BMA

Principles for Artificial Intelligence (AI) and its application in healthcare

British Medical Association
bma.org.uk

Contents

Executive summary 2

Introduction 3

1 AI usage in UK health services 4

1.1 Defining Artificial Intelligence 4

1.2 Examples of the application of Artificial Intelligence to healthcare 4

2 Impact of AI on healthcare – benefits and risks 7

2.1 How does AI have the potential to improve healthcare? 7

2.2 AI brings both benefits and risks 8

2.2.1 Health outcomes and patient experience 8

2.2.2 Clinicians' experience of work and system efficiency 11

2.2.3 The benefits and risks of AI are interrelated 13

2.3 The key issue is how AI is implemented 14

3 BMA principles for AI policy and implementation 19

Executive summary

AI in healthcare includes a range of applications aimed at enhancing efficiency, diagnosis, and treatment. Though AI lacks a single definition, it broadly refers to technologies simulating human intelligence to perform complex tasks, in the context of healthcare this includes: healthcare administration, clinical decision-making, improving diagnostics, personalised treatment, providing digital therapies, population health data analysis, and biomedical research.

AI holds promise for transforming healthcare by improving precision, efficiency, and preventive measures. It can enhance diagnostic accuracy, personalise treatments, and streamline administrative tasks, potentially reducing healthcare demand and improving outcomes. However, the success of AI depends on its implementation, including proper testing, integration into workflows, and addressing issues of liability, regulation, and data governance. Risks include potential harms to patient health, exacerbation of health inequalities, and impacts on doctor-patient relationships and productivity. Effective AI use requires careful management to maximise benefits and mitigate risks, ensuring it complements and enhances existing healthcare systems.

AI should be expected to transform, rather than replace, healthcare jobs by automating routine tasks and improving efficiency. This could enhance job quality and reduce burnout by minimising administrative burdens and allowing staff to focus on complex, value-added activities. However, the implementation of AI including effective deployment and long-term evaluations, must be managed carefully to avoid increasing workload and pressures.

Implementing AI in healthcare requires careful consideration beyond merely introducing new technologies. Key issues include ensuring AI tools are rigorously tested for safety and efficacy, avoiding reliance solely on lab-based evaluations. Effective integration into clinical practice is crucial, as poor real-world performance can arise from inadequate training or data quality. Governance and regulation must adapt to manage AI's evolving nature and protect patient safety. Involving staff and patients in development and implementation, along with continuous training, is essential. Moreover, robust IT infrastructure and clear legal liability frameworks are needed to support successful AI adoption in healthcare.

The BMA advocates, as set out in the principles at the end of this paper, for AI in healthcare to prioritise safety, efficacy, ethics, and equity. Each AI implementation must be rigorously assessed in real-world settings and continuously monitored to ensure it improves care quality and job satisfaction without exacerbating inequalities. Strong governance and up-to-date regulation are essential to protect patient safety. Involving staff and patients in AI development and providing them with the choice to opt out or dispute AI decisions is crucial. Training for healthcare professionals and robust IT infrastructure are necessary for effective AI integration. Legal liability must be clear, ensuring developers are accountable and doctors can challenge AI decisions.

Introduction

In recent years, policy makers and journalists have made much of the potential impact of Artificial Intelligence (AI) technologies to enhance and disrupt healthcare provision in the NHS.

The idea that AI might be useful for healthcare is not new. Jess Morley from the Oxford Internet Institute notes that interest in AI for healthcare '*dates to at least the 1940s, but has gained enthusiasm since the 1960s*'. New forms of health data, as well as rapid developments in the AI field in general (including large language models such as ChatGPT) and in the health sphere (notably applications of AI to radiology and imaging) mean that interest amongst developers and policy makers in healthcare AI is at an all-time high. The list of potential uses for AI are numerous and it has, at one point or another, been proposed that AI could theoretically help with almost any clinical task.

In 2019, then-NHS England chief executive Simon Stevens set out a call for the NHS in England to become a world leader in the use of AI. Since then, dedicated funding pots have been made available to foster and develop AI for use, and at the Spring Budget 2024 the Chancellor announced further funding of over £1 billion to invest in AI to improve the productivity of the NHS in England. NHS Scotland's Data Strategy for Health and Social Care (2023) and NHS Wales' Digital and Data Strategy for health and social care in Wales (2023) also explicitly mention AI as an important area of opportunity.[a]

Proponents of AI argue that it has the potential to integrate large amounts of clinical and epidemiological data to develop a more detailed understanding of the prognosis, diagnosis and treatment of diseases – this information can then be made available to health providers and clinicians to provide more effective care. It could also reduce the administrative burden of healthcare delivery by automating tasks. However, despite the potential benefits, there are significant limitations to many current technologies, which may be limited in robustness, have excessive computational requirements, or are biased in their findings. Furthermore, there are significant ethical, regulatory, and implementation challenges in integrating AI into existing workflows and clinical settings, which must be carefully considered before any AI interventions are introduced.

In fact, except for some limited success stories, real-world implementation of AI technologies in front-line healthcare – particularly UK health services - have been much more limited than the rhetoric suggests. It is not known how many doctors use AI in their day-to-day working lives, but existing studies suggest that awareness of AI is not widespread amongst doctors and medical students, and although there are increasing opportunities for doctors to become involved with AI, there are still limited structured pathways for doctors to engage with AI.

This briefing sets out how AI is being used in healthcare delivery currently or in the near future and examines both the potential benefits and drawbacks with respect to patients, clinicians, and the efficiency of the overall healthcare system. We conclude that AI in and of itself is only a tool – if implemented appropriately and effectively, it can improve outcomes; if not, it can worsen them. Therefore, we also set out some principles the BMA believe must be adhered to for AI to safely, effectively, ethically, and equitably improve healthcare outcomes.

[a] HSC NI Digital Strategy (2022 – 2030) does not explicitly mention Artificial Intelligence.

1. AI usage in UK health services

Although AI is a term that has entered common usage by academia, industry, and the media, there is no widely accepted single definition of 'artificial intelligence' and what it involves. For the purposes of this paper, we focus on AI technologies currently in use by UK health services or in other healthcare settings. Current use cases include diagnostics and clinical decision-making; healthcare administration; analysis of population health risks; and biomedical research. AI is currently most used in diagnostics, and while the many other potential use cases are numerous, they are not implemented widely yet.

1.1 Defining Artificial Intelligence

In general, AI refers to the simulation of human intelligence by machines. In its National AI Strategy, the UK government gives a working general definition of "*Machines that perform tasks normally performed by human intelligence, especially when the machines learn from data how to do those tasks*". Similarly, the NHS Transformation Directorate defines AI as "*the use of digital technology to create systems capable of performing tasks commonly thought to require human intelligence*".

However, 'tasks normally performed by human intelligence' is not well defined and may well change over time – for example, a pocket calculator can do simple mathematics that are even beyond the reach of many intelligent humans. But we do not tend to consider the average pocket calculator as AI. What we consider to be AI covers an ever-changing set of capabilities as new technologies are developed. In fact, it is an often-cited maxim amongst AI researchers and practitioners that as soon as machines have conquered a task that previously humans can do, it is no longer considered a mark of 'intelligence' or AI.

More confusingly, AI is often used to refer to both current technologies performing tasks commonly thought to require human intelligence, as well as hypothetical future technologies that could perhaps do anything humans can currently do – or exceed human cognition. If such a technology is ever created, AI would have the ability to adapt to new tasks for which they have not been trained – even to do anything a doctor or other healthcare professional could do. However, it is highly uncertain whether this will ever happen. This means some of the discussion around AI focusses on future technologies that may or may not be realised.

For the purposes of this paper, we focus on AI technologies currently in use by the NHS or in other healthcare settings, or technologies that have been developed and may be used soon. These include: machine learning (a statistical technique for fitting models to data and 'learn' by training models with data); natural language processing (including speech recognition, simulation of human conversation); and vision/image recognition.

1.2 Examples of the application of Artificial Intelligence to healthcare

AI can be categorised by solution type or use case. Key use cases currently in healthcare include:

- Healthcare organisation and administration
- Diagnosis and treatment (clinical decision making and service delivery)
- Population health and prevention
- Biomedical research

Further detail on each type of use case is set out in Figure A.

Figure A: Descriptions of AI use cases in healthcare

Category	Description	Examples
Healthcare organisation and administration	AI functions, such as natural language processing, are used to automate and optimise 'back-end' processes, including (1) staff rostering and appointment scheduling; (2) repetitive admin, such as note taking and transcribing during consultations, filing of an individual's electronic patient health record and patient letter writing, printing and posting; and (3) analysis of patient feedback to inform quality improvement.	In Hong Kong, the HK Health Authority is using an AI-based tool to produce monthly or weekly nursing staff rosters that satisfy a set of constraints, such as staff availability, preferences, working hours, ward operational requirements, and hospital regulations. It has been deployed across 40 public hospitals and is responsible for 4,000 staff schedules. At Imperial College Healthcare NHS Trust, a pilot tested the use of Natural Language Processing to analyse patient feedback in real-time, which led to responses to feedback being implemented more quickly than without the tool.
Clinical decision-making	AI pattern recognition of CT, MRI or ultrasound scans, for instance, and AI analysis of clinical data, genomic data, health records, personal and family histories, speech patterns and voice modulations, clinical guidelines, best practice and medical research are used to support, optimise, personalise and/or automate decisions about triage, diagnostics, prognosis and subsequent care pathways at the point of care via decision support systems.	Moorfields eye hospital have trialled the use of optical coherence tomography (a non-invasive diagnostic technique that renders a cross-sectional view of the retina) to pick up retinal diseases through AI tagging of 'urgent' cases in need of referral. IBM's Watson can parse millions of pages of medical literature in seconds and generate diagnostic insights based on a patient's symptoms. Ethos is a tool using AI to target radiotherapy in cancer treatment – currently in use at the Beatson West of Scotland Cancer Centre. Traditionally, clinicians must draw up and continuously adjust treatment plans as tumours and surrounding tissue typically change as the disease and treatment progresses; this tool aids clinicians to make these decisions more quickly and effectively.
Service delivery	AI based apps, bots, personal, wearable and smart devices are used to interact directly with patients to (entirely or mostly) deliver therapies, provide health information, and/or support patients to stick to prescribed interventions and manage health conditions.	Computerised Cognitive Behavioural therapy (CBT), for example, has a relatively long history in the NHS, but a new generation of digital therapies aims to deliver CBT at scale with better engagement. Sleepio is one example: a six-week tailored programme delivered online that is designed to treat insomnia. The therapy is personalised using AI that tailors the intervention to patient data.
Population health	Population-level applications of AI analysis using a large amount of new data forms for novel, real-time insights into (1) epidemics, disease spread and the drivers of illness; and (2) individuals/groups at risk of developing particular diseases that could gain from proactive, early intervention.	During the pandemic, NHSX launched the Covid-19 Data Store, bringing together data from several sources within the health and social care system as part of a project to use AI to build a predictive model to inform the government's response to Covid-19.
Bio-medical research	AI analysis applied to new forms of data, including genomic data and patient records, to inform new drug and treatment discoveries.	Recently, researchers at MIT discovered a new class of compounds capable of killing drug resistant bacteria by using AI screening of millions of potential chemical compounds to narrow down those with high predicted ability to kill bacteria and low toxicity to living tissue, leading to several hundred new compounds that were worth testing. Upon empirically testing those, several were capable of killing drug resistant bacteria.

It is important to note that currently AI products approved for use in the UK or comparable countries are usually for the purpose of relatively narrowly defined tasks, and tend to be deployed as some form of clinical support tool with the aim of improving the accuracy of diagnosis, tests or therapies; or improving the reliability of decision making.

A survey of AI technologies in the NHS in England from 2021 found the most common category of products in use (50%) were diagnostic solutions, most commonly used in radiology and cardiology – for example, classifying X-ray, MRI or CT scans.

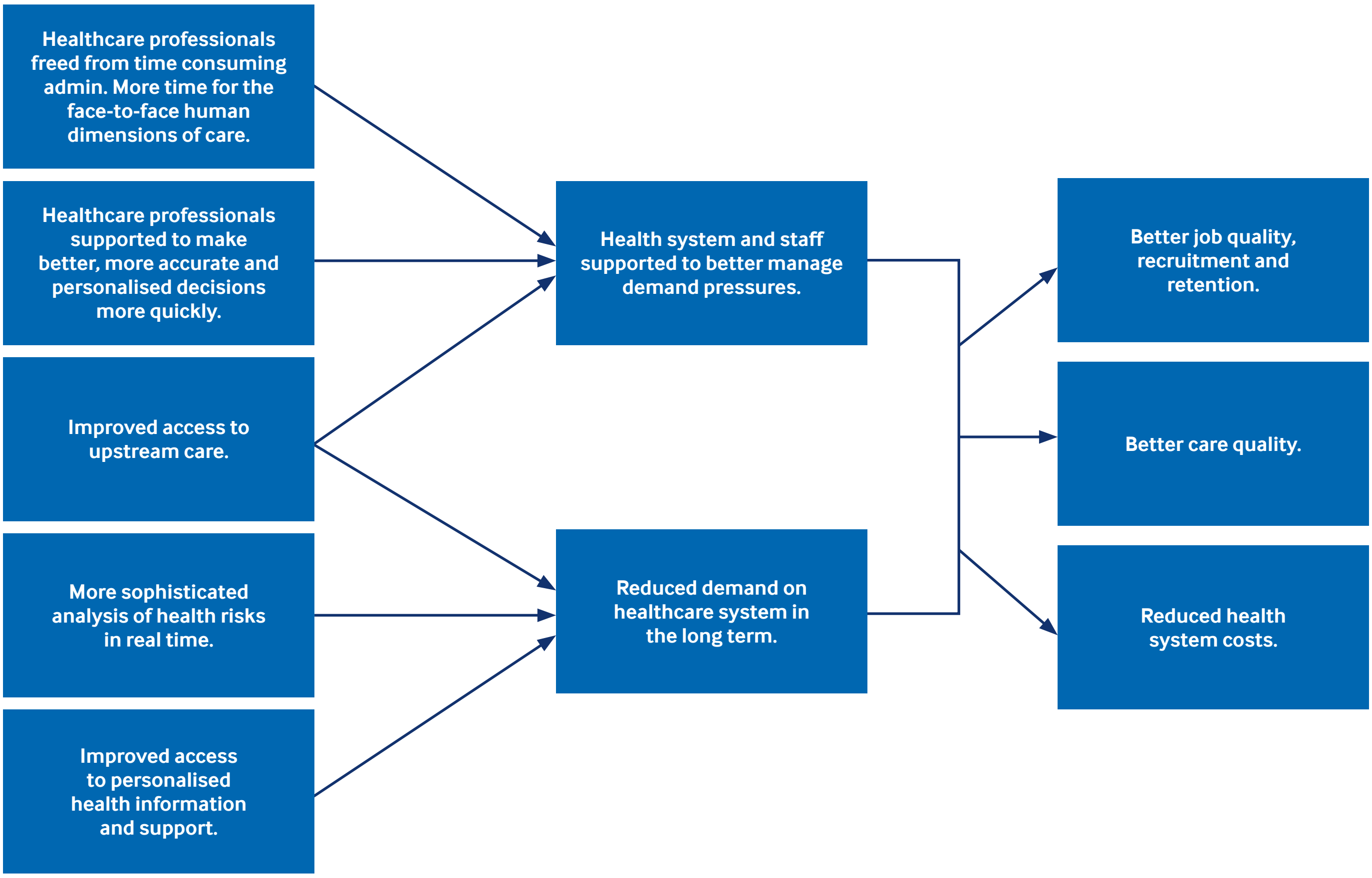
2. Impact of AI on healthcare – benefits and risks

AI has the potential to improve the way healthcare is delivered. However, its usage also comes with risks that must be proactively managed. In addition, for these benefits to be realised, AI must be implemented appropriately. This section sets out how AI can provide improvements to healthcare, and then discusses how the outcomes achieved could result in both benefits and harms, across patient health, health inequalities, doctor experience and the doctor-patient relationship, and productivity, resource allocation and cost. Whether AI leads to benefits or harms in large part depends on how it is implemented – how AI is tested and evaluated, how it is integrated into workflows and how people are trained in its usage, issues of liability and regulation, and IT and data governance.

2.1 How does AI have the potential to improve healthcare?

As AI systems become better at sorting data, finding patterns and making predictions, these technologies are likely to take on an expanded role in healthcare. Proponents of healthcare AI pinpoint a range of potential benefits to AI expansion in healthcare. It is often claimed that AI will reduce demand on healthcare, enable health professionals to better manage existing demand for care and, in turn, improve both healthcare outcomes and employment quality[b]. Figure B sets out these potential benefits.

Figure B: Potential healthcare AI outcomes map



Healthcare professionals freed from time consuming admin. More time for the face-to-face human dimensions of care.

Healthcare professionals supported to make better, more accurate and personalised decisions more quickly.

Improved access to upstream care.

More sophisticated analysis of health risks in real time.

Improved access to personalised health information and support.

Health system and staff supported to better manage demand pressures.

Reduced demand on healthcare system in the long term.

Better job quality, recruitment and retention.

Better care quality.

Reduced health system costs.

[b] For example, see https://topol.hee.nhs.uk/.

Outcomes are thought to be achieved through three primary mechanisms:

i) Precision and personalisation

AI can match or outperform humans at performing certain healthcare tasks with accuracy. One well-evidenced area is accuracy in imaging diagnostic processes.

AI may also be able to develop more detailed understanding of disease and more effective medications and treatments, by integrating and analysing large amounts of personal, social, clinical, genomic, and epidemiological data. This information can then be made available at the click of a button to healthcare providers at the point of care to support diagnosis and prognosis – including drug prescription. This holds the potential to make medical interventions more accurate and precise, personalised to the individual and based on the most up-to-date knowledge, guidance, and best practice.

ii) Efficiency and productivity

AI imaging and clinical decision making allow the possibility for quicker and more efficient healthcare interventions. There are also quality and efficiency gains to be made from applying AI technologies to more routine, everyday tasks such as dealing with letters or appointment scheduling. The NHS England long-term workforce plan cites McKinsey research arguing that routine administrative tasks can take up to 70% of a health practitioner's time.

iii) Prevention and early intervention

AI algorithms have the potential to examine factors such as population demographics, disease prevalence, and geographical distribution to identify patients or patient groups at higher risk of certain conditions.

If AI can provide such benefits as population health risk prediction, surveillance, a source of health information, and therapy, this can have positive implications for the efficacy of preventative medicine. Where upstream health and wellbeing support is prioritised, demand on healthcare can be mitigated, allowing resource to be deployed in other areas of need.

2.2 AI brings both benefits and risks

Whilst AI has the potential to produce many benefits for patients, clinicians, and healthcare systems, there are also significant limitations and risks to current AI technologies. As Jessica Morley puts it: '*the problem is that enthusiasm for all these potential benefits is currently masking the very real limitations, implementation challenges, ethical considerations, patient safety risks and regulatory hurdles that are – in the vast majority of cases – preventing many of the … theorised benefits from being realised'*. Potential impacts of the use of AI, positive and negative, relate to health outcomes and patient experience; clinician's experience of work; and healthcare system efficiencies.

2.2.1 Health outcomes and patient experience

AI could improve health outcomes or cause harms

The greater precision AI potentially offers in diagnosis and treatment could lead to better outcomes for patients, by reducing the risk of errors made and aiding clinical decision making. For instance, lab-based studies have shown that specificity and sensitivity – i.e. correctly identifying positive and negative results, increase when AI and the radiologist interpret the imaging, supporting AI to be a tool to enhance radiologists at diagnosis breast cancer. Studies have also reported the accuracy of AI in determining if lung nodules seen on CT scans are cancerous. Similar claims are made of AI retinal scanning and eye diseases. Incorporating AI into imaging diagnostic processes could, therefore, improve accuracy.

AI can also help aid clinicians in differential diagnoses. Tools are being developed with the aim of providing a list of potential diagnoses and their likelihoods to doctors based on patient health data and symptoms to assist in the process. This holds the potential to make medical interventions more accurate and precise, personalised to the individual and based on the most up-to-date knowledge, guidance and best practice.

It can potentially also speed up diagnosis and treatment, reducing risks to patients. For example, AI has been shown to aid analysis of brain tumours in cancer patients. To safely remove a brain tumour without damaging surrounding healthy brain tissue, neurosurgeons typically need to gather information once the patient is on an operating table. However, AI tools can help neurosurgeons understand what specific sub-type of tumour a patient has and where they must operate – before the patient is in the operating theatre – potentially reducing the time patients spend in an operation.

In another example, one AI model has been shown to accurately predict a woman's five year risk of getting endometrial cancer solely based on personal health data without the use of invasive tests.

However, although there are many potential ways AI can aid in diagnosis and treatment, it also has the very serious potential to cause harms. A review by the Panel for the Future of Science and Technology (STOA) describes three main patient safety risks from AI technologies in healthcare:

- False negatives in the form of missed diagnoses of disease.
- Unnecessary treatment due to false positives.
- Unsuitable interventions due to imprecise diagnoses, or incorrect prioritisation of interventions in emergency departments.

These risks are driven by a number of factors, including noise and artefacts in AI's clinical inputs and measurements, data shifts between AI training data and real-world data, and variation in clinical context and environments. Though it must be acknowledged such risks exist in healthcare without AI, the underlying reasons behind such harms could be unique to AI. Unsuitable interventions may be made due to a lack of capability in understanding the patient empathically and as a whole person. These issues are discussed further in section 2.3.

One of the primary concerns is the potential reduction in human interaction. Face-to-face communication with healthcare providers plays a significant role in patient care, particularly in mental health. These interactions offer empathy, understanding, and emotional support, which are difficult to replicate with AI. If AI tools and automation lead to fewer opportunities for personal interaction, patients might feel isolated, misunderstood, or neglected, potentially worsening their mental health, damaging the doctor-patient relationship, and their willingness to engage in future treatment.

AI can lead to discrimination against marginalised groups and widen inequalities, but could also improve access to healthcare

AI has the potential to improve access and reduce barriers to treatment. Research by Healthwatch, National Voices, and the Good Things Foundation all find that digital services can be more accessible for some, including those with caring responsibilities, those with reduced mobility, and those who are immunosuppressed or shielding. There is also some evidence that digital mental health services bypass some of the access barriers to traditional services, including stigma.

Societal inequalities and biases permeate all walks of life and healthcare is no exception. There is evidence that clinicians can think women exaggerate the pain they are experiencing and that those from ethnic minorities are able to withstand more pain, which can be detrimental to their treatment. AI could be developed in such a way to limit its exposure to such preconceptions and therefore help address inequalities in healthcare.

However, there is a growing body of evidence that AI and data driven health technologies can lead to discrimination against underserved or marginalised groups, exacerbating existing bias and systemic health and healthcare inequalities.

Firstly, the data that AI is trained on presents a problem. If the datasets used to train AI models are not representative, then the resulting output will be inequitable, whether in prediction, prevention, triage, diagnostics or prognostics, and decision support.

A key source of data for many AI models is based on electronic patient record systems (EPRs); however, this data often excludes groups that may not seek healthcare as frequently. A well-documented example of this is the way in which skin cancers and other skin lesions in patients from ethnic minorities are less well picked up by AI informed diagnostic tools, because these tools are trained on data from white patients. One study from 2021 examined the representation of ethnic minorities in images of skin lesions that were used to train AI systems. Researchers found that of 2,436 pictures where skin colour was stated, only 10 were of brown skin and only one was of dark brown or black skin. Among the 1,585 pictures with information on ethnicity, none were from people with African, Afro-Caribbean, or South Asian backgrounds.

Other examples of algorithmic bias can be seen in patients with neurodevelopmental disorders like autism and ADHD, as models often fail to account for the diverse needs of these groups due to their reliance on data from "average" patients. To address this, AI should target comorbid conditions, integrate physical and mental health information, expedite mental health service use, and support treatment adherence through social care integration.

Digitally mature hospitals with electronic health records also disproportionately serve more privileged segments of the population, risking over-representation. 10% of UK hospitals currently have no EPRs. Even when they do, case notes and health records can contain subjective judgements of health professionals, making them open to prejudiced interpretation.

Secondly, digital exclusion is a considerable cause for concern. As the Joseph Rowntree Foundation have said, "*AI is a seismic shift in the goalposts of what it means to be digitally included.*" Digital exclusion is about not having the access, skills, and confidence to use the internet and benefit fully from digital technology in everyday life. In 2023, Ofcom estimated that 7% of the UK population (nearly 5 million people) did not have home internet access and are therefore at risk of digital exclusion. Poverty is the most reliable indicator of internet access and use among adults, with those in the lowest socio-economic groups more than three times as likely as those in the highest to not use the internet. Those most likely to face digital exclusion, therefore, are likely to be at the sharp end of unequal health outcomes. Digital exclusion can also come from technology-illiteracy, which can be many patients who did not grow up with and learn to use technology or do not have the dexterity to use it regardless.

Population groups who have traditionally had worse experiences in healthcare due to structural racism, ableism, and/or misogyny, have as a result developed lower levels of trust in healthcare. This was borne out most clearly in the vaccine hesitancy shown in some ethnic minority communities during the Covid-19 vaccine rollout. The risk, therefore, is that such mistrust is replicated in AI technology and exacerbates health inequalities further (the groups most mistrusting of healthcare being the ones who have worse health outcomes and therefore need it the most).

AI could impact the doctor-patient relationship

AI has the potential to improve or damage the doctor-patient relationship. The greater precision that AI could potentially bring may improve outcomes for patients, which could heighten patients' trust in clinical interventions and further develop the doctor-patient relationship.

AI could free up clinicians' time by taking over certain bureaucratic activities, which would allow clinicians to spend more time with each patient and develop a positive relationship with each one. This, in turn, could improve communication and trust between patients and doctors, and give more opportunities for continuity of care. However, freed up time may instead be used for seeing a higher volume of patients rather than spending more time with each patient, depending on implementation.

Furthermore, even if productivity gains free up more time for face-to-face care, certain forms of healthcare automation through AI could result in the loss of effective communication, empathy and one-to-one communication which are critical elements of effective patient care. AI cannot replace the continuity of care possible with general practice.

If the use of AI forms a barrier between a doctor and patient, and thus limits the latter's access to the former, the patient may feel alienated, resulting in distrust. In a situation, for example, in which a patient has been given a diagnosis of a life-limiting illness, to speak to a doctor to ask questions could make a significant difference to a patient's wellbeing. Whilst these questions could be answered by AI, speaking with a doctor could be reassuring and comforting, particularly if a rapport has already been built. Being unable to do so might alienate a patient from health services, depriving them of essential care. Furthermore, health inequalities could be entrenched by a changed doctor-patient relationship. For groups who are known to have less trust in health services because of discrimination past and present, a good doctor-patient relationship could help instil that trust that has been lost. From a clinician's perspective, developing a relationship and connection based on shared humanity with a patient can be essential to understanding their needs. AI could threaten this, and thus impact on quality of care.

AI supplanting healthcare professionals is nowhere near a practical reality, but forms of digital triage and service delivery – such as AI based digital therapeutics for insomnia and anxiety - are already here and in use now. A Health Foundation survey of both staff and the public found that '*the prospect of health care becoming more impersonal with less human contact was ranked the biggest risk of automation and artificial intelligence*'.

2.2.2 Clinicians' experience of work and system efficiency

AI is unlikely to replace doctors but may transform the nature of work

Workforce displacement from AI receives a lot of attention. One study suggests that 35% of UK jobs, economy wide, could be lost to AI automation over a 20-year period. However, PWC predict the risk of job displacement in health and social care from AI and related technologies to be lower than in other sectors, with technology largely proving 'complementary'. Fewer healthcare roles consist of wholly automatable tasks and given the backdrop of rising demand for care - health and social care is predicted to see the largest net employment increase of any sector over the next two decades.

A Health Foundation-funded study into the potential of automation in primary care found that, while there were a small number of roles that were likely to be heavily impacted by automation, no single occupation could be entirely automated.

Similarly, research by Deloitte concludes that '*the healthcare jobs most likely to be automated would be those that involve dealing with digital information, radiology and pathology, for example, rather than those with direct patient contact*' but even here '*there are several reasons radiology jobs, for example, will not disappear soon*'. AI systems in radiology perform a single task – reading and interpreting images – but radiologists do much else besides this. They also '*consult with other physicians on diagnosis, treat diseases and perform image-guided medical intervention, define the technical parameters of imaging examinations to be performed (tailored to the patient's condition), relate findings from images to other medical records and test results, discuss procedures and results with patients, and many other activities*'.

Instead of replacing staff in healthcare, technology is likely to transform the nature of work. The Health Foundation argue that '*Rather than job displacement… in many cases automation and AI in health care will enable staff to switch their time and attention to tasks that cannot be automated, and to focus on activities where humans add more value*'.

AI could improve job quality or increase risk of burnout

Burnout is recognised as a huge problem in UK health services, with the 2023 NHS England staff survey noting that about 30% of NHS staff experience it. AI could address this by reducing the workload NHS staff face, through efficiency and productivity mechanisms as outlined above. This could have knock-on effects on staff retention, as fewer staff feel the need to leave or retire due to health reasons.

Increased productivity and efficiency also mean AI could reduce the bureaucratic burden on doctors, allowing them to pursue areas of their job they find more satisfying, as well as giving space for career development.

However, AI could also create new pressures. Routine work can act as a 'buffer' to more intensive or challenging work, and staff may find themselves working more often or continuously at the limit of their skillset – which could conversely increase the chance of burnout.

It is worth noting that these outcomes will not necessarily follow from any productivity gains – it depends on how the productivity gains are realised and how the time saved is distributed. If, for instance, gains are used to keep care costs down by delivering more care with a similar number of staff, doctors could end up seeing more patients rather than having more time for each patient.

Furthermore, for AI to improve job experience and efficiency of workflow, AI solutions must be implemented effectively, and the challenges of this should not be underestimated (see barriers section below). The wrong technologies, poorly implemented, can needlessly add to workloads with negative effects on productivity.

AI may improve efficiency but could also increase healthcare costs

AI could reduce pressures on staff time and healthcare system finances, particularly bureaucratic time, allowing them to reallocate resources to where they are needed. The NHS planned workforce expansion set out in NHS England's Long-term Workforce Plan, for instance, is predicated on '*stretching productivity ambitions*', with innovations including AI '*one of the most important ways of delivering*' them.

This can also present as a form of cost-saving, which would allow the saved funds to be reinvested in other areas, such as addressing staff shortages. Whether such productivity improvements lead to cost savings, depends on how those productivity savings are realised, as discussed above.

The Topol Review estimates a minimal saving of one minute saved per patient from new technologies such as AI, equating to 400,000 hours of emergency department consultation time and 5.7 million hours of GP time. In 2018, the Institute for Public Policy Research estimated that AI and automation could save the NHS in England £12.5 billion per year by freeing up staff time.

This can be directly, as a result of tools developed to improve back-office efficiency, streamlining administrative tasks and optimising allocation of resources. Research by Oxford University found that 44% of all admin work in General Practice can now be mostly or fully automated – freeing up staff time for other activities.

AI can also lead to quicker and more efficient healthcare interventions. In radiology, for instance, AI can be used to identify and prioritise findings that need early attention. Emerging evidence from Somerset NHS Foundation Trust has shown how AI software can speed up the diagnostics pathway for patients, reducing the wait for a CT scan following a chest x-ray from seven to less than three days.

However, AI may not always reduce healthcare costs. A recent simulation study of AI in colonoscopy in the USA found that it may increase costs in the short term by increasing the number of detected abnormalities leading to the need for intensive surveillance colonoscopies; although it may contribute to the reduction of colorectal cancer in the long term which could ultimately lead to cost reduction. Similarly, a study of glaucoma screening in China found it would be able to reduce disease progression risks, but the excess costs of screening would be unlikely to offset by this. Part of the problem in understanding the true impact of AI on costs is a lack of real-world studies. When and if AI is implemented on a large scale in health services, robust large-scale studies with long-term follow up will be required to understand if its usage truly contributes to cost savings and/or improved patient outcomes.

2.2.3 The benefits and risks of AI are interrelated

The potential benefits and risks of AI do not exist in isolation. For example, a doctor no longer suffering from burnout may be willing to give more time to develop the doctor-patient relationship. Furthermore, better patient outcomes could increase the trust between patients and clinicians and thus improve the doctor-patient relationship. There is, therefore, the potential for a multiplier effect with the positive outcomes of AI.

Similarly risks have a potential to multiply. For example, if AI is leading to poor outcomes, doctors will avoid using it, potentially disrupting workflows and productivity gains.

However, it is also important to be realistic about what AI can achieve. One AI product or technology may not be able to realise the triple aims of improving patient outcomes, job quality, and cost reduction all at once – there may be a trade-off between these outcomes.

2.3 The key issue is how AI is implemented

The dominant policy approach to AI in healthcare has been to support the development of technology, with the aim of dragging and dropping newly developed technologies into the health system. Several system barriers, including those listed below, mean that such an approach may be ineffective without further consideration of how to integrate and implement AI. As the Health Foundation argue, government need to '*fund the change, not just the technology'*.

AI must be robustly assessed for safety and efficacy in clinical settings

The current evidence for AI healthcare tools is often poor. Robust studies have shown that AI models are able to outperform human clinicians in certain isolated healthcare tasks (see section 2.2.1), but there have been several systematic reviews that showcase the limitations with the existing evidence base for overall AI performance. One such review describes a '*paucity of robust evidence*' for claims for the benefits of AI in advancing clinical outcomes, '*where [there are] only a handful of RCTs comparing AI-assisted tools with standard-of-care management in various medical conditions'*.

A large part of the problem is that much of the literature focuses on technical evaluations in a lab setting, rather than evaluation of clinical efficacy in the 'real world' - including how these technologies impact patient care in actual practice. As the Oxford Internet Institute note, '*it is important to remember that building an accurate or high-performing AI model and writing about it in an academic publication is not the same as building an AI model that is ready for deployment in a clinical system. Moving from 'the lab' to 'the clinic' is a key part of the transition and yet very few AI models have successfully made the leap across this 'chasm"*.

'*A Google retinal disease detection system was found to behave poorly when deployed in several hospitals in Thailand, despite performing as accurately as a human specialist during development. This was because retinal scans taken in practice were of worse quality than those on which it had been trained.*' The deep learning system was trained on high quality scans, as most models are, and would reject any images below a certain quality. However, in practice, the hospitals were scanning dozens of patients in limited time and with poor lighting – more than a fifth of images ended up rejected and consequently patients needed to rebook and staff time was squandered.

Rushing the deployment of AI technologies within the healthcare system, without this high-quality, clinical and real-world evaluation, risks introducing ineffective technologies, using limited resources poorly (especially where opportunity cost is not adequately considered) and increasing strain on the healthcare system in the process. Worse still, it risks patient safety arising from errors or unforeseen consequences. The tech mantra of 'move fast and break things' could have disastrous consequences when applied to patient care.

There are also issues around human interaction with AI systems. Health professionals' cognitive biases can cause them to place undue trust or distrust in an automated decision ('automation bias'), especially when short on time or not provided with the skills and knowledge to understand the workings of an AI technology.

Governance and regulation to protect patient safety is vital

Governance, including regulation, is key to protecting safety and ensuring that AI is trusted, and therefore ultimately used, by staff and patients alike.

AI regulation is a reserved matter – i.e. taken at a UK level, and is regulated by the Medicines and Health products Regulatory Agency (MHRA). UK medical device regulation, including AI as medical devices, is currently in a transitional phase following the UK's exit from the EU.

In general, the government has been tentative in its approach to AI regulation. Last year, the AI white paper set out processes by which existing regulators should regulate its use through a patchwork of existing legislation. In healthcare, AI is governed by a complex tapestry of law, covering, for example, medical device law, liability laws, data protection law, and intellectual property law; all will be challenged by AI. Given AI's (current and prospective) ability to learn and optimise as they receive new input data, one major issue is how to regulate an AI device that can adapt and change over time, both in terms of function and performance. Some are calling for regulation to expand beyond the development phase to include an element of continued monitoring to ensure they remain safe and effective.

There are also questions about the capacity of existing regulators to adequately keep up to date with, and enforce regulatory compliance, given the extra duties and expectations placed on them by the AI bill without additional resource.

Staff and patient involvement throughout the development and implementation process is necessary

Successful adoption of AI will be dependent on the involvement of and buy-in from staff, patients and the public as end users of these technologies. Without this, there is a danger that the wrong technologies will be adopted, adopted technologies won't be taken up and used, or technologies will be implemented poorly. Vulnerable groups in society must be considered when buy-in is being achieved, failure to specifically consider marginalised populations risks perpetuating mistrust among these groups in healthcare structures and will worsen health inequalities. Public communication around AI, similarly to other health messaging, must be appropriately targeted to the audience.

In a recent Health Foundation survey, a sizable section of the public (29%) reported that they had '*heard, seen or read nothing at all about [healthcare AI]*'. This was also true of 24% of NHS staff. However, opinions were overall closely balanced with 40% of both public and staff reporting positive feelings towards the potential of AI, whilst 37% and 36% reported negative feelings respectively. Those reporting greater familiarity with the topic tended to communicate more positively about AI.

There is also a question surrounding how those legally recognised to have reduced decision-making capabilities would interact with AI. Young children, and adults who lack capacity, could potentially be in a vulnerable position when interacting with AI, especially regarding sensitive information. This can be due to limited understanding, cognitive impairments, and difficulty comprehending privacy implications. These groups are more susceptible to exploitation, emotional manipulation, and reliance on AI for decision-making, which can lead to the unintentional sharing of sensitive information. This would need to be addressed in regulations. Patient consent is crucial when using AI in healthcare to respect individuals' autonomy and ensure transparency in medical decision-making. Clear public communication on this is vital to help patients make informed choices, fostering a collaborative healthcare environment where ethical standards are upheld, and patient rights are protected.

Staff must be trained on new technologies (initially and continuously) and they must be integrated into workflows

For the safe and effective roll out of AI technologies in the health system, staff will need to know both how to use these new technologies safely and how they work at a base level. Without this, and the time to interrogate their outputs, there is a danger that their ability to oversee these technologies will be undermined and there will be a greater risk of automation bias (see above). The 2019 Topol Review argued there will need to be an increase in digital literacy among healthcare professionals; training and retraining are a critical part of successful implementation.

Relatedly, embedding technology successfully will require new organisational routines, ways of working and behaviours. Healthcare organisations will need to implement continuous training and AI literacy programs for staff, ensure robust data governance and quality assurance, and foster interdisciplinary collaboration. They must develop ethical guidelines and comply with regulatory standards while securing leadership support and engaging stakeholders in the change management process. Workflow integration, performance monitoring, and promoting an innovation mindset focused on patient care are all also essential for successfully embedding AI technology. The challenges associated with this, including the need to establish the implications of new technology for workflows and the need to coordinate new ways of working accordingly, are recurring themes of evaluations of technological interventions. Oxford Internet Institute note that '*Accurate AI models implemented into robust clinical systems may still fail to be useful if they necessarily disrupt existing workflows, add additional burden to already complex and already stretched care pathways or do not adequate[ly] replicate the steps involved in human decision making.*'

A 2021 poll asked NHS England staff for their views on implementation changes of automating technologies, including AI, finding almost 40% listing that '*staff shortages or inadequate equipment might make it difficult to use these technologies properly*' as their primary concern. Over 30% listed the '*large shift in culture and ways of working*'.

Successful AI requires a robust and functioning NHS to be effective

Introducing AI into a broken healthcare system is not a panacea; it requires a robust and well-functioning health service to be effective. AI can enhance healthcare delivery, but without foundational improvements such as adequate staffing, funding, and infrastructure, its potential benefits may be undermined. A strong healthcare system is essential to integrate AI successfully, ensuring it complements rather than exposes existing weaknesses and ultimately provides better patient outcomes.

The success of AI in healthcare hinges on the human element – the healthcare professionals who use and interact with these technologies. Continuous training and education are vital to ensure that staff are competent and comfortable in using AI tools. In a well-supported NHS, where staff have manageable workloads and access to ongoing professional development, the integration of AI can be smooth and beneficial. However, in a strained system where healthcare workers are already stretched thin, introducing complex AI systems without adequate support can exacerbate stress and burnout, potentially leading to resistance or misuse of the technology.

Furthermore, investment in AI should not come at the expense of basic healthcare needs. Prioritising AI development without addressing core issues such as patient access, wait times, and primary care services can result in a lopsided healthcare system where technological advancements overshadow fundamental care requirements. A balanced approach that ensures the NHS is equipped with the necessary human and material resources can enable AI to enhance, rather than detract from, healthcare delivery. When the foundation is strong, AI can help streamline processes, improve efficiency, and facilitate better clinical decision-making, thereby achieving its full potential to transform patient care.

Existing IT infrastructure and data must be improved

The successful implementation of AI relies on existing technological readiness, as well as large, high-quality training datasets. The state of NHS IT infrastructure, however, is notoriously bad – as evidenced in a BMA paper published in 2022. With patient data being ingested from tens or hundreds of different sources in one trust and stored on systems with little to no interoperability with one another, capturing and sharing high quality data can be next to impossible.

Ultimately, without wholesale root & branch digital transformation of NHS hospital IT systems, any attempts to fully harness patient data will be limited. There is hope that £3.4bn of fresh funding in England, ringfenced for technology will go some way to closing this gap – ensuring that all trusts use digital record systems and improving the platforms that they sit on.

In addition to some of the more practical problems, there exist a range of barriers from an information governance (IG) perspective. All patient data collected within the NHS is collected primarily for direct care. If the data is then used for a secondary purpose such as research or planning and if it is used in a way that is identifiable or potentially re-identifiable, consent may be required – either implicitly via 'opt-out' systems or explicitly via active permission.

This presents two key challenges for how the health services implement AI. Firstly, there is an immediate issue with how data is collected and shared. Notwithstanding individual arrangements between ICBs and commercial/research organisations where a relatively rich data set encompassing primary and secondary care is made available, data drawn from more siloed parts of the NHS automatically excludes whole areas of healthcare. For example, AI trained to identify patterns from CT scans will rely far more heavily on existing CT scans than it will on the rich qualitative free-text entries in GP records. Therefore, a bespoke agreement will need to be put in place to ensure that companies, who develop CT scan AI can access real scans captured in the NHS.

The second challenge relates to the development of synthetic data sets (datasets that are artificially generated via simulations). At present, GPs in England are data controllers for the GP record, hospitals are controllers for hospital data and NHSE are controllers for whatever GPs and hospitals share with them for the purpose of research and planning – GPs via GPES and Hospitals via Hospital Episode Statistics (HES) data. This is the case across the UK, with the data controller/data subject model – put in place under GDPR, remaining the predominant model for information processing in all four nations. If data is shared with third parties, it is (mostly) on the express permission that the data cannot then be disseminated further by said third party. Although coming into possession of that data makes the third party a de-facto controller, they do not gain the right to share it themselves. However, it is unclear how synthetic datasets fit into this process. If real data is used to create synthetic datasets, then who do these synthetic data actually belong to – given that they won't hold any identifiable information and have not been generated as a direct result of engagement with a patient, there is an argument that – at least in terms of IG - there exists no obligation to treat the data as anything other than a tool – with no protections.

Legal liability must be clarified

The increased desire to implement AI in a healthcare setting poses a number of complex legal challenges and risks. One area of particular concern is the attribution of legal liability in instances where patient safety is compromised (and AI has been used in the provision of the patient's care and treatment). When attributing liability, a 'legal personality' is required to bear legal responsibility. Simply, this means that the responsibility for the acts and omissions of an AI system will fall on its human or corporate creators, suppliers and users.

Existing laws establish that a doctor treating a patient has a legal duty to provide reasonable care. Where care falls below that reasonable standard and causes harm, the patient can bring a claim for damages. If a doctor misuses equipment (for instance negligently misreads a scan) the doctor is held to account if the error causes harm. Equally, an employer can be held vicariously liable for the actions of their staff.

Efforts are being made to provide certainty around liability when using AI. In 2019, the EU Commission issued AI liability guidelines titled *Liability for artificial intelligence and other emerging digital technologies*. The document identifies numerous challenges in this area and explains that some applications of AI will warrant strict liability (responsibility for misconduct with liability irrespective of fault), also noting that manufacturers of products that incorporate emerging digital technology — including AI — should, as with other products, be liable for damage caused by defects in their products. In February 2020, the EU Commission published *AI, White Paper on Artificial Intelligence: a European approach to excellence and trust* explaining that additional compliance requirements would apply to "*high-risk AI applications*" such as healthcare, transport, and energy. These additional requirements include, among other items, keeping records concerning the algorithm used in AI. The EU's AI Liability Directive (if enacted) seeks to adapt the current EU liability framework to make it easier for individuals to bring claims for harms caused by AI.

The UK government has also published a policy paper in August 2023 titled *A pro-innovation approach to AI regulation*. However, it does not provide substantive clarity on liability stating '*… it is too soon to make decisions about liability as it is a complex, rapidly evolving issue which must be handled properly to ensure the success of our wider AI ecosystem*'.

While the legal principle surrounding attribution of liability is well-developed, its application in contexts involving AI is less defined – this is an area that will continue to develop and will require close monitoring and scrutiny in cases where AI has been utilised in the provision of medical care.

3. BMA principles for AI policy and implementation

As discussed, what matters is how AI is implemented and applied within healthcare. AI is intrinsically neutral, though, as this paper has set out, many AI applications incorporate values into the systems themselves, and their use can have both harmful and beneficial consequences. It has the potential to considerably improve the healthcare system but it also, if utilised poorly or with the wrong intentions, could be seriously harmful to both patients and doctors. It is not possible, therefore, to make sweeping generalisations but instead each implementation of each AI technology must be considered on a case-by-case basis. The BMA believes that the wellbeing of doctors and patients must be at the forefront of any such consideration. It is with this in mind that we have developed the following principles for the use of AI in healthcare.

The BMA supports the adoption of new technologies that:

- Are safe, effective, ethical, and equitable.
- Support doctors to deliver the best possible care and improve care quality.
- Improve job quality.

AI must be robustly assessed for safety & efficacy in clinical settings

- Technology must be well evidenced to be effective, safe, ethical, and equitable before it is adopted widely. Evidence must be derived from real world settings, with interventions evaluated for clinical efficacy. Where technology is intended to improve resource allocation/improve cost effectiveness, there must be a thorough review into the evidence base for this before implementation. There must also be mechanisms for continuous monitoring and evaluation of AI system performance and efficacy, with checkpoints to iteratively improve AI systems based on monitoring and feedback.
- Risks and harmful consequences must be properly understood and accounted for, including bias and the risk of widening inequalities. Where relevant, technology should support those with the worst health to access care, information and support, rather than lock them out; aiming to reduce inequities.
- Vulnerable patient cohorts, such as those with reduced decision-making capabilities, should not be disadvantaged or suffer harms through AI healthcare technologies.

Governance and regulation to protect patient safety is vital

- Regulation, standards, and guidance must be strengthened to safeguard patient safety, with regulators empowered to stay up to date with developments and enforce legislation.

Staff and patient involvement throughout the development and implementation process is necessary

- Staff and patients need to be involved in decision making processes about technology and its use from the outset and empowered to scrutinise new innovations.
- Patients should have the right to choose to have an AI tool used in their care and be able to opt-out of AI involvement. In addition, patients need to have the ability to dispute the recommendations of an AI algorithm if they believed they were in error.
- The doctor-patient relationship needs to be protected: technology should be used to complement, rather than replace, face-to-face care and human decision making.

Staff must be trained on new technologies (initially and continuously) and they must be integrated into workflows

- Doctors must continue to be trained in skills that AI might replace to ensure they maintain a high level of expertise and adaptability in their practice. Relying solely on AI could lead to a deskilled generation of doctors, who may struggle to provide quality care if AI usage is discontinued or fails.

Successful AI requires a robust and functioning NHS to be effective

- Health services must fund the implementation and change necessary to make the most of AI, not just its development, by ensuring that staff have the necessary training, time, and infrastructure to make use of, and safely oversee, new technologies.

Existing IT infrastructure and data must be improved

- A comprehensive and interoperable data infrastructure must be established. This involves the wholesale digital transformation of NHS IT systems to enable the seamless exchange of high-quality data across all trusts.
- The implementation of robust information governance practices must safeguard patient consent and address the complexities surrounding the use of synthetic datasets.

Legal liability must be clarified

- Clear lines of legal liability must be established, including the AI developers.
- Doctors must have the ability to challenge decisions made by AI.

BMA
British Medical Association, BMA House,
Tavistock Square, London WC1H 9JP
bma.org.uk

BMA 20240640