AI & Healthcare

30th September – 1st October 2019

REPORT

What and who is 'healthcare'?
The term 'healthcare' embraces many, varied sectors and entities and their subdivisions. For the purposes of discussion, broadly, these were identified as:

- public healthcare providers: e.g., NHS services, trusts, GPs
- private enterprise: e.g., creators of healthcare apps, drug manufacturers, hospices, carers
- individuals: patient support groups, carers for family members, people treating themselves

The existing role of the healthcare provider (whether public or private) is necessarily challenged by innovations in healthcare provision, including AI.

Healthcare provision is much broader than 'doctoring'. However, as an example, examining the misalignments of incentives and motives between doctors, employers, and the public can help illustrate some of the kinds of gaps which proponents of artificial intelligence (AI) healthcare solutions would have it fill. Patients describe a doctor's (ideal) role in terms of accessibility, availability and convenience. However, doctors themselves may wish to pursue a more rewarding career through specialisation. Meanwhile, employers want doctors who integrate across silos and have insight into a broad range of medical issues.

Definition and distinction of AI in the healthcare context
An AI algorithm can be defined as software, which functions to replace human judgment. More specifically, the type of AI called machine learning (ML) is software that can learn from its informational environment and structure its own learning to reach different or new conclusions from those it would have reached before the learning period. This learning capacity distinguishes machine learning from other software or technology. Many AI systems currently in use or being developed are characterised by their ability to learn subtle patterns, sometimes not apparent to human experts, in large quantities of data in order to make inferences about disease or its treatment.

'AI' and 'ML' are often used interchangeably. This discussion of AI uses the generally recognised popular term but focusses on machine learning techniques.

The difference between 'tech' and 'AI' is significant for healthcare providers in a number of ways:

- public perception and trust or mistrust of AI *as* 'different'
- practitioners' perceptions and trust or mistrust of AI
- hype, political traction; funding
- the ability (and need, in many cases) of AI to crunch big data including personal health data
- acknowledged areas of performance in which machines do better than humans, e.g., image recognition in dermatology, radiography

Trust and bias

Trust is earned by demonstrating integrity, credibility and reliability i.e. trustworthiness. It is often engendered in a healthcare context by a sense of shared experience; for example, contributors to patient forums may provide information to others affected by similar conditions, and this information is trusted on the basis of shared experience rather than demonstrated medical expertise or accredited qualifications. The democratisation of AI production and an increase in open-source software with which to do so will see more AI health applications produced by patients for patients.

Trust must also be extended from practitioner to patient. This may mean (for example) that an expert clinician who participates at key points in the individual journey of a patient can also at times step back and allow patients to support each other. In an information-saturated age in which various forms of AI offer more and more opportunities for self-assessment or diagnosis, patients often arrive having educated themselves (as far as possible) about their symptoms or condition. A collaborative approach to the relationship between patient and provider may be key in future use of AI in healthcare.

Trust of healthcare providers by patients is usually high. Sometimes it may be inappropriately high in cases where, in fact, the quality of any care element is poor or inconsistent, or is not validated, or where healthcare providers do not have adequate means to measure whether human medical judgment is 'good' (for example, in conditions exhibiting as clusters of suggestive symptoms rather than definite, readily identifiable and observable characteristics such as a pathogen, tumour, etc.).

Trust/trustworthiness is a personal commodity. What forms must AI integrity, credibility and reliability take and to what standards in order for us to learn to trust AI agents in healthcare? And what happens when that trust is betrayed? (No AI agent is likely to be perfect).

In a system, perhaps trustworthiness is expressed through risk management. The question of 'who is responsible' (software manufacturer? health system deploying AI? practitioner who utilised AI output?) will likely determine liability in such a system. This is a significant question in managing AI for two connected reasons:

- AI reaches conclusions in ways which are often not easily explained (the 'black box' phenomenon), which makes it hard to evaluate the means by which it has reached its conclusions (human beings too may find it hard to explain how they have reached a conclusion but we have ways of rationalising after the event; can and should AI agents do likewise?)
- AI may well provide conclusions upon which life-changing or life-threatening decisions are made, and establishing the boundaries of liability for wrong medical decision-making between humans and machines is new legal territory

To help make decisions about determining liability in healthcare AI settings, it is important to separate 'drawing conclusions' (which AI could well do) from 'deciding to act on the conclusions, and how' in healthcare (which could and probably should remain a human preserve). Whether people trust more in humans or machines, our legal systems are better geared to the consequences of significant health decisions being made by humans. This places a heavy burden on human healthcare providers, however. In cases where machine conclusions very often prove right, it is easier and may seem better simply to follow the machine recommendation than to differ. But when machines do then err, humans are likely to be held liable for relying on 'flawed machine conclusions' in making decisions about healthcare.

It is hard to know how a human health professional in future will be able to comment professionally on an AI-generated diagnosis. Trust inheres not only in the system itself but in the surrounding ecosystem and in the reputations of those who can be held responsible.

Any AI healthcare applications or systems will rely on data provided from existing populations and repositories of information about people's health. These inevitably contain whatever bias has shaped the information entered, and these biases will enter (and may be magnified by) any AI process upon which decisions may be made that affect patients' or populations' health. This is a significant problem.

Further difficult questions around data protection, data privacy, trust, and public interest may not all be answered by existing legal structures, although GDPR presents clear rules on usage for data controllers vs data processors. Not only are there disagreements about appropriate/desirable use of personal data, but people change their point of view according to experience.

However, some existing procedures and regulation for medical processes and data use can be adapted for use around AI and refined over time. An important basic principle is to invest in infrastructure to give people information about how data is used, backed up by a credible audit trail, and encourage feedback and engagement with the process including follow-up. Trust in a system, which includes AI, can also be engendered by establishing patient leads and testing groups in parallel with the clinical, instituting patient-led AI projects, advisors, panels, and research, and thinking carefully about language used. Ethical frameworks such as the models used by Citizens' Councils may be valuable in guiding complex decision-making.

Hype

Ignorance about how AI works, its limitations and strengths, makes it hard to explain its applications to the general public and to practitioners, and therefore also difficult to confer greater agency in healthcare decision-making through greater knowledge. This 'difficult to understand' characteristic of AI tends to promote simplistic hype such as the projection of a workless future when experience shows that technological advances may well bring more work with them than before ('fauxtomation'). Hype can also be positive.

Timescales are also subject to hype: AI advances and possibilities are advertised long before they can reasonably be tested and implemented in a national health system. Finally, consumer products using even non-learning health software are promulgated as being 'AI-based'.

Data (commercial and non-commercial)

Data security, data protection, and privacy are distinct from one another. The monetary value of data (and who is entitled to that value) is much debated. If there is clear patient benefit, partnerships which allow money to be made outside the NHS using patient data may be of value. However, the emotive aspects of AI and personal data make this a hard trust issue, and conclusions which may apply to a more privileged group may look very different (and threatening) to, for example, members of immigrant or minority groups.

Data breach is largely viewed as an inevitability at some point for any organisation holding personal data. The question is what to do about it. Organisations should not over-promise on data security, should be honest about risks and balance them against benefits, and state candidly the available options, consequences and strategies for individuals submitting or withholding data.

A tension arose during the meeting between the idea that 'people do not trust AI in part because they are concerned not to lose control of personal health-related data' (such as having one's email address leaked from a gender-change clinic) and 'people are bored by the issue of personal data, and have to be persuaded to value and look after it'.

There is much debate around the value of data for AI and healthcare in general. Value is a concept that brings together both financial value and a set of non-financial values that might be created. Value is generally created from data by processing it for some purpose. However, the value of data for use in AI varies according to how it's been prepared; it has to be cleaned and sorted, and also 'prospected' for potential value. As AI continues to develop, so new uses will be found for different parts of the same data sets.

In commercial settings, if not in all pure research, it will probably remain true that the less anonymised the data is, the more valuable it will prove to analysing entities. A corollary is that, in many analytical settings, data cannot be completely anonymised and still remain useful data, and that much 'anonymised' data can in fact already be correlated with other information and analysed to identify individuals.

The degree to which an individual could or should own rights over every aspect of their personal health data is debatable: as with vaccination or tracking notifiable disease, there is a strong argument that some health data informs the public good. The NHS is a communal effort; we accept that some people pay more in, some people take more out; should the same not be true of personal information? It will not be easy to argue that individuals should yield up indefinitely to various agencies large quantities of personally identifiable, possibly sensitive health information for unknown use and application, although that is probably the ideal scenario for healthcare AI development. One consequence is that those individuals who contribute data will benefit others like themselves. This has implications for equitable healthcare.

In any discussion of AI in healthcare it should be established whether the AI is patient-facing, is intended to assist a healthcare provider directly (for example, in diagnosis) or is focused on internal administrative or logistical systems to assist in healthcare delivery more broadly; and also whether it is part of a consumer product or device, or is intended to have a place in the NHS. Answering these questions helps to focus goals and parameters, and (not least) regulatory requirements.

Strengths

As AI is relatively new, its strengths in healthcare applications remain largely unexplored. The NICE PICO table (population, intervention, comparators, outcomes) is a likely way to gauge AI strengths in particular cases and will be a necessary hurdle for any app to clear in order to be recommended for NHS use. In addition to the above, it is important to evaluate the regulatory status of any healthcare-related AI, its safety trajectory in real use, and information governance issues around consent and data sharing.

- In some areas of healthcare, particularly certain diagnostic processes involving image recognition, AI already produces better results than human beings: well-known examples include dermatological and radiographical analyses.
- AI requires and thrives upon vast quantities of data to train its learning capabilities. While a typical UK patient may spend on average 10–15 hrs/year with medical practitioners, an AI can evaluate data from the remaining 8000 hours for greater insights.
- Benefits to the patient as an individual: AI can be highly portable and responsive, delivering more of a sense of safety and reassurance to its users, which can help individuals to have more confidence in dealing independently with ongoing health issues.

- Benefits to NHS staff; in some areas, time savings are likely (e.g. automated image analysis); in other cases, it may create more work (e.g. excessive alerts; operating two, parallel systems)
- One strength of the healthcare sector is that it is cautious in adopting innovations. This can also be viewed as a weakness in that the sector is not geared to being 'leading-edge'.
- The hope is of an equitable, reliable, universal new set of tools to help shape better healthcare.

Weaknesses

- AI may be unreasonably over-trusted (automation bias) or mistrusted ('killer bots') by practitioners or patients. However, this may equally be said of the existing healthcare system relying on human conclusions and judgments. As with counselling over genetic information, it is advisable for models of AI healthcare analysis and prediction to include patient consultation with an expert human interpreter who can help with the trade-offs of knowledge vs action and risks associated with data being known (or not).
- AI may be inaccessible to sectors of the population and may exacerbate inequality of access. Those who already do not engage will not provide data, meaning that their cases will not be taken into account by the system.
- AI does not replace human judgment, partly owing to issues around trust and transparency and partly owing to inherent unpredictabilities in its decision-making.
- It can be hard to evaluate AI performance in the absence of a validated baseline (and the accuracy of human judgment can be inconsistent, flawed and hard to prove).
- AI does not replace human contact/company.
- Data carries inherent bias and incompleteness. Eliminating this requires complete data sets for the population.
- Human medical diagnoses at present include many 'fuzzy' categories of health-condition (such as 'pre-diabetes') which may not lend themselves to AI analysis.
- Scalability of AI applications per se is skewed by the speed of adoption of consumer products; deployable AI technologies can (and do) reach market long before they can have undergone medical trials and value testing for their place in the NHS. The consumer-led approach is fast and messy.
- Scalability of any new technology through the NHS faces challenges of fragmentation through 150 acute trusts, 220 clinical commissioning groups and 40 Mental Health Trusts.

Opportunities

Making AI more about opportunities than threats requires grappling with issues of equality and community and taking advantage of AI's strengths. Worthwhile opportunities address real pain-points in the NHS and healthcare provision more generally. At the delivery end:

- better diagnoses
- making up staff shortfalls in the NHS and social care (approx. 100,000 in each)

- savings in money (but this cannot be the sole end of using AI, or trust is lost) including potentially reducing future costs
- more control over their own health for individuals (but this must not result in loss of access to the NHS) and, therefore, fewer emergency hospital beds being used for e.g. dementia cases
- more opportunities for individuals to learn from, say, layered AI explanations with which to explore one's diagnosis
- shared decision-making between healthcare providers and patients – AI adds value around set data and set outputs (rather than everyday interactions)
- more assistance for a large ageing population
- potential intervention in areas where a large difference could be made across the population – for example, in improving cardiovascular health

In the NHS, AI can be used as a pathway or a point intervention; these require different approaches. Possibly, AI can assist at a systems level by facilitating existing processes in healthcare. Future roles such as Chief Nursing Information Officer should be envisaged.

Healthcare providers are often frustrated by the lack of time to see patients. Although AI is (like other innovations) sometimes touted as a time-saver, it's unlikely in practice that it will result in longer appointments or more leisure to cultivate leadership skills, study new research etc.

At least one collaboration between the NHS, industry and academia for tech-improved healthcare management (TIHM) has used AI to deliver tailored health services to (in this example) dementia patients and family carers in their own homes, helping fill the gap between hospitals and society. These results are reported to have prompted interest from other trusts. However, the difficulty of working across SMEs, hospitals and universities, and questions about value for money, market-readiness and scalability diminish the likelihood that such collaborations can characterise broad applications of AI in the NHS in the immediate future.

Threats
Data bias and incompleteness of data can result in poorer health outcomes for some sectors of the population than could or should be the case in a 'neutral' system. Nudging towards a white, affluent, Western set of standards may disadvantage others. The partiality of datasets can be generated unwittingly or deliberately engineered to target or omit sectors of the population.

Lack of trust in AI (as outlined above) is an issue for both healthcare provider and recipients of healthcare. Humans too are fallible but less visibly and more acceptably. For AI to be used effectively, it must be possible to:

- show *how* decisions are reached, by identifying stages and processes used and citing evidence and other experts
- subordinate AI conclusions to human judgment for any final say that affects people

- make consistent predictions

Provision of large quantities of personal data for use by AIs brings the threat of loss of control of that data by individuals or authorised users of the data.

Some jobs may be lost where AI is more effective or efficient than humans in e.g., diagnoses involving reference to very many images or lengthy behavioural analysis.

Businesses providing healthcare AI as consumer products, like tech applications more generally, may be overly driven by their ability to supply a sellable 'solution' to a 'problem' that is not really a problem ('a hammer looking for nails'). Health outcomes can therefore be skewed by business goals – for example, the widely believed dictum that one should walk 10,000 steps per day was driven largely to begin with by marketing efforts to sell an app, which was capable of measuring steps. Secondly, the personalisation of health care could result in atomisation of healthcare; compare, for example, the rise in 'genetic' diagnoses and treatments.

As noted above, AI shifts liability for mistakes away from a system containing indemnifications and towards the patient/carer using the AI. It is hard to see how to resolved questions of liability around who is responsible for poor outcomes if a doctor disregards AI 'advice' (or takes it). The example of responsibility in the case of accidents in self-driving cars is not entirely analogous.

There is a risk that by deferring or relegating health-management tasks to individuals through AI applications, the individual will engage in self-surveillance, loss of patient perspective through the 'digital medical gaze', turning healthcare into an aspect of 'lifestyle choice', and will be manipulated through nudging according to others' agenda. This is a question of who takes (or is given, and by whom) the responsibility for maintaining people's health.

Challenges

- How to advance the health sector at a pace concomitant with the speed of consumer technology advances, while remaining thoughtful and measured when regulation is either too great or lacks nuance.
- How to find credible experts who can advise on AI health technology for regulation and use.
- How to gather data on communities and individuals who are less engaged with the health system in order to train AIs on these for greater access vs disempowerment
- How to match the communal nature of healthcare with the personalisation inherent in consumer apps and the personal data at risk; some factors in health interventions are social (eg vaccination)
- Different data rules and standards in use apply in different parts of the world, while health data can be used more broadly than the boundaries of these rules.

- We must identify and use respected methodologies in developing AI health applications so that they can move fast and securely to market; and check value, as tech solutions tend to come in expensive on the costs pyramid, which NICE views unfavourably; to contemplate a new product or method (to benefit the system as a whole or patients) would require either better health benefits that might cost £20–£30K per quality-adjusted life year, or else a non-inferior clinical performance with less overall cost.

What is needed

More and broader conversation and communications on the topic

- Reconvene for a more focussed discussion face-to-face. The Institute of Engineering and Technology has agreed to co-host an event with St George's House in February 2020; more details to come soon through Guy Gross. **Participants are asked** to suggest senior representatives who will strengthen conversations going forward, especially in areas not represented so far (MHRA, NIHR, CQC, wider industry representation, etc.)
- Extend the conversation to the wider community through a position paper in a peer-reviewed journal.
- Build a formal network for 'AI in Healthcare', linked to SGH, to meet informally and regularly (perhaps every two months in London) to continue the conversation and sustain a place for thought leadership; open, honest, considered public debate to address issues round privacy and intrusion in relation to health data, along with attention to the role of the media in articulating issues in a responsible way.
- Bring this position paper to government attention and, where government will spend money, help make sure it is spent wisely, through conversation with government, industry and colleagues.
- Seek DataLab funding to help maintain this conversation and momentum.

Broader education in AI

- More understanding and literacy re standards, models, levels of evidence from different sources; evidence-based AI benefits and a knowledge base; good use cases, examples, case studies for AI specifically which can be understood by the public too.
- Encourage more training work including key concepts and concerns; issue basic AI-in-healthcare training videos to first-year PhD students?
- Build on existing collaboration and recognise that a shared-information network can help coordinate our individual activities; exploit this network to share information; those of us whose day-job includes a remit to coordinate can take this on.

Collaboration and shared standards

- More shared decision-making with patients.
- More, successful partnerships to help identify sweet spots where technology can be adopted to make a significant difference.
- Insurance companies must discuss liability in AI-healthcare scenarios.

- Doctors and nurses must raise the standards of their knowledge and find opportunities to converse with each other, technological experts and patients in discussing risks/benefits of AI in healthcare.
- Find out who owns or recommends what metrics for 'good' in various areas of healthcare (including procedures and processes) and how/whether, these can be redeployed for use in measuring AI: consult NICE standards, Medical Development Register.
- A set of 'Fair Trade'-style standards for data? – ethically-sourced, accredited, secure, the values and biases known, returning benefits to the community whence it issued?

Realising NHS AI

- A coherent strategy for AI applications in healthcare in the UK.
- Public showcased examples of AI health applications, which can be shown reliably to work: perhaps three to ten topics could be supported as the right topics for value propositions towards trust in the right technology.
- Solutions, rather than technology for technology's (or cost-cutting's) sake.
- Keep what works.

Participants

Mr Jorge Armanet	CEO / Founder, HealthUnlocked
Dr Natalie Banner	'Understanding Patient Data' Lead, Wellcome Trust
Professor Christopher Bishop	Technical Fellow & Lab Director, Microsoft
Mr Guy Boersma	Managing Director, Kent Surrey Sussex AHSN
Mr Phil Booth	Coordinator, medConfidential
Mr David Cole	Head of Innovation and Business Development, EMEA, IBM Watson Health
Dr Aldo Faisal	Director, UKRI Centre in AI for Healthcare, Imperial College London
Dr Matthew Fenech MD PhD	Medical Safety Lead, Ada Health
Dr Guy Gross	Head of Research / Consultant, LaingBuisson
Dr Mark Halling-Brown	Head of Scientific Computing, Royal Surrey County Hospital
Mr Matthew Honeyman	Researcher, Policy, The King's Fund
Mrs Mirella Marlow	Programme Director, National Institute for Health and Care Excellence (NICE)
Dr Tahreema Matin	Clinical Advisor for National Education Reform & Consultant Academic Radiologist, Health Education England
Ms Niamh McKenna	Managing Director, Accenture UK Limited
Dr Pritesh Mistry	Head of Innovation, Royal College of GPs
Professor Hugh Montgomery	Professor of Intensive Care Medicine; Director, UCL Centre for Health and Performance, University College London (Consult for Deepmind Health)
Dr Ben Panter	CEO, Blackford Analysis
Professor Reza Razavi	Vice President & Vice-Principal of Research, King's College, London Director of Research, King's Health Partners Professor of Paediatric Cardiovascular Science, King's College, London Consultant Cardiologist, Guy's and St Thomas' NHS Foundation Trust
Professor Geraint Rees	Pro-Vice-Provost (AI), UCL
Mrs Tricia Roberts	Director, KMR Healthcare Consulting Ltd

Mr Mark Salmon	Deputy Director Evidence Services, Programme Director Information Resources, NICE
Ms Sue Saville **(Chair)**	Moderator, Sue Saville Media
Dr Katharine Scarfe-Beckett	Rapporteur
Mr Ross Scrivener	eHealth Lead, Royal College of Nursing