

Blog

 [Back to blog](#)

Involving patients in AI research to build trustworthy systems

Soumya Banerjee

08 November 2022

Categories: Machine learning

Tags: News



Machine learning and artificial intelligence (AI) are helping researchers and clinicians make great strides in healthcare. However, just half of people think AI will bring more benefits than drawbacks, according to a World Economic Forum [survey](#), so communicating its beneficial medical applications, and engaging the public with healthcare research is critical.

better AI systems in healthcare projects. I believe that integrating patient and public involvement (PPI) in AI projects may help in adoption and acceptance of these technologies.

Building better AI systems

I specialise in explainable AI, which is a set of processes and methods that allow users to comprehend and trust the results and outputs created by ML algorithms. Useful explanations will be different for computer scientists, clinicians and patients alike. One way to identify which explanation is most useful in a given context is to let people play with models and help generate their own explanations.

In a paper published in a [Cell Press Journal](#), we argue that AI algorithms for healthcare should be co-designed with patients and healthcare workers. In the case study we presented, a patient with a severe mental illness came to us to see whether we could use AI to analyse electronic hospital records to better understand the chance of side effects associated with coming off the medication he was taking.

Having a patient approaching us with a clear hypothesis to set the research agenda was profound. This patient-led approach prevents study participants from becoming mere datapoints and potentially becoming marginalised in the research process. We are still waiting for some of the data, but were able to build simple models to get the process started. Initially we built linear regression models and explained them to the patient. We added layers and in monthly meetings, took steps to explain and validate them with everyone involved.

In the study, we laid out what we think is the gold standard for this collaborative approach. The process we describe ideally involves actively including patients with lived experience of a disease, as well as creating a research advisory group (RAG) to walk patients through the process of AI model building, starting with simple models.

Perhaps the most important benefit of this methodology is that co-developing models transparently, and marrying the technical with the human builds trust. It enables patients to not only help shape the models, but also develop a relationship of trust with the team of researchers behind them, because they listen to people and try to understand their needs and allay their concerns.

Opening up AI models to investigation not only enables rich interaction, but pinpoints errors and areas of improvement too. It can also highlight unconscious biases.

Unfortunately, such a participatory approach does not happen often enough because it can be difficult. It's much easier to get access to data and write papers and it can be hard to overcome logistical challenges like recruiting patients and finding time for everyone to meet. There are also few incentives for researchers to work in this way as we typically gain tenure for publishing impactful papers rapidly, rather than experimenting with new engagement techniques. I believe scientists need more training in engaging effectively with stakeholders to facilitate successful collaborations. Resources that can be shared with the public – such as the repository we created – are helpful. I would love to see the incentive structure modified so we could start to see ripple effects that would benefit patients and researchers alike.

A collaborative future

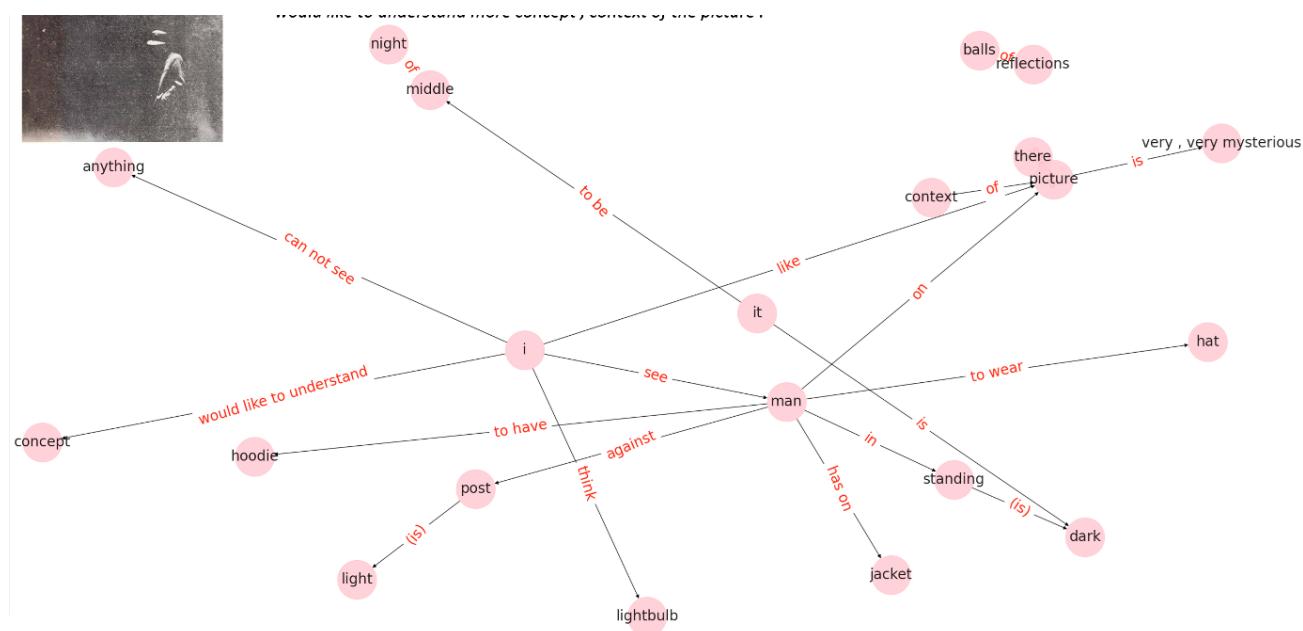
We wrote in our paper that a collaborative process ideally necessitates the creation of a research advisory group to involve all stakeholders in the process of building AI models and I believe this is how all AI research should be conducted; where everyone, including end users and domain experts like clinicians and data scientists can work together effectively with a shared goal, and collectively benefit from the results.

An understanding of what AI can and cannot do will build trust and allay fears, while a realistic appraisal of risks and benefits may help in adoption and



Cambridge quick links

Related news



How can we... use natural language processing to learn more about schizophrenia?

23 March 2023

[READ MORE](#)



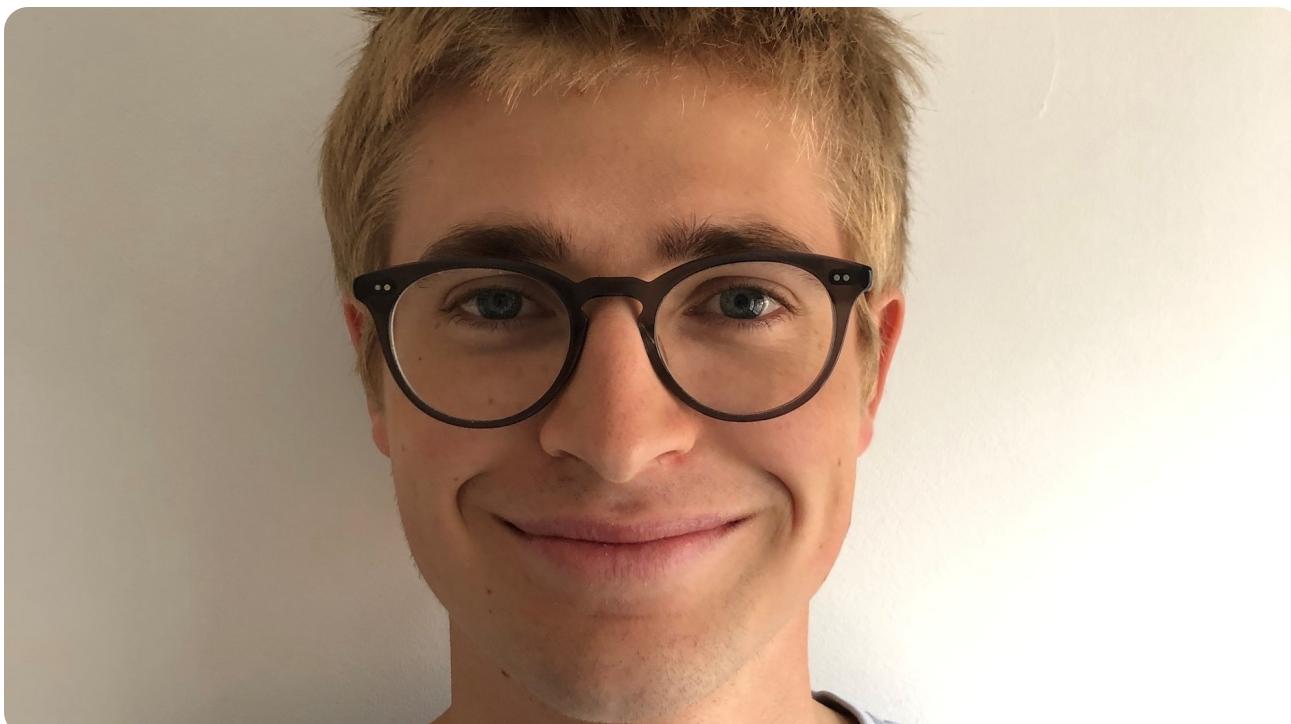


[Cambridge quick links](#)

immunotherapy?

20 March 2023

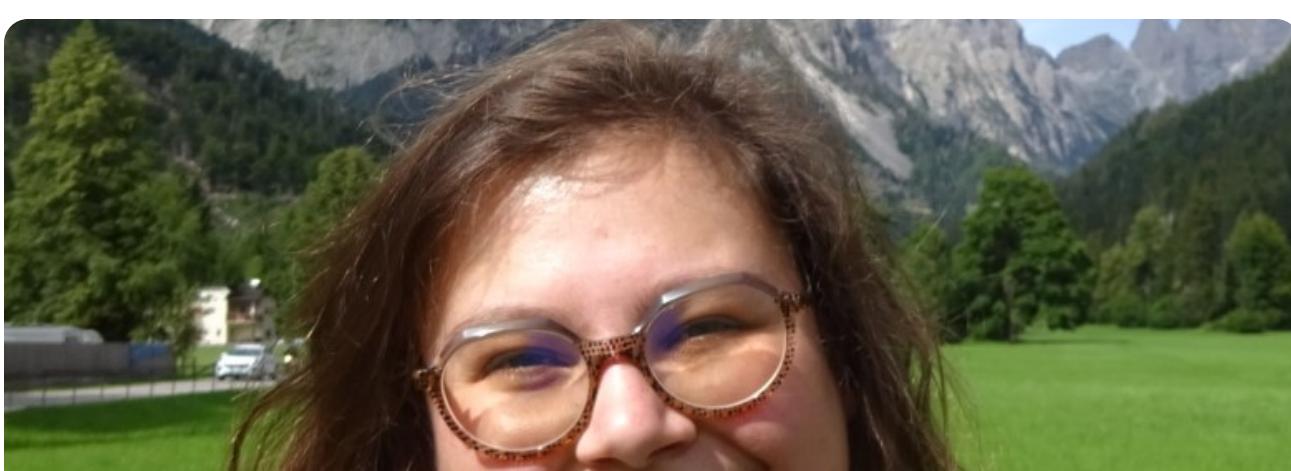
[READ MORE](#)



Using AI to aid causal methods in environmental science

27 February 2023

[READ MORE](#)





[Cambridge quick links](#)

ACCELERATE SPARK DATA SCIENCE RESIDENCY

How can we ... use AI to understand how we read?

15 February 2023

[READ MORE](#)

NAVIGATION /

[About Us](#)

[Projects](#)

[News](#)

[Events](#)

[Software Engineering](#)

[Resources](#)

[Get Involved](#)

[Privacy](#)

CAMBRIDGE UNI /

[Study at Cambridge](#)

[About the University](#)

[Research at Cambridge](#)

CONTACT US /

Department of Computer Science and Technology

William Gates Building



Cambridge quick links

CB3 OFD

Contact: accelerate-science@cst.cam.ac.uk

SOCIAL /



Designed and built by The Lonely Pixel