Ethics of AI in global health research

Cape Town, 29&30 November 2022

Governance paper
Responsible research and development in AI for healthcare: what we are learning from establishing a national collaborative platform in the UK

Dr Kate Devlin, King's College London, UK

Context: National good practice/guidance/frameworks for research – setting up a government-funded research hub on trusted and trustworthy systems.

Introduction

Today, easy access to large amounts of data and of computing power means machine learning is embedded and used in many of the applications we take for granted. However, power inequalities in AI at a global level – the dominance of corporate control in the US, UK and Europe – mean that AI systems and their creators are designing and deploying products that inherently benefit them and often actively harm outgroups.

In healthcare, AI shows clear promise in automating diagnoses and personalise treatments. This includes the use of autonomous systems: software and machines that are able to take actions with little or no human supervision. For example, during the Covid-19 pandemic, autonomous systems such as mobile robots were able to significantly reduce the risk of infectious disease transmission[1]. But automation is not without its social impact downstream: autonomous systems can threaten clinicians' agency, could lead to reduced funding in staffing, and could result in job loss, as well as facing scepticism and concern from patients.

As the social repercussions of AI emerge, the ethics community has paid closer attention to methods of mitigation, such as guidelines, principles and frameworks[2-5] and a wider call for attention to responsible research and innovation. Actual regulation, however, is not easy given issues around jurisdiction, corporate pushback, and lack of algorithmic transparency[6].

While there is much agreement on the need for ethical AI, the tangible steps – the ethics in action – are very much a work in progress. There is no quick software fix for AI's flaws, nor can we engineer out deep-rooted and systemic biases overnight. However, now is the time to implement practical guidelines and approaches from the ground up. An additional responsibility is to earn the trust of the end user via transparent development, clear communication, and informative engagement.

The Trusted Autonomous Systems Hub

With the above in mind, three UK universities – Southampton, Nottingham, and King's College London – applied to establish a national network on trust in AI that foregrounded ethical, responsible and inclusive development. The resulting UK's Trusted Autonomous Systems Hub[7] is the world's largest research programme in Trustworthy AI and Autonomous Systems. It is the focal point of the £33m UKRI TAS Programme, involving six TAS Nodes dedicated to the topics of functionality, resilience, security, governance and regulation, verifiability, and trust. TAS has over 100 international industry partners. Health and AI is a large part of our remit.

From the start, we actively sought to create a programme that would begin to address these issues by making our work contingent on adhering to fairness and equity in all that we do. Our goal is to deliver world-leading best practices for the design, regulation and operation of socially beneficial autonomous systems which are both trustworthy in principle and trusted in practice.

TAS is built around the core principles of responsible research and innovation (RRI) with equal attention to equality, diversity and inclusion (EDI) – ethical by design. These are issues that are key our research, and any projects we fund must centre these criteria in named, practical ways before we will consider supporting them. We fund projects, set up networks, advise on policy, and invite researchers, industry, NGOs and the public to engage and contribute use-cases/datasets or collaborate on research projects, tech transfer, and training activities.

TAS carries out research internally and also awards open grants to UK academics and industry partners. For our pump-priming funding – pilot studies to explore areas of AI and trust – our co-investigators, network members and industry partners worked on the following health research:

Trustworthy light-based robotic devices for autonomous wound healing – Robotic technologies have the potential to guide wound healing at the cellular level. Machine learning allows us to tailor the control to individual cellular dynamics on the go, enabling personalised solutions. This raises questions about how to ensure these systems are trustworthy and safe.

Identifying conflict and confluence in stakeholder imaginaries of autonomous care systems – Identifying the conflicts and confluences in the imaginaries of robotic and autonomous systems in the health-social care ecosystem by conducting of expert interviews and workshops with stakeholders and members of the public.

A participatory approach to the ethical assurance of digital mental healthcare – Developing a novel approach to assurance through participatory methodology, to underwrite the responsible design, development, and deployment of autonomous and intelligent systems in digital mental healthcare.

Co-designing Trustworthy Autonomous Diabetes Systems – Designing algorithms for diabetes management during life transitions using co-design, provenance and explainable AI, bringing together clinicians, data scientists, and people with type-1 diabetes.

Trustworthy autonomous systems to support healthcare experiences – How trustworthy autonomous systems embedded in devices in the home can support decision-making about health and wellbeing.

Diagnostic AI System for Robot-Assisted A&E Triage – Prototyping a robot-assisted A&E triage solution for reducing patient waiting time and doctor workload.

TAS Hub is in its second year now and we are already gaining valuable insight in how to set up a national framework for conducting ethical, responsible research in AI.

Conclusion and recommendations
This practical governance paper focuses on the bigger picture: the move from theory to practice in terms of AI ethics and responsible research development. There are two aspects to our work which we seek to highlight at GFBR:

- First, *trustworthy in principle*, where we share what we have learned from setting up a large, nationally-funded, multidisciplinary programme on safe AI – one that is novel in its approach by centering ethics and ethical development ahead of results and outputs. This involves the successes and also the failures, and the course corrections we have had to take. This includes:
    o devising criteria for grant reviews that reward projects which centre stakeholder engagement;
    o promoting early career leadership opportunities;
    o ensuring tangible ethical approaches;
    o writing actionable equality, diversity and inclusion strategies – and using them;
    o forming an operational framework;

- o collaborating with industry (and researchers' varying reactions to who we work with – some people have strong views on who invests in our work); and
- o the choices we made when setting up our Board, our Strategy Advisory Network and our International Scientific Committee.

- Second, *trusted in practice*, where we are embarking in collaboration not only with industry and academia but also with end-users and the public. This includes such aspects as participatory and collaborative project design, as well as the resulting education and outreach, as we strive to be transparent about a future with autonomous systems. This includes:
- o working with stakeholders, such as charities and mental health service-users, to explore ideas around trust in digital mental healthcare systems;
- o providing resources such as video conversations, podcasts, and teaching materials; and
- o commissioning and developing interactive creative artwork that leads to thought-provoking encounters – for example, our Cat Royale art project that involves pets being cared for and played with by a robot arm[8].

We will describe what is working well for us as we establish a national platform, and what struggles we have faced, particularly when operating in a sector which often prioritises results over method, and one where industry collaboration may mean differing expectations or priorities. We hope that this information can be shared widely and freely to encourage similar initiatives, and we seek collaboration on this type of work from existing or nascent programmes. We are particularly keen for anything people find useful from TAS to be reused, adopted and adapted as needed by those working towards the same goal.

References
1. Tavakoli, M., Carriere, J., & Torabi, A. (2020). Robotics, smart wearable technologies, and autonomous intelligent systems for healthcare during the COVID-19 pandemic: An analysis of the state of the art and future vision. Advanced Intelligent Systems, 2(7), 2000071.
2. Bartoletti, I. (2019). AI in Healthcare: Ethical and Privacy Challenges. In: Riaño, D., Wilk, S., ten Teije, A. (eds) Artificial Intelligence in Medicine. AIME 2019. Lecture Notes in Computer Science(), vol 11526. Springer, Cham. https://doi.org/10.1007/978-3-030-21642-9_2
3. Dignum, V. (2019). Responsible artificial intelligence: how to develop and use AI in a responsible way. Springer.
4. Winfield, A. (2019). Ethical standards in robotics and AI. *Nature Electronics*, 2(2), 46–48.
5. Floridi, L., Cowls, J., King, T. C., & Taddeo, M. (2020). How to Design AI for Social Good: Seven Essential Factors. Science and Engineering Ethics, 26(3), 1771–1796.
6. Hoffmann-Riem, W. (2020). Artificial intelligence as a challenge for law and regulation. In *Regulating artificial intelligence* (pp. 1-29). Springer, Cham.
7. Trusted Autonomous Systems Hub web site: https://www.tas.ac.uk/ (accessed Sept 2022)
8. Cat Royale by Blast Theory https://www.youtube.com/watch?v=WLAy3GWBvzQ (accessed Sept 2022)

This paper was prepared for GFBR 2022. Further details on the meeting are available at www.gfbr.global.