

## What we still need to use AI safely and quickly in the healthcare

**Rachel Dunscombe and Jane Rendall examine what needs to happen to make sure extremely important algorithms can be introduced rapidly, and in a way that is clinically safe for populations.**

When one NHS trust in the North of England started to introduce artificial intelligence several years ago, hospital clinicians needed to sit postgraduate data science courses in order to understand how algorithms worked.

Like most healthcare organisations, the Trust didn't have a uniformed approach to onboarding algorithms and applying necessary supervision to how they performed.

It became a manually intensive operation for clinicians to carry out the necessary clinical safety checks on algorithms, requiring a huge amount of overhead and in turn significantly limiting the organisation's ability to scale the use of AI.

### **AI needs supervision**

AI in many ways needs to be managed like a junior member of staff. It needs supervision. Hospitals need to be able to audit its activity, just as they would a junior doctor or junior nurse, and they need sufficient transparency of how an algorithm works in order to provide necessary oversight and assess if and when intervention is needed to improve its performance and ensure it is safe.

So, how can we do this in a scalable way? Expecting doctors to do a master's degree in data science isn't the answer. But developing a standard approach to managing the lifecycle of algorithms could be. In the UK, organisations like NHSX are making progress. But the real opportunity is to develop an internationally accepted approach.

If we are to adopt AI at the pace and scale now needed to improve care, and to address widening workforce and capacity gaps, we need to address the current absence of international standards on AI adoption. This could help to inform developers before they start to produce algorithms, and inform the safe application of those algorithms to specific populations.

Put simply, this is about what we need to do in order to make sure we adopt AI with similar diligence that we apply to safely adopting new medicines, but without having to wait the years it can take to get important medicines to patients.

### **A starter for 10 – thinking about an international approach to AI**

Arriving at that international consensus will mean a lot of rapid progress and dialogue – and will most likely involve sharing lessons from across different sectors beyond healthcare.

But here are six suggestions of some of the components that could underpin a model and help healthcare to safely accelerate adoption:

- 1. Clinical safety.** We need to embed AI into tools that can allow hospitals to examine the clinical safety of an algorithm. Healthcare organisations already have tools for clinical safety in their organisation – systems that gather data on the performance of doctors and nurses. Interfaces from AI algorithms should feed those same systems. We should report on AI in the same way as a doctor or nurse. There has been a lot of work from the Royal College of Radiologists about supporting junior colleagues to evolve in their career. Similar mechanisms could help to peer review the work done by the AI. This is about creating the same feedback cycles that we have for humans to understand where AI may have faltered or misinterpreted, so that we know where improvement is needed.

- 2. Bias detection.** This is about examining demographics based on age, gender, ethnicity, other factors and determining where bias might exist. Hospitals need to understand if there are people for whom an algorithm might work differently, or not work as effectively. It might not be suitable for paediatrics for example. Skin colour, and a great many other factors can also potentially be significant. If a bias is detected – two options then exist: training that bias out of the algorithm, or creating a set of acceptable pathways for people with whom it won't work and continue to use it for groups where a bias isn't present. This could mean answering some big practical and ethical questions around access and equity. For example, is it appropriate to have a manual pathway for someone if the algorithm doesn't work safely for them, and to use the AI for the remainder of the population? But to even get to those questions requires transparency. Algorithm developers need to be transparent on the cohorts used to train the algorithm. As a healthcare provider you can consider if this matches your cohort, or if there is a mismatch you should be aware of. You can then choose to segment your cohorts or your population, or capacity accordingly, or choose a different algorithm.
- 3. New demographic validation.** One local geography might have two demographic minorities. Another, only a few miles away, might have a significant mix of ethnic minorities making up around half the population. Healthcare systems, like the NHS in the UK for example, usually buy technology before extending it over other geographies. This requires looking at new demographic validation. If the population in question changes – for example through immigration an extension of services, or something else happening: an algorithm needs to be validated against a new dataset. Something that can operate safely in the UK, might not operate safely in parts of South America, or China. Bias detection has allowed for validation in your original population, but you can't test it on day one against every set of demographics where it might be used. There are so many ethnicities and groups on this planet that this has to be done in stages. So, as you extend the algorithm across new demographics, you need to validate. If a service in Mersey extended out to Manchester, then it would need to be tested again.
- 4. Explainable un-blackboxing.** Having to send doctors on data science degrees isn't practical. But we don't have a standard way of drawing pictures or writing words to say what an algorithm is doing at the moment. If you think about a packet of food, you get an ingredient list. We need a similar standardised approach for AI. We need to work towards explainable un-blackboxing that will include clinical terminology, but it will also include common measures we find across different industries in terms of performance. If you are going to get a CE mark or certification – it could be standard across health, nuclear, aviation and other sectors. The EU is early in its thinking on how that can work, but discussion has started.
- 5. Clinical audit.** We need a clinical audit capability in algorithms. If a case is taken to a coroner's court, if there has been an untoward incident, we will have to show how an algorithm contributed towards care. This is something we already do with human doctors and nurses. We need to do it with algorithms.
- 6. Pathway performance over-time.** In areas like radiology there is an opportunity to examine the performance of an algorithm compared with human reporting. This isn't about AI replacing humans, but it can help healthcare organisations to make decisions about where and how to make best use of the human in the pathway. For disciplines like radiology this is key, given the significant human resource challenge faced in some countries. We also need to think about this from the perspective of the patient. If algorithms can report a lot faster than humans, could humans delay the diagnosis, particularly when humans are being used for double reading? Could that impact the surgery or treatment? Are there opportunities to change that pathway, or to potentially use AI to help free up the human resource to focus on diagnosing more complex cases more quickly? This is about looking at the performance of the pathway and measuring outcomes where AI can make a difference. Playing that back to citizens at a time when trust

issues are still prevalent around algorithms, can help to demonstrate how AI is being used to improve healthcare.

Healthcare organisations are looking to AI to help to address a significant number of matters – from the ongoing pandemic to long established challenges. Not bringing AI will mean that we will otherwise hit crisis points – especially in areas like radiology, where in some countries demand continues to grow by around 10% year on year, whilst the number of trainees continues to decline.

But the situation is more complex than simply acquiring algorithms. A standard approach to managing algorithm lifecycle could make all the difference for successful adoption at the pace required.

**Rachel Dunscombe is a director for Tektology**

**Jane Rendall is UK managing director for Sectra**