C, Raffle J, Aisyah DN, Sartain F, Kozlakidis Z. Big Data Analytics, Infectious Diseases and Associated Ethical Impacts. Philos Technol [Internet]. 2019 Mar [cited 2019 Apr 27];32(1):69–85. Available from: http://link.springer.com/10.1007/s13347-017-0278-y

20. Kleinpeter E. Four Ethical Issues of "E-Health". IRBM [Internet]. 2017;38(5):245–9.

21. Racine E, Boehlen W, Sample M. Healthcare uses of artificial intelligence: Challenges and opportunities for growth. Healthc Manage Forum [Internet]. 2019;

22. Mann SP, Savulescu J, Sahakian BJ. Facilitating the ethical use of health data for the benefit of society: Electronic health records, consent and the duty of easy rescue. Philos Trans R Soc Math Phys Eng Sci [Internet]. 2016;374(2083).

23. Department of Health and Social Care. Code of conduct for data-driven health and care technology [Internet]. GOV.UK. 2019 [cited 2019 Apr 15]. Available from: https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology

24. Taylor L, Floridi L, van der Sloot B, editors. Group privacy: new challenges of data technologies. Switzerland: Springer; 2017. 237 p. (Philosophical studies series).

25. Flahault A, Geissbuhler A, Guessous I, Guérin P, Bolon I, Salathé M, et al. Precision global health in the digital age. Swiss Med Wkly [Internet]. 2017 Apr 7 [cited 2019 Apr 28];147(1314). Available from: http://doi.emh.ch/smw.2017.14423

26. Maher NA, Senders JT, Hulsbergen AFC, Lamba N, Parker M, Onnela J-P, et al. Passive data collection and use in healthcare: A systematic review of ethical issues. Int J Med Inf [Internet]. 2019;129:242–7.

27. Floridi L. The Method of Levels of Abstraction. Minds Mach [Internet]. 2008 Sep [cited 2019 Apr 1];18(3):303–29. Available from: http://link.springer.com/10.1007/s11023-008-9113-7

28. Machado C, Burr C, Morley J, Taddeo M, Floridi L. The Debate on the Ethics of AI in Health Care: a Reconstruction and Critical Review. Forthcoming.

29. Berg J, Björck L, Dudas K, Lappas G, Rosengren A. Symptoms of a first acute myocardial infarction in women and men. Gend Med [Internet]. 2009 Sep [cited 2019 Aug 8];6(3):454–62. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1550857909000990

30. Vollmer S, Mateen BA, Bohner G, Király FJ, Ghani R, Jonsson P, et al. Machine learning and AI research for Patient Benefit: 20 Critical Questions on Transparency, Replicability, Ethics and Effectiveness. ArXiv181210404 Cs Stat [Internet]. 2018 Dec 21 [cited 2019 Apr 18]; Available from: http://arxiv.org/abs/1812.10404

31. Popejoy AB, Fullerton SM. Genomics is failing on diversity. Nature [Internet]. 2016 Oct [cited 2019 Aug 8];538(7624):161–4. Available from: http://www.nature.com/articles/538161a

32. Diakopoulos N. Algorithmic Accountability: Journalistic investigation of computational power structures. Digit Journal [Internet]. 2015 May 4 [cited 2019 Mar 9];3(3):398–415. Available from: http://www.tandfonline.com/doi/full/10.1080/21670811.2014.976411

33. Binns R. An Overview of the Auditing Framework for Artificial Intelligence and its core components [Internet]. ICO. Available from: https://ai-auditingframework.blogspot.com/2019/03/an-overview-of-auditing-framework-for_26.html

34. Morley J, Floridi L, Kinsey L, Elhalal A. From What to How. An Overview of AI Ethics Tools, Methods and Research to Translate Principles into Practices. ArXiv190506876 Cs [Internet]. 2019 May 15 [cited 2019 May 17]; Available from: http://arxiv.org/abs/1905.06876

35. Floridi L. AI opportunities for healthcare must not be wasted. Health Management. 2019;19.