UKRI Trustworthy Autonomous Systems Hub

The University of Nottingham

University of Southampton

The Policy Institute

King's College London

Trusted autonomous systems in healthcare

A policy landscape review

Rachel Hesketh

October 2021

About the TAS Hub

The UKRI TAS Hub assembles a team from the Universities of Southampton, Nottingham and King's College London. The Hub sits at the centre of the £33M Trustworthy Autonomous Systems Programme, funded by the UKRI Strategic Priorities Fund.

The role of the TAS Hub is to coordinate and work with research nodes to establish a collaborative platform for the UK to enable the development of socially beneficial autonomous systems that are both trustworthy in principle and trusted in practice by individuals, society and government. Read more about the TAS Hub here.

Acknowledgement

The author would like to thank Dr. Saba Hinrich-Krapels, Assistant Professor at TU Delft, for kindly reviewing the draft manuscript. Any errors or omissions are of course the author's responsibility. This publication acknowledges funding from Engineering and Physical Sciences Research Council (EP/V00784X/1).

Contact

Rachel is a Research Associate in the Policy Institute, working on research projects spanning spanning social and economic policy and a particular interest in local economic development and industrial strategy.

The Series Editor is Professor Mark Kleinman, Professor of Public Policy at the Policy Institute.

Citation: Hesketh R. (2021) 'Trusted autonomous systems in healthcare – a policy landscape review. Available from: https://doi.org/10.18742/pub01-062

Abbreviations

AAMI	Association for the Advancement of Medical Instrumentation
AI	Artificial Intelligence
AoMRC	Academy of Medical Royal Colleges
BEIS	Department for Business, Energy and Industrial Strategy
BSI	British Standards Institution
DHSC	Department of Health and Social Care
GDPR	General Data Protection Regulation
HCPs	Healthcare professionals
ICO	Information Commissioner's Office
MHRA	Medicines and Healthcare products Regulatory Agency
NAO	National Audit Office
NHS	National Health Service
POST	Parliamentary Office of Science and Technology
TAS-Hub	Trustworthy Autonomous Systems Hub

Executive summary

Autonomous systems, which are able to take actions with little or no human supervision (Trustworthy Autonomous Systems Hub, n.d.), are believed to hold huge promise for transforming health and care systems – improving patient outcomes, reducing costs and enabling new medical discoveries.

Despite their very wide range of potential applications and high levels of development activity, these technologies are, as yet, little used in health and care settings, and early applications are likely to embody the simplest of these technologies. This situation presents policymakers with both questions and opportunities. Questions around whether the adoption of these technologies is being impeded by the presence of barriers, and opportunities to fully consider, while there is time to do so, the potential risks or drawbacks associated with their application.

In both cases, issues of trust are likely to be very relevant. Are there features of autonomous systems in health and care that undermine their trustworthiness in the eyes of the medical profession, patients, and the public, and how can these be addressed – for example, through design or regulation? Are there other reasons why, in practice, trust in these systems may be limited? In this report, we draw on both the policy and the academic literature to provide an overview of the issues and challenges identified around the utilisation of autonomous systems in health and care, focusing on issues that are likely to impact the trustworthiness of, and trust in, these systems.

We look to span a range of potential applications of these technologies, including those that are likely to be deployed in the near term, such as for diagnosis and screening, and those that are unlikely to be in mainstream use for some years, such as autonomous care and autonomous surgical robots.

We follow the Nuffield Council on Bioethics (2018) in classifying the issues associated with the application of artificial intelligence (AI)/autonomous systems in health and care under eight broad headings.[1] In our view, each heading relates to some facet of the trustworthiness, or trustedness of these systems in health and care settings and requires the attention of policymakers.

Reliability and safety: Safety is always of paramount concern with the introduction of new technologies in health settings. Systems can make mistakes and algorithms can contain errors, which may be difficult to spot and could be replicated at scale. This underlines the importance of effective validation. The risk of automation bias, where busy healthcare professionals (HCPs) do not critically assess the outputs of autonomous systems, has also been raised.

Transparency and accountability: Some autonomous systems produce their outputs in opaque ways that cannot be interpreted by humans. These so-called "black box" systems pose questions around how to ascribe accountability and liability for errors, as HCPs are influenced by their outputs, but do not know how they were reached. There is a lack of legal precedent for how such cases would be resolved.

1 The Nuffield analysis focuses on AI in healthcare, rather than autonomous systems explicitly. We make one adaptation to the Nuffield framework, renaming the category they term "trust" as "public acceptance"

Data bias, fairness, and equity: There is concern that AI models may embody biases that mean they do not deliver accurate predictions for some groups. This highlights the necessity of using high quality, representative datasets to develop algorithms. There are also questions around how the introduction of autonomous systems in healthcare could affect inequalities in access to care.

Public acceptance: Polling points to mixed public opinion around the use of autonomous systems in healthcare, with some people comfortable with their use and others less so. What seems to matter for the public is that AI technologies do not fully replace the clinician-patient relationship. Issues of public trust in data sharing are also relevant when considering the use of autonomous systems, with opinion research generally pointing to wariness of sharing data with commercial organisations.

Effects on patients: There are concerns that, if autonomous systems begin to replace some patient-clinician interactions, some insights into patient health and wellbeing could be missed. The use of robots in care settings also raises a variety of questions for patient wellbeing, such as whether their use could impinge on the autonomy of those being cared for, whether it could lead to coercion or deception of vulnerable people, and whether reduced human contact could enhance feelings of social isolation.

Effects on healthcare professionals: While it has been suggested that autonomous systems could lead to job displacement in healthcare, this is judged to be unlikely in the context of the NHS, with new technologies seen as augmenting rather than replacing HCPs, freeing up time to devote to contact with patients. There is, however, the possibility that clinicians' roles could be changed in undesirable ways if the use of autonomous systems leads to the deskilling or side-lining of HCPs.

Data privacy and security: Access to relevant data is clearly essential to develop AI technologies for use in health and care settings, but past incidents, such as the collaboration between the Royal Free NHS Trust and Google DeepMind, and the Care.data experience, may have dented public and HCP confidence in this. A lack of public engagement with data safeguarding measures has also been identified.

Malicious uses of AI: There is the risk that AI technologies could be used for surveillance or to gather information on people's health without their knowing, and the vulnerability of autonomous systems to adversarial attacks and data breaches has also been raised.

It is also important to consider at the outset that the understanding of "autonomy" may be different in health to in other settings. While an autonomous vehicle clearly has wide scope to make its own decisions and act on them, AI technologies in health are seen more as just one of many inputs into decision-making by clinicians, with responsibility remaining firmly in human hands. Fully autonomous systems, where humans are removed from the decision-making process, are likely only to materialise in the distant future in healthcare contexts.

Introduction

Autonomous systems are seen as having transformational potential in health and care. The independent Topol Review, commissioned by the government to examine the impact of digital technologies on the National Health Service (NHS), sees technologies such as AI and robotics as providing healthcare professionals (HCPs) with the "gift of time" – relieving them of mundane tasks to allow them to devote more attention to patients (Topol, 2019). At the other end of the spectrum, these technologies can undertake analytical tasks that are beyond human ability, analysing vast datasets for patterns to aid drug discovery and the realisation of precision medicine (OECD, 2020).

"Autonomous systems are seen as having transformational potential in health and care"

Roles are also envisioned for autonomous systems in interacting directly with patients. It has been proposed that autonomous systems can help to meet rising social care needs among ageing populations (Tan and Taeihagh, 2020), with robots providing companionship, physical assistance, and cognitive support (POST, 2018). Robots with autonomous capabilities could also help in the delivery of mental health services (Fiske et al. 2019), and in making surgery safer and more accurate (Royal College of Surgeons, 2018).

With such promise, it is absolutely vital for policymakers and researchers to consider how these technologies can be deserving of the trust of patients, HCPs and the wider public, and be trusted in practice. This means, most obviously, ensuring safety and reliability, but also involves considering wider social and ethical concerns, such as how these technologies might affect health inequalities, and how they could change the roles of those who work in the health service.

Given the very early stage we are at in the development of many of these technologies, and their testing in health and care settings, we have an excellent opportunity to assess these issues, and ensure that trustworthiness is built into the design and deployment of autonomous systems in health and care from the outset.

Our approach

In this report we seek to map the issues of relevance to policymakers in developing and deploying trustworthy autonomous systems in health. We identify these issues via a scoping of the relevant policy literature, including publications from UK government and overseas governmental organisations; relevant policy and regulatory bodies; professional associations and thinktanks. We have also drawn on the academic literature, particularly reviews of the policy and ethical issues around the use of AI/ autonomous systems in health.

This does not pretend to be a systematic review of the literature given the sheer body of research in this field. It is also not possible in the scope of this review to cover the topic of autonomous systems in health and care comprehensively, given the range of relevant technologies and the number of different contexts in which they might be applied. Instead, it aims to provide an overview of the landscape and synthesis of the key themes, rather than focusing on any single context or technology in depth.

The rest of the report proceeds as follows. We first provide an overview of what we mean by autonomous systems, both generally and in health, their range of

applications and the current extent of their deployment. We then discuss the associated policy issues grouped under eight themes. These largely mirror the eight categories of ethical and social issue associated with the use of AI in healthcare and research identified by the Nuffield Council on Bioethics (2018). We found that these categories represented a particularly effective structure for classifying the salient issues identified in the literature.

Autonomous systems in health and care

Definining an autonomous system

In defining our central concepts of autonomous systems and trust in these systems, we follow the definitions advanced by the Trustworthy Autonomous Systems Hub (TAS-Hub). For the Hub, an autonomous system is "a system involving software applications, machines, and people, that is able to take actions with little or no human supervision" (TAS-Hub, 2020).

This definition makes clear that while autonomous systems may not require human control or supervision to perform their task, removing the human from the loop is also not a prerequisite to consider a system autonomous. Others argue that in the future, AI developments are likely to be towards keeping humans in the loop, rather than displacing them entirely (Dwivedi et al. 2019).

While the TAS-Hub definition does not elaborate on the relationship between autonomous systems and AI (terms that are commonly used together or interchangeably), others do so. NASA, for example, emphasise that an autonomous system may make use of AI, but that the two are not the same (NASA, 2020; Fong, 2018).

Turning to issues of trust, the TAS-Hub defines autonomous systems as trustworthy when: "their design, engineering, and operation ensures they generate positive outcomes and mitigates potentially harmful outcomes" (TAS-Hub, 2020). This depends on factors including:

- Their **robustness** in dynamic and uncertain environments.
- The **assurance of their design** and operation through verification and validation processes.
- The **confidence they inspire** as they evolve their functionality.
- Their **explainability, accountability, and understandability** to a diverse set of users.
- Their **defences** against attacks on the systems, users, and the environment they are deployed in.
- Their **governance** and the **regulation** of their design and operation.
- The consideration of **human values and ethics** in their development and use.

(TAS-Hub, 2020).

We draw on this conception of trustworthiness in our consideration here of the policy issues associated with trust in, and the trustworthiness of, autonomous systems.

Background

The term "autonomous system" is not one that appears to be used frequently in the health literature. Instead, it seems to be more common to refer to artificial intelligence (AI) in the context of new technologies in health and care, and to then consider the degree of autonomy these technologies may have. The British Standards Institution (BSI, 2019) observes that the degree of autonomy varies across AI technologies in health, and Challen et al. (2019) predict this to increase over time, from decision support tools today to autonomous equipment, such as ventilators and insulin pumps, over the long term. It is also argued that, at least in the near term, AI systems in health will be aids to clinician decision making, rather than being decision-makers in their own right (AoMRC, 2019). This degree of autonomy may be less than that observed in other sectors.

Following the literature, we will use both the terms "AI" and "autonomous system" in this report. While acknowledging that these terms are not synonyms, some of the policy-related discussion pertaining to AI is relevant to our purposes here. Hence, while autonomous systems are the term of interest to us, we will refer to AI where this is the term employed by the source in question.

Applying autonomous systems in health and care

As a general-purpose technology, AI has a very wide range of potential applications across health and care (OECD, 2020). NHSX, the body responsible for the digital transformation of the NHS, sets out five broad areas where it could be deployed in the health service (see Figure 1): diagnostics (eg image recognition); knowledge generation (eg drug discovery); public health (eg epidemiology); system efficiency (eg optimising care pathways and assessing staffing requirements); and P4 medicine (eg predictive, preventive, personalised and participatory medicine) (Joshi & Morley, 2019).

FIGURE 1: AREAS OF CARE WHERE AI CAN BE APPLIED

Diagnostics	Knowledge generation	Public health	System efficiency	P4 medicine
• Image recognition, eg symptoms checkers and decision support • Risk stratification	• Drug discovery • Pattern recognition • Greater knowledge of rare diseases • Greater understanding of causality	• Digital epidemiology • National screening programmes	• Optimisation of care pathways • Prediction of Do Not Attends • Identification of staffing requirements	• Prediction of deterioration • Personalised treatments • Preventative advice

Source: Joshi & Morley (2019)

Autonomous systems in health take different forms – they may be embodied in robots, for example patient support robots or autonomous surgical robots, or in information systems, for example those that interpret medical images or carry out administrative tasks.

There are also important distinctions between autonomous systems that communicate directly with patients, and those that are mediated by a HCP. For example, chatbots that will interact directly with patients are being developed for use in mental health services (Topol, 2019; Fiske et al. 2019), while many diagnostic uses of autonomous systems will support and assist decision-making by HCPs (AoMRC, 2019; Future Advocacy, 2018). Also relevant is that while some technologies will be employed in care settings, such as hospitals, others will be used by patients in their own homes, while a further subset will be deployed at the health system level, for administrative or public health purposes. As we will explore, where an autonomous system is employed and its relationship with the patient can have policy-relevant implications.

The ambition for the use of autonomous systems in health and care

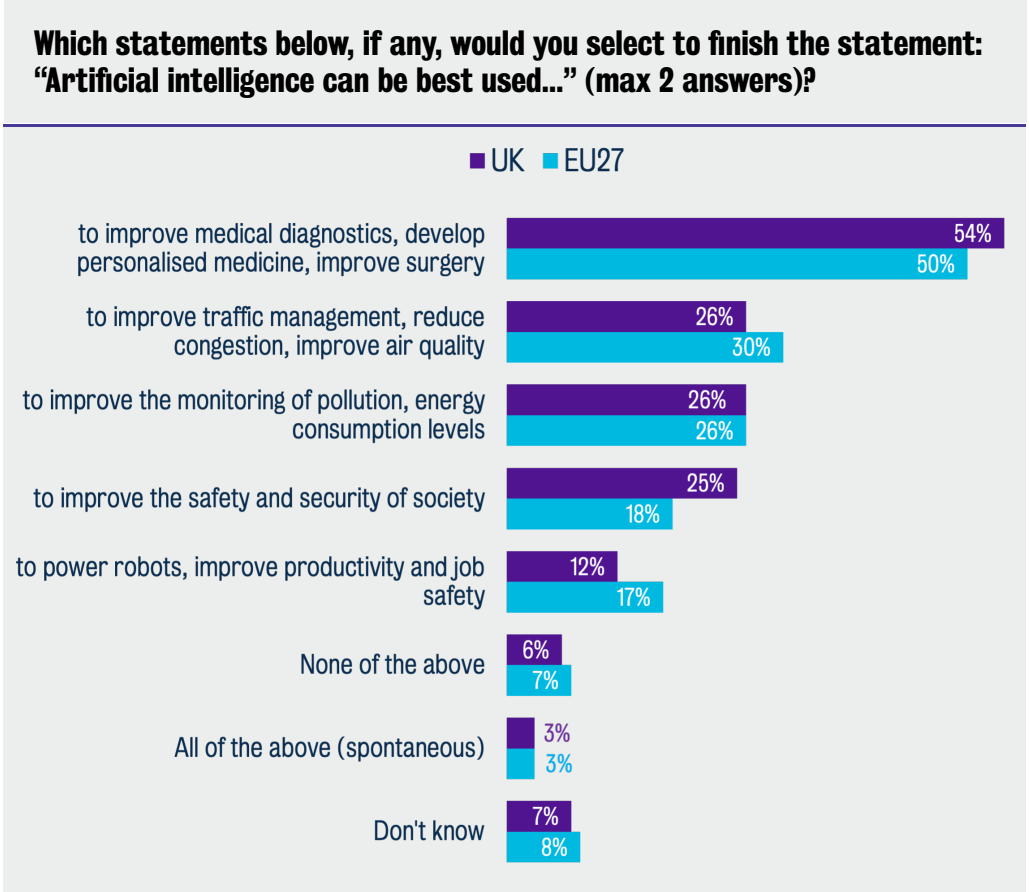
The UK government and the NHS have asserted clear ambitions around the application of artificial intelligence to health and care. In 2018, the government's Industrial Strategy articulated four "Grand Challenges" (BEIS, 2018), one of which, the AI and Data Grand Challenge, set out the goal to "[u]se data, Artificial Intelligence and innovation to transform the prevention, early diagnosis and treatment of chronic diseases by 2030" (BEIS, 2021).

"The UK government and the NHS have asserted clear ambitions around the application of artificial intelligence to health and care"

Similarly, the NHS Long-Term Plan, published in 2019, articulated the intention to: "Use decision support and artificial intelligence (AI) to help clinicians in applying best practice, eliminate unwarranted variation across the whole pathway of care, and support patients in managing their health and condition" (NHS, 2019 p. 92).

There also appears to be public endorsement of the application of artificial intelligence in healthcare, with potentially more support for its use in this sector than in other areas of the economy and society. Polling by Eurobarometer in 2019 asked Europeans where they thought AI could be best used (European Commission, 2019). Respondents in both the 27 EU member states and in the UK were most likely to select medical applications (improving diagnosis and surgery and developing personalised medicine), choosing this option over uses in traffic management, pollution monitoring, security and to improve productivity and safety in the workplace (see Figure 2). It is worth noting, however, that the use of AI in healthcare only just attracted majority support in the UK (54 per cent of respondents), and marginally lower support in the EU27 countries.

FIGURE 2: VIEWS ON THE POTENTIAL BENEFITS OF AI



Source: European Commission (2019)

The current status of the application of these technologies

Despite these ambitious intentions, the deployment of autonomous systems in health and care is currently very limited. Harwich and Laycock (2018) point to some examples of virtual assistants and chatbots in use in the health service, but for the most part, AI technologies for health and care are in their infancy (Marr, 2018).

Topol predicts the large-scale adoption of AI and robotics in the NHS by

2040

When more than

80%

of the workforce will be affected by these technologies

It is also not straightforward to build a granular timeline for the roll-out of autonomous systems in health and care in the future. Topol (2019) sketches out broad timescales for the at-scale adoption of AI and robotic technologies by the NHS. Of the four categories they look at (speech recognition and natural language processing; automated image interpretation using AI; interventional and rehabilitative robotics; and predictive analytics using AI), they anticipate that all except robotics will be in routine use in the NHS by 2040 (with more than 80 per cent of the workforce affected by them). While their timelines are understandably indicative, they suggest that natural language processing and speech recognition technologies in particular are likely to be widely deployed sooner, perhaps from 2030.

Others emphasise the readiness of the field of medical image interpretation for the mainstream application of artificial intelligence (eg King et al. 2018). The efficacy of these technologies has already been demonstrated (Topol, 2019; van der Schaar &

> While the deployment of autonomous systems to assist clinician decision-making is on the horizon, the possible replacement of the judgement of medical professionals by AI technologies is seen to be a long way in the future "

Zame, 2018), and their development has benefited from the availability of digitised data from radiology, pathology and ophthalmology to train deep learning systems (Topol, 2019).

In general, the use of autonomous robots is seen as being a far more distant prospect, both in surgery and in care settings. the Royal College of Surgeons does not believe the use of fully autonomous surgical robots is likely within the next 20 years, though they are likely to begin performing some simple tasks, such as suturing, before then (Royal College of Surgeons, 2018). The development of robots for social care purposes is still in its very early stages (POST, 2018), though some countries, for example Japan and Singapore, have gone further in trialling these technologies (POST, 2018; Tan & Taeihagh, 2020).

Barriers to the widespread deployment of autonomous systems in health

While there is considerable optimism about the potential to deploy these technologies in routine healthcare in the coming decades, others point to the existence of important barriers to this. Panch et al. (2019) identify two key obstacles to the mainstream adoption of AI algorithms. Firstly, existing ways of working in healthcare organisations, which are shaped by complex forces and incentives that are unlikely to respond to new innovations; and secondly the lack of the necessary data infrastructure within healthcare organisations to train algorithms to represent the local population and to check them for biases.

A review by Dwivedi et al. (2019) points also to additional financial and organisational barriers – the high cost of medical AI technologies, the need to ensure technologies are developed specifically to meet the needs of health systems, and shortages of the skills and talent needed to utilise these technologies.

It is also important to note that, while the deployment of autonomous systems to assist clinician decision-making is on the horizon, the possible replacement of the judgement of medical professionals by AI technologies is seen to be a long way in the future (AoMRC, 2019; King et al. 2018).

The trustedness and trustworthiness of autonomous systems in health

Through an assessment of the policy and academic literature we identified a range of issues for policymakers to consider around the use of autonomous systems in health and care, particularly focusing on issues related to their trustworthiness. We utilise the framework set out by the Nuffield Council on Bioethics (2018), which classifies the ethical and social issues associated with the application of AI in health and care under eight broad headings[2]. While this framework was not specifically designed with trust as the overarching issue, it succinctly captures relevant issues for policymakers, all of which link to some facet of the trustedness or trustworthiness of these systems.

It is important to recognise that other frameworks for classifying the ethical, social and trust issues associated with use of AI/autonomous systems do exist. Blobel et al. (2020) point to several different sets of ethical frameworks for autonomous and intelligent systems. Dwivedi et al. (2019), meanwhile, set out a framework to guide public policy practitioners in assessing the "safety and social desirability of any AI system" (p.30). These frameworks are not specific to the health context, however. It is also notable that all frameworks tend to share common elements, emphasising, for example, the importance of transparency, accountability, data protection and fairness. We selected the Nuffield framing for its tractability, specificity to health and care and the ease with which relevant literature fit under its headings.

Where possible we also include information about current regulatory and policy responses to the issues raised and highlight key outstanding questions for policymakers to consider.

Reliability and safety

> "There is a risk that these systems may produce incorrect predictions or diagnoses, or that robots may malfunction... errors within algorithms can [also] lead to the occurrence of harms at scale and may be difficult to identify"

As with all new technologies, there are concerns about the safety of autonomous systems developed for health and care settings, though the risks are likely to vary according to the technology in question and the context in which it is applied. Most obviously, there is a risk that these systems may produce incorrect predictions or diagnoses (Nuffield Council on Bioethics, 2018; AoMRC, 2019; Morley et al. 2020), or that robots may malfunction (Fiske et al. 2019). It is also pointed out that errors within algorithms can lead to the occurrence of harms at scale and may be difficult to identify (Nuffield Council on Bioethics, 2018; AoMRC, 2019; Morley et al. 2020). The need for the effective validation of algorithms is frequently noted (eg, see BSI, 2019; King et al. 2018).

There are some specific safety concerns associated with machine learning systems in healthcare contexts. Challen et al. (2019) point to the risks of a changing environment on the effectiveness of machine learning systems. For example, a change in disease patterns over time may damage the performance of an algorithm as the data it is required to analyse increasingly differ to that it was trained on. Ordish et al. (2019) focus on the subset of machine learning systems that update themselves, pointing out that even small changes in the model can lead to very different outcomes, which could have important safety implications. They suggest that these systems do not fit comfortably into the current regulatory regime for medical devices.

2 We make one change to the Nuffield headings, renaming the category "Trust" as "Public acceptance".

Other technical safety issues raised in the literature include the need for machine learning systems, particularly those that are not readily interpretable by clinicians, to have a "fail safe" mode. This would prevent the system from making a prediction when it has low confidence in that prediction, in situations of inadequate information, or where analysing data different to that it was trained on (Challen et al. 2019). There is also the longer-term concern that, over time, autonomous systems could find ways to "game" the measure they are targeting, where the target is achieved in a way that is not best for the patient's overall health and welfare. This is termed "reward hacking" (Challen et al. 2019).

How humans interact with autonomous systems in health settings is of central importance from a safety perspective. The risk of "automation bias", where there is a tendency to uncritically accept the outputs of machines, is frequently highlighted. HCPs may become complacent or unwilling to critically assess and question the output of a technology, meaning errors are less likely to be spotted (AoMRC, 2019; Challen et al., 2019). BSI (2019) suggest the need for close monitoring of how HCPs interact with AI technologies in their early stages of deployment.

There are specific concerns in the cases where autonomous systems interact directly with patients. The Academy of Medical Royal Colleges (AoMRC) suggests that there is a risk of people misunderstanding the information they are given, for example around the severity of a condition or the size of a health risk, and that vulnerable groups may be more susceptible to errors or bad advice from AI systems that communicate directly with patients (AoMRC, 2019).

Existing regulatory response

There are a number of organisations with responsibilities for the safe deployment of autonomous systems in healthcare. NHSX was established in 2019 with a remit that includes setting standards for the use of technology in the NHS, and hosts an 'AI Lab', which promotes the safe application of AI in health and care.[3] The NHS AI Lab is working in partnership with regulatory agencies and other health organisations to develop the regulatory framework for AI in health and care (DHSC, 2021b).

The Department of Health and Social Care (DHSC) has also developed a "guide to good practice for digital and data-driven health technologies" which was most recently updated in January 2021. This sets out guidance for the developers of data-driven health technologies, covering issues such as safety, regulation, effectiveness and data protection (DHSC, 2021a).

More recently, the DHSC published a draft data strategy, Data saves lives: reshaping health and social care with data (DHSC, 2021b), which sets out further measures around the regulation of AI in health and care. The department commits to developing "unified standards for the efficacy and safety testing of AI solutions, working with MHRA and NICE" by 2023, as well as assisting regulators in reviewing AI legislation as part of amendments to the Medical Devices Regulations 2002, following the UK's departure from the European Union.

3 See NHSX, About the NHS AI Lab: https://www.nhsx.nhs.uk/ai-lab/about-nhs-ai-lab/

> The line between what qualifies as a medical device and what constitutes a lifestyle or wellbeing device is increasingly blurred"

Healthcare regulators play an important role in the regulation of AI in health and care. For example, the Medicines and Healthcare products Regulatory Agency (MHRA) regulates medical devices/ in-vitro diagnostic medical devices that use AI technologies (Joshi & Morley, 2019; POST, 2021), though it has been pointed out that it is "unclear in many cases whether or not 'algorithms' count as medical devices" (Joshi & Morley, 2019 p. 22). Similarly, Ordish et al. (2019) note that "the line between what qualifies as a medical device and what constitutes a lifestyle or wellbeing device is increasingly blurred" (p. 37). Given these grey areas, developers may not be aware that their technology needs to be regulated as a medical device (Joshi & Morley, 2019).

While there is clearly an active regulator presence around the application of AI systems in health, it can be perceived by developers as a confusing landscape. NHSX point to the absence of an overall responsible body who can coordinate the various regulators; instances of regulators with overlapping remits; and the absence of a regulator with responsibility for the quality of the data used in developing algorithms (Joshi & Morley, 2019). This may be why the DHSC has pledged to support regulators in developing "a multi-agency service for innovators seeking advice on their regulatory journey in getting their product to market", with plans to pilot the service in 2021, and roll it out by 2023 (DHSC, 2021b).

There are some international initiatives looking at developing common standards for AI in health, for example the joint work of the International Telecommunications Union and World Health Organisation (see International Telecommunications Union, n.d.), and collaboration between the British Standards Institution (BSI) and the Association for the Advancement of Medical Instrumentation (AAMI) in the US (see BSI, 2019).

Transparency and accountability

A crucial feature of some (but not all) autonomous systems in healthcare is the opaque way in which they produce their outputs – the so-called "black box". This may be because the algorithm used to generate it is proprietary, or, in the case of some machine learning models, it may be that it is simply too complex for humans to understand and interpret (Nuffield Council on Bioethics, 2018; Smith, 2020; Ordish et al. 2019).

There are likely to be difficulties accepting black box autonomous systems in health settings. In particular, there is an outstanding issue over who is accountable for the decisions of opaque autonomous systems, and who is liable in the case of things going wrong. On the one hand, some stress that the technology is simply a decision-making aid, and thus responsibility for decisions resides with the clinician (Smith 2020). On the other, it is argued that if a clinician cannot fully understand and explain how a decision was reached by an autonomous system, they cannot easily be considered accountable or responsible for it. In this argument, the developers of the technology should be ascribed some responsibility (Smith 2020; Habli et al. 2020).

It is suggested that, as things currently stand, liability for adverse events in which the output of an algorithm was used would lie with the clinician given that the

> There is an outstanding issue over who is accountable for the decisions of opaque autonomous systems, and who is liable in the case of things going wrong... Given this uncertainty, the question has been raised as to whether opaque decision-making systems should be used in healthcare at all"

system simply informs the clinician and does not make the decision itself (Morley et al. 2020). However, many sources stress the uncertainty around this, and the lack of legal precedent for how issues of liability would be resolved (eg, see Smith 2020; POST, 2021; AoMRC, 2019). The Royal College of Surgeons has expressed concerns about the potential for increased litigation arising from the use of AI technologies (Royal College of Surgeons, 2018).

Given this uncertainty, the question has been raised as to whether opaque decision-making systems should be used in healthcare at all (Future Advocacy, 2018; Smith, 2020). Morley et al. (2020) suggest that clinicians will be unlikely to want to use algorithmic decision-making tools until issues of liability are clear. Such a choice is further complicated by evidence indicating a trade-off between model interpretability and accuracy (Ordish et al. 2019; van der Schaar & Zame, 2018).

The proliferation of health apps, which give advice direct to the user, raises a separate set of questions around accountability. Particularly, who would be liable for adverse outcomes resulting from the guidance issued by an app (Ada Lovelace Institute, 2020)?

Existing regulatory response

The DHSC's Guide to good practice for digital and data-driven health technologies includes guidance around the transparency of the algorithm used. This encourages developers to be clear about their algorithm's learning methodology, the data used to develop it, and its limitations (DHSC, 2021a). The government has said it is developing tools to help technology developers comply with these guidelines (BEIS, 2021).

The 2018 General Data Protection Regulation (GDPR) includes some provisions for transparency around "the logic involved" in automated decision making (Future Advocacy, 2018), which some have interpreted as a "right to explanation" of the decisions made by AI systems (Future Advocacy, 2018, Wachter et al. 2017). Others suggest that, in practice, this legislation is likely to be relatively ineffectual, providing access to only limited information, rather than the full transparency that some have assumed (Wachter et al. 2017).

> [There is] a risk of bias in algorithms, either in their design or from training algorithms on biased datasets that don't fully represent the population they are designed to be applied to"

Data bias, fairness, and equity

There is concern that AI models in health could contribute to existing health inequalities. This stems from the risk of bias in algorithms, either in their design or from training algorithms on biased datasets that don't fully represent the population they are designed to be applied to (Murphy et al. 2021). This means that when the model is applied in the real world, it may generate inaccurate predictions for some groups (OECD, 2020; Harwich and Laycock, 2018; King et al. 2018).

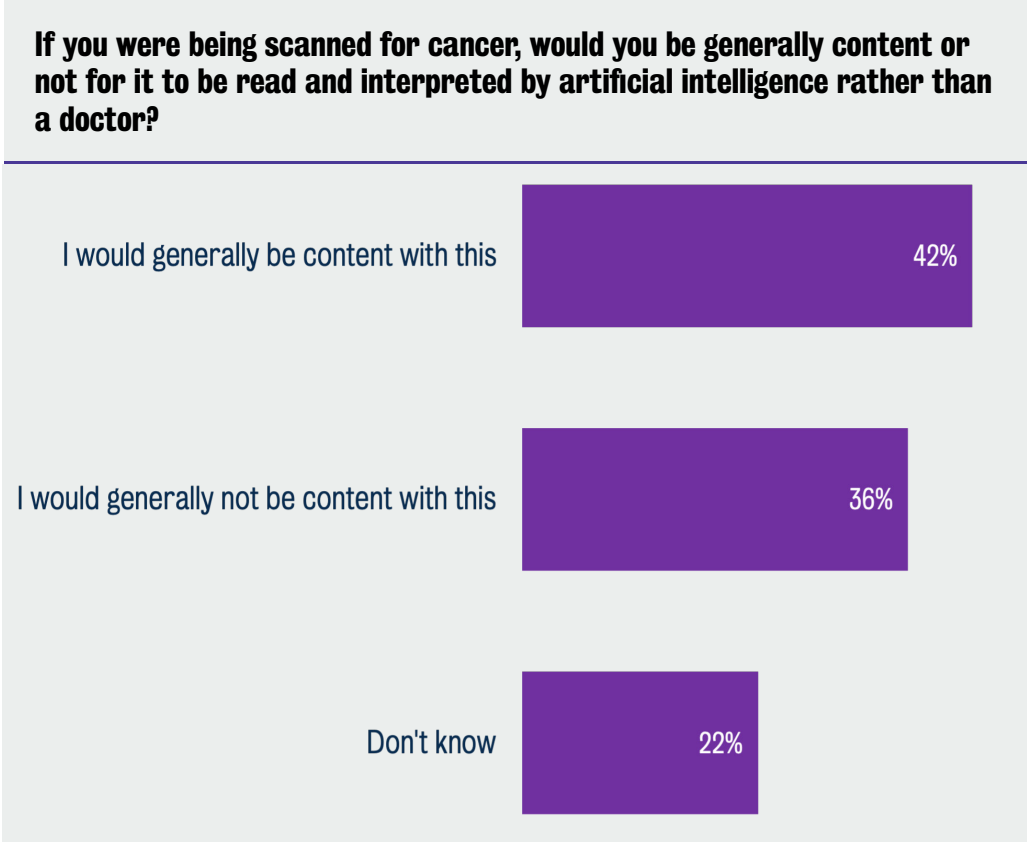
There is also debate about how the introduction of autonomous systems could affect inequalities in access to care. In some ways it is possible that they could exacerbate inequalities; vulnerable people may struggle to use digital healthcare systems (DHSC, 2018), or it could be that those with more money are able to pay for the form of

diagnosis/care that is judged to be superior (human-led or machine-led) (AoMRC, 2019). On the other hand, some autonomous systems, for example mental health chatbots or community-level diagnostic services, could help to reduce inequalities in access to and quality of care, particularly those driven by geography (King et al. 2018).

Public acceptance

The public acceptability of autonomous systems in health and care appears to be relatively mixed and unclear. One-off polling points towards divided public opinion on the use of AI diagnostic devices and receiving health advice from AI systems. For example, in recent YouGov polling, 42 per cent of respondents said they would be comfortable with a scan for cancer being read and interpreted by an AI system, while 36 per cent expressed discomfort with this (see Figure 3) (YouGov, 2020). Importantly, almost a quarter of respondents said that they didn't know, pointing to a substantial degree of uncertainty, and possibly scope to shape public opinion.

FIGURE 3: VIEWS ON THE USE OF AI FOR CANCER SCANS

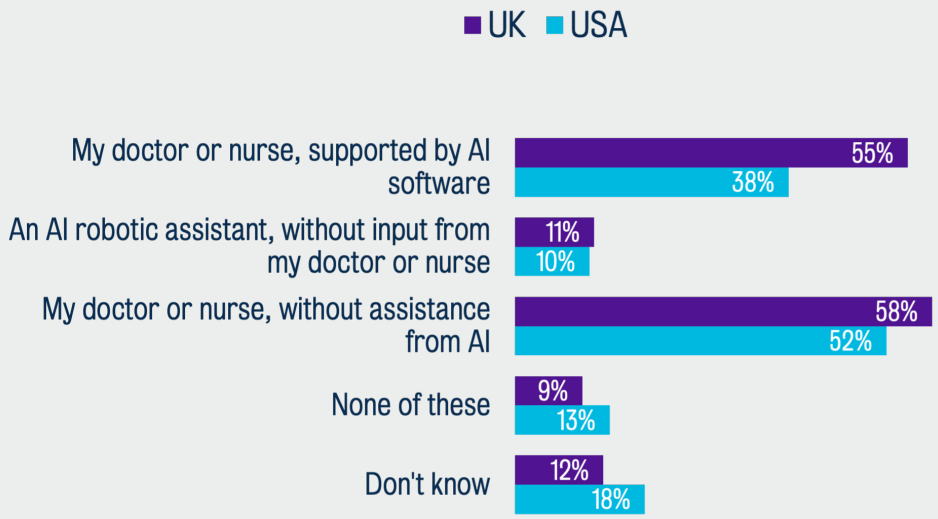


Source: YouGov (2020). 1737 GB adults surveyed

People may be particularly uneasy about the use of these systems when they replace doctors or other health practitioners entirely. In polling by YouGov for Ghafur et al. (2020), only 10 per cent of respondents in the US and 11 per cent in the UK said they would be happy to receive healthcare advice from an AI robotic assistant without input from a medical professional (see Figure 4) (Ghafur et al. 2020).

FIGURE 4: VIEWS ON THE USE OF AI FOR HEALTHCARE DELIVERY

From which, if any, of the following would you be happy to receive advice on your healthcare (eg diagnoses, treatment plans, lifestyle advice etc.)? Please select all that apply



Source: Ghafur et al. (2020). YouGov online survey of UK adults (n=2080) and US adults (n=1114)

More in-depth, qualitative research in the UK by Ipsos MORI points to public optimism about the use of data-driven technologies, including AI, machine learning and natural language processing, in healthcare. There is concern though that the application of these technologies should not replace the patient-clinician relationship or restrict patient choice in healthcare (Castell et al. 2018). Others have suggested that the more transparent are AI systems, both in terms of clarity around how their outputs are derived and wider processes around their development and commissioning, the more likely they are to gain people's trust (BSI, 2019; Future Advocacy, 2018).

A related issue highlighted in the Ipsos MORI study is public trust in data sharing, essential to enable the development of data-driven technologies. The study identified a lack of knowledge about data safeguards, and the expectation that the NHS would take responsibility for keeping people's data safe (Castell et al. 2018). Studies tend to point towards a wariness around sharing health data with commercial organisations (see for example Genomics England, 2019; Ghafur et al 2020), though the Ipsos MORI study nuanced this somewhat. They found that willingness to share data with commercial organisations depended on the benefits that would be realised and to whom, and the nature of the data (identifiable or depersonalised) that would be shared.

Effects on patients

The potential impacts on patients of the increased use of autonomous systems in health and care tend to be associated with patients' direct interactions with these systems, and the removal of the clinician or carer presence. There is a strong appreciation among the public for the patient-clinician relationship, and a desire for

this not to be compromised by new technologies (Castell et al, 2018). Clinicians too express concerns about the loss of these interactions, asserting the importance of person-to-person consultations for picking up on non-verbal cues, and issues such as safeguarding or loneliness (AoMRC, 2019).

> “There is a strong appreciation among the public for the patient-clinician relationship, and a desire for this not to be compromised by new technologies”

Relatedly, Fiske et al. (2019) raise questions around the safety of chatbot technologies in mental health care, specifically whether and how they would be able to connect those who need them with in-person services.

It is also suggested that autonomous systems, eg robot carers, could impede the autonomy of the person they are caring for, preventing them from doing harmful activities such as smoking (POST, 2018; Murphy et al. 2021). There are also questions over how those with dementia or intellectual disabilities, and children, could consent to the involvement of robots in their care (POST, 2018, Fiske et al. 2019), and the risk that vulnerable people may be coerced by robots (Fiske et al. 2019). The possibility for deception, if those being cared for are unaware that they are not interacting with a “real” carer or companion, is also raised (Nuffield Council on Bioethics, 2018; Murphy et al. 2021).

It is also suggested that the use of robots for care purposes could have implications for the dignity of those being cared for and may increase feelings of social isolation if they reduce human contact (Nuffield Council on Bioethics, 2018; Murphy et al. 2021). Over the longer term, Fiske et al. (2019) consider the possibility that people could become dependent on their engagement with robots and suggest that this could have implications for their relationships with other humans, and personal sense of identity.

An important debate therefore remains about what the acceptable uses of autonomous systems in health are – which tasks and decisions should be delegated to them, and which should remain with humans (Di Nucci, 2019; Morley et al. 2020).

Effects on healthcare professionals

The 2019 Topol Review, commissioned by the Secretary of State for Health, assessed the impact of digital technologies, including AI and robotics, on the NHS. It emphasised the potential for these technologies to improve the jobs of those working in the health service, providing them with the “gift of time” to spend on interacting with patients (Topol, 2019).

While the possibility that autonomous systems could lead to job displacement in healthcare has been mooted (eg, see Future Advocacy, 2018), the Topol Review concluded that “[o]ur review of the evidence leads us to suggest that these technologies will not replace healthcare professionals, but will enhance them” (p. 9). It did, however, stress the need for workforce training in the ethics of autonomous systems, the critical assessment and interpretation of AI outputs and the management of health data. Further, it recommended the creation of new roles within the health service for data scientists, technologists and knowledge specialists, and the establishment of an industry exchange scheme (Topol, 2019).

“Even if the total number of roles in healthcare does not diminish with the increased use of autonomous systems, there remains the possibility for the deskilling of clinical roles”

Even if the total number of roles in healthcare does not diminish with the increased use of autonomous systems, there remains the possibility for the deskilling of clinical roles (Joshi & Morley, 2019; Morley et al. 2020). The AoMRC raises the questions of whether the doctor could be relegated to a “second opinion”, and whether reducing the need for face-to-face consultations and problem solving could reduce job satisfaction (AoMRC, 2019).

Data privacy and security

Access to large, high-quality datasets is vital for the development of autonomous systems, meaning issues of data privacy and security are forefront. As we discussed when considering public trust, there is some public wariness of data sharing, though this depends on the type of data being shared, who it is being shared with and for what purpose. Also important is that people do not tend to be well-informed about measures to safeguard data, instead putting their trust unquestioningly in the NHS to protect data privacy (Castell et al. 2018).

There are also high-profile examples of health data sharing missteps, further underlining the need for consideration of how privacy can be protected. One such case is that of the agreement between the Royal Free NHS Trust in London and Google DeepMind, which involved the sharing of identifiable patient data without explicit consent to develop an app for the management of acute kidney injury (Powles and Hodson, 2017). The Information Commissioner’s Office concluded in 2017 that the Royal Free breached the Data Protection Act through its sharing of patient data (ICO, 2017).

The Care.data scheme to centralise patient data held within the NHS is another such example. Presser et al. (2015) conclude that patient anonymity did not receive suitable protections under the Care.data scheme, and that clearer consent processes, better communication with the public and more effective oversight of how patient data are used are important lessons for future schemes. Harwich and Laycock (2018) meanwhile suggest that the experience of the Care.data project has made clinicians more reticent towards data sharing initiatives.

A further complication when it comes to the privacy and security of health data is that it is increasingly unclear what actually constitutes health data. The Ada Lovelace Institute (2020) points out that more and more data are being collected via devices such as phones and wearables from which insights about health can be inferred. Indeed, one of the strengths of machine learning technologies are their ability to analyse large amounts of data from diverse sources to shed more light on the risk factors for disease and deliver a more personalised approach to health (van der Schaar & Zame, 2018).

Of concern though is that these data are being held by organisations outside of the medical profession, who may not treat it as sensitive data, and could use it for purposes that people are unaware of and would be uncomfortable with (Ada Lovelace Institute, 2020).

Finally, the development of AI technologies for health and care in the home, for example care robots, raises an additional set of concerns about data privacy and security (POST, 2018). Public opinion research suggests that people may be uncomfortable with technologies that surveil them in private settings (Castell et al. 2018).

Existing regulatory response

Health data are protected under the European GDPR legislation, which is embedded in UK law via the 2018 Data Protection Act (DHSC, 2018). While this clearly covers data such as medical records, there is concern that novel types of data from which health could be inferred may not be covered by this legislation, and thus would not be regulated as health data (Ada Lovelace Institute, 2020).

> “There is the risk that AI technologies could be used for surveillance or to gather information on people’s health without their knowing ”

The UK also has the National Data Guardian for Health and Social Care, an independent organisation that “advises and challenges the health and care system to help ensure that citizens’ confidential information is safeguarded securely and used properly” (National Data Guardian, n.d.). The organisation has set out the eight “Caldicott Principles” to guide the appropriate use of people’s confidential information (National Data Guardian, 2020).

The DHSC has recently proposed a set of measures to increase public knowledge of and confidence in how their health data are used. These include publishing a transparency statement on how data have been used across the health and care sector and giving the public the opportunity to see who has access to their data, and for what research purposes it has been used (DHSC, 2021b).

Malicious uses of AI

There is the risk that AI technologies could be used for surveillance or to gather information on people’s health without their knowing (Nuffield Council on Bioethics, 2018). The vulnerability of autonomous systems to cyber-attacks and data breaches has also been raised. The 2017 WannaCry attack exposed the NHS’s vulnerabilities to such an incident, with the National Audit Office (NAO) judging that it was the failure to follow straightforward cyber security good practices that exposed affected NHS organisations (NAO, 2018). It isn’t necessarily the case that new measures are required to prevent autonomous systems from cyber-attack, however. BSI concludes that “information security for AI solutions [does] not pose any known additional or distinct challenges when compared to other types of software” (BSI, 2019, p. 7).

Finlayson et al. (2019) address the vulnerability of medical deep learning systems to adversarial attacks, which involve introducing inputs to machine learning models that force it to make errors. They judge that “medicine may be uniquely susceptible to adversarial attacks, both in terms of monetary incentives and technical vulnerability” (p. 1), though it is important to note that they focus their analysis on the US healthcare system, which has a very different structure of financial incentives to the NHS.

Dwivedi et al. (2019) view AI systems as inherently vulnerable to adversarial attacks by hackers given their inability to use human-style intelligence in processing the information they are given. "As the programmes do not understand the inputs they process and outputs they produce, they are susceptible to unexpected errors and undetectable attacks" (p. 6).

Conclusion

The impact of autonomous systems in health and care is potentially enormous as they are deployed over the coming decades. Given our current position at the start of the process of developing, testing, and adopting these technologies, there is an invaluable opportunity to ensure trustworthiness is built in from the start, and to address, through policy and research, the issues that are likely to undermine trust in these technologies.

“The impact of autonomous systems in health and care is potentially enormous as they are deployed over the coming decades... [and] there is an invaluable opportunity to ensure trustworthiness is built in from the start

This landscape review has pointed to the need for policymakers and researchers to consider a range of issues pertinent to trust in the deployment of autonomous systems. While some issues are highly relevant to systems that are likely to be deployed in the near term, such as questions of safety and bias, data privacy and cyber security, others may be more relevant over the longer term, such as accountability with “black box” systems and humans’ interactions with robot carers. Regardless, beginning to consider our response to these issues today can ensure that the benefits of autonomous systems in health are shared by all, and that the risks associated with these innovations are managed.

This high-level review points to several areas where further research would be valuable. First, more precision around the degree of autonomy of technologies in health and care contexts would be helpful, along with further work understanding how trust and trustworthiness are likely to vary with autonomy. Second, the use of autonomous systems in health and care constitutes a very wide range of potential applications and settings. It seems highly likely that judgements about trustworthiness are likely to be very different depending on whether these technologies are being deployed in research contexts, for public health purposes, in hospitals to inform doctors or directly to patients in their homes. More research into what constitutes trustworthiness in different contexts would therefore be of value. Finally, as pointed out by Di Nucci (2019), there is a need for more deliberation about the types of task and decision that should be allocated to AI systems, and which should remain in human hands.

References

Ada Lovelace Institute (2020). The data will see you now: Datafication and the boundaries of health. https://www.adalovelaceinstitute.org/wp-content/uploads/2020/11/The-data-will-see-you-now-Ada-Lovelace-Institute-Oct-2020.pdf

AoMRC (2019). Artificial Intelligence in Healthcare. https://www.aomrc.org.uk/wp-content/uploads/2019/01/Artificial_intelligence_in_healthcare_0119.pdf

Blobel, B., Ruotsalainen, P., Brochhausen, M., Oemig, F., & Uribe, G. A. (2020). Autonomous Systems and Artificial Intelligence in Healthcare Transformation to 5P Medicine–Ethical Challenges. *Studies in Health Technology and Informatics*, *270*, 1089-1093. DOI: 10.3233/shti200330

BSI (2019). The emergence of artificial intelligence and machine learning algorithms in healthcare: Recommendations to support governance and regulation. Position paper. https://www.bsigroup.com/globalassets/localfiles/en-gb/about-bsi/nsb/innovation/mhra-ai-paper-2019.pdf

Castell, S., Robinson, L. and Ashford, H. (2018). Future data-driven technologies and the implications for use of patient data: Dialogue with public, patients and healthcare professionals. https://acmedsci.ac.uk/file-download/6616969

Challen, R., Denny, J., Pitt, M., Gompels, L., Edwards, T., & Tsaneva-Atanasova, K. (2019). Artificial intelligence, bias and clinical safety. *BMJ Quality & Safety*, *28*(3), 231-237. Doi: http://dx.doi.org/10.1136/bmjqs-2018-008370

BEIS (2021). The Grand Challenge missions. https://www.gov.uk/government/publications/industrial-strategy-the-grand-challenges/missions

BEIS (2018). Industrial Strategy: Building a Britain fit for the future. https://www.gov.uk/government/publications/industrial-strategy-building-a-britain-fit-for-the-future

DHSC (2021a). A guide to good practice for digital and data-driven health technologies. https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology

DHSC (2021b). Data saves lives: Reshaping health and social care with data (draft). 28 July 2021. https://www.gov.uk/government/publications/data-saves-lives-reshaping-health-and-social-care-with-data-draft/data-saves-lives-reshaping-health-and-social-care-with-data-draft#helping-developers-and-innovators-to-improve-health-and-care

DHSC (2018). The future of healthcare: our vision for digital, data and technology in health and care. https://www.gov.uk/government/publications/the-future-of-healthcare-our-vision-for-digital-data-and-technology-in-health-and-care

Di Nucci, E. (2019). Should we be afraid of medical AI? *Journal of Medical Ethics*, *45*(8), 556-558. Doi: http://dx.doi.org/10.1136/medethics-2018-105281

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., ... & Williams, M. D. (2019). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002

European Commission (2020). Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0064&from=en

European Commission (2019). Standard Eurobarometer 92: Europeans and Artificial Intelligence. https://europa.eu/eurobarometer/surveys/detail/2255

Finlayson, S., Chung, H. W., Kohan, I. and Beam, A. (2019). Adversarial Attacks Against Medical Deep Learning Systems. arXiv preprint arXiv:1804.05296.

Fiske, A., Henningsen, P., & Buyx, A. (2019). Your robot therapist will see you now: ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. J Med Internet Res 2019; 21(5): e13216. doi: 10.2196/13216

Fong, T. (2018). Autonomous Systems: NASA Capability Overview. NASA. https://www.nasa.gov/sites/default/files/atoms/files/nac_tie_aug2018_tfong_tagged.pdf

Future Advocacy (2018). Ethical, social and political challenges of artificial intelligence in health. https://wellcome.org/sites/default/files/ai-in-health-ethical-social-political-challenges.pdf

Genomics England (2019). A public dialogue on genomic medicine: time for a new social contract? https://www.genomicsengland.co.uk/public-dialogue-report-published/

Ghafur, S. et al. (2020). Public perceptions on data sharing: key insights from the UK and the USA. *The Lancet,* 2(9) doi: https://doi.org/10.1016/S2589-7500(20)30161-8

Habli, I., Lawton, T. and Porter, Z. (2020). Artificial intelligence in healthcare: accountability and safety. *Bulletin of the World Health Organization*, 98: 251-256. doi: http://dx.doi.org/10.2471/BLT.19.237487

Harwich, E. and Laycock, K. (2018). Thinking on its own: AI in the NHS. *Reform*. https://reform.uk/sites/default/files/2018-11/AI%20in%20Healthcare%20report_WEB.pdf

Information Commissioner's Office (2017). Royal Free – Google DeepMind trial failed to comply with data protection law. 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/

International Telecommunications Union (no date). Focus Group on "Artificial Intelligence for Health". https://www.itu.int/en/ITU-T/focusgroups/ai4h/Pages/default.aspx)

Joshi, I., Morley, J. (eds) (2019). Artificial Intelligence: How to get it right. Putting policy into practice for safe data-driven innovation in health and care. NHSX. https://www.nhsx.nhs.uk/ai-lab/explore-all-resources/understand-ai/artificial-intelligence-how-get-it-right/

King, D., Karthikesalingam, A. and Rees, G. (2018). Chapter 12: Emerging technologies in healthcare. *In Annual Report of the Chief Medical Officer 2018: health 2040 - better health within reach*. https://www.gov.uk/government/publications/chief-medical-officer-annual-report-2018-better-health-within-reach.

Marr, B. (2018). How Is AI Used In Healthcare - 5 Powerful Real-World Examples That Show The Latest Advances. Forbes. https://www.forbes.com/sites/bernardmarr/2018/07/27/how-is-ai-used-in-healthcare-5-powerful-real-world-examples-that-show-the-latest-advances/?sh=387ab8f75dfb

Morley, J., Machado, C. C., Burr, C., Cowls, J., Joshi, I., Taddeo, M., & Floridi, L. (2020). The ethics of AI in health care: A mapping review. *Social Science & Medicine*, 260. doi: https://doi.org/10.1016/j.socscimed.2020.113172

Murphy, K., Di Ruggiero, E., Upshur, R. et al. (2021). Artificial intelligence for good health: a scoping review of the ethics literature. *BMC Med Ethics* 22,14. Doi: https://doi.org/10.1186/s12910-021-00577-8

NASA (2020). Autonomous Systems. SE Research Consortium. https://www.nasa.gov/consortium/AutonomousSystems

National Audit Office (2018). Investigation: WannaCry cyber attack and the NHS. https://www.nao.org.uk/wp-content/uploads/2017/10/Investigation-WannaCry-cyber-attack-and-the-NHS.pdf

National Data Guardian (2020). The Caldicott Principles. https://www.gov.uk/government/publications/the-caldicott-principles

National Data Guardian (n.d.). About us. https://www.gov.uk/government/organisations/national-data-guardian/about)

NHS (2019). The NHS Long Term Plan. https://www.longtermplan.nhs.uk/wp-content/uploads/2019/08/nhs-long-term-plan-version-1.2.pdf

Nuffield Council on Bioethics (2018). Artificial intelligence in healthcare and research. https://www.nuffieldbioethics.org/publications/ai-in-healthcare-and-research

OECD (2020). Trustworthy AI in Health. http://www.oecd.org/health/trustworthy-artificial-intelligence-in-health.pdf

Ordish, J., Murfet, H. and Hall, A. (2019). Algorithms as medical devices. PHG Foundation. https://www.phgfoundation.org/documents/algorithms-as-medical-devices.pdf

Panch, T., Mattie, H. & Celi, L.A. (2019). The "inconvenient truth" about AI in healthcare. *npj Digit. Med.* 2, 77. https://doi.org/10.1038/s41746-019-0155-4

POST (2021). AI and healthcare. https://post.parliament.uk/research-briefings/post-pn-0637/

POST (2018). Robotics in Social Care. https://post.parliament.uk/research-briefings/post-pn-0591/

Powles, J., & Hodson, H. (2017). Google DeepMind and healthcare in an age of algorithms. *Health and Technology*, 7, 351-367. https://doi.org/10.1007/s12553-017-0179-1

Presser, L., Hruskova, M., Rowbottom, H. and Kancir, J. (2015). Care.data and access to UK health records: patient privacy and public trust. *Technology Science*. 2015081103. https://techscience.org/a/2015081103/

Royal College of Surgeons (2018). Future of Surgery. https://futureofsurgery.rcseng.ac.uk/?_ga=2.63604391.986526945.1547226199-1970392652.1547226199)

Smith, H. (2020). Clinical AI: opacity, accountability, responsibility and liability. *AI & Society*, 1-11. https://doi.org/10.1007/s00146-020-01019-6

Tan, S. Y., & Taeihagh, A. (2020). Governing the adoption of robotics and autonomous systems in long-term care in Singapore. *Policy and Society*, 1-21. https://doi.org/10.1080/14494035.2020.1782627

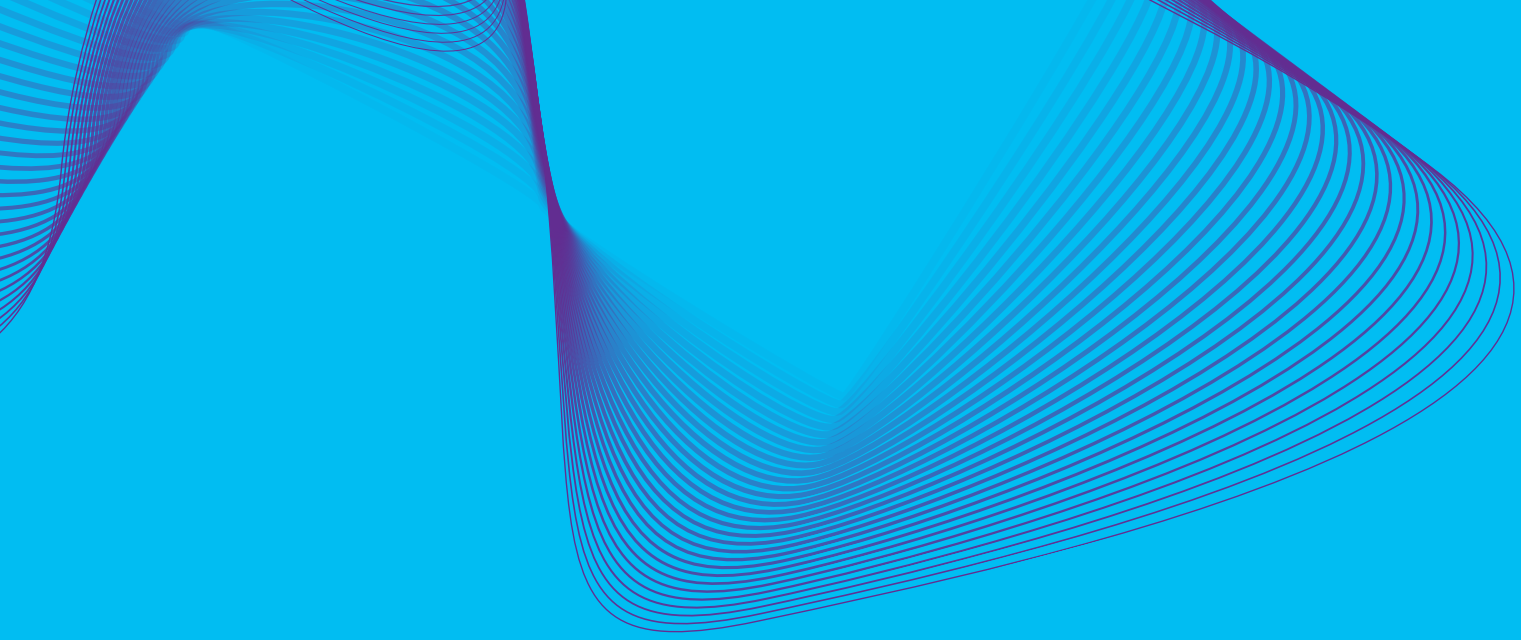
Topol, E. (2019). The Topol Review. Preparing the healthcare workforce to deliver the digital future. https://topol.hee.nhs.uk/

TAS-Hub (2020). Our Definitions. https://www.tas.ac.uk/our-definitions/

van der Schaar, M. and Zame, W. (2018). Chapter 10: Machine learning for individualised medicine. In *Annual Report of the Chief Medical Officer 2018: health 2040 - better health within reach*. https://www.gov.uk/government/publications/chief-medical-officer-annual-report-2018-better-health-within-reach

Wachter, S., Mittelstadt, B. and Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the general data protection regulation. *International Data Privacy Law*, *7*(2), 76-99. doi:10.1093/idpl/ipx005

YouGov (2020). If you were being scanned for cancer, would you be generally content or not for it to be read and interpreted by artificial intelligence rather than a doctor? https://yougov.co.uk/topics/health/survey-results/daily/2020/01/02/67595/1



UKRI Trustworthy Autonomous Systems Hub
The University of Nottingham
UNIVERSITY OF Southampton
THE POLICY INSTITUTE
KING'S College LONDON