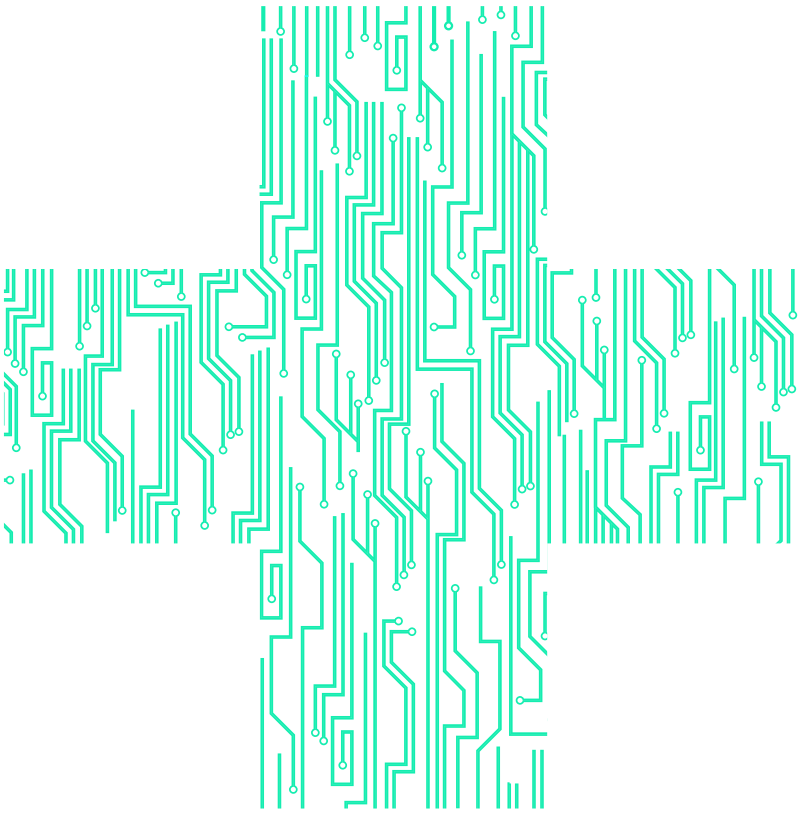
FUTURE ADVOCACY



ETHICAL, SOCIAL, AND POLITICAL CHALLENGES OF ARTIFICIAL INTELLIGENCE IN HEALTH

A report with

wellcome

Future Advocacy is a think tank and consultancy working on some of the greatest challenges that humanity faces in the 21st century. We advocate for smart, forward-thinking policies that will allow us to capitalise on the opportunities and mitigate the risks presented by artificial intelligence.

www.futureadvocacy.org
@FutureAdvocacy

Written and researched by Matthew Fenech, Nika Strukelj, Olly Buston for the Wellcome Trust April 2018

CONTENTS

ACKNOWLEDGEMENTS

We are very grateful to all the interviewees and participants in our roundtables held in February and March 2018, who very kindly gave their time to share their insights, expertise and experiences with us. A full list of contributors is available in the Appendices.

Particular thanks go to Eleonora Harwich, Head of Digital and Technological Innovation at Reform, who made excellent suggestions for potential interviewees for this project.

We are grateful to Jillian Hastings-Ward of the 100,000 Genomes Project, Mariana Campos of Genetic Alliance UK, Adam Cross of the Royal College of Physicians' Patient and Carer Network, and Sinduja Manohar of the British Heart Foundation's Patient Data Panel, who helped with recruiting patients and members of the public to participate in this project.

Finally, we are very grateful to Dr Stephen Cave and the team at the Leverhulme Centre for the Future of Intelligence for their review of a draft of this report and their helpful comments and suggestions.

EXECUTIVE SUMMARY

Artificial intelligence (AI) is everywhere these days. By taking an inclusive definition of intelligence as 'problem-solving', we can consider 'an artificially intelligent system' to be one which takes the best possible action in a given situation. Such AI systems already filter our spam, decide what we see on social media, provide legal advice, and may even determine whether we're paid a visit by the police.

As AI systems become better at sorting data, finding patterns, and making predictions, these technologies will take on an expanded role in health and care, from research, to medical diagnostics, and even in treatment. This increasing use of AI in health is forcing nurses, doctors and researchers to ask: "How do long-standing principles of medical ethics apply in this new world of technological innovation?" In order to address this question, we have undertaken a detailed review of existing literature, as well as interviewing more than 70 experts all round the world, to understand **how AI is being used** in healthcare, **how it could be used in the near future**, and **what ethical, social, and political challenges these current and prospective uses present**. We have also sought the views of patients, their representatives, and members of the public.

We have categorised the current and potential use cases of AI in healthcare into 5 key areas:

- *Process optimisation* e.g procurement, logistics, and staff scheduling
- *Preclinical research* e.g drug discovery and genomic science
- *Clinical pathways* e.g. diagnostics and prognostication
- *Patient-facing applications* e.g delivery of therapies or the provision of information
- *Population-level applications* e.g. identifying epidemics and understanding non-communicable chronic diseases

Across these use cases, a number of ethical, social, and political challenges are raised and the 10 most important are:

01 What effect will AI have on human relationships in health and care?

02 How is the use, storage and sharing of medical data impacted by AI?

03 What are the implications of issues around algorithmic transparency/explainability on health?

04 Will these technologies help eradicate or exacerbate existing health inequalities?

05 What is the difference between an algorithmic decision and a human decision?

06 What do patients and members of the public want from AI and related technologies?

07 How should these technologies be regulated?

08 Just because these technologies could enable access to new information, should we always use it?

09 What makes algorithms, and the entities that create them, trustworthy?

10 What are the implications of collaboration between public and private sector organisations in the development of these tools?

In this report, we explore these and the other challenges raised by our research and make recommendations for further study in this complex and sensitive field. We also find that there are overarching ethical themes, namely *consent*, *fairness* and *rights*, that cut across the challenges we identify. We ask how users can give meaningful consent to an AI where there may be an element of autonomy in the algorithm's decisions, or where we do not fully understand these decisions. Ensuring fairness through preventing and eliminating health inequality, and providing value to stakeholders is another critical issue. Finally, the right to health may well be expanded to encompass questions such as "do people have a right to know how much AI is used in their care?" and "do people have a right not to have AI involved in their care at all?"

We recommend a multidisciplinary approach to dealing with these issues. This refers not only to galvanising a broad range of experts, many of whom will use and be impacted by these tools, but also to the active participation of patients, their relatives, and the public in their development. It is equally important that these technologies are developed with a view to sharing their benefits as widely as possible. This is the best way to ensure that real-world challenges are addressed, that the needs of patients beyond their clinical care are considered, and that these technologies are accepted by patients and practitioners alike.

SUMMARY OF ETHICAL, SOCIAL, AND POLITICAL CHALLENGES AND SUGGESTIONS FOR FURTHER RESEARCH

01 What effect will AI have on human relationships in health and care?

- What effect will these technologies have on relationships between patients and healthcare practitioners?
- What effect will these technologies have on relationships between different healthcare practitioners?
- What do healthcare practitioners think about the potential for these technologies to change their jobs, or to lead to job displacement?
- How do these tools fit into the trend of enabling patients to have greater knowledge and understanding of their own conditions? How different are they from looking up one's symptoms on a search engine before going to see a healthcare practitioner?
- Given that AI is trained primarily on 'measurable' data, does reliance on AI risk missing non-quantifiable information that is so important in healthcare interactions?
- If AI systems become more autonomous, how should transitions between AI and human control be incorporated into care pathways?

02 How is the use, storage, and sharing of medical data impacted by AI?

- How is medical data different from other forms of personal data?
- What is the most ethical way to collect and use large volumes of data to train AI, if the consent model is impractical or insufficient?
- How do we check datasets for bias or incompleteness, and how do we tackle these where we find them?
- Should patients who provide data that is used to train healthcare algorithms be the primary beneficiaries of these technologies, or is it sufficient to ensure that they are not exploited?

03 What are the implications of issues around algorithmic transparency and explainability on health?

- Are expert systems or rule-based AI systems more suitable for healthcare applications than less interpretable machine learning methods?
- What do patients and healthcare practitioners want from algorithmic transparency and explainability?

- Are improved patient outcomes, efficiency and accuracy sufficient to justify the use of 'black box' algorithms? If such an algorithm outperforms a human operator at a particular healthcare-related task, is there an ethical obligation to use it?

- Could 'explanatory systems' running alongside the algorithm be sufficient to address 'black box' issues?

04 Will these technologies help eradicate or exacerbate existing health inequalities?

- Which populations may be excluded from these technologies, and how can these populations be included?

- Will these technologies primarily affect inequalities of access, or of outcomes?

05 What is the difference between an algorithmic decision and a human decision?

- How do we rank the importance of a human decision as compared to an algorithmic decision, particularly when they are in conflict?

- Do human and algorithmic errors differ simply in degree, or is there an essential, qualitative difference between a machine 'giving the wrong answer' and a human making a mistake?

- How will patients and service users react to algorithmic errors?

- Who will be held responsible for algorithmic errors?

06 What do patients and members of the public want from AI and related technologies?

- How do patients and members of the public think these technologies should be used in health and medical research?

- How comfortable are patients and members of the public with sharing their medical data to develop these technologies?

- How do patients and other members of the public differ in their thinking on these issues?

- What is the best way to speak to patients and members of the public about these technologies?

07 How should these technologies be regulated?

- Are current regulatory frameworks fit for purpose?

- What does 'duty of care' mean when applied to those who are developing algorithms for use in healthcare and medical research?
- How should existing health regulators interact with AI regulators that may be established?
- How should we regulate online learning, dynamic systems, as opposed to fixed algorithms?

08 Just because these technologies could enable access to new information, should we always use it?

- What would the impact of ever-greater precision in predicting health outcomes be on patients and healthcare practitioners?
- What are the implications of algorithmic profiling in the context of healthcare?

09 What makes algorithms, and the entities that create them, trustworthy?

10 What are the implications of collaboration between public and private sector organisations in the development of these tools?

- What are the most ethical ways to collaborate?
- How do we ensure value for both the public sector and for the private sector organisation, for example in the use of data? In publicly-owned/taxpayer-funded healthcare systems, such as the UK NHS, how do we ensure that citizens receive value too?
- What are the implications of the concentration of intellectual capacity in private sector organisations?

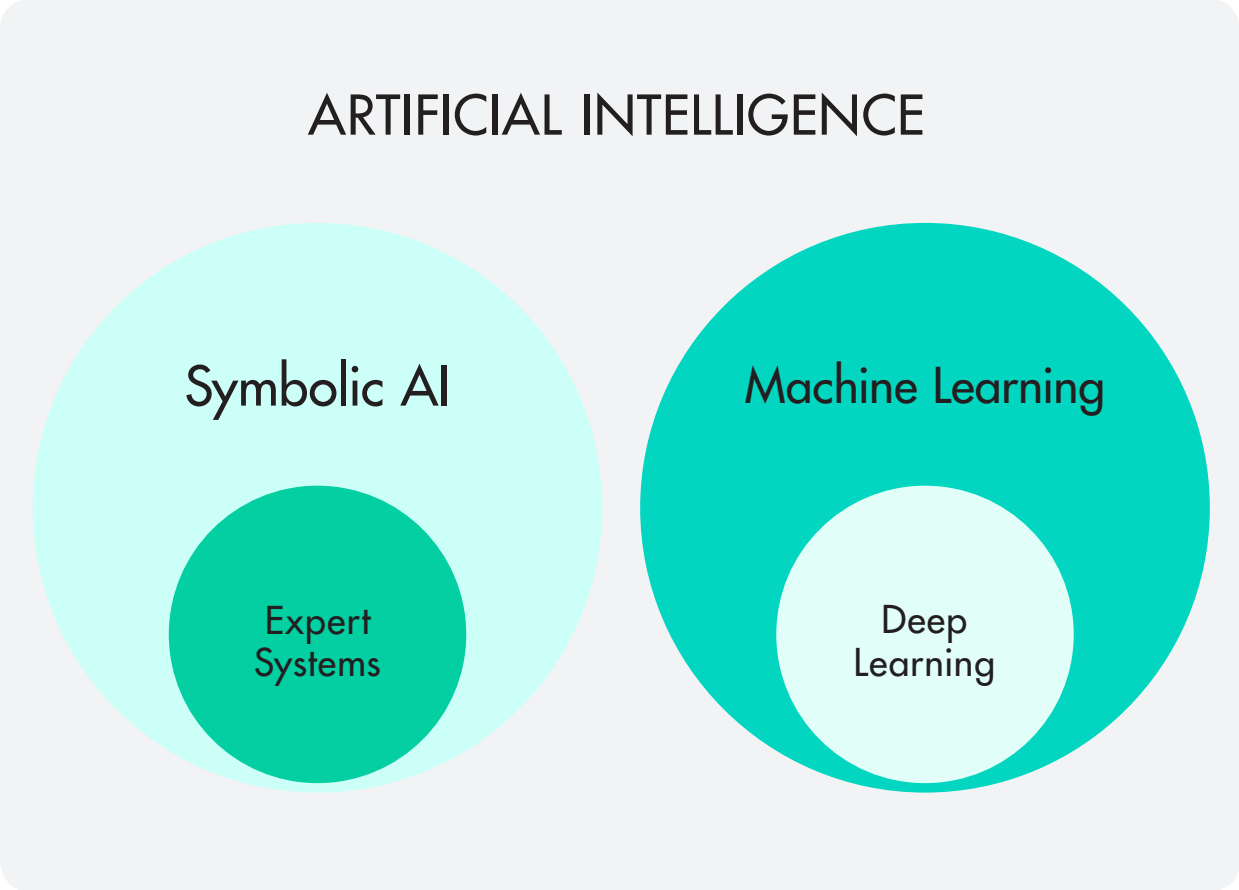
INTRODUCTION

It can feel like artificial intelligence (AI) has come out of the blue and is everywhere all of a sudden, but the concept of creating machines that can perform tasks that require intelligence is not new. Indeed, the term 'artificial intelligence' was coined in 1955 by John McCarthy, then an Assistant Professor of Mathematics at Dartmouth College.[1] The field has since gone through several hype cycles, followed by disappointment and criticism ('AI winters'), followed by funding cuts, followed by renewed interest years or decades later.

The most recent renewal of interest occurred in 2012 and 2013, with the publication of a series of highly influential papers.[2,3,4] Since then, considerable progress has been made in areas such as speech recognition, image recognition, and game playing, coupled with considerable enthusiasm in the mainstream media. These advances are at least in part due to increases in computing power available for use in artificial intelligence development, but also owe a great deal to the huge quantities of data that are being generated in the internet age. It has been estimated that 2.5 quintillion bytes of data are generated daily, and that more than 90% of the data in the world today has been created in the last four years alone.[5] The major developments in AI technologies that are exciting so much interest at the moment could not have been made without big data.

Defining AI is difficult, not least because 'intelligence' itself is so difficult to define. At Future Advocacy, we use an inclusive definition of intelligence as 'problem-solving' and consider 'an intelligent system' to be one which takes the best possible action in a given situation.[6] The phrase '**artificial intelligence**' is an umbrella term comprising a number of techniques (Figure 1). 'Symbolic AI', which is also known as 'good old-fashioned AI' and relies on human-readable representations of problems and logic, was the dominant paradigm of AI research until the 1980s.[7] The majority of the current excitement around AI is focused on machine learning (ML) techniques such as deep learning and neural networks, which rely on complex statistical methods to recognize patterns in data, learn from these patterns, and subsequently make predictions based on these data. The 'learning' aspect of these algorithms raises the prospect of 'dynamic, online learning' systems that optimise their ability to tackle a problem on the fly (see 'Appendix A: Glossary' for full list of definitions used in this report). Other terms that fall under the 'artificial intelligence' umbrella include predictive analytics and data analytics.

1. Nilsson, N. (2010) "The Quest for Artificial Intelligence", Cambridge University Press
2. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). "Imagenet classification with deep convolutional neural networks", University of Toronto, available at https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf
3. Cireşan, D., Meier, U., Masci, J., & Schmidhuber, J. (2012). "Multi-column deep neural network for traffic sign classification", Neural networks, available at https://www.sciencedirect.com/science/article/pii/S0893608012000524
4. Zeiler, M. (2013) "Visualizing and understanding convolutional networks", New York University, available at https://arxiv.org/pdf/1311.2901v3.pdf
5. IBM 'Bringing big data to enterprise', available at https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html
6. Russell, S. J., and Norvig, P., (1995) "Artificial Intelligence: A Modern Approach", Englewood Cliffs, NJ: Prentice Hall.
7. Haugeland, J. (1989) "Artificial Intelligence: The Very Idea", MIT Press; New Ed edition



▲ **Figure 1:** A simple classification of the major classes of artificial intelligence. As outlined in the main text, in this report we focus on 'narrow' or 'weak' artificial intelligence. (Adapted from 'Artificial intelligence: The Road Ahead in Low and Middle-Income Countries', Web Foundation, 2017)

As AI systems become better at sorting data, finding patterns, and making predictions, these algorithms are undertaking an ever-increasing range of tasks, from filtering email spam, to delivering takeaways, to tackling more sophisticated problems such as providing legal advice or deciding whether you are visited by the police.[8,9,10,11,12] It is clear that these technologies will also take on an expanded role in medical diagnostics and treatment. This is because of the reliance of modern medicine on ever-increasing amounts of data derived from imaging, histopathological, biochemical, and other investigations, as well as the fact that many modern management pathways follow strict, semi-algorithmic protocols. The private sector is pouring money into this field - market research firm Frost & Sullivan predicts an annual growth rate in the

8. Mitchell, T. (1997) Machine Learning. London, UK: McGraw-Hill Education.
9. Janakiram, MSV (2017) "In The Era Of Artificial Intelligence, GPUs Are The New CPUs", Forbes, available at https://www.forbes.com/sites/janakirammsv/2017/08/07/in-the-era-of-artificial-intelligence-gpus-are-the-new- cpus/#6f8728b55d16
10. Legal advice being provided by AI algorithms ranged from suggesting strategies for appealing parking tickets to guiding asylum applications. From "Chatbot that overturned 160,000 parking fines now helping refugees claim asylum", (2017) The Guardian, available at https://www.theguardian.com/technology/2017/mar/06/chatbot-donotpay- refugees-claim-asylum-legal-aid
11. The Chicago police department have used predictive policing to visit those at a high risk of committing an offence to offer them opportunities to reduce this risk, such as drug and alcohol rehabilitation or counseling. See Saunders, J., Hunt, P., & Hollywood, J. S. (2016). Predictions put into practice: a quasi-experimental evaluation of Chicago's predictive policing pilot. Journal of Experimental Criminology, 12(3), 347-371 and Stroud, M. (2016, 19 August) "Chicago's predictive policing tool just failed a major test." The Verge (available at https://www.theverge.com/2016/8/19/12552384/chicago-heat-list-tool-failed-rand-test). Areas of the UK, such as Kent, are beginning to use predictive policing. For example, see O'Donoghue, R. (2016) 'Is Kent's Predictive Policing project the future of crime prevention?' KentOnline, available at http://kentonline.co.uk (accessed on 10 March, 2017).
12. Waugh, R., (2017) "Robots are already delivering people's food in London – here's how to summon one", Metro, available at http://metro.co.uk/2017/07/26/robots-are-already-delivering-peoples-food-in-london-heres-how-to-summon-one- 6808269/

global AI market for healthcare of 40% between 2014 and 2021, reflecting an increase from US$ 634m to US$ 6.662bn.[13]

Throughout this report, we refer only to '**narrow**' forms of AI, whose learning is limited to one task or domain of activity only, as opposed to '**broad**', 'general', or 'human-level' AI, which most experts agree is still many decades away.[14] Various philosophers and computer scientists, including Nick Bostrom, Ray Kurzweil and David Chalmers, have written about the potential for an 'intelligence explosion' - that is, the recursive self-improvement that will follow the development of artificial general intelligence (AGI), leading to the exponentially rapid emergence of artificial superintelligence (ASI). Achieving what is referred to as the 'technological singularity' will have unimaginable consequences for all of human civilisation. Although it may seem like the preserve of science fiction, some work has already begun on considering the ethical responsibilities and consequences of work to develop such technology, led by institutions such as the Future of Life Institute in Cambridge, Massachusetts, and the Leverhulme Centre for the Future of Intelligence in Cambridge, UK.

Even without considering the possibility of AGI and ASI, we trust that this report makes a convincing case that there are many challenges that follow from the increased use of artificial narrow intelligence in health and medical research, and that require urgent attention. As healthcare practitioners (HCPs) and biomedical researchers recognise the immense potential for AI technologies, they must revisit long-standing principles of medical ethics and consider how their understanding of these principles will be impacted, as well as reflecting on whether new ethical, social, and political questions are raised. We hope that this report will contribute to this discussion by providing a clear outline of how AI is being used in healthcare and biomedical research today, what AI is likely to be used for in the next few years, and an accessible discussion of the ethical, social, and political issues that these uses raise, and those that remain unaddressed.[15] We also hope this report will inspire research into ways of ensuring that as many people as possible benefit from this application of AI.

13. Frost & Sullivan (2016) "Transforming healthcare through artificial intelligence systems. AI Health and Life Sciences", available at http://ai-healthandlifesciences.com/hypfiles/uploads/2016/08/AI-Healthcare-Research- Insights-KisacoResearch.pdf
14. Marcus, G., (2017) 'Artificial general intelligence is stuck. Here's how to move it forward', New York Times, available at https://www.nytimes.com/2017/07/29/opinion/sunday/artificial-intelligence-is-stuck-heres-how-to-move-it- forward.html
15. We are being deliberately vague with respect to the proposed timeframe for suggested new uses of AI in healthcare to come onstream, but 2-5 years seems a reasonable time period.

CURRENT AND POTENTIAL USES OF ARTIFICIAL INTELLIGENCE IN HEALTH

We have identified five types of use case for artificial intelligence technologies in health and medical research (Table 1).

Use Cases	Examples	
	Current	Future
1. Process optimisation *Using AI to optimise 'back-end' processes in healthcare, such as procurement, logistics, and staff scheduling*	• Rota/staff schedule management, e.g. Hong Kong Health Authority • Emergency services dispatch management e.g. Corti	• Data-driven optimisation of logistics, procurement • Automated analysis/ completion of medical notes and other documentation • Patient experience analysis e.g. Alder Hey
2. Preclinical research *Using AI in preclinical applications such as drug discovery and genomic science*	• Candidate molecule screening, e.g. BenevolentAI, AtomNet • Repurposing drugs, e.g. Teva Pharmaceuticals • Predicting potential side effects, e.g. Cloud Pharmaceuticals • Analysis of large -omics datasets to gain insights.[16]	• Determining targets for gene editing, e.g. CRISPR

16. The '-omics' fields are those fields of study that end in -omics, such as genomics, proteomics, lipidomics, and metabolomics. They are concerned with the collective characterisation and quantification of whole pools of biological molecules, with a view to understanding biological structure and function.

3. Clinical pathways *Fitting AI into existing and new clinical workflows, such as in diagnostics and prognostication*	• Analysis of optical coherence tomography (OCT) images, e.g. DeepMind-Moorfields collaboration • Analysis of radiological imaging, e.g. Viz.ai • Analysis of clinical conversations e.g. Corti • Prognostication e.g. prediction of all-cause mortality [Stanford, KenSci], prediction of cardiovascular risk [University of Nottingham]	• Radiologists' assistants, e.g. suggesting best imaging modality in particular clinical situation, improved image acquisition processes leading to radiation dose reduction • Management decision-support for healthcare practitioners, suggesting best treatment for particular patient • Automated transcription of clinical interactions • Automated completion and submission of investigation requests/referrals
4. Patient-facing applications *Using AI to interact directly with patients and other service users, including in the delivery of therapies or the provision of information*	• Chatbots, e.g. Oli [Alder Hey], AVA [Arthritis Research UK], Lark Weight Loss Coach • Autonomous (closed-loop) insulin pumps • Personalised health advice and interventions, e.g. CareSkore, Viome, DayTwo	• Smart homes and wearables • Robot carers • Robot surgeons
5. Population-level applications *Using AI to gain insights into population health, such as identifying epidemics and monitoring disease spread.*	• Prediction of infectious disease outbreaks, e.g. Dengue fever app in Malaysia • Better targeting of public health spending and other interventions, e.g. University of Southern California tool • Better understanding of risk-factors for non-communicable disease, e.g. childhood obesity [Indiana University tool]	

▲ **Table 1:** An outline of the five main uses cases for AI in health and medical research, with examples of current and potential future applications in each category.

Process optimisation

Effective delivery of healthcare services relies on the strategic deployment of resources, both physical and human. Even as spending on healthcare continues to outpace broader economic growth, it remains difficult to completely meet a community's or a country's health needs, particularly as people are living longer worldwide, with an attendant increase in complex chronic conditions.[17,18,19] Although many of our interviewees indicated that optimising use of limited resources could be a major use of AI technologies in healthcare systems worldwide, examples of this use case category currently in practice are few and far between.

One area where AI is being applied in healthcare systems is rostering. The Hong Kong Health Authority, for example, is using an AI-based tool developed at the local City University of Hong Kong to produce monthly or weekly staff rosters that satisfy a set of constraints, such as staff availability, staff preferences, allowed working hours, ward operational requirements and hospital regulations.[20] This tool has been deployed across 40 public hospitals, and is responsible for the rostering of over 40,000 staff. Since being introduced, the Hospital Authority reports increased productivity, improved staff morale, and improved quality of service, as the system is seen to be fair, frees up managers' time, and can provide management with insights into working patterns and resource utilisation.[21] This feedback contrasts with reports about the use of scheduling software in other sectors. Companies in industries such as hospitality and retail respond to real-time analysis of factors such as sales and weather and modify their staffing accordingly. However, this increased efficiency when it comes to the use of staff resources could lead to significant disruption to the lives of low-wage employees, who may receive as little as a day's notice of their changed timetable.[22] More broadly, the trend towards increased use of 'people analytics' - the comprehensive collection of data about employees' behaviour, which is then used to inform managerial decisions - has been criticised as having a dehumanising effect on work, and may not even be effective in increasing productivity or optimising working practices.[23]

Another 'back-end' application of AI in a healthcare system is found in Denmark. In 2016, the Copenhagen-based start-up Corti partnered with the city's emergency medical service (EMS) to provide an AI assistant to augment human operators receiving emergency calls on the 112 emergency number. Besides helping with triage (see 'Clinical Pathways' section, below), Corti's technology is also being used to oversee and optimise the whole dispatch process, for example by identifying and alerting human operators to errors in the address any emergency response is sent to.[24]

An innovative approach to using AI will be to apply it to supporting quality and service improvement. This is currently in the final rounds of a research grant application for investigation by the team at Alder Hey Children's NHS

17. Drouin, J. P., Hediger, V. and Henke, N. (2008) "Health care costs: A market-based view", The McKinsey Quarterly, available at https://www.mckinsey.com/~/media/mckinsey/dotcom/client_service/healthcare%20systems%20and%20services/pdf s/healthcare-costs-a-market-based-view.ashx
18. "Emerging trends in healthcare", PwC (2017), available at https://www.pwc.com/gx/en/industries/healthcare/emerging-trends-pwc-healthcare.html
19. "Healthcare Challenges and Trends", CGI (2014) available at https://www.cgi.com/sites/default/files/white- papers/cgi-health-challenges-white-paper.pdf
20. Chun, H.W., Chan H.C., Lam P.S., Tsang M.F., Wong J. and Yeung W.M., (2000) "Nurse Rostering at the Hospital Authority of Hong Kong," Proceedings of the Twelfth Conference on Innovative Applications of Artificial Intelligence, Austin, available at https://www.cs.cityu.edu.hk/~hwchun/research/PDF/IAAI2000HA.pdf
21. http://www.cs.cityu.edu.hk/~hwchun/AIProjects/stories/hrrostering/harostering/
22. Kantor, J. "Working anything but 9 to 5", The New York Times, available at https://www.nytimes.com/interactive/2014/08/13/us/starbucks-workers-scheduling-hours.html
23. Gal, U. (2017) "Why algorithms won't necessarily lead to utopian workplaces", The Conversation, available at https://theconversation.com/why-algorithms-wont-necessarily-lead-to-utopian-workplaces-73132Interview with Dr Iain Hennessey and Carol Platt, 25th January 2018
24. Peters, A. (2018) "Having a heart attack? This AI helps emergency dispatchers find out", Fast Company, available at https://www.fastcompany.com/40515740/having-a-heart-attack-this-ai-helps-emergency-dispatchers-find-out

Foundation Trust. One application of the methodology involves medical students collecting data during a surgical day case visit where they observe and record patient and family reactions as they progress through the clinical pathway, including anxiety levels and overall experience. This data is currently manually evaluated in order to identify aspects of the patient and family experience that require improvement or that are working well, in order to provide feedback to staff. Digitisation of this process and application of AI technologies such as sentiment analysis will greatly enhance the improvement cycle.[25]

A clear potential use of AI in optimising the delivery of healthcare is in logistics. Just as large retailers such as Amazon and Zara are using predictive analytics to anticipate demand for particular products, so too could healthcare systems use similar tools to help with procurement, logistics, and distribution.[26,27]

AI-based technologies could also help healthcare practitioners (HCPs) with administrative tasks, which constitute a major part of their day-to-day work, freeing them up to redirect their energy and expertise directly to the patient. For example, nurses in the UK's National Health Service (NHS) report spending an estimated 2.5 million hours a week on clerical tasks, according to a poll carried out in 2013 for the Royal College of Nursing, equating to more than an hour per day for every nurse, which could otherwise be spent on direct patient contact.[28] One potential use of natural language processing (NLP) technologies may be to analyse medical notes, digesting the contents for speedy access by HCPs.[29] A major hurdle for wider use of NLP to parse medical notes and automatedly complete forms, such as investigation requests and referrals to other HCPs, is the difficulty experienced worldwide with the deployment and use of electronic health record (EHR) systems. For example, McKinsey reported that 40% of US doctors are not currently using them.[30] The unequal adoption of EHR across the NHS has been extensively described, with universal use in primary care, and only patchy use in secondary/hospital-based care.[31] Obstacles to adopting EHRs include lack of resources, institutional/practitioner inertia, regulatory constraints, and the unstructured nature of medical data.

Even where EHR systems have been adopted, those in different clinics or hospitals do not necessarily interface well with each other, meaning that the communication of healthcare information from one EHR system directly to another is impossible. As a result, various healthcare systems worldwide continue to use outdated communication technology such as the fax machine. Healthtech company athenahealth is using AI to import data from a fax into a patient record about twice as fast as a human does, thanks to which more than 3 million hours of work were saved in 2017.[32]

25. Interview with Dr Iain Hennessey and Carol Platt, 25th January 2018
26. LaRiviere, J., McAfee, P., Rao, J., Narayanan, V. K., and Sun, W. (2016) "Where Predicitve Analytics is having the biggest impact", Harvard Business Review, available at https://hbr.org/2016/05/where-predictive-analytics-is-having-the-biggest-impact
27. Wills, J. (2016) "7 ways Amazon uses big data to stalk you", Investopedia, available at https://www.investopedia.com/articles/insights/090716/7-ways-amazon-uses-big-data-stalk-you-amzn.asp
28. Owen, J. (2013) "Patients denied care as nurses fill in forms", the Independent, available at http://www.independent.co.uk/life-style/health-and-families/health-news/patients-denied-care-as-nurses-fill-in-forms-8581573.html
29. Bughin, J. et al. (2017) "Artificial Intelligence: The next digital frontier?", McKinsey Global Institute, available at https://www.mckinsey.com/~/media/McKinsey/Industries/Advanced%20Electronics/Our%20Insights/How%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/MGI-Artificial-Intelligence-Discussion-paper.ashx
30. Manyika, J. et al. (2015) "Digital America: A tale of the haves and the have mores", McKinsey Global Institute, available at https://www.mckinsey.com/~/media/McKinsey/Industries/High%20Tech/Our%20Insights/Digital%20America%20A%20tale%20of%20the%20haves%20and%20have%20mores/Digital%20America%20Full%20Report%20December%202015.ashx
31. Harwich, E., Laycock, K. (2018) "Thinking on its own: AI in the NHS", Reform, available at http://www.reform.uk/wp-content/uploads/2018/01/AI-in-Healthcare-report_.pdf
32. Bush, J. (2018) "How AI is taking the scut work out of healthcare", Harvard Business Review, available at "https://hbr.org/2018/03/how-ai-is-taking-the-scut-work-out-of-health-care?utm_campaign=hbr&utm_source=twitter&utm_medium=social"

Preclinical research

Drug development is a costly and time-consuming process. It can take up to 15 years for a new drug to reach the market, following testing tens of thousands of compounds and three phases of clinical trials.[33] AI is being used to provide insights that may be able to reduce the time it takes to develop a drug and get it to market. For example, BenevolentAI uses neural networks to make predictions about potential drug candidates, which are then used to decide which candidates to take forward. This approach has been used to identify promising new therapies for motor neurone disease - of 5 new candidate molecules identified, one has shown excellent results in vitro. The next stage is assessing viability for a clinical trial.[34]

Similarly, AtomNet uses deep learning on 3D models of molecules to predict the likelihood of two molecules interacting, and can screen 1 million compounds per day. Its prediction of how well a compound can be used to treat an illness is then used by researchers to narrow down the options. Potential drugs for Ebola and multiple sclerosis have been identified by AtomNet using this screening process; these are now in the development pipeline and one has already been licensed to a UK pharmaceutical company.[35]

Besides difficulties in identifying potentially-useful molecules to take forward through the drug development process, predicting potential side effects is a major issue for the pharmaceutical industry. Cloud Pharmaceuticals are using a predictive tool to identify molecules able to cross the blood-brain barrier based on their chemical properties (which might therefore be expected to have neurological side effects), with an accuracy rate of 80%.[36] AI tools can also help with predicting other effects, whether physiological or metabolic.

Machine learning has also been put to use for repurposing drugs, which is a much cheaper alternative to starting from scratch. For example, IBM and Teva Pharmaceuticals are working together to identify potential new uses for established drugs.[37, 38]

Other fields of biomedical research, including the '-omics' fields (e.g. genomics, proteomics, lipidomics, and metabolomics) are increasingly characterised by huge, complex datasets. Sophisticated tools relying on AI are being used to analyse these datasets more quickly than human analysis would allow. For example, Deep Genomics is using AI to try to predict the consequences of a specific mutation when this is unknown. Thus far, they have a database suggesting potential clinical consequences for more than 300 million genetic variations.[39,40]

Another potential application for machine learning is gene editing, a complex process where specific alterations are made to DNA at the cellular level. Developments in CRISPR-Cas9 technology have raised hopes that gene editing could one day be used to treat human genetic disease.[41,42] Machine learning could be used to assist with the process of target identification, a

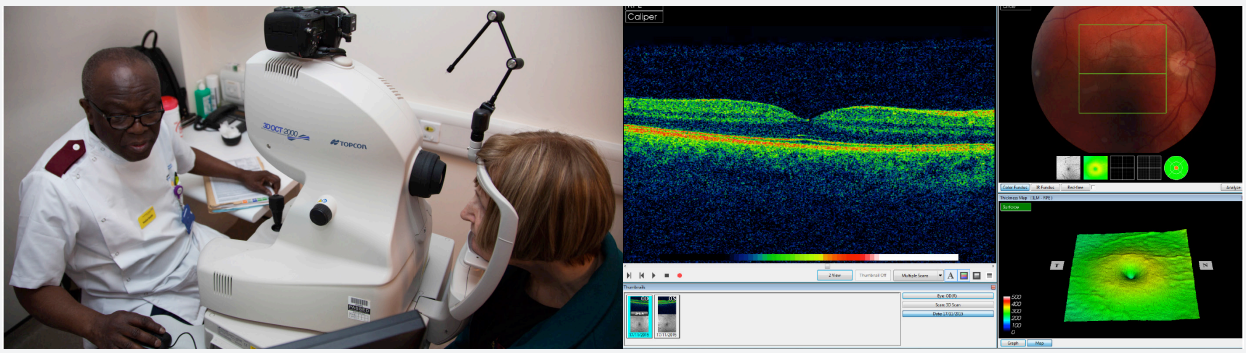
33. Bates Ramirez, V. (2017) "Drug discovery AI can do in a day what currently takes months", SingularityHub, available at https://singularityhub.com/2017/05/07/drug-discovery-ai-can-do-in-a-day-what-currently-takes-months/#sm.00001fq62zzxo7evwz32qyl5fzmdw
34. Hirschler, B. (2017) "AI hunts for new ALS treatments", Scientific American, available at https://www.scientificamerican.com/article/ai-hunts-for-new-als-treatments/
35. Dormehl, L. (2017) "Artificial intelligence can invent new drugs far faster than any human could", Digital Trends, available at https://www.digitaltrends.com/cool-tech/artificial-intelligence-inventing-drugs/
36. Stone, P. et al. (2016) "Artificial intelligence and life in 2030", Stanford University, available at https://ai100.stanford.edu/sites/default/files/ai_100_report_0831fnl.pdf
37. ibid
38. "Teva Pharmaceuticals and IBM Expand Global Partnership to Enable Drug Development and Chronic Disease Management with Watson", IBM (2016), available at https://www-03.ibm.com/press/uk/en/pressrelease/50969.wss
39. Loria, K. (2015) "This man thinks he can crack one of the greatest mysteries about what makes us sick", Business Insider, available at http://uk.businessinsider.com/how-deep-genomics-is-using-ai-to-solve-genetic-mysteries-2015-9?r=US&IR=T
40. Xiong, H. et al. (2015) "The human splicing code reveals new insights into the genetic determinants of disease", Science, available at http://science.sciencemag.org/content/347/6218/1254806
41. Kang, X. (2017) "Addressing challenges in the clinical applications associated with CRISPR/Cas9 technology and ethical questions to prevent its misuse", Springer Link, available at https://link.springer.com/article/10.1007/s13238-017-0477-4
42. (2018) "What are genome editing and CRISPR-Cas9?", Genetics Home Reference, available at https://ghr.nlm.nih.gov/primer/genomicresearch/genomeediting

tricky and unpredictable part of the CRISPR-Cas9 process. London-based Desktop Genetics is a software company that is doing this, and has reported improved results in the ability of an algorithm to predict CRISPR activity following an ML-based process of algorithmic training.[43]

Clinical pathways

Given the advances in machine vision technology, it is perhaps unsurprising that most progress has been made in developing AI to perform image-related tasks in clinical pathways. These include the analysis of radiological and histopathological images. The increased reliance on imaging and other types of diagnostic investigations in healthcare systems around the world is leading to a situation where healthcare practitioners are faced with an ever-increasing volume of imaging data to interpret and act on. Robust systems to highlight 'high-risk' cases, for example images that are likely to show a particular diagnosis that requires urgent action, are in high demand. Rather than replacing human operators currently undertaking these image analysis tasks, most applications being developed are envisaged as assisting healthcare practitioners - what many contributors to our report described as 'augmented intelligence'. An example is the smartphone app developed by Viz.ai, that uses machine vision to detect signs of a stroke in brain scans and alert specialists via their phones when these are seen. The pathway of alerting specialists in this way has recently been approved for clinical use by the United States Food and Drug Administration (FDA).[44]

Another example of this is provided by the ongoing work at Moorfields Eye Hospital, London. Pearse Keane, Consultant Ophthalmologist at Moorfields Eye Hospital and NIHR Clinician Scientist at University College London, spoke to us about the increasing use of optical coherence tomography (OCT), a safe and relatively straightforward way of imaging the retina to pick up retinal diseases (Figure 2). More than 1000 OCT scans are performed every day at Moorfields Eye Hospital. Furthermore, an estimated 5-10% of high street optometrists are beginning to offer this service, increasing the potential number of people who are scanned using this technique. In 2016, there were 7,000 referrals to the Outpatients Clinic that were tagged as 'urgent'. Of these, 800 patients turned out to have wet age-related macular degeneration (AMD), a serious sight-threatening condition that requires urgent treatment. Because



▲ **Figure 2:** Optical coherence tomography (OCT) is an imaging technique that is increasingly used when investigating eye complaints. On the left is a patient having an OCT scan performed, while on the right is an OCT scan. This imaging technique can reveal problems with the retina, the layer of light-sensitive cells at the back of the eye. (Images used with permission of Moorfields Eye Hospital NHS Foundation Trust)

43. Sennaar, K. (2018) "Machine Learning in genomics – current efforts and future applications", Tech emergence, available at https://www.techemergence.com/machine-learning-in-genomics-applications/
44. Simonite, T. (2018) "Using AI to Help Stroke Victims When 'Time Is Brain'", Wired, available at https://www.wired.com/story/using-ai-to-help-stroke-victims-when-time-is-brain

of the sheer volume of referrals, there is a risk that the diagnosis of wet AMD in some of those affected may be delayed. Keane therefore approached DeepMind Health in 2015, and in collaboration with them has been developing a tool that uses deep learning to analyse OCT images and flag up those that are more likely to show significant pathology, acting as a useful triage system. Initial results from a 'real-world' clinical study of this tool have been submitted to a major peer-reviewed journal, and are due for publication soon.[45] There are many other examples of tools being developed to undertake clinical imaging tasks. For example:

- A collaboration between IBM and the University of Alberta has developed an AI tool that reviews functional magnetic resonance imaging (fMRI) scans to diagnose schizophrenia, with an initial accuracy rate of 74% when tested on a 95-member dataset.[46]

- At the last Association for Computing Machinery conference in 2017, a team from Louisiana State University presented a deep neural network that had been trained to identify suspicious regions on mammography images, and then classify them as cancerous or benign, in one step. The authors claim a detection accuracy of up to 90% and a classification accuracy of 93.5%.[47]

- A team based in Oxford has developed Ultromics, an AI system that analyses echocardiograms to diagnose coronary heart disease.[48] The results of an initial study have not been published in a peer-reviewed journal yet, but the Ultromics team claim that their algorithm outperforms expert cardiologists in this task.[49]

- Multiple systems have been developed to review photographs of skin moles to diagnose malignant melanoma. In one Nature-published study, a team led by renowned computer scientist Sebastian Thrun developed a tool that correctly categorised moles as benign or malignant as accurately as a panel of 21 board-certified dermatologists.[50,51]

Besides image analysis, Dr Hugh Harvey, Clinical Lead at Kheiron Medical, a deep learning startup, and a member of the Royal College of Radiologists Informatics Committee, pointed out that AI tools could be developed to assist radiologists and other clinicians to choose the ideal imaging modality to use in a particular clinical scenario. This would reduce unnecessary imaging and potentially speed up diagnostic and treatment pathways.[52]

AI tools have been developed that can review other clinical data sources to assist diagnostic pathways. The Corti algorithm used in Copenhagen (see 'Process optimisation' section, above) is being trained on audio data from emergency calls. By analysing the caller's speech patterns as well as background noise, it is hoped that the algorithm will flag up cases that are more likely to go into cardiac arrest and thus require a more urgent dispatch.[53]

Another data source is the medical literature. It is estimated that it would take at least 160 hours of reading a week just to keep up with the publication of new medical knowledge.[54] IBM's Watson provides an example of the potential of AI to parse the medical literature and generate useful insights. In 2016, doctors at the University of Tokyo's Institute of Medical Science were baffled by a particular patient's presenting symptoms. When these were inputted into Watson, it took just 10 minutes to come up with the diagnosis of a rare secondary leukemia.[55]

45. Interview with Mr Pearse Keane, 7th February 2018
46. Tarantola, A. (2017) "IBM's AI can predict schizophrenia by looking at the brain's blood flow", engadget, available at https://www.engadget.com/2017/07/20/ibms-ai-can-predict-schizophrenia-by-looking-at-the-brains-blo/
47. Platania, R., Shams, S., Yang, S., Zhang, J., Lee, K., and Park, S. (2017). "Automated Breast Cancer Diagnosis Using Deep Learning and Region of Interest Detection (BC-DROID)", Proceedingso f ACM-BCB '17, Boston, MA, USA. DOI: http://dx.doi.org/10.1145/3107411.3107484
48. http://www.ultromics.com/technology/
49. Ghosh, p. (2018) "AI early diagnosis could save heart and cancer patients", BBC, available at http://www.bbc.co.uk/news/health-42357257
50. Esteva, A. et al. (2017) "Dermatologist-level classification of skin cancer with deep neural networks", Nature, available et https://www.nature.com/articles/nature21056
51. Waltz, E. (2017) "Computer diagnoses skin cancers", IEEE Spectrum, available at https://spectrum.ieee.org/the-human-os/biomedical/diagnostics/computer-diagnoses-skin-cancers
52. Interview with Dr Hugh Harvey, 17th January 2018
53. Peters, A. (2018) "Having a heart attack? This AI helps emergency dispatchers find out", Fast Company, available at https://www.fastcompany.com/40515740/having-a-heart-attack-this-ai-helps-emergency-dispatchers-find-out
54. Steadman, I. (2013) "IBM's Watson is better at diagnosing cancer than human doctors", Wired, available at http://www.wired.co.uk/article/ibm-watson-medical-doctor
55. Rohaidi, N. (2016) "IBM's Watson detected rare leukemia in just 10 minutes", Asian Scientist, available at https://www.asianscientist.com/2016/08/topnews/ibm-watson-rare-leukemia-university-tokyo-artificial-intelligence/

It is important to point out that this was a one-off use of Watson's capability, and that attempts to apply Watson in cancer diagnostics at scale have had mixed results.[56,57,58]

Patient medical records can also provide a rich source of data for AI systems to mine for insights. Researchers at the Massachusetts Institute of Technology (MIT) and Harvard Medical School trained an algorithm using patient data including demographics, family history, and pathology reports from breast biopsies, in order to develop a decision support tool that would make recommendations as to whether surgery should be performed on so-called 'high-risk lesions' identified on breast biopsies. When tested on 335 high-risk lesions, the model correctly diagnosed 97 percent of the breast cancers as malignant and reduced the number of benign surgeries by more than 30 percent compared to existing approaches.[59] Similarly, researchers at the University of Pennsylvania trained an algorithm on over 160,000 records of discharged patients over three years, after which it could monitor hundreds of variables and detect severe sepsis a full 12 hours ahead of onset, early enough to avoid deterioration.[60]

There has been a lot of attention devoted to training AI algorithms not just to diagnose, but to prognosticate based on a particular patient's data and parameters at presentation. Examples of this use case include a system developed at Stanford, that uses deep learning to predict all-cause mortality over the next 3 to 12 months, and one being developed by KenSci, which aims to predict 6- to 12-month mortality risk.[61,62] The Stanford researchers are now attempting to use this system to see if it can assist palliative care teams by diverting their limited resources to the patients that need them most.[63] Similarly, a machine learning tool developed at the University of Nottingham has demonstrated better accuracy than traditional guidelines at determining cardiovascular disease risk, in part because it can consider more data points, including ethnicity, arthritis, and kidney disease. Consequently, more patients who could benefit from preventive treatment are identified, while others avoid unnecessary intervention.[64,65] Other prognostic tools using AI that have been developed include one that predicts mortality following major cardiac surgery, and mortality in patients newly-diagnosed with pulmonary hypertension, based on cardiac MRI data.[66,67]

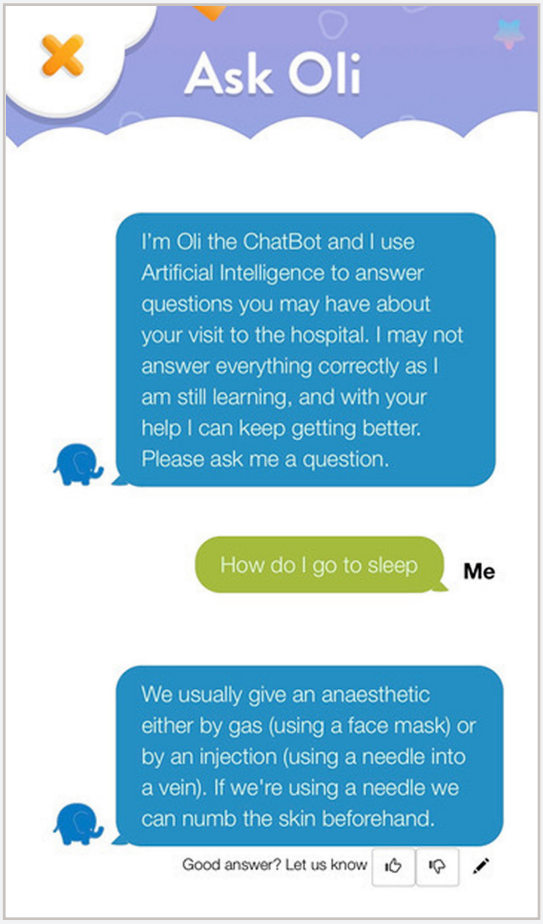
It is important to note that, in spite of the mooted superhuman or par-human accuracy figures for these diagnostic and prognostic tools, the vast majority have not been extensively tested in 'real-world' clinical settings to see how they would materially augment an HCP's practice, and ultimately improve patient outcomes (the forthcoming Moorfields-DeepMind publication is a notable exception). Many of our clinical interviewees stressed this point to us, and reminded us that the ultimate measure of these tools must be clinical efficacy.[68]

56. Herper, M. (2017) "MD Anderson benches IBM Watson in setback for Artificial Intelligence in Medicine", Forbes, available at https://www.forbes.com/sites/matthewherper/2017/02/19/md-anderson-benches-ibm-watson-in-setback-for-artificial-intelligence-in-medicine/#493978a73774
57. Mulcahy, N. (2017) 'Big Data Bust: MD Anderson-Watson Project Dies', Medscape, available at https://www.medscape.com/viewarticle/876070
58. Freedman, D. H. (2017) "A Reality Check for IBM's AI Ambitions", MIT Technology Review, available at https://www.technologyreview.com/s/607965/a-reality-check-for-ibms-ai-ambitions/
59. Conner-Simons, A. (2017) "Using artificial intelligence to improve early breast cancer detection", MIT News, available at http://news.mit.edu/2017/artificial-intelligence-early-breast-cancer-detection-1017
60. Gaskell, A. (2017) "Using AI to prevent sepsis", Huffington Post, available at https://www.huffingtonpost.com/entry/using-ai-to-prevent-sepsis_us_5927f8f6e4b0d2a92f2f42e3
61. Dignan, L. (2018) "Death and data science: How machine learning can improve end-of-life care", ZDNet, available at http://www.zdnet.com/article/death-and-data-science-how-machine-learning-can-impact-hospice-referrals-improve-last-days-of-life/
62. Avati, A. et al. (2017) "Improving palliative care with Deep Learning", Cornell University Library, available at https://arxiv.org/pdf/1711.06402.pdf
63. ibid
64. Weng, S. et al. (2017) "Can machine-learning improve cardiovascular risk prediction using routine clinical data?", PLoS ONE, available at http://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0174944&type=printable
65. Hutson, M. (2017) "Self-taught artificial intelligence beats doctors at predicting heart attacks", Science, available at http://www.sciencemag.org/news/2017/04/self-taught-artificial-intelligence-beats-doctors-predicting-heart-attacks
66. Allyn, J. et al. (2017) "A comparison of a machine learning model with EuroSCORE II in predicting mortality after elective cardiac surgery: a decision curve analysis", PLoS ONE, available at http://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0169772&type=printable
67. Dawes, T. et al. (2016) "Machine Learning of three-dimensional right ventricular motion enables outcome prediction in pulmonary hypertension: a cardiac MR imaging study", Radiology, available at http://pubs.rsna.org/doi/pdf/10.1148/radiol.2016161315
68. A review of ClinicalTrials.gov (the United States National Institutes of Health and National Library of Medicine's database of publicly- and privately-funded clinical studies worldwide) in April 2018 revealed 21 active clinical studies assessing the use of artificial intelligence in a variety of conditions, such as glaucoma, chronic kidney disease and type 1 diabetes. Of these, 8 were based in China, 4 in the United States, 3 in France, 3 in Spain, 2 in Israel and 1 in Switzerland.

Patient-facing applications

Alder Hey Children's Hospital in Liverpool began developing an AI-enabled chatbot in collaboration with IBM and the STFC (Science & Technology Facilities Council) Hartree Centre in 2016. Over 2,000 patient and carer questions have been collected to date, and subsequently classified, answered and used to train the algorithm. The result is a chatbot named 'Ask Oli', whose objective is to ease the inevitable anxiety brought on by a hospital visit, whether by guiding parents to the parking lot or advising on what to expect during procedures or their visit (Figure 3). The chatbot uses natural language processing to classify intents and respond appropriately. It is also able to recognise a specific entity from a previous question. For example an initial question asking "What is a blood test?" followed up by "Will it hurt?", will respond with the appropriate answer around whether a blood test will hurt. Where questions become too intrusive or difficult, for instance those addressing mental health issues, the chatbot closes down the conversation and directs the patient towards seeking a parent or member of staff for advice and help. Users currently access the chatbot via the Alder Play app as guests, but there are plans to make it personalised in the future by linking it up with the individual patient's medical records.[69] A similar technology has also been used by IBM in partnership with Arthritis Research UK, to develop an Arthritis Virtual Assistant (AVA) to answer questions set to it by arthritis patients. A beta version of this tool is currently available on the Arthritis Research UK website.[70]

Another role for AI may be to enable patients to self-manage their conditions at home once discharged from hospital. Patients from the University of Pittsburgh Medical Center (UPMC) transfer their health information back to UPMC through a mobile device through a programme that uses machine learning technology from Vivify Health.



▲ **Figure 3:** Screenshot from 'Ask Oli', the chatbot app that uses natural language processing to answer questions by patients and their relatives about their hospital stay. In the example above, Oli correctly 'understands' that "sleep" in this context means anaesthesia for general surgery, and gives relevant information. (Image used with permission of Alder Hey Children's NHS Foundation Trust, The STFC Hartree Centre and Alder Hey Children's Charity)

69. Interview with Dr Iain Hennessey and Carol Platt, 25th January 2018
70. Arthritis Research UK, available at https://www.arthritisresearchuk.org/Arthritis%20information.aspx

It monitors symptoms of conditions such as heart failure or diabetes, as well as blood pressure, weight and oxygen levels and flags patients at risk of ending up in the emergency department, allowing a healthcare practitioner to intervene earlier by phone or a visit. During its first year of enrollment, UPMC reported Medicare beneficiaries using the remote monitoring system were 76% less likely to be readmitted within 90 days of discharge.[71]

Examples of autonomous systems delivering treatment to patients are scarce. Perhaps most progress has been made in the case of autonomous insulin pumps for type 1 diabetes (the 'artificial pancreas'). These closed-loop systems regularly sample patient glucose levels, and automatedly adjust insulin delivery rate accordingly. Furthermore, they can predict when a patient's blood sugar is likely to decrease, and suspend insulin delivery 30 minutes in advance of the predicted hypoglycaemic episode, reducing the risk of debilitating hypoglycaemia.[72] Although potentially revolutionary, access to these closed-loop insulin pumps is currently limited by cost and availability - although the MiniMed 670G pump system by Medtronic has been approved by the FDA, such systems are not widely used in the UK NHS.[73] This limited access has prompted some patients to set up the OpenAPS community, which campaigns using the #WeAreNotWaiting hashtag. They are developing solutions and reverse-engineering existing products to make basic closed loop APS technology more widely available to anyone with compatible medical devices who is willing to build their own system.[74]

Another type of AI-enabled tool that interfaces directly with patients and other users is the chatbot that provides health advice. For example, the Lark Weight Loss Health Coach is an automated text-based mobile coaching service that combines a chatbot interface with additional clinician support, to help users reach their weight goals. The bot gives either positive reinforcement or constructive criticism, for example on how to eat better following the user's description of their breakfast. Users are also able to report feelings such as guilt, and receive advice from the app in return. A longitudinal observational study showed that users lost an average of 2.4% of their baseline weight and the number of healthy meals consumed increased by 31%.[75] User acceptability was high; participants gave the app a satisfaction rating of 7.9 out of 10, suggesting the potential for AI to drive positive lifestyle changes.

The needs of an aging population have also attracted the attention of AI developers. At the Toronto Rehabilitation Institute, scientists are developing 'smart home' systems to support older adults in their own home, by providing step-by-step prompts for daily tasks, responses to emergency situations, and means to communicate with family and friends.[76,77,78] AI is being used to understand speech commands and non-verbal communication made by the user, and provide appropriate support. The goal of this research is to develop spaces that contain embedded systems connected to mobile assistive robots. Similar assisted living tools for patients with cognitive impairment are being developed by researchers at the University of

71. Johnson, S."The future is now?", Modern Healthcare, available at http://www.modernhealthcare.com/indepth/how-ai-plays-role-in-population-health-management/ Overweight and Obese Adults", JMIR Diabetes, available at http://diabetes.jmir.org/2017/2/e28/
72. Available at https://www.medtronicdiabetes.com/products/minimed-670g-insulin-pump-system
73. Sennaar, K. (2018) "AI in medical devices – three emerging industry applications", tech emergence, available at https://www.techemergence.com/ai-medical-devices-three-emerging-industry-applications/
74. Available at https://openaps.org/
75. Stein, N. & Brooks, K. (2017) "A Fully Automated Conversational Artificial Intelligence for Weight Loss: Longitudinal Observational Study Among Overweight and Obese Adults", JMIR Diabetes, available at http://diabetes.jmir.org/2017/2/e28/
76. Rudzicz, F. et al. (2014) "Speech recognition in Alzheimer's disease with personal assistive robots", University of Toronto, available at http://www.cs.toronto.edu/~frank/Download/Papers/rudzicz_slpat14.pdf
77. Mihailidis, A. et al. (2008) "The COACH prompting system to assist older adults with dementia through handwashing: An efficacy study", BMC Geriatrics, available at https://bmcgeriatr.biomedcentral.com/track/pdf/10.1186/1471-2318-8-28?site=bmcgeriatr.biomedcentral.com
78. Lee, T. and Mihailidis, A. (2005) "An intelligent emergency response system: preliminary development and testing of automated fall detection", Journal of Telemedicine and Telecare, available at http://journals.sagepub.com/doi/pdf/10.1258/1357633054068946

South Wales, including the 'Multi Agent Systems for Smart Home Environments' project and the 'Smart Dementia Wales' project.[79]

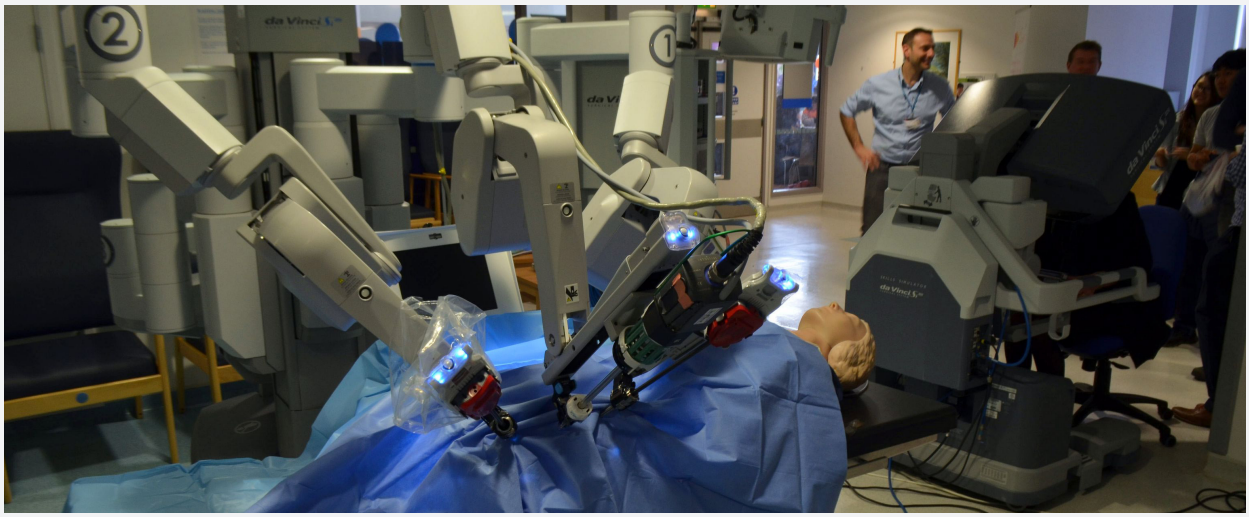
The use of robotics may provide another set of tools to assist older or vulnerable patients. Japan is faced with a rapidly aging population and its Government has launched various initiatives to tackle the consequences. Its Robot Strategy has the explicit aim that four in five care recipients will accept having some support provided by robots by 2020, for example.[80] Thus, robotics engineers are looking into the possibility of augmenting robotic tools with AI to provide more intelligent support to these patients, reminding them when to take their medicine, for instance.[81] Moreover, significant investment is being put into high-risk, high-impact research and development, via the 'ImPACT' innovation programme. Projects within its scope aim to maximise the independence of nursing care recipients and minimise the burden of the caregiver, to support knowledge acquisition and prevent cognitive decline in older people, and to accelerate medical research and development.[82]

Devices such as intelligent walkers and wheelchairs may also augment safety and independence. Hasbro, the toymaker, is developing a collection of animatronic cats and dogs, to support older adults in need of reminders, such as those with early dementia. In addition to locating objects, the project aims to further identify user needs; the toys will be equipped with sensors, so that purrs and growls can direct the owner to their medication or desired object. A prototype is slated for release in three years' time.[83] In addition to compensating for a shortage of staff, robots can take over physical tasks such as lifting patients from their beds. Japanese-developed Robobear is a newer iteration of RIBA (Robot for Interactive Body Assistance).[84] The bear-shaped robot can turn patients, lift them into a wheelchair and help them stand.

Robots are already assisting surgical procedures (Figure 4). In 2016, ophthalmologist Professor Robert MacLaren used a remotely controlled robot to lift a membrane 100th of a millimetre thick from the patient's retina.[85] Such precision could facilitate procedures that humans are currently unable to perform due to factors such as tremors in the surgeon's hand. Professor Ferdinando Rodriguez y Baena, Professor of Medical Robotics at Imperial College London, envisages robots in healthcare more as smart assistants than autonomous tools, by giving access to additional information from sensors located at the interface with the patient, which can then in turn be used to train surgeons or to improve surgical techniques.[86] Professor Rodriguez y Baena's research has included developing minimally invasive knee replacement surgery with a hands-on robot. The Acrobot system, which helps the surgeon align the replacement knee parts with the existing bones, consists of surgical planning software and a surgical arm and has been successfully proven in clinical practice, with higher accuracy than conventional surgery.[87] In future, the combination of such robots with automating algorithms could lead to the development of powerful tools for use in healthcare, but many commentators agree that we are some way away from this becoming reality.

AI could also be used to provide personalised health advice and interventions. In Chicago, CareSkore uses AI to learn from historic patient data. The tool aims to provide actionable insights

79. Interview with Dr Bertie Muller, 26th January 2018
80. The Headquarters for Japan's Economic Revitalization (2015) "New Robot Strategy", Ministry of Economy, Trade & Industry, available at http://www.meti.go.jp/english/press/2015/pdf/0123_01b.pdf
81. Stone, P. et al. (2016) "Artificial intelligence and life in 2030", Stanford University, available at https://ai100.stanford.edu/sites/default/files/ai_100_report_0831fnl.pdf
82. ImPACT pamphlet, available at https://www.jst.go.jp/impact/download/data/ImPACT_p_en.pdf
83. Muoio, D. (2017) "Researchers adding AI, medication reminders to companion robots for seniors", mobi health news, available at http://www.mobihealthnews.com/content/researchers-adding-ai-medication-reminders-companion-robots-seniors
84. Szondy, D. (2015) "Robear robot care bear designed to serve Japan's aging population", New Atlas, available at https://newatlas.com/robear-riken/36219/
85. Press Release (2016) "World first for robot eye operation", Oxford University, available at http://www.ox.ac.uk/news/2016-09-12-world-first-robot-eye-operation#
86. Interview with Ferdinando Rodriguez y Baena, 26th January 2018
87. Press Release (2006) "Robot assisted surgery more accurate than conventional surgery", Imperial College London, available at http://www.imperial.ac.uk/college.asp?P=7449



▲ **Figure 4:** A da Vinci Robot Surgical System on display at Addenbrooke's Treatment Centre, Cambridge, during the 2015 Cambridge Science Festival. Although autonomous surgical robots are not currently in use, robots are already assisting surgical procedures, and the data being gathered from the use of this technology can be used to train surgeons or to improve surgical techniques. (Image used under the Creative Commons Attribution-Share Alike 3.0 Unported license)

and informs hospitals of a patient's various risks, such as the risk of readmission, falls, and sepsis. For example, a recently-discharged heart patient with a history of depression may be flagged and consequently provided with a tailored care plan. The platform also recommends the best way to remind patients about upcoming appointments, choosing to send a single text reminder to some patients, or suggesting phone calls to others, to try to reduce no-shows at clinics. Readmission rates at the Methodist Hospital of Chicago dropped from 12% to 4% through CareSkore's identification of primarily social determinants and the simple solution of calling patients to check on progress.[88]

Similarly, wellness monitoring services are springing up which claim to use AI to provide personalised diet and lifestyle recommendations based on the results of blood, urine, saliva and stool samples that users send in. For example, clients of companies such as DayTwo and Viome send their stool samples in to be analysed for their gut microbiome - the collection of bacterial organisms in our intestine - and then receive tailored diet recommendations.

Looking to the future, the global predictive genetic testing & consumer/wellness genomics market could be worth an estimated $4.6 billion by 2025.[89,90] Sequencing company Veritas Genetics is developing an AI-powered platform with data from millions of genomes, with the aim of allowing customers to request interpretations of their own DNA in terms of their risk of conditions such as cancer and cardiovascular disease.[91]

88. Lee, K. (2017) "Population health management platform uses AI, machine learning", TechTarget, available at http://searchhealthit.techtarget.com/feature/Population-health-management-platform-uses-AI-machine-learning
89. Grand View Research, Inc. (2017) "Predictive genetic testing & consumer/wellness genomics market worth $4.6 billion by 2025", PR Newswire, available at https://www.prnewswire.com/news-releases/predictive-genetic-testing–consumerwellness-genomics-market-worth-46-billion-by-2025-grand-view-research-inc-612533583.html
90. Frey, B. (2015) "Taking the genome further in healthcare", Tech Crunch, available at https://techcrunch.com/2015/12/17/taking-the-genome-further-in-healthcare/
91. Taylor, P. (2017) "Veritas buys Curoverse to boost genomic data capabilities", Fierce Biotech, available at https://www.fiercebiotech.com/medtech/veritas-buys-curoverse-to-boost-genomic-data-capabilities

Population-level applications

Such applications include the identification of groups likely to require interventions to prevent the onset of disease, which can be particularly effective for managing chronic illness.[92,93] AI-based tools could be useful given the ability to derive insights from large volumes of data, discover patterns, and uncover predictive trends.

In combination with the powerful analytical ability at scale of AI, tools are being developed to use non-traditional clinical data sources such as mobile phone activity to forecast the progression of epidemics, and thus divert the necessary resources to where they are most needed at the correct time. Mobile phone data has already been used to model the spread of cholera in Haiti in 2010, and of dengue fever in Pakistan in 2013.[94,95,96] In 2016, Malaysia became the first country in the world to use an app to predict dengue outbreak - the Dengue Outbreak Prediction Platform. AI is used to analyse parameters including geography, weather and symptoms of dengue cases to predict hotspots, where preventative actions such as the elimination of mosquito larvae are then performed.[97] The platform is able to predict outbreaks three months ahead with an accuracy of 86%.[98] Similarly, Researchers from the University of Southern California have developed an AI algorithm that can slow the spread of infectious diseases by making data-driven suggestions on how to allocate limited public health resources, such as funds for information campaigns. In two real world cases - communities with tuberculosis in India and with gonorrhea in the US - outcomes were better when this tool was used than with traditional health outreach policies.[99]

AI could also be used to model the changing incidence of non-communicable diseases. For example, a machine learning model has been developed that predicts childhood obesity, and suggests non-standard risk factors for obesity that healthcare practitioners could start to take increased notice of.[100]

92. Hibbard, J. et al. (2017) "Improving Population Health Management Strategies: Identifying Patients Who Are More Likely to Be Users of Avoidable Costly Care and Those More Likely to Develop a New Chronic Disease", NCBI, available at https://www.ncbi.nlm.nih.gov/pubmed/27546032
93. Kini, P. (2017) "Why population health is an AI problem", Hospital Review, available at https://www.beckershospitalreview.com/population-health/why-population-health-is-an-ai-problem.html
94. Bengtsson, L. et al. (2015) "Using mobile phone data to predict the spatial spread of cholera", Nature, available at https://www.nature.com/articles/srep08923.pdf
95. Wesolowski, A. et al. (2015), "Impact of human mobility on the emergence of dengue epidemics in Pakistan", PNAS, available at http://www.pnas.org/content/112/38/11887.full.pdf
96. Bengtsson, L. et al. (2015) "Using mobile phone data to predict the spatial spread of cholera", Nature, available at https://www.nature.com/articles/srep08923.pdf
97. San, M. (2016) "Malaysia is first in the world to use a mobile app to predict dengue outbreak", Asia News Network, available at http://annx.asianews.network/content/malaysia-first-world-use-mobile-app-predict-dengue-outbreak-33938
98. Brown, V. and Page, T. (2016) "AI and big data joins effort to predict deadly disease outbreaks", CNN, available at https://edition.cnn.com/2018/03/06/health/rainier-mallol-tomorrows-hero/index.html
99. PTI (2018) "Artificial intelligence may help check tuberculosis spread in India: Study", The Indian Express, available at http://indianexpress.com/article/technology/science/artificial-intelligence-may-help-check-tuberculosis-spread-in-india-study-5074489/
100. Dugan, T. et al. (2015) "Machine learning techniques for prediction of early childhood obesity", Applied Clinical Informatics, available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4586339/pdf/ACI-06-0506.pdf

ETHICAL, SOCIAL, AND POLITICAL CHALLENGES

The various uses of artificial intelligence raise interesting and pressing ethical, social, and political considerations. Our understanding of these issues has been developed by undertaking a detailed review of existing literature, interviewing and organising roundtable discussions with more than 70 experts all round the world, and seeking the views of patients, their representatives, and members of the public. Some of these can be seen as challenging existing ethical issues in new ways, while others can be seen to constitute whole new areas of discourse. We outline and discuss the major ethical, social, and political issues that we have identified below.

01 What effect will AI have on human relationships in health and care?

- What effect will these technologies have on relationships between patients and healthcare practitioners?
- What effect will these technologies have on relationships between different healthcare practitioners?
- What do healthcare practitioners think about the potential for these technologies to change their jobs, or to lead to job displacement?
- How do these tools fit into the trend of enabling patients to have greater knowledge and understanding of their own conditions? How different are they from looking up one's symptoms on a search engine before going to see a healthcare practitioner?
- Given that AI is trained primarily on 'measurable' data, does reliance on AI risk missing non-quantifiable information that is so important in healthcare interactions?
- If AI systems become more autonomous, how should transitions between AI and human control be incorporated into care pathways?

Healthcare is built on relationships between individuals. Apart from the central interaction between the healthcare practitioner (HCP) and the patient, many other types of relationship have been identified as essential to the delivery of good healthcare: between the HCP and caregivers/relatives; between caregivers/relatives and patients; between different HCPs; between HCP, inpatient facility, and patient; between "front of house" administrative staff and the patient; between top-level administrators and HCPs 'on-the-ground'; and between patients and wider society.[101] AI technologies have the potential to modify all of these relationships.

Our interviewees and the participants in our patient roundtable described the issue of AI, or health data, acting as "third participant" in the previously binary patient-HCP dynamic. Dr Brent Mittelstadt (Research Fellow at the Oxford Internet Institute) indicated that AI could be used to combine different data sources, including non-medical data such as from wearables, to make inferences about the patients that HCPs and patients would not have had access to previously.[102] Whereas the relationship between patient and HCP was characterised by a two-way flow of information between these individuals, an autonomous decision-making algorithm, or a cache of health data collected by the patient themselves, introduces a new set of interactions, including that between the AI and the HCP, and between the AI and the patient. This raises a number of interesting questions. How are patients and HCPs to respond when the inference from the patient's data contradicts the patient's own account - especially if the

101. Adapted from Selinger, H. A. Fam Pract Manag. 2013;20(5):40, available at https://www.aafp.org/fpm/2013/0900/p40.html
102. Interview with Dr Brent Mittelstadt, 1st February 2018

patient has no control over the inferences the algorithm makes? How are patients and HCPs to respond when the automated 'decision' made by an algorithm contradicts that HCP's thinking and recommendations? How is the "additional information" derived from the algorithm, that the patient may be bringing into the relationship between them and their HCP, different from the increasing tendency for patients to research their own conditions on the internet as part of their care pathway?

Many of our interviewees discussed issues around how autonomously-operating algorithms hand decision-making control back to human operators, which are relevant when considering the relationship between HCPs and their patients, and between different HCPs. Although perhaps this is more of a future consideration in the context of healthcare, it is a very live issue in some other industries. For example, in the Air France Flight 447 crash in the Atlantic Ocean on May 31, 2009, a key factor that led to the disaster was a failure of the human pilots to take over safely when the automated 'fly-by-wire' system shut itself off, as it was programmed to do, when a pressure probe on the outside of the plane iced over, and the system could no longer tell how fast the plane was going.[103] Similarly, in the autonomous vehicle industry, trials show that significant issues with processes of switching from autonomous to manual vehicle control are yet to be addressed.[104] In the context of healthcare, if autonomous algorithms only handover to human operators in complex situations that they are not designed to handle, how will human practitioners keep up their skills sufficiently to be able to address these situations? Furthermore, should we flag this transition from algorithmic control to human control clearly to patients, and if so, how?

As part of their project on Machine Learning, the Royal Society asked Ipsos MORI to study public perceptions towards these technologies. With respect to their application in health, a major concern that was raised across the 978 interviews conducted with members of the public is the risk of loss of human interaction - that is, that AI technologies could encroach or in some way degrade the patient-HCP relationship.[105] Our own poll, conducted by YouGov, also suggests the importance patients place on the relationship with their HCP in health and care (see Question 6). However, many of our interviewees emphasised the potential for algorithms to perform routine, repetitive tasks, freeing up HCPs to spend more time interacting with their patients. It is therefore unclear whether these algorithms will negatively or positively affect HCP-patient relationships.

Another perspective on the effect of algorithms on relationships in healthcare was raised by Professor Margaret Boden, Research Professor of Cognitive Science at the University of Sussex. Algorithms are trained on data that is 'measurable', such as images, health records, and blood test results - there is a definite skew towards using quantitative data. However, Professor Boden highlighted that many healthcare interactions depend on more than just this 'measurable' data, including non-verbal communication between individuals, and their social and other circumstances.[106] How do we capture the value to healthcare of this data that is much harder to measure? Or, as Professor Boden put it, "If we only measure what we can measure, what do we miss out on?"[107]

Professor Stefan Schulz, Professor of Medical Informatics at Medical University Graz, Austria, raised a further issue that relates to the

103. Mars, R. (2015) "Air France Flight 447 and the Safety Paradox of Automated Cockpits", Slate, available at http://www.slate.com/blogs/the_eye/2015/06/25/air_france_flight_447_and_the_safety_paradox_of_airline_automation_on_99.html
104. Masters, J. (2017) "Autonomous vehicles: 'handover' process crucial say researchers", Infrastructure Intelligence, available at http://www.infrastructure-intelligence.com/article/jun-2017/autonomous-vehicles-handover-process-crucial-say-researchers
105. Ipsos MORI (2017) "Public views of Machine Learning: Findings from public research and engagement conducted on behalf of the Royal Society", available at https://royalsociety.org/~/media/policy/projects/machine-learning/publications/public-views-of-machine-learning-ipsos-mori.pdf
106. An interviewee who contributed to the Royal Society/Ipsos MORI survey on Machine Learning put this issue well: "[The] computer may notice that speech is slurred, but the computer won't notice that the person may smell like brandy when they come in the room." Citation as above.
107. Interview with Margaret Boden, 29th January 2018

relationship between the algorithms and HCPs. If the thresholds for alerts raised by the AI agents is set too low, there is a risk that the "AI agents [...] flood the workplace with lots and lots and lots of alerts, with the effect that then nobody takes the alerts seriously, the same as if the fire alarm [goes off] every day in your office - then if there's really a fire, then nobody takes it seriously." These 'false positives' would be disruptive, and could bring about an "adverse reaction" towards AI on the part of HCPs.[108] In order for the relationship between HCPs and algorithms to function effectively, practitioners need to trust the algorithms (see Question 9, below).

Another relatively unexplored issue is the effect that AI-driven automation may have on healthcare practitioners' jobs, both in terms of displacement, or more broadly in terms of changing the nature of jobs. As outlined in the 'Current and potential uses of artificial intelligence in health' section above, automation may free up HCP time that is currently occupied by routine administrative tasks, allowing them to spend more time interacting with patients. However, as the technology improves and more tasks become automatable, it is increasingly possible that fewer 'human practitioners' will be required to run healthcare systems worldwide. Another trend that may gain increasing traction is that of 'Uber-isation' of the healthcare workforce - that is, increased reliance on platform or 'gig' workers. This may impact the relationship between top-level administrators and workers at the frontline of the health service, and is an important issue in the study of current employment models. It was a particular focus of the 'Taylor Review of Modern Working Practices', commissioned by the UK Government in October 2016 and published in July 2017.[109] To our knowledge, there is no systematic analysis of how HCPs view this potential impact of automation and any preparatory measures that are being taken by their professional bodies.

The relationship between patients and wider society may also be affected by the increasing use of AI and algorithms in healthcare. As outlined below (Question 8: 'Just because these technologies could enable access to new information, should we always use it?') the use of AI may enable the assignment of people to newly-created subgroups and categories, separating them from the wider community they consider themselves to be a part of. Moreover, one of the proposed benefits of these technologies is the potential for increased personalisation of therapies and treatments, possibly basing these on an individual's genome or other unique attributes. Many have argued that society as a whole is becoming more individualistic, so it's perhaps unsurprising that healthcare is not immune to this trend.[110,111] The concept of 'solidarity' has been invoked in bioethical discussions on the balance between the good of the community and that of the patient, when these are seen to be in conflict. Barbara Prainsack and Alena Buyx, in their work for the Nuffield Council on Bioethics, have suggested that solidarity implies "a collective commitment to carry 'costs' (financial, social, emotional or otherwise) to assist others".[112] It remains to be seen if and how solidarity is redefined and reevaluated if AI-driven personalisation becomes an increasing feature of healthcare.

Lastly, the use of technology to incentivise, or 'nudge', better health behaviours by users could also be seen to impact the relationships between patients, their HCPs and wider society. Patients are used to receiving lifestyle advice from their doctors or nurses and may consent to this as part of their care pathway. Would receiving

108. Interview with Stefan Schulz, 30th January 2018
109. Taylor, M., Marsh, G., Nicol, D. and Broadbent, P. (2017) ''Good work: The Taylor Review of Modern WorkingPractices", RSA, available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/627671/good-work-taylor-review-modern-working-practices-rg.pdf
110. Santos, H. C., Varnum, M. E. W., and Grossmann I. (2017) "Global Increases in Individualism" Psychol Sci. 28(9):1228-1239. doi: 10.1177/0956797617700622.
111. Douthat, R. (2014) "The Age of Individualism" The New York Times, available at https://www.nytimes.com/2014/03/16/opinion/sunday/douthat-the-age-of-individualism.html
112. Prainsack, B., and Buyx, A. (2011) "Solidarity: reflections on an emerging concept in bioethics", Nuffield Council on Bioethics, available at http://nuffieldbioethics.org/wp-content/uploads/2014/07/Solidarity_report_FINAL.pdf

this advice via an AI-enabled tool be identical to receiving it from an HCP, or could it lead to behavioural 'manipulation' if the advice is not clearly flagged as such? This question is made all the more pertinent when one considers that AI-enabled 'newsfeeds' on social media have been implicated in the promotion of fake news and the creation of 'filter bubbles', with potential political consequences.[113,114] Would it be justified for governments or other authorities to use such tools to 'nudge' whole populations into healthier behaviours on a large scale, on the grounds that it will lead to improved population health? How different would this be from existing public health campaigns?

02 How is the use, storage, and sharing of medical data impacted by AI?

- How is medical data different from other forms of personal data?
- What is the most ethical way to collect and use large volumes of data to train AI, if the consent model is impractical or insufficient?
- How do we check datasets for bias or incompleteness, and how do we tackle these where we find them?
- Should patients who provide data that is used to train healthcare algorithms be the primary beneficiaries of these technologies, or is it sufficient to ensure that they are not exploited?

We now live in a data-sharing world and artificial intelligence relies on large volumes of it. Issues around the use of data clearly go beyond the fields of health and medical research. A lot of attention is being paid to data governance in various jurisdictions, with the General Data Protection Regulation (GDPR) coming into force in all EU member states in May 2018, and the UK Parliament currently considering an updated Data Protection Bill that will replace the Data Protection Act of 1998. There is growing awareness and concern among the public and media too. Recent headlines have been dominated by the news that Cambridge Analytica, a British political consulting firm, gained unauthorised access to the data of potentially millions of Facebook users, in order to influence voter behaviour. This is leading to calls for greater regulation of personal data handling by social media companies, much of which is done using AI.

The GDPR concerns itself with a wider swathe of data types than has previously been the case with data protection legislation. According to the GDPR, 'personal data' that falls under its scope is "any information relating to an identified or identifiable natural person ('data subject')".[115] Clearly, many forms of data can be considered sensitive, where their abuse or misuse could result in harm to the individual concerned. And yet, is there something 'special' about health data or medical information?

Our interviewees provided various different answers to this question. Professor Eduardo Magrani, Professor of Law, Technology and Intellectual Property at Fundação Getúlio Vargas Law School, Brazil, highlighted the lack of control one has over one's medical information. If, for example, a car insurer fitted a recorder to a client's car and based their insurance premiums on their driving data, the driver could take action to improve their behaviour and cheapen their premiums, for example by driving more slowly. On the other hand, Prof. Magrani argued, there are certain aspects of a patient's medical record that are out of their control, such as their genetic sequence, or their past medical

113. Hern, A. (2017) "How social media filter bubbles and algorithms influence the election", The Guardian, available at https://www.theguardian.com/technology/2017/may/22/social-media-election-facebook-filter-bubbles
114. El-Bermawy, M. M. (2016) "Your Filter Bubble is Destroying Democracy", Wired, available at https://www.wired.com/2016/11/filter-bubble-destroying-democracy
115. An "identifiable natural person" is "one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person". More information is found at Information Commissioner's Office, "Guide to the General Data Protection Regulation (GDPR): Key Definitions", available at https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/key-definitions/

history. This lack of control may make us more sensitive about our medical data.[116]

Dr Benedict Rumbold, a research fellow in the Department of Philosophy at University College London, spoke about the context where medical data is usually shared, such as the confidential setting of a consultation with a healthcare practitioner. The fact that this data has traditionally been given in confidence may explain why in today's discussions on medical data sharing, confidentiality remains a major concern, but it is unclear why medical information is something we have traditionally wanted to keep private in the first instance.[117] The stigmatising effect of certain conditions throughout history may be relevant here.

Professor Rhema Vaithianathan, Co-Director of Centre for Social Data Analytics at Auckland University of Technology (NZ), told us about the interesting work she has been doing as a member of the Data Futures Partnership (DFP), an independent group established by the NZ government to develop solutions for data-use issues. Together with her colleagues at the DFP, she found that there are cultural differences in service users' approaches to data governance. Moreover, the DFP did not find major differences between attitudes to health data and other types of sensitive data - for example, many subjects were just as concerned about their children's education data (i.e. their grades and progress reports) as they were about their medical data. After consulting with thousands of New Zealanders about their levels of comfort on data use, in 2017 the DFP published the *NZ Guidelines for Trusted Data Use*, for use by public and private sector organisations wanting to establish trusted use of data.[118]

In the UK, bodies such as the National Data Guardian, the Information Commissioner's Office and the Wellcome Trust's Understanding Patient Data initiative have led the way in gauging the public's attitude to health/medical data and its use. Key issues identified by these projects include:

- There is a lack of public understanding of how patient data is used and there is an appetite both on the part of patients and of healthcare practitioners to be educated about this.
- People seek transparency about the type of data shared, who it is used by and for what purpose, as well as data security.
- There is a clear hierarchy of trust. The public trusts the NHS, universities and to a certain degree pharmacies, to have access to data for research purposes, as these types of organisations are perceived to work in the public interest. Those that are not perceived to work in the public interest, such as insurance or marketing companies, do not have the public's trust (See Question 9, below).
- People want the option to opt out from personal confidential data being used beyond their own direct care.
- There is a strong desire for data users to be held accountable for any data misuse, for example by receiving a large fine.

The Understanding Patient Data (UPD) project is already working towards developing tools to address the issues above. In collaboration with the Academy of Medical Sciences and Ipsos MORI, UPD is developing a programme of public dialogue, with the aim of exploring public, patient, researcher and healthcare professionals' views about "new and emerging data-driven technologies that use patient data in healthcare and research."[119,120] The aim is to use this research to develop policy around the use of patient data in developing new technologies, such as AI for medical use.

116. Interview with Eduardo Magrani, 15th January 2018
117. Interview with Benedict Rumbold, 21st February 2018
118. Interview with Rhema Vaithianathan, 19th February 2018
119. Interview with Natalie Banner and Nicola Perrin, 19th January 2018
120. "Use of patient data in healthcare and research", Academy of Medical Sciences, available at https://acmedsci.ac.uk/policy/policy-projects/use-of-patient-data-in-healthcare-and-research

Much of medical practice operates on the basis of consent. This is especially the case in medical research, where various frameworks exist to regulate the use of patient tissue or information in research.[121] The sheer size of the datasets used to train AI algorithms for medical (or indeed other) use makes consent a potentially impractical framework to operate under - it may be impossible to get specific informed consent from each and every patient whose data is in a particular training dataset. It is argued by many that retrospective, historical data can be used in an anonymised fashion for research without seeking specific consent from the individuals concerned, but what about when data is collected prospectively to use in AI development? Similarly, as AI excels at finding patterns and correlations in data that may not be obvious on human analysis, it is impossible to state, at the point of collection, exactly how an algorithm will use a particular data point from a particular person, and whether this will be important for the algorithm as a whole. These points and others mean that frameworks may need to be developed and tested that bypass informed consent as a legal and ethical basis on which to conduct data collection for use in AI research, but that still protect patient autonomy and privacy. The National Data Opt-Out programme, which is being run by NHS Digital and launches in May 2018 to coincide with GDPR coming into force, may provide one such framework. Under this programme, patients and the public who decide they do not want their personally identifiable data to be used for planning and research purposes will be able to set their national data opt-out choice online or via a 'non-digital alternative'.[122] This approach demonstrates similarities with opt-out approaches to organ donation, which are in force in Wales, planned in Scotland, and under consultation in England.[123]

Data bias - that is, the use of datasets that are not fully representative of the population they seek to typify - is a concern in artificial intelligence that goes beyond its use in healthcare applications. The maxim 'garbage in, garbage out' was repeated to us many times over the course of our interviews, underlining the fact that algorithms trained on biased datasets will provide biased outputs. In their landmark report on 'Data management and use: governance in the 21st century', the Royal Society and British Academy clearly explain this issue:

> "To this aim, it is crucial for policymakers to recognise that there is no simple technological fix for monitoring the social impact of data use. Computational tools for data tracking and monitoring continue to improve at breathtaking speed, and yet they unavoidably rely on human decisions about what counts as data in the first place and how data should be ordered, labelled and visualised. These decisions are particularly significant given that not all data are equally easy to digitally collect, disseminate and link through existing algorithms, resulting in a highly biased data pool that does not accurately reflect reality (and in some cases actively distorts it). Far from being purely technical, data management decisions therefore affect what kinds of uses data can be put towards, and its implications."[124]

121. Examples include the General Medical Council's 'Consent to Research' guidelines (https://www.gmc-uk.org/guidance/ethical_guidance/5993.asp) and the Medical Research Council's 'Guidance on Patient Consent' (https://www.mrc.ac.uk/research/policies-and-guidance-for-researchers/guidance-on-patient-consent/)
122. NHS Digital, 'National Data Opt-Out Programme', available at https://digital.nhs.uk/national-data-opt-out
123. UK Government (2017) 'Government announces consultation on organ donation opt-out system', available at https://www.gov.uk/government/news/government-announces-consultation-on-organ-donation-opt-out-system
124. The Royal Society and the British Academy (2017) 'Data management and use: governance in the 21st century', available at https://royalsociety.org/~/media/policy/projects/data-governance/data-management-governance.pdf

It is unclear what the implications of such data bias are when applied to health and medical research, but it is worth noting that it is not a new problem in health either. Research conducted by the US Food and Drug Administration (FDA) shows that African-Americans comprise less than 5% of clinical trial participants and Hispanics just 1%, in spite of the fact that they represent 12% and 16% of the total US population respectively.[125] Besides dealing with underrepresentation on the grounds of ethnicity, datasets for use in AI need to ensure balance in other parameters, including gender, age, sexual orientation, educational status and employment status (for example undocumented workers and their families), as well as a factor which perhaps is not normally considered in health contexts: digital literacy.

Another data-related issue that emerged from our interviews and roundtable discussions is that of the value of data, both to patients themselves, and to citizens in a publicly-funded healthcare system. Many are concerned that the current model of public-private partnerships used to develop AI algorithms, where private sector organisations partner with bodies such as hospitals and universities to develop algorithms using data held by the latter group of institutions (see Question 10, below), may not be a good way to ensure that the value of data is adequately recouped for patients and citizens, and may even enable exploitation of patients. Some have argued that patients should be the primary beneficiaries of technologies developed using their data, whereas others have suggested that it is sufficient to ensure that they are not exploited, even if they derive no direct benefit. Sir John Bell, Regius Professor of Medicine at the University of Oxford and lead on a recent review of the Life Sciences Industry for the UK Government, has mooted a number of different options to ensure value is captured from NHS data, including charging fees to access the databank, or a licence system that pays the UK Treasury royalties from sales of products developed using NHS data.[126] How such initiatives could be implemented is an area that requires further exploration.

03 What are the implications of issues around algorithmic transparency and explainability on health?

- Are expert systems or rule-based AI systems more suitable for healthcare applications than less interpretable machine learning methods?
- What do patients and healthcare practitioners want from algorithmic transparency and explainability?
- Are improved patient outcomes, efficiency and accuracy sufficient to justify the use of 'black box' algorithms? If such an algorithm outperforms a human operator at a particular healthcare-related task, is there an ethical obligation to use it?
- Could 'explanatory systems' running alongside the algorithm be sufficient to address 'black box' issues?

Modern machine learning algorithms, particularly neural networks, have often been referred to as 'black boxes'.[127] Such decision-making systems are often deployed as a background process, unknown and unseen by those they impact. The use of this technology in this way raises significant and justifiable concerns. A notorious example is provided by the COMPAS algorithm, which was used by American courts to assess the likelihood of an individual re-offending. It was found to be two times less likely to falsely flag white people and two times more likely to falsely flag black people as likely to reoffend.[128] Even worse, when challenged,

125. Buch, B. D. (2016) "Progress and Collaboration on Clinical Trials", FDA Voice, available at https://blogs.fda.gov/fdavoice/index.php/tag/fda-sia-section-907/
126. Aldrick, P. (2018) "Data could be a huge source of funding for the NHS and we are about to give it away", The Times, available at https://www.thetimes.co.uk/article/7ffc0130-1e4a-11e8-95c3-8b5a448e6e58
127. Castelvecchi, D. (2016) 'Can we open the black box of AI?' Nature News, available at http://www.nature.com/news/can-we-open-the-black-box-of-ai-1.20731
128. Angwin, J., Larson, J., Mattu, S., and Kirchner, L. (2016) 'Machine Bias', Pro Publica, available at https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

the manufacturers claimed that the algorithm was protected under intellectual property law and was therefore not open to scrutiny.[129]

Chapter 3 of the General Data Protection Regulation (GDPR; described above in Question 2) deals with transparency in cases of automated decision-making, and provides for "meaningful information about the logic involved", which can be translated as the right to explanation. Although the issue of algorithmic explainability is one that comes up in almost any discussion of applications of artificial intelligence, it is currently unclear how regulatory frameworks such as the GDPR will interact with applications of AI in healthcare.

One suggestion that was made by some of our contributors was to restrict the type of algorithms used in healthcare applications to explicitly-programmed, rule-based expert systems. These are more interpretable than machine learning techniques. However, advances in machine vision, powered by exactly the type of deep learning algorithms that raise concern due to their impenetrability, have underpinned the superhuman performance shown by some algorithms. If this technology continues to get better and eventually consistently outperforms humans in, for example, image analysis tasks, should the opacity of these algorithms be a bar to their widespread application? Some of our contributors argued that it is in fact unethical to withhold these algorithms from medical practice if they clearly outperform human practitioners.

It is unclear what patients and practitioners want in terms of understanding how these algorithms work. Professor Burkhard Schafer, Professor of Computational Legal Theory at the University of Edinburgh's School of Law, emphasised that different users and different situations will require different things in terms of explainability. The explanation of an output by a medical algorithm that a patient wants and deserves is almost certainly different from the explanation demanded by, say, a student who wants to understand why an automated marking system has failed their last paper.[130]

Our interviewees also recognised that human explanations of their own behaviour, which we have lived with in the context of healthcare for millenia, are overwhelmingly based on a post-hoc rationalisation of the decision taken, rather than an exhaustive understanding of our brain's decision-making process. Is this human 'black box' very different from the algorithmic black box? Some of our interviewees, including some of the patients and members of the public who contributed to our research, in fact adopted a very pragmatic approach to the issue of algorithmic explainability, pointing out that ultimately what matters is that the algorithm is clinically efficacious and improves patient outcomes, therefore justifying its use.

Various potential solutions have been suggested to the problem of algorithmic explainability in other contexts, including having 'explanatory systems' running in parallel with the main algorithm. Sandra Wachter and colleagues at the Oxford Internet Institute, for example, have put forward the concept of 'counterfactual explanations' to be provided with all decisions made by an algorithm. These 'counterfactuals' would be the minimal bit(s) of information that would have changed the outcome of the model to the desired one for the user. For example, in the context of an algorithm determining creditworthiness, a counterfactual explanation could be "You were denied a loan because your annual income is £30,000. If your annual income was £45,000, your loan application would have been approved." Such explanations would inform and help the individual understand why a particular decision was reached, provide grounds to contest the decision if the outcome is undesired, and to understand what would need to change in order to receive a desired result in future - these principles could be applied to the healthcare context.[131]

129. Angwin, J., Larson, J., Mattu, S., and Kirchner, L. (2016) 'Machine Bias', ProPublica, available at https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
130. Interview with Burkhard Schafer, 16th January 2018
131. Wachter, S., Mittelstadt, B,. and Russell, C. (2017) "Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR", Harvard Journal of Law & Technology, available at https://ssrn.com/abstract=3063289 or http://dx.doi.org/10.2139/ssrn.306328

Another angle to algorithmic transparency was touched on by Dr Geraint Lewis, Chief Data Officer at NHS England and Honorary Clinical Senior Lecturer at University College London. He highlighted the fact that almost all decisions are based on a combination of information and preferences. For example, the algorithm running a satellite navigation system in a car uses information such as current position, geographical distances and current traffic conditions to determine the recommended route. Depending on the user's preferences, however, the device will provide either the fastest route, the greenest route, or a route that avoids toll roads or low bridges and so on. In many health care algorithms, these user preferences are hidden. Dr Lewis gave the example of an algorithm that recommends the suggested dose of the blood-thinning drug warfarin based on the results of a blood test. Preferences are 'baked into' the algorithm but could potentially be adjusted to include how often the patient minds having blood tests performed, and whether they and their clinician would rather err on the side of over-anticoagulation ('blood thinning') or under-anticoagulation. Dr Lewis argued that the default settings for these preferences should be made transparent because a lack of transparency risks the development of 'loaded' algorithms that promote the interests of the manufacturer rather than the interests of the end-user.[132]

04 Will these technologies help eradicate or exacerbate existing health inequalities?

- Which populations may be excluded from these technologies, and how can these populations be included?
- Will these technologies primarily affect inequalities of access, or of outcomes?

Unequal health outcomes persist worldwide, both between different countries and within countries. The seminal work of Sir Michael Marmot, for example, underscored the relationship between lower socioeconomic class and poor health.[133] In the UK, a report by the Longevity Science Panel found that the gap in expected lifespan between boys in the richest areas of the country and those in the poorest has increased to 8.4 years.[134] Other non-medical determinants of health outcomes include educational level, an individual's physical environment (for example whether they live in crowded conditions), and access to good quality health care.[135]

The increased use of AI and other technologies in healthcare is likely to have a complex effect on health inequality. Some of our contributors argued that many of these technologies are empowering, with wearable tech, for example, giving us all a deeper insight into our behaviours and health data. Armed with the combination of this data and the power of algorithmic insights, the patient could enter a healthcare situation as an equal partner with their healthcare practitioner, rather than as a passive recipient of information and advice.

However, others have made the point that the use of wearables and apps presupposes digital literacy, and that access to these tools may be expensive. This may limit access not only to poorer individual users in advanced economies, but potentially to whole healthcare systems in low- and middle-income countries. The type of national healthcare system may also be relevant, in that certain systems, such as the US model, are known to perpetuate, or at least fail to tackle, health inequalities to a greater degree than others.[136,137]

132. Interview with Geraint Lewis, 1st February 2018
133. Marmot, M. (2017) "Social justice, epidemiology and health inequalities", European Journal of Epidemiology, available at https://doi.org/10.1007/s10654-017-0286-3
134. "Life expectancy gap between rich and poor widens", BBC News (2017), available at http://www.bbc.co.uk/news/health-43058394
135. "Social Determinants of Health", CDC, available at https://www.cdc.gov/nchhstp/socialdeterminants/definitions.html
136. Johnson, C. Y. (2017) "America is a world leader in health inequality", The Washington Post, available at https://www.washingtonpost.com/news/wonk/wp/2017/06/05/america-is-a-world-leader-in-health-inequality/?noredirect=on&utm_term=.db2f735f5f0d
137. Samuel L Dickman, MD et al. (2017) "Inequality and the health-care system in the USA", The Lancet, available at http://www.thelancet.com/journals/lancet/article/PIIS0140-6736(17)30398-7/fulltext

It is not always intuitive who is most likely and able to use these tools, and who might be excluded from them. It might be assumed, for example, that older people may be reluctant to engage with digital tools due to a lack of digital literacy. Jacob Lant, Head of Policy and Public Affairs at Healthwatch England, however pointed to research suggesting that older people are more familiar with the health service and therefore more comfortable with using a range of tools, compared to younger people who use healthcare services less, and therefore seek human interaction when they do.[138]

It is arguable that what is most important in health, in the final analysis, is equality of outcome - this doesn't necessarily follow from equality of access. It is unclear how AI-based tools will help or hinder equality of outcome in the health context.

As with other discussions on inequality, it is important to understand who the most vulnerable people that may be impacted by a particular system are. It is unclear who would be most vulnerable with wider use of AI-based systems in health and medical research.

05 What is the difference between an algorithmic decision and a human decision?

- How do we rank the importance of a human decision as compared to an algorithmic decision, particularly when they are in conflict?
- Do human and algorithmic errors differ simply in degree, or is there an essential, qualitative difference between a machine 'giving the wrong answer' and a human making a mistake?
- How will patients and service users react to algorithmic errors?
- Who will be held responsible for algorithmic errors?

Decision support was one of the earliest suggested uses of algorithms in healthcare.[139] Many of our contributors raised the issue that might arise as decision support tools are increasingly used, namely that algorithmic decisions will start to be considered as separate from the decisions of their human operators. Even if the tools are purely 'supporting' the human healthcare practitioner and have no autonomous capability to institute actions based on a decision, what happens when an algorithm and a human 'disagree' on their decision? Which will be ranked higher in terms of importance by both patients and other healthcare practitioners?

Part of the discussion around medical decisions focuses on the situation that arises when a decision is made by an algorithm or a human that is objectively 'wrong', and leads to a harmful outcome. Medical error is a major cause of morbidity and mortality.[140,141] It is therefore unsurprising that much of the discourse around the use of AI in health centres around 'algorithmic error'. Many of our contributors suggested that we appear to hold algorithms to a higher standard than we hold humans, perhaps as a result of the fact that we don't understand algorithms and cannot empathise with them the way we would with fellow humans. Dr Debra Mathews and Dr Travis Rieder, both of the Johns Hopkins Berman Institute of Bioethics, argued that trust is a significant factor in this phenomenon - an algorithmic error could be "more distressing" than a human error, because we still trust human interactions in healthcare over ones with computers (see Question 9 below for a more detailed discussion on trust). Over the course of our interviews, parallels were drawn with the situation with autonomous vehicles - although many expect them to be safer than human drivers, the bar for widespread acceptability on the roads has been set very high.[142]

Professor Burkhard Schafer suggested three differences between human and algorithmic error:

1. Transparency: most of our legal and ethical frameworks depend on the ability of a human to give an explanation if their decision was wrong and is challenged. This ability to explain helps determine whether the operator should apologise, and whether they are negligent

138. Interview with Jacob Lant, 18th January 2018
139. Interview with John Fox, 17th January 2018
140. Allen, M., and Pierce, O, (2016) "Medical Errors Are No. 3 Cause Of U.S Deaths, Researchers Say", NPR, available at https://www.npr.org/sections/health-shots/2016/05/03/476636183/death-certificates-undercount-toll-of-medical-errors
141. Laurance, J (2015) "Doctors' basic errors are killing 1000 patients a month", The Independent, available at http://www.independent.co.uk/life-style/health-and-families/health-news/doctors-basic-errors-are-killing-1000-patients-a-month-7939674.html
142. Hook, L., (2017) "For driverless cars, how safe is safe enough?", FT, available at https://www.ft.com/content/70924ace-cf0d-11e7-b781-794ce08b24dc

or malicious. This is an essential part of taking those who have been harmed by an erroneous decision seriously and respecting them, and it is unclear how a parallel process would operate with algorithms.

2. Perception: because we can empathise with humans, we know that they make mistakes and are biased. We understand algorithms less well, and as a result understand their 'mistakes' less well.

3. Unequal impact: some algorithmic errors can systematically burden certain groups over and above others, without this problem necessarily being visible if we look only at the overall performance. Thus, it is not sufficient to say that an algorithm is 'safer' or 'makes fewer mistakes' than human counterparts overall; we need to take a granular look at the effects on specific subgroups - what should we do e.g. if we found an algorithm was better in detecting a genetic illness across the entire population than a human doctor, but nonetheless made significantly more errors in the diagnosis when the patient comes from an ethnic minority (which may have been underrepresented in the training samples).[143]

Professor Enrico Coiera, Director of the Centre for Health Informatics at the Australian Institute of Health Innovation and Professor of Medical Informatics at Macquarie University, reminded us that "the technologies that we use today, medical records etcetera, are often designed poorly, they can impede workflow, and they may generate errors that harm and sometimes kill patients." He said this in the context of current health information technology (IT) systems, which are largely "passive" - there is definitely a potential for harm with more 'active' tools that are supporting decision-making, or eventually taking decisions autonomously. This is all the more so when considering the phenomenon of 'automation bias', where incorrect machine-triggered guidance is prioritised and followed by humans, leading to potential harm.[144] He and his colleagues have in fact developed a framework for classifying harmful outcomes related to technology.[145,146] It is foreseeable that such frameworks will need to be updated to take into consideration the ramifications of the use of AI and related technologies.

With respect to liability for error, Dr Dominic King, Clinical Lead at DeepMind Health and Honorary Clinical Lecturer in Surgery at Imperial College London, pointed out that:

> "When it comes to liability and compensation, these will be critically important issues if ever artificial intelligence technology were to replace the expert opinion of a medical professional. However, at the moment it is important to note that the efforts in AI that are currently most likely to lead to use in clinical practice – such as using deep learning to analyse and classify medical images like eye scans much more efficiently than current techniques allow – will augment, not replace, a clinical expert's judgement. Final responsibility for diagnosis and treatment would continue to rest with the clinician, as it would with any healthcare process involving an assistive technology. But DeepMind and other organisations would of course be responsible for the safe and effective functioning of our contributing technologies."[147]

143. Interview with Burkhard Schafer, 16th January 2018
144. Interview with Enrico Coiera, 23rd January 2018
145. Magrabi, F. et al, "An analysis of computer-related patient safety incidents to inform the development of a classification", J Am Med Inform Assoc. 2010; 17(6): 663–670. doi: 10.1136/jamia.2009.002444
146. Kim, M. O., Coiera, E. and Magrabi, F., "Problems with health information technology and their effects on care delivery and patient outcomes: a systematic review", J Am Med Inform Assoc. 2017; 24(2):246–250. doi:10.1093/jamia/ocw154
147. Interview with Dominic King, 28th January 2018

Once again, this situation will be challenged by algorithms that function more autonomously, rather than as an assistive technology, although many of our contributors argued that the responsibility should lie firmly with the algorithms' manufacturers in the case of harmful outcomes.

06 What do patients and members of the public want from AI and related technologies?

- How do patients and members of the public think these technologies should be used in health and medical research?
- How comfortable are patients and members of the public with sharing their medical data to develop these technologies?
- How do patients and other members of the public differ in their thinking on these issues?
- What is the best way to speak to patients and members of the public about these technologies?

There is a risk that these technologies are developed without the input of patients and those who use them - that is, the people who will be most impacted by these technologies. We came across many examples of good practice, such as Rhema Vaithianathan's work in the US where predictive risk modelling tools were developed in conjunction with the communities they were going to affect. DeepMind Health has addressed the ruling of the Information Commissioner's Office, which found that the Royal Free NHS Foundation Trust failed to comply with the Data Protection Act, and its own failure to incorporate patients and the public in its work.[148,149] It has since expressed its commitment "to meaningful public and patient involvement", manifested in regular meetings with patients and participants in studies.[150,151,152]

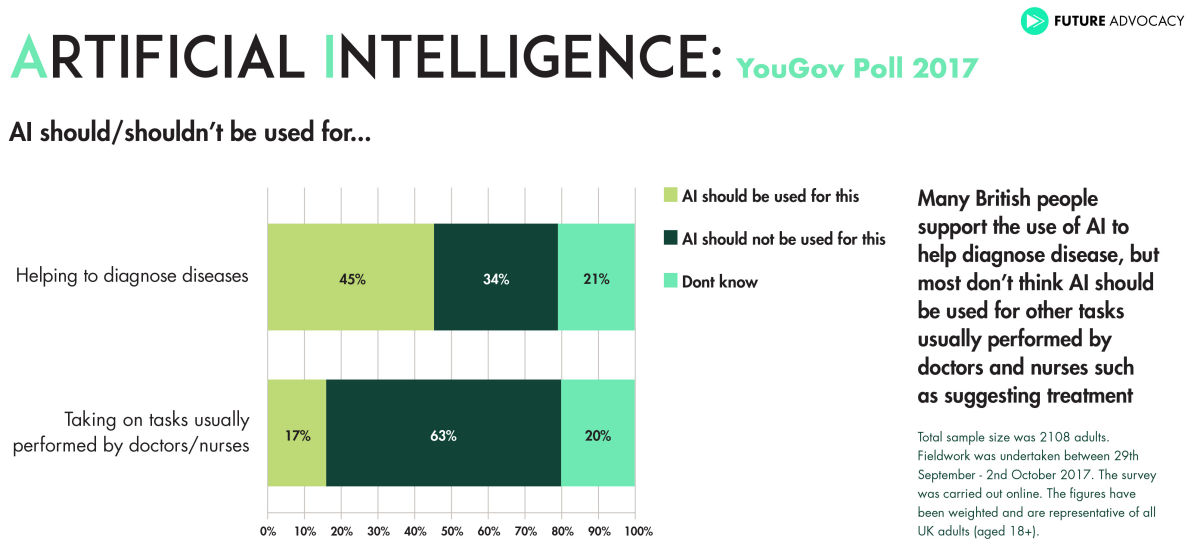
Nevertheless, our work leads us to conclude that not enough is known about what patients and members of the public want from these tools in healthcare, or indeed how this may change if the use of these tools becomes more widespread. It is also unclear how the needs and concerns of these two groups (people who identify as 'patients', and other members of the public) differ. We and others have attempted to begin addressing this gap in knowledge.[153] We asked YouGov to conduct a survey, in which 2108 adults (a weighted sample representative of all UK adults aged over 18) were asked for their perspectives on the use of AI in health and the use of data to develop healthcare algorithms (Figure 5). Respondents made a clear distinction between the use of AI in diagnosis of disease (where 45% said AI should be used for this) and in other tasks normally performed by doctors and nurses, such as answering medical questions, and suggesting treatments (where only 21% thought AI should be used for this, while 63% said AI should not be used for this). With respect to the use of healthcare data, a relative majority (49%) said that they would not be comfortable for their medical data to be used to develop algorithms that could improve healthcare, but it is noticeable that a significant proportion (40%) were comfortable with this, even after it was explained that data security could not be 100% guaranteed.

In contrast to the relative apprehension our poll's respondents expressed with regards to sharing medical data, a majority of the patients and members of the public at our roundtable underlined their eagerness to share medical data.[154] Of those who were concerned about the use and sharing of their medical data, a common theme was the fear that their data could be used by insurance companies to deny them or their children insurance, or to raise premiums. On the other hand, Elaine Manna, who is a patient at

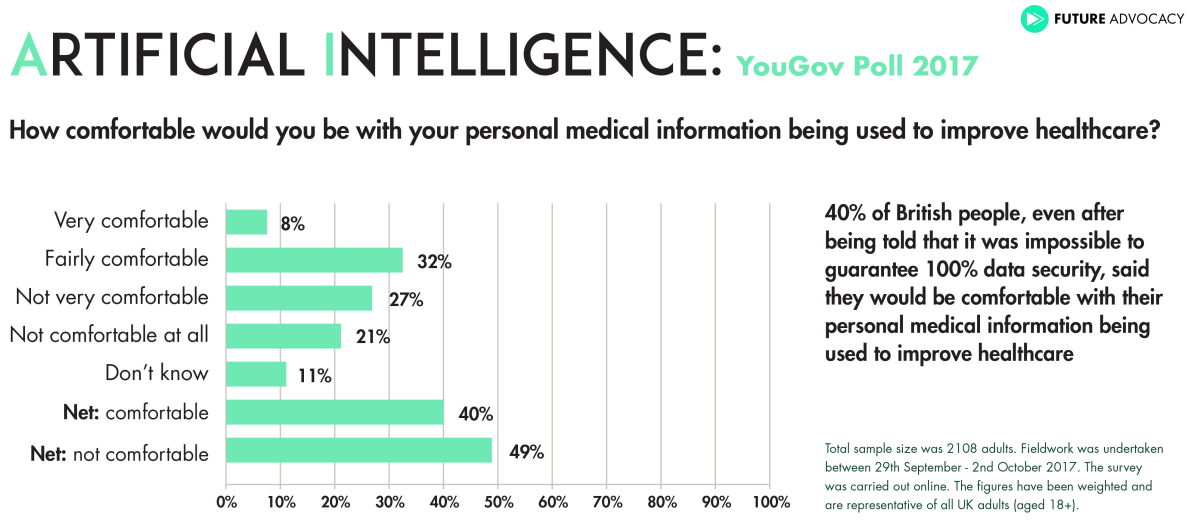
148. Information Commissioner's Office (2017), "Royal Free - Google DeepMind trial failed to comply with data protection law", available at https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2017/07/royal-free-google-deepmind-trial-failed-to-comply-with-data-protection-law/
149. Suleyman, M. & King, D. (2017), "The Information Commissioner, the Royal Free, and what we've learned", available at https://deepmind.com/blog/ico-royal-free/
150. Interview with Rhema Vaithianathan, 19th February 2018
151. Interview with Dominic King, 28th January 2018
152. Interview with Pearse Keane, 7th February 2018
153. Ipsos MORI (2017) "Public views of Machine Learning: Findings from public research and engagement conducted on behalf of the Royal Society", available at https://royalsociety.org/~/media/policy/projects/machine-learning/publications/public-views-of-machine-learning-ipsos-mori.pdf
154. Please see Appendix F ('Methodology by which patients/public contributors were recruited', page 60) for more details on how patients/members of the public came to participate in our roundtable discussion. Briefly, this was a small self-selecting group of people who are highly interested in issues related to the benefits and risks of medical data sharing. We make no claim that their views are representative of the views of all patients, who of course do not constitute a homogenous group.

Moorfields Eye Hospital in London and participant of the DeepMind collaboration outlined earlier (see 'Introduction'), described her willingness be named in this study in order to be able to tell her story and for people to relate to her experience. She believes that such 'humanisation' of AI applications, through the telling of stories of the people involved in the development of an algorithm, could be a powerful way to fostering the acceptability of data sharing and the use of AI in health and medical research.

a)



b)



▲ **Figure 5:** Results of our survey of public attitudes towards artificial intelligence in health (conducted by YouGov). A weighted, representative sample of UK adults was asked about (a) the use of AI in various health applications, and (b) how comfortable they are with their data being used to develop medical algorithms.

What our participants agreed on was the need for better education. For example, they were keen that discussions around data sharing make the benefits - to individual members of the public, their relatives, and to society as a whole - much clearer. Patients who have joined the 100,000 Genomes Project, for example, are willing to share their data even if they will not personally benefit from the results of the study. Many of our interviewees also echoed this point. Rhema Vaithianathan, for example, found as part of her DFP work on data use in New Zealand, that service users weighed up 'value' and 'comfort with sharing data': if the perceived value to them and their community was high, then they would be more comfortable with sharing their data, even if the perceived trustworthiness of the organisation was low. On the other hand, if the perceived value is low, then the organisation needed to meet a much higher bar of trustworthiness in order for service users to be comfortable with sharing their data.[155]

The need for better education and information around data sharing was put in the context of previous failed initiatives, such as care.data. This programme, launched in 2013, aimed to bring together patients' medical records from primary and hospital care into a single unified database, providing a tool that many thought would be invaluable for medical research.[156] However, a failure to address concerns by patient groups, privacy campaigners, politicians and clinicians - including, amongst others, around whether and how this data would be shared with private sector organisations - led to the programme being initially delayed, and eventually closed altogether in 2016. A failure to clearly outline the benefits of data sharing in this regard has been highlighted as one of the reasons for the collapse of this initiative.[157,158] There was no general agreement amongst our contributors about who should be responsible for educating the public on the benefits of data sharing. In the UK, there is a lot expected from the NHS in this regard. However, the example of care.data, as well as the feeling that different parts of the NHS are poor at communicating with each other (for example, primary care and secondary care, or hospitals and mental health services), meant that many were not sure that the NHS could undertake this task effectively.

Another issue raised by the patients and members of the public we spoke to was the language used to discuss these issues. For example, contributors questioned whether the term 'data' was helpful or not, and whether alternatives such as 'personal health information' would be better understood and have fewer negative connotations.

Lastly, this discussion needs to be put in the context of more general trends in healthcare. In many countries (perhaps predominantly in richer economies), there is a drive to increase patient empowerment and autonomy, which patients and members of the public contributing to our research strongly agreed with. Addressing how people can feel left out in decision-making around their own health is a long-standing issue. Professor John Fox, Professor of Engineering Science at Oxford University and Chairman and co-founder of OpenClinical, told us that he and his colleagues were actively considering these issues back in 1985, as they developed and trialled a variety of simple AI systems using a cognitive 'human like' approach. He stressed that teaching clinicians and members of the public more mathematics and statistics in order to be able to use these tools better is not a viable solution, particularly as AI tools become increasingly complex.[159] It would appear to us, therefore, that better ways of communication between those who make the tools and those who use them require further development.

155. Interview with Rhema Vaithianathan, 19th February 2018
156. NHS England News (2013) "NHS England sets out the next steps of public awareness about care.data", available at https://www.england.nhs.uk/2013/10/care-data/
157. Triggle, N. (2014) "Care.data: How did it go so wrong?", BBC News, available at http://www.bbc.co.uk/news/health-26259101
158. Evenstad, L., (2016) "NHS England scraps controversial Care.data programme", Computer Weekly, available at http://www.computerweekly.com/news/450299728/Caldicott-review-recommends-eight-point-consent-model-for-patient-data-sharing
159. Interview with John Fox, 17th January 2018

07 How should these technologies be regulated?

- Are current regulatory frameworks fit for purpose?
- What does 'duty of care' mean when applied to those who are developing algorithms for use in healthcare and medical research?
- How should existing health regulators interact with AI regulators that may be established?
- How should we regulate online learning, dynamic systems, as opposed to fixed algorithms?

Healthcare and medical research are highly regulated in many countries. Many of our interviewees discussed how AI and related technologies will be regulated to ensure patient safety. Many felt that existing regulatory frameworks, including the need to submit new drugs and devices for clinical trials, should be extended to include algorithms. However, many spoke of the need to develop new regulatory frameworks to deal with AI algorithms. John Wilkinson, Director of Devices at the MHRA, told us that "regulating software is a step into an area, particularly the AI-related stuff, for which the tools aren't well developed."[160] The US Food and Drug Administration (FDA) is currently creating a regulatory framework for software that aids healthcare providers in diagnosing and treating diseases and conditions.[161] Bakul Patel, Associate Center Director for Digital Health at the FDA, confirmed the Administration is looking into ways for the regulatory oversight of AI to allow maximum benefit to be derived from these technologies while maintaining confidence when used in healthcare. The main anticipated challenges include understanding how recommendations generated by AI algorithms differ from those made in current medical practice, how to deal with dynamic learning systems, and issues around algorithmic transparency.[162] Whether new frameworks need to be developed, or existing ones adapted, it is undeniable that the pace of development of these algorithms is much faster than regulators used to dealing with drugs and medical devices are used to, meaning that new or amended regulatory processes need to be agile and flexible to account for this speed.

It was suggested to us that the concepts of patient safety aren't firmly entrenched in the tech industry, and indeed may come into conflict with the tendency of tech entrepreneurs to want to 'move fast and break things'.[163] It is unclear how traditional notions of 'duty of care', held by healthcare practitioners and upheld by their own regulatory bodies such as the GMC and the NMC in the UK, apply to software developers and those purchasing software tools on behalf of a healthcare system, for example.

One of the interesting discussions around regulation that came up frequently in our interviews centred around the use of 'fixed' as opposed to 'dynamic' algorithms in health and medical research. 'Fixed' algorithms are those that regulators, clinicians and other users are perhaps more used to, as they do not change over time, whereas newer technologies could allow the use of 'dynamic' algorithms that 'learn online' - that is, algorithms that in the course of normal operation, use new data that is presented to them to improve their ability to reach their preset goal (such as making a prediction).[164] Some argued that it is easier to regulate fixed algorithms as compared to dynamic ones, with some going as far as saying that dynamic regulations should not be used at all for healthcare applications. However, others disagreed. Dr Joanna Bryson,

160. Interview with John Wilkinson, 15th January 2018
161. "Statement from FDA Commissioner Scott Gottlieb, M.D., on advancing new digital health policies to encourage innovation, bring efficiency and modernization to regulation" (2017) FDA Statement, available at https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm587890.htm
162. Interview with Bakul Patel, 6th April 2018
163. "Move fast and break things" was Facebook's unofficial motto, attributed (probably apocryphally) to Mark Zuckerberg himself. It was even written on their initial IPO paperwork. More recently, Facebook has publicly moved away from this position, preferring to "move fast and fix things" and "move fast with stable infra". See https://www.cnet.com/news/zuckerberg-move-fast-and-break-things-isnt-how-we-operate-anymore/ and https://mashable.com/2014/04/30/facebooks-new-mantra-move-fast-with-stability/
164. Bottou, L. (1998) "Online Algorithms and Stochastic Approximations", in Online Learning and Neural Networks. Cambridge University Press, Cambridge, UK

Reader in Computer Science at the University of Bath and Affiliate of the Centre for Information Technology Policy, Princeton University, told us that the idea that learning changes regulatory issues is "false". She drew comparisons with human doctors, who learn all the time, and with the process of auditing: "when you audit a company, you audit accounts, not the accountant's brain structure". Thus, a process of continuous certification for AI (parallel to regulated continuous professional development and licensing for doctors and nurses) that focuses on the outputs and outcomes of these algorithms could potentially be developed to cater for online learning algorithms. Facebook, for example, already use a process of 'continuous release' to update their website and apps, which relies on separate processes continuously monitoring these systems to ensure they are functioning as expected.[165]

08 Just because these technologies could enable access to new information, should we always use it?

- What would the impact of ever-greater precision in predicting health outcomes be on patients and healthcare practitioners?
- What are the implications of algorithmic profiling in the context of healthcare?

At the population level, it would seem that more information is desirable, particularly if it allows new insights that would otherwise be inaccessible to health planners and health practitioners. Brent Mittelstadt, for example, highlighted how social media and other forms of big data could potentially be used for public health surveillance.[166] It is worth highlighting that previous attempts at this have failed to meet expectations, such as the case with Google Flu Trends in 2013, but there is still active interest in this potential use of AI.[167] At the individual level, however, fears have been raised of the social harms that may follow from improved predictions of people's outcomes. We have already discussed the use of algorithms to predict mortality, for example, and Enrico Coiera asked whether it is ethical to use such algorithms to make a decision to move from active care to palliative care in individuals who are flagged up as likely to die, and whether this would be acceptable to the patient him or herself.[168] It is clear that if 'algorithmic predictions' come to be equated with 'destiny', then this could have a hugely disempowering effect on patients. In many ways, this is similar to the ongoing discussions around the value of insights from genomics in individual people's health. Having one gene or another may not necessarily give any more information about a particular clinical condition, so will patients and their relatives want to know about it? In the case of algorithmic predictions, a perception of futility and diminishment of hope may have a negative impact, where for instance an individual fears that they may not have access to certain interventions. Furthermore, not everyone would like to discover that they are at high risk, especially if the treatment or cure options are limited.

This issue of 'profiling by algorithm' is not unique to healthcare. Luciano Floridi and colleagues have discussed how "algorithmic activities, like profiling, reontologise the world by understanding and conceptualising it in new, unexpected ways, and triggering and motivating actions based on the insights it generates."[169] Algorithms can therefore create new categories and subgroups within existing populations, and assign people to these groups, leading to inferences and choices being made about them, possibly without their knowledge. Increased reliance on AI algorithms may therefore accelerate a process that is already apparent within modern medicine, which is the classification of

165. Interview with Joanna Bryson, 7th February 2018
166. Interview with Brent Mittelstadt, 1st February 2018
167. Lazer, D. and Kennedy, R. (2015) "What we can learn from the epic failure of Google Flu Trends", Wired, available at https://www.wired.com/2015/10/can-learn-epic-failure-google-flu-trends/
168. Interview with Enrico Coiera, 23rd January 2018
169. Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S. and Floridi, F. "The ethics of algorithms: Mapping the debate", Big Data and Society 2016;3(2). doi:10.1177/2053951716679679

individuals into 'risk categories' and suggesting interventional templates based on membership of this category. A controversial example of this process is the creation of the term 'prediabetes' to describe individuals who are at risk of developing type 2 diabetes mellitus, based on criteria such as results of blood sugar tests.[170] Proponents of the use of this categorisation claim that it allows early intervention in those most at risk to prevent the development of a condition associated with significant morbidity and early mortality; opponents argue that creation of this category risks 'overmedicalising' otherwise healthy people, causing anxiety, and making them a target for marketing by pharmaceutical companies.[171] In addition to this risk of overmedicalisation, attributing individuals to certain ategories of disease can be stigmatising; for instance, the label 'prediabetes' may suggest the person leads an unhealthy lifestyle.

It appears unclear where the balance between fully making use of the insights that AI could give us, and mitigating the risks of categorisation of societies and communities, should be struck. An interesting line of inquiry may be to explore how the concept of solidarity, as discussed earlier in this report, might provide a remedy for the increasing individualisation and atomisation of cohorts of patients, by reinforcing the importance of the collective 'good'.[172]

09 What makes algorithms, and the entities that create them, trustworthy?

It was extremely interesting to hear different answers to this question from our various interviewees and participants in our patient and public roundtable. In the case of the latter, the patients we spoke to suggested that their ability to trust an algorithm depended on the answers to certain questions, such as:

- What is the AI's success rate?
- Where does the AI come from, who developed it?
- What kind of data was the AI trained on? If I am a member of a minority group, will the AI work well for me?

With respect to the second question, it appears that branding is a powerful means for gaining the user's trust. Our entirely UK-based group of patient and public participants all agreed that having an app branded with the the NHS logo would go a long way to increase trust, as it would suggest to them that some kind of overview of the process by which the tool was developed has occurred. These participants did stress, however, that however the product is branded, they would require assurance that it has undergone rigorous testing and complies with strict regulation.

Debra Mathews and Travis Rieder of the Johns Hopkins Berman Institute of Bioethics drew a distinction between 'trustedness' and 'trustworthiness'. The former refers to whether, in a one-on-one interaction between a human and a tool, the human will trust it. In this case, studies have shown that the ability to form a relationship with the tool allows us to trust it, and in this regard, anthropomorphising is a major factor.[173] This is certainly a feature that roboticists keep in mind when designing robotic tools, such as those used in the care of the elderly and other vulnerable people (see Introduction, above).

On the other hand, when considering whole systems, the concept of trustworthiness comes into play. In this regard, Mathews and Rieder argued that the more transparent the system is, the more we will trust it. This point was also

170. Buysschaert, M., Medina, J.L., Buysschaert, B., and Bergman, M. (2016) "Definitions (and Current Controversies) of Diabetes and Prediabetes." Curr Diabetes Rev 12(1):8-13.
171. LaMattina, J. (2016) "Is Prediabetes An Epidemic Or A Creation Of Drug Companies?" Forbes, available at https://www.forbes.com/sites/johnlamattina/2016/08/04/is-prediabetes-an-epidemic-or-a-creation-of-drug-companies/#2138a6979c28
172. Prainsack, B., and Buyx, A. (2011) "Solidarity: reflections on an emerging concept in bioethics", Nuffield Council on Bioethics, available at http://nuffieldbioethics.org/wp-content/uploads/2014/07/Solidarity_report_FINAL.pdf
173. Interview with Debra Mathews and Travis Rieder, 26th February 2018

made by Dr Julian Huppert, Director of the Intellectual Forum at Jesus College, Cambridge and Chair of DeepMind Health's Independent Review Panel. He agreed that the more open a system is, including in terms of the development, commissioning and procurement of algorithmic tools, then the more protected users will feel from the risks of 'capture' by a particular organisation or body.[174]

10 What are the implications of collaboration between public and private sector organisations in the development of these tools?

- What are the most ethical ways to collaborate?
- How do we ensure value for both the public sector and for the private sector organisation, for example in the use of data? In publicly-owned/taxpayer-funded healthcare systems, such as the UK NHS, how do we ensure that citizens receive value too?
- What are the implications of the concentration of intellectual capacity in private sector organisations?

In many healthcare systems, health and research institutions such as hospitals and universities are partnering with private sector organisations to deliver technological solutions. We have given various examples of this model throughout this report, including the collaboration between Moorfields Eye Hospital and DeepMind Health, and between IBM Watson and Alder Hey Children's Hospital. There is no doubt that positive outcomes have flowed from such collaborations. Many interviewees emphasised that the expertise and funds necessary to develop such tools reside almost exclusively in the private sector, meaning that if it wasn't for these collaborations, these tools may not be developed at all. Nevertheless, many concerns regarding these collaborative arrangements remain. We have already referred to the question of how to source value for patients and citizens from publicly-owned data being used to develop these algorithms (see Question 2, above). The situation is complicated by the fact that there is no universally-accepted methodology for quantifying 'value' in these contexts. Other contributors mentioned the risk for technologists to "fall in love with the solution, rather than with the problem". This aphorism refers to the tendency for technologists to develop and advance technological solutions, without having a clear idea of the clinical or other problem that is being addressed by these technologies. Interviewees emphasised the importance of having clinicians and patients involved from the earliest stages and throughout the development process.

One approach that may prove fruitful could be to examine the differing contributions of public sector, private sector and other organisations in terms of the incentives that drive them when developing these tools. Are they driven by a desire to improve the health of individuals or of populations? Is the primary motive a financial one, such as delivering cost savings to taxpayers or dividends to shareholders? Or are they keen to demonstrate success with using a particular type of AI in a novel way? Understanding the incentives that motivate the various actors, and how these relate to each other, will be crucial in mapping collaborations between these different organisations and determining whether the terms of collaboration are likely to be acceptable to patients, healthcare practitioners and the wider public.[175]

The issue of consent and accountability of both public and private sector organisations was also raised. Dr Sobia Hamid, Digital Service Development Officer at University Hospitals of Leicester NHS Trust and Founder & CEO of Data Insights Cambridge, mentioned the agree-

174. Interview with Julian Huppert, 8th February 2018
175. The issue of how value is derived for patients and the public, for example from their data, is one of the main issues discussed in 'AI in the UK: Ready, willing and able?', a report by the House of Lords Select Committee on Artificial Intelligence. This report was published after the research for our report had been completed, but we recommend it to the reader, particularly Chapter 7 on 'Healthcare and artificial intelligence'. The report can be found at https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf

ment struck between IBM Watson Health and the Italian Government in 2016.[176] Under the terms of this agreement, IBM gains access to the anonymised health data of 61 million Italian citizens, including prescriptions and electronic health data, without their explicit consent, in return for a $150 million investment in the building of a new research center in Milan for its Watson Health division. Not only is IBM gaining access to the data, but the Italian government is contributing €60 million of taxpayers' money to this research, with IBM alone having the rights to the results and the ability to license these to third parties.[177]

Julian Huppert spoke to us at length about the model adopted by DeepMind Health in setting up an Independent Review Panel. He told us that "[the Panel's] brief is to review [DeepMind Health] ... effectively in the public and patient interest, rather than to give them advice." In order to fulfil this function, Independent Reviewers have the ability to investigate and review any aspect of DeepMind Health's operations at any time, and have no confidentiality obligations to DeepMind, meaning they are not constrained in what they can say to the media. Moreover, the Panel is provided with a budget of £100,000 per year by DeepMind to be able to undertake these reviews meaningfully. The Independent Review Panel produces an annual report, a public document of which DeepMind Health is given 5 days' notice.[178] This arrangement appears unprecedented in the context of technology companies developing tools for the public good, and may well provide a model for other private sector organisations to follow.

Finally, Dr Claudia Pagliari, a Senior Lecturer in Primary Care and Informatics and Director of Global eHealth at Edinburgh University, raised another issue that has become apparent with the increased involvement of private sector companies, namely that of 'brain drain'. It is becoming apparent that bodies such as universities cannot compete financially against technology companies, leading to experts such as data scientists being increasingly concentrated in the private sector.[179] This risks a creating a monopoly of intellectual capacity, the consequences of which on the development of these AI tools are unknown.

176. Interview with Sobia Hamid, 7th February 2018
177. Moody, G., (2017) "Detailed medical records of 61 million Italian citizens to be given to IBM for its "cognitivecomputing" system Watson" Privacy News Online, available at https://www.privateinternetaccess.com/blog/2017/05/detailed-medical-records-61-million-italian-citizens-given-ibm-cognitive-computing-system-watson/
178. Interview with Julian Huppert, 8th February 2018
179. Interview with Claudia Pagliari, 8th February 2018

CONCLUSION

It is clear that the use of artificial intelligence in health and medical research across the five use cases we have identified (process optimisation, preclinical research, clinical pathways, patient-facing applications and population-level applications) raises important ethical, social, and political challenges that require further research.

From an ethical perspective, a number of overarching themes have emerged. Firstly, the issue of **consent** runs through the entirety of this work. This is unsurprising given the crucial importance the concept of consent has in biomedical ethics, and its interaction with the central principle of personal autonomy. The challenges we outline with respect to human relationships in healthcare, the use of patient data, the consequences of a lack of algorithmic transparency, responsibility for error and the definition of trust all touch on consent in some way. A high-level question that can be asked is "how do we give meaningful consent to the use of AI to deliver services, where there may be an element of autonomy in the AI's decisions, or where we do not fully understand these decisions?"

Another major theme that the challenges we identify touch upon is that of **fairness**. This is particularly relevant to the issues we discuss around health inequality, what patients and the public want from these technologies, and of ensuring value to stakeholders throughout the processes of development and deployment of AI algorithms. Are the three general principles of distributive fairness (responsibilities, capabilities, and needs) a useful guide to help address these issues? These principles are highly open to interpretation, meaning that new approaches to fairness may need to be considered, particularly given the rapidly-changing nature of these technologies.[180]

Closely related to the concept of fairness is that of **rights**. This is clearly an area that will need to be considered when regulatory oversight of these technologies is being discussed and developed. Various international frameworks refer to a minimum standard of health to which all individuals are entitled, including Article 25 of the United Nations' Universal Declaration of Human Rights, Article 12 of the United Nations' International Covenant on Economic, Social and Cultural Rights, and the preamble of the World Health Organization's Constitution. With the addition of AI technologies to care pathways, or potentially, the increased reliance on aspects of care being delivered autonomously by AI, a new discourse around rights may emerge, asking "do people have a right to know how much AI is used in their care?" and "do people have a right not to have AI involved in their care at all?" At its core, this issue centres on whether the 'right to health' equates to a 'right to human delivery of healthcare'.

It is essential that research focusing on these ethical, social, and political challenges is multidisciplinary, drawing on the expertise of those who develop AI tools, those who will use and be impacted by these tools, and those who have knowledge and experience of addressing other major ethical, social, and political challenges in health. Most importantly, it is vital that the voices of patients and their relatives are heard, and that their needs - clinical, pastoral, spiritual and more - are kept in mind at all stages of such research. It is only by developing tools that address real-world patient and clinician needs and that tackle real-world patient and clinician challenges that the opportunities of artificial intelligence and related technologies can be maximised, while the risks are minimised.

180. Underdal, A., and Wei, T. (2015) "Distributive fairness: A mutual recognition approach" Environmental Science & Policy, 51:35-44. DOI: https://doi.org/10.1016/j.envsci.2015.03.009

APPENDICES

A: Glossary of Terms

Algorithm: A set or sequence of step-by-step operations that need to be carried out to perform a calculation, to process a set of data, or to test a logical statement.

Artificial intelligence (AI, or machine intelligence): A field of study that combines computer science, engineering and related disciplines to build machines capable of behaviour that would be said to require intelligence were it to be observed in humans. Such behaviour includes solving problems.

Automation: The use of automatic processes and equipment in manufacturing or other settings.

Big data: Large structured or unstructured datasets that are so complex that traditional data processing application software is inadequate to deal with them. The term has also been applied to the discipline of data analytics that has emerged to extract value from and identify patterns in these data.

Biomedicine: A branch of medical science that applies biological and physiological principles to clinical practice.

Black box: A device, system or object which can be viewed in terms of its inputs and outputs, without any knowledge of its internal workings.

Chatbot: A computer programme that conducts a conversation via auditory or textual methods.

Deep learning: A branch of machine learning that involves algorithms that analyse data through multiple layers of complex processing. Each layer's output becomes the input to the next layer to carry out pattern analysis and classification and to establish hierarchical relationships for both supervised and unsupervised learning.

Deep neural networks: A kind of deep-learning architecture based on artificial neural networks that uses multiple layers of processing units that loosely mimic human brain structure and can model complex nonlinear relationships.

Histopathology: A branch of pathology concerned with the tissue changes characteristic of disease.

Machine learning: A type of artificial intelligence that has risen to recent prominence. It refers to the ability of computers to learn without being explicitly programmed. Algorithms use complex statistical methods to recognize patterns in data, learn from these patterns, and subsequently make predictions based on these data. Various techniques allow the algorithm to continuously improve its pattern-finding and predictive abilities.

Machine vision: These AI technologies enable a computing device to inspect, evaluate and identify still or moving images, using automated image capturing, evaluation and processing capabilities.

Natural language processing: An area of computer science and artificial intelligence concerned with the analysis and synthesis of natural language and speech.

Neural networks: Artificial neural networks are an architecture of computing used in machine learning. Inspired by the organization and processing mechanisms of biological neural networks, artificial neural networks have been used in speech recognition, image recognition, and other areas involving machine learning.

Online machine learning: A method of machine learning in which data becomes available and is used to update the model continuously, such that the algorithm changes as it is used. It is used in situations where it is necessary for the algorithm to dynamically adapt to new patterns in the data, or when the data itself is generated as a function of time.

Parse: Breaking a data block into smaller chunks by following a set of rules, so that it can be more easily interpreted, managed, or transmitted by a computer.

Predictive analytics: The range of statistical techniques (including machine learning, predictive modelling and data mining) used to estimate, or 'predict', future outcomes. Also known as 'data analytics'.

Primary care: Healthcare that is delivered in the community, such as by general practitioners, district nurses and community midwives.

Robotics: The design, construction, operation, and application of robots.[181]

Secondary and tertiary care: Healthcare that is delivered in hospitals, usually organised around specialities and subspecialties such as surgery (with neurosurgery and cardiothoracic surgery as examples of subspecialties), medicine (with endocrinology and gastroenterology as subspecialties), and obstetrics, amongst many others.

Triage: The assignment of degrees of urgency to wounds or illnesses to decide the order of treatment.

181. British Automation & Robot Association, 'Definition of robots', available at http://www.bara.org.uk/definition-of- robots.html

B: List of abbreviations

AGI: artificial general intelligence

AI: artificial intelligence

AMD: age-related macular degeneration

ANI: artificial narrow intelligence

ASI: artificial superintelligence

CT: computerised tomography

EHR: electronic health record

EMS: emergency medical service

FDA: Food and Drug Administration (United States)

GDPR: General Data Protection Regulation

HCP: healthcare practitioner, e.g. nurses, therapists, doctors, and other professionals involved in direct patient care

IT: information technology

ML: machine learning (see Glossary above)

MIT: Massachusetts Institute of Technology

MRI: magnetic resonance imaging

NHS: National Health Service (United Kingdom)

NLP: natural language processing (see Glossary above)

OCT: optical coherence tomography

C: List of interviewees

- **Dr Natalie Banner**, Policy Adviser at Understanding Patient Data and the Wellcome Trust
- **Professor Margaret Boden**, Research Professor of Cognitive Science at the University of Sussex
- **Dr Ben Bray,** Clinical Fellow for Big Data at the Royal College of Physicians and Honorary Senior Clinical Lecturer at King's College London
- **Dr Joanna Bryson**, Associate Professor in the Department of Computing at the University of Bath
- **Simon Burall**, Senior Associate at Involve, Programme Director at Sciencewise and Co-Chair of the RSA Advisory Group on AI and Ethics
- **Sophie Castle-Clarke**, Senior fellow in Health Policy at the Nuffield Trust
- **Victoria Cetinkaya**, Senior Policy Officer at the Information Commissioner's Office
- **Hannah Chalmers**, Policy and Public Affairs Lead at National Voices
- **Andrew Chapman**, Sector Lead of Digital Health at Digital Catapult
- **Professor David Clifton**, Associate Professor in the Department of Engineering Science of the University of Oxford
- **Professor Enrico Coiera**, Director of the Centre for Health Informatics at the Australian Institute of Health Innovation and Professor of Medical Informatics at Macquarie University
- **Dr Genya Dana**, Head of Precision Medicine at the World Economic Forum
- **Dr Jeanette Dickson**, Vice-President for Clinical Oncology at the Royal College of Radiologists and Clinical Oncologist at Mount Vernon Cancer Centre (MVCC)
- **Virginia Dignum**, Associate Professor at the Faculty of Technology, Policy and Management, Delft University of Technology
- **Professor Murali Doraiswamy**, Professor of Psychiatry and Behavioral Sciences at Duke University
- **Dr Anat Elhalal**, Artificial Intelligence and Machine Learning Lead at Digital Catapult
- **Dr Ari Ercole**, Consultant in Neurointensive Care at Cambridge University Hospitals NHS Foundation Trust
- **Dr Jon Fistein**, Associate Professor of Clinical Informatics at the University of Leeds
- **Dr Tom Foley**, Senior Clinical Lead, Domain H: Data Outcomes at NHS Digital
- **Professor John Fox**, Professor at the Department of Engineering Science at the University of Oxford
- **Rose Gray,** Senior Policy Adviser at Cancer Research UK
- **Andreas Haimböck-Tichy**, Director of Healthcare and Life Sciences at IBM
- **Dr Sobia Hamid**, Founder of Data Insights Cambridge and Digital Service Development Officer at University Hospitals of Leicester NHS Trust
- **Dr Hugh Harvey,** Clinical Lead for Kheiron Medical and Royal College of Radiologists Informatics Committee Member

- **Jillian Hastings Ward**, Chair of the Participant Panel for the 100,000 Genomes Project
- **Professor Sabine Hauert**, Assistant Professor in Robotics at the University of Bristol
- **Eleonora Harwich**, Head of Digital and Technological Innovation at Reform
- **Mr Iain Hennessey**, Theme Lead for Paediatric Surgical Technologies at the National Institute for Health Research (NIHR), and Clinical Director of Innovation at Alder Hey Children's Hospital
- **Imogen Heywood**, Engagement Manager at the Centre for Information Sharing
- **Matthew Honeyman**, Policy Researcher at The King's Fund
- **Nigel Houlden**, Head of Technology Policy at the Information Commissioner's Office
- **Dr Julian Huppert**, Director of the Intellectual Forum at Jesus College, Cambridge and Chair of DeepMind Health's Independent Review Panel
- **Dr Mona Johnson**, Senior Clinical Lead, Doman A: Self-care & Prevention at NHS Digital
- **Professor Jeffrey Kahn**, Director of the Johns Hopkins Berman Institute of Bioethics
- **Dr Pearse Keane,** NIHR Clinician Scientist, Institute of Ophthalmology, UCL and Moorfields Eye Hospital NHS Foundation Trust
- **Dr Dominic King**, Clinical Lead at DeepMind Health and Honorary Clinical Lecturer in Surgery at Imperial College London
- **Jacob Lant,** Head of Policy and Public Affairs at Healthwatch England
- **Dr Geraint Lewis,** Chief Data Officer at NHS England and Honorary Clinical Senior Lecturer at University College London
- **Dr Harry Longman**, Founder and Chief Executive of GP Access
- **Maxine Mackintosh**, PhD candidate at University College London's Farr Institute of Health Informatics and Co-founder of One HealthTech
- **Professor Eduardo Magrani**, Professor of Law and Technology at Fundação Getulio Vargas Law School
- **Christopher Markou**, PhD candidate in the Faculty of Law at the University of Cambridge
- **Dr Ben Maruthappu**, Co-founder and CEO of Cera Care
- **Dr Debra Mathews**, Assistant Director for Science Programs for the Johns Hopkins Berman Institute of Bioethics, and Associate Professor in the Department of Pediatrics, Johns Hopkins University School of Medicine
- **Dr Brent Mittelstadt**, Research Fellow and British Academy Postdoctoral Fellow at the Oxford Internet Institute
- **Ben Moody**, Head of Health and Social Care at techUK
- **Dr Bertie Müller,** Senior Lecturer in Computing at the University of South Wales
- **Michaela Muruianu**, Innovation Co-ordinator at Digital Catapult

- **Dr Luke Oakden-Rayner**, Radiologist and PhD candidate with the School of Public Health at the University of Adelaide
- **Dr Claudia Pagliari,** Senior Lecturer in Primary Care and Informatics and Director of Global eHealth at the University of Edinburgh
- **Imogen Parker**, Head of Justice, Citizens and Digital Society Programmes at The Nuffield Foundation
- **Dr Ali Parsa**, Founder and CEO of Babylon Health
- **Bakul Patel**, Associate Center Director for Digital Health at the Food and Drug Administration (FDA)
- **Nicola Perrin**, Head of Understanding Patient Data
- **Carol Platt**, Innovation Associate at Alder Hey Children's Hospital
- **Professor Nasir Rajpoot**, Professor in Computational Pathology at the Department of Computer Science, University of Warwick
- **Professor Daniel Ray,** Director of Data at NHS Digital
- **Professor Geraint Rees**, Dean of the UCL Faculty of Life Sciences and Professor of Cognitive Neurology at University College London
- **Dr Travis Rieder**, Assistant Director for Education Initiatives, Director of the Master of Bioethics degree program and Research Scholar at the Berman Institute of Bioethics
- **Professor Renato Rocha Souza**, Professor at the Applied Mathematics School, Fundação Getulio Vargas
- **Professor Ferdinando Rodriguez y Baena**, Professor of Medical Robotics in the Department of Mechanical Engineering at Imperial College London
- **Dr Caroline Rubin**, Vice-President for Clinical Radiology at the Royal College of Radiologists and Consultant Radiologist at the University Hospital Southampton NHS Foundation Trust
- **Dr Benedict Rumbold**, Research Fellow in the Department of Philosophy at University College London
- **Professor Burkhard Schafer**, Professor of Computational Legal Theory at the University of Edinburgh's School of Law
- **Professor Stefan Schulz,** Professor of Medical Informatics at Medical University Graz, Austria
- **Allan Tucker**, Senior Lecturer of Computer Science at Brunel University
- **Professor Rhema Vaithianathan**, Co-Director of the Centre for Social Data Analytics at the University of Auckland
- **Jenny Westaway**, Head of the Office of the National Data Guardian
- **Hugh Whittall**, Director of the Nuffield Council on Bioethics
- **John Wilkinson**, Director of Devices at the Medicines and Healthcare products Regulatory Agency (MHRA)
- **Professor Stephen Wilkinson**, Professor of Bioethics

D: Patients and members of the public who contributed to this report

- Alex Brownrigg
- Mariana Campos
- Ann Cawley
- Annabel Dawson
- Ruth Day
- Eric Deeson
- Fran Husson
- Elaine Manna
- John Marsh
- Richard Melville Ballerand
- Dave McCormick
- Kath Pollock
- Bob Ruane
- Edward Sherley-Price
- Chris Warner
- Marney Williams

E: List of attendees at expert roundtable

- **Professor Richard Ashcroft**, Professor of Bioethics at Queen Mary University of London
- **Shirley Cramer CBE**, Chief Executive of the Royal Society for Public Health
- **Professor Bobbie Farsides**, Professor of Professor of Clinical and Biomedical Ethics at the University of Sussex
- **Professor John Fox**, Professor at the Department of Engineering Science at the University of Oxford
- **Professor Nina Hallowell**, Associate Professor at the Nuffield Department of Public Health, University of Oxford
- **Dr Hugh Harvey,** Clinical Lead for Kheiron Medical and Royal College of Radiologists Informatics Committee Member
- **Eleonora Harwich**, Head of Digital and Technological Innovation at Reform
- **Dr Geraint Lewis**, Chief Data Officer at NHS England and an Honorary Clinical Senior Lecturer at University College London
- **Maxine Mackintosh**, PhD candidate at University College London's Farr Institute of Health Informatics and co-founder of One HealthTech
- **Dr Benedict Rumbold**, Research Fellow in the Department of Philosophy at University College London
- **Professor Ilina Singh**, Professor of Neuroscience & Society at the Department of Psychiatry at the University of Oxford and Co-Director of the Wellcome Trust Centre for Ethics
- **Dr Nicola Strickland**, President of the Royal College of Radiologists and Consultant Radiologist at the Imperial College Healthcare NHS Trust
- **Professor Stephen Wilkinson**, Professor of Bioethics

F: Methodology by which patient/public contributors were recruited

Patients and members of the public that were interviewed or that participated in our roundtable on the 22nd February 2018 were recruited via one of two methods. Firstly, we placed an advert on the 'People in Research' website, which is run by the National Institute of Health Research (NIHR)'s INVOLVE programme, which aims to support active public involvement in the NHS, public health and social care research.[182] Secondly, we recruited from ongoing projects and bodies with established groups of patients and members of the public that are involved in their research. These included the 100,000 Genomes Project, Genetic Alliance UK, the Royal College of Physicians' Patient and Carer Network, and the British Heart Foundation's Patient Data Panel. Administrators for these bodies/ research kindly circulated the notice regarding our roundtable through their networks.

182. More information is available at http://www.invo.org.uk/

G: Scenarios illustrating ethical, social, and political challenges of AI in health and care

SCENARIO 01 A tool to check your skin moles at home

Mark is 43. He is a travel photographer and often visits hot countries. His variable schedule means he is often away from home during the week and so it is difficult for him to schedule GP appointments. He uses a handful of healthcare and monitoring apps to give him peace of mind and to save him the hassle of seeking a doctor in unknown locations. He looks out for apps with an NHS logo as he trusts these more.
Mark is well aware of the dangers of skin cancer, especially as he becomes older and as a result of his regular sun exposure. He uses 'SkinDeep', an imaging app which analyses photos of moles or lesions on the human body. All he needs to do is take a photo with his smartphone, and the app comes back with one of two results in a matter of minutes.

Result 1: "Your mole is normal, there is no need to seek further medical advice."
Result 2: "Your mole may be abnormal. Please seek further medical advice."

One day, Mark is showering and notices a mole on the left of his abdomen. He believes it to be new and it looks dark, so Mark takes a photo for analysis by 'SkinDeep'. 4 minutes later, he is surprised to get result number 2. He has been using the app for over a year and this is the first such outcome, so he takes it seriously. He calls his GP to schedule an appointment, but the earliest available appointment is in a fortnight's time. He is confident his careful self-monitoring means that even if the diagnosis is malignant, they will have caught it in its very early stages. Nevertheless, Mark feels unsettled right the way up to his appointment.

Two weeks later, Mark goes to see his GP, Dr Fontana, who after careful examination, determines the lesion does not look abnormal and would not recommend a referral. Despite the favourable conclusion, Mark is not completed satisfied and wishes to understand what it was about the mole that identified it as potentially worrying. Dr Fontana is unable to help him with this, so Mark contacts 'SkinDeep' through its chatbot, which assures him that the algorithm has been through rigorous testing and meets all regulatory requirements. However, Mark is unable to get clear answers, as he is told that despite its accuracy rate of 97%, the algorithm is not able to demonstrate the individual steps or reasoning it takes to reach a certain conclusion.

The following few weeks are very busy with work, so Mark doesn't think much of the mole until one evening when he notices some spots of blood on his shirt. He is very concerned to realise that his mole is bleeding. He immediately books an appointment with a private dermatologist, who sees him the following day, and on examination of the mole, says that it looks suspicious and needs to be removed.

SCENARIO 02 Taking part in a research project using AI

Lisa is 29. On Saturdays she plays netball with her work colleagues in a mini league. An awkward landing causes Lisa to sprain her ankle badly, resulting in a visit to the A&E, where it is decided that she needs to have an X-ray of her foot and some painkillers. While sitting in the waiting room, she is approached by a doctor and her research assistant.

Dr O'Hara explains to Lisa that she is developing an AI system that will analyse heart rhythms to identify a specific inherited problem with the heart that can cause it to beat irregularly. To develop this algorithm, she is asking patients visiting the hospital for unrelated reasons to have a heart tracing (ECG), which will then

be fed into the algorithm to see if this abnormal heart rhythm can be detected. The project is in collaboration with the well-known tech giant, Boggle, which is gathering and storing the data, and developing the algorithm. Lisa's participation will simply involve the research assistant attaching electrodes to her chest and the reading being entered into the dataset. Lisa's data will be anonymised, meaning that no one will be able to identify her from the data. Her father is terminally ill and she knows the value sharing data can bring, especially for research purposes, so she is happy to go ahead. Dr O'Hara also informs her that the results of the ECG will be mailed to her by post.

Following the ECG, Lisa's foot is bandaged, she is given firm instructions to rest up and is discharged from the hospital. The following week, she receives a letter generated by the AI system, thanking her for participating in the trial and confirming that her heart tracing did not show any abnormalities, meaning all is well with her heart. Lisa is pleased that despite her injury, she was able to contribute to medical research in some way.

One month later, Lisa receives a phone call from Dr O'Hara with the bad news that following a human review of her heart tracing, Lisa does in fact have the rare heart problem the study was looking for. The doctor asks Lisa to come in for a consultation in order to discuss the next steps, which will likely involve an operation. Lisa spends most of the time leading up to her consultation with Dr O'Hara feeling by turns angry, vulnerable and confused, as she is uncertain whether to be pleased that this potentially fatal heart rhythm has been diagnosed, or whether it would have been better not to know at all.

SCENARIO 03 **Companion pets for the elderly**

Jacintha is 83. Her husband passed away a few years ago. Her daughter lives an hour's drive away and her son lives abroad. Both have their own families. Jacintha has no physical impairments and has always been a sprightly woman. However, over the last few months she has started to experience symptoms of early dementia, which manifest mainly as short-term memory loss. Her children have made the decision to move her into an elderly care home, as they are afraid their mother may come to harm living alone with the illness.

At the Home, director Carl Biden has recently purchased a number of Puppionics, an animatronic dog, whose primary function is to offer companionship and stimulation to the Home's inhabitants, many of whom previously owned pets but are now unable to care for a cat or dog. The toy barks gently when stroked and even produces licking noises and gestures; its principal benefit is that it never runs away from the patient! Mr Biden and his staff are well- aware of the potential health benefits of pet ownership and have noticed that many of the patients seem to be just as happy with a Puppionic in their lap as when a member of the family or a care worker joins them for a chat.

Jacintha's grandchildren really enjoy playing with the toy together with her during their first few visits, but soon grow bored of the toy's limited capabilities. Jacintha's two children and their spouses feel disappointed and almost jealous of their mother's apparent preference for petting a Puppionic than talking to them. In the next months, Jacintha's symptoms deteriorate and her attachment to the toy dog increases. As a result, her family's visits dwindle.

The Home has applied for further government funding to invest in more sophisticated robots to help in the care of its patients with dementia. The plan for next year is to give patients like Jacintha their own animatronic pet to keep in their room, which through the use of sensors will be able to detect changes in mood or falls, and alert both staff and family accordingly. Although the addition of such devices will no doubt improve patient care, Mr Biden is concerned that the increased presence of robots may lead to a degree of dehumanization, especially as some families feel that their parent or grandparent responds much more to these devices than to them. Getting updates remotely may discourage family from visiting, as they may feel

close to their loved one without having to visit in person. However, with the number and demands of elderly patients ever-growing, Mr Biden feels he has no choice but to invest in these kinds of technologies.

SCENARIO 04 **The AI has missed my cultural background**

Luther is 32. He has a demanding job in the financial sector and has generally paid little attention to his health. Following weeks of intermittent nausea and vomiting, he finally schedules an appoint with his GP. Dr Suarez notices that Luther's skin seems jaundiced, and immediately refers him to a specialist.

The following day, Luther arrives at the hospital and is met by the specialist. Dr Cole is assisted by an AI system called Pan MD, which has been fed Luther's electronic health record, including the transcript from his recent appointment with Dr Suarez. It has already suggested a treatment plan based on this information and the diagnosis, which is stage 4 pancreatic cancer. It is now Dr Cole's job to break the news to Luther, inform him of the chosen treatment plan and admit him into the oncological ward, where the chemotherapy will begin right away.

Although Luther feels that Dr Cole is taking his time and offering a high level of empathy, he still feels overwhelmed and rushed with the decision to undergo chemotherapy. He imagines himself confined to a hospital bed, robbed of his hair and dignity. The AI has failed to account for Luther's cultural background, which has highly patriarchal traditions that have little tolerance for male weakness, and he knows family hospital visits would only bring discomfort to everyone involved. He asks Dr Cole for further treatment options, but the doctor assures him this is the best course. Luther does not explain the reasons for his reluctance, as he knows the doctor's background is vastly different to his and that he would not understand.

Arrangements are made for Luther to be admitted in two days' time, but he does not show up.

This project was funded by

www.futureadvocacy.org
@FutureAdvocacy

www.wellcome.ac.uk

APRIL 2018